

融合视觉语言模型的柚子分形树图像生成增强方法

赖力潜^{1,2} 段洁利¹ 杨 洲^{1,3} 袁浩天¹

(1. 华南农业大学工程学院, 广州 510642; 2. 嘉应学院计算机学院, 梅州 514015;
3. 广东海洋大学机械工程学院, 湛江 524088)

摘要: 为了降低柚子等水果目标检测对大量标注数据的依赖,本文提出了一种融合视觉语言模型的柚子分形树图像生成增强方法。该方法仅需3~5幅无标注真实图像,即可在无训练条件下生成大规模带标注的训练数据集。首先利用基于文本提示的零样本分割模型(Grounded segment anything model, Grounded SAM)提取柚树组件,然后结合稳定扩散模型 Stable Diffusion 使用文本提示生成随机背景,最后使用改进的分形树算法生成柚树以提升多样性及真实感。试验采用 YOLO v10 轻量化版本进行验证,在自建的非结构化环境柚子目标检测数据集上,当训练集真实图像数量分别为0、8、16、32、64幅时,使用本文方法后模型多阈值平均精度均值(Mean average precision at intersection over union thresholds from 0.50 to 0.95, mAP50-95)提升率依次达到662.3%、24.9%、13.7%、8.8%、1.8%。当训练集中真实图像数量为221幅,生成图像数量为512幅时,模型达到最优性能:精确率为76.9%,召回率为62.7%,mAP50为70.3%,mAP50-95为38.4%。迁移到橙子目标检测任务,相同数据规模下的性能提升分别为212.9%、16.5%、14.0%、5.2%、4.1%。当训练集中真实图像数量为1302幅,生成图像数量为512幅时,模型同样达到最优性能:精确率为90.3%,召回率为87.8%,mAP50为94.0%,mAP50-95为54.0%。试验结果表明,该图像生成增强方法在零样本和少样本学习场景中能够有效扩展训练数据,提高YOLO v10轻量化版本目标检测的性能,并展现出良好的泛化能力。

关键词: 柚子目标检测; 生成式数据增强; 少样本学习; 视觉语言模型

中图分类号: S24 **文献标识码:** A **文章编号:** 1000-1298(2026)01-0311-08

OSID:



Pomelo Fractal Tree Image Generative Data Augmentation Method
Using Vision-language Models

LAI Liqian^{1,2} DUAN Jieli¹ YANG Zhou^{1,3} YUAN Haotian¹

(1. College of Engineering, South China Agricultural University, Guangzhou 510642, China

2. College of Computer Science, Jiaying University, Meizhou 514015, China

3. School of Mechanical Engineering, Guangdong Ocean University, Zhanjiang 524088, China)

Abstract: Aiming to address the heavy reliance on large amounts of annotated data in fruit object detection tasks such as pomelo, a pomelo tree image generative data augmentation method was proposed based on vision-language models. The approach required only 3~5 unlabeled real images to generate a large-scale labeled dataset, which can be used to train object detection models and enhance their performance in zero-shot and few-shot scenarios. The method consisted of the following three main stages. Firstly, real pomelo tree components (including fruits, leaves) were extracted from unlabeled images by using the grounded segment anything model (Grounded SAM). Secondly, stable diffusion was used to create diverse background images based on textual descriptions, increasing the complexity and variability of the training data. Thirdly, a modified fractal tree algorithm was employed to construct structurally diverse pomelo trees, integrating real components with synthetic backgrounds to produce a variety of tree images and corresponding automatic annotations. Experimental results on pomelo object detection by

收稿日期: 2025-05-16 修回日期: 2025-07-31

基金项目: “十四五”广东省农业科技创新十大主攻方向“揭榜挂帅”项目(2024KJ27)、广州市重点研发计划项目(2024B03J1355)和2025年度嘉应学院科研项目(325E0317)

作者简介: 赖力潜(1989—),男,博士生,嘉应学院高级工程师,主要从事农业人工智能研究,E-mail: lqlai@stu.scau.edu.cn

通信作者: 杨洲(1972—),男,教授,博士,主要从事水果生产机械化与信息化研究,E-mail: yangzhou@scau.edu.cn

using the YOLO v10 model (Nano version) showed that the proposed method improved mAP50 - 95 performance by 662. 3% , 24. 9% , 13. 7% , 8. 8% , and 1. 8% when the number of real training images was 0 , 8 , 16 , 32 , and 64 , respectively . With 221 real and 512 generated images , the model achieved optimal performance : precision was 76. 9% , recall was 62. 7% , mAP50 was 70. 3% , and mAP50 - 95 was 38. 4% . When transferred to orange detection tasks under the same data conditions , performance gains were 212. 9% , 16. 5% , 14. 0% , 5. 2% , and 4. 1% . With 1 302 real and 512 generated images , the model achieved the best overall performance : precision was 90. 3% , recall was 87. 8% , mAP50 was 94. 0% , and mAP50 - 95 was 54. 0% , demonstrating strong generalization ability . Compared with tree images generated with blank backgrounds , the proposed method consistently outperformed across all training set sizes , whereas the blank-background approach only excelled in the zero-shot setting . Against traditional data augmentation techniques such as mosaic , this method performed better under low-shot conditions in pomelo detection , and although not the best in orange detection for every individual case , it achieved the best overall results under the default configuration of Ultralytics YOLO . In summary , the proposed method effectively mitigated the limitations caused by insufficient labeled data in fruit object detection model training and offered promising practical value and scalability .

Key words: pomelo object detection; generative data augmentation; few-shot learning; vision-language models

0 引言

柚子是中国南方和东南亚地区的重要经济作物。由于农村劳动力的不断流失,智能采收逐渐成为当前柚子产业发展趋势^[1]。目标检测是柚子智能采收的基础环节。近年来,基于深度学习的水果目标检测方法取得了较大进展^[2-5]。深度学习模型的成功往往依赖于采用大量带标注图像^[6-10]来训练模型。水果数据标注工作因数据量大、任务复杂,整体上是一个高度依赖人力投入的过程,特别是部分柚子品种如沙田柚采收时颜色与背景相近、遮挡较多,更加大了该过程的难度。

为了解决此问题,多项研究^[11-16]采用旋转、平移、缩放、翻转、裁剪、颜色空间变换等数据增强方法,用于扩展和增强水果目标检测数据集,减少对大量标注数据的依赖,同时提高模型的鲁棒性,减少过拟合情况^[17]。这些基于变换的图像增强技术在提升模型性能方面取得了良好效果,但其对原始数据的数量与质量仍具有较高依赖性。

生成式数据增强通过合成带标注的新图像,有效缓解了数据不足对水果目标检测性能的影响^[18]。相关方法^[19-23]在多类作物上均提高了检测精度,显示出良好应用前景。然而,这些方法多数对原始数据需求依然较大^[20-23],部分已有研究所处理的图像相对简单^[19],同时基于生成对抗网络的生成方法对算力资源的要求较高,应用推广仍面临挑战。

为降低非结构化场景下柚子检测对标注数据的依赖,本文提出基于 Grounded SAM^[24]和 Stable Diffusion^[25]的图像生成方法,通过组件提取、背景生

成、分形树生成与自动标注,生成多样带标注图像以增强训练。

1 材料与方法

1.1 数据集来源

(1) 柚子目标检测数据集

数据集包含 306 幅 1 920 像素 × 1 080 像素图像,共标注 9 628 个柚子实例,采集于梅县金柚冬季管理示范基地。图像采集未限定时间,农场未进行深度农艺改造,图像中常见曝光过度、运动模糊、遮挡严重、目标远近不一、柚子与背景颜色接近等问题,反映了非结构化环境下柚子检测的难点。

(2) 橙子目标检测数据集

数据集由 Roboflow 于 2024 年发布,包含 1 864 幅经方向校正、缩放至 640 像素 × 640 像素、对比度增强的图像^[26]。该数据集用于农业自动化中的橙子检测,本文用于评估方法的泛化能力。

1.2 研究方法

Grounded SAM 是一种融合开放词汇目标检测与图像分割的通用视觉模型,其核心架构整合了 Grounding DINO (DETR with Improved Denoising Anchor Boxes)^[27]和“分割一切”模型 (Segment Anything Model, SAM)^[28];前者是一种支持开放词汇的目标检测模型,能够基于文本提示在无需额外训练的情况下在一定程度上识别图像中的目标位置;后者则是一种高性能的图像分割模型,可针对任意图像区域生成精确的分割掩码。通过将 Grounding DINO 提供的目标定位结果输入至 SAM,可在无需额外训练或人工标注下,实现从自然语言提示到像素级分割的完整流程。

Stable Diffusion 是一种高效的文本生成图像模型,其核心潜在扩散模型 (Latent diffusion model, LDM)降低了计算资源消耗,使其在普通消费级图

形处理单元 (GPU) 上也能高效运行。此外,如图 1b 所示,Stable Diffusion 可以根据文本描述生成风格多样、构图多变的图像,有效缓解样本背景匮乏问题。

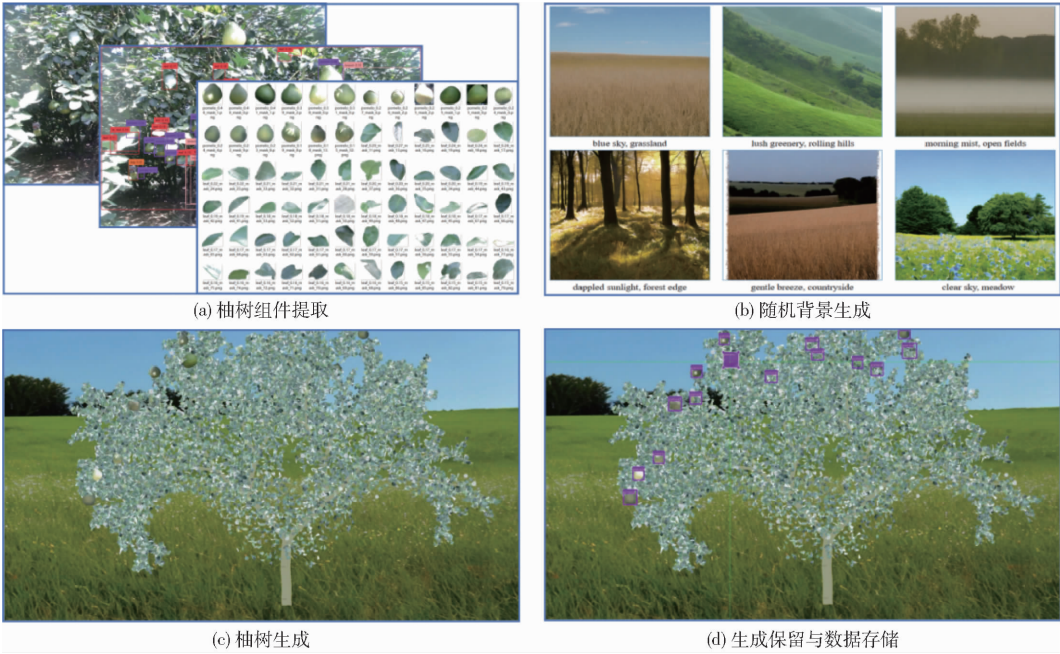


图 1 柚树图像生成方法流程图

Fig. 1 Flowchart of pomelo tree image generation method

传统的分形树 (Fractal tree) 算法通过递归方式构建具有层次结构分支的树形图像^[29]。通过调整分支角度、比例和数量,可以生成复杂和逼真的树形结构。

此外,YOLO 系列是主流的目标检测框架之一,因其速度快、精度高被广泛应用于实时检测场景^[30]。YOLO v10 进一步优化了模型结构与推理效率,特别适用于资源受限环境下的部署需求^[31]。在考虑农业生产的实际需求及限制后,本文选择使用 YOLO v10 的轻量化 (Nano) 版本对图像生成方法的有效性进行验证。

1.3 柚树图像生成方法

柚树图像生成方法包括柚树组件提取、随机背景生成、柚树生成、坐标保留及数据存储等模块,其整体流程如图 1 所示。准备工作包括使用 Grounded SAM 模型从数幅无标注图像中提取柚子、树叶等柚树组件,建立柚树组件库。待柚树组件库建立后,启动柚树图像生成流程:首先调用 Stable Diffusion 生成图像背景;然后使用改进的分形树算法生成树的枝干、覆盖粘贴变换后的树叶及柚子;最后将柚子的坐标转换成 YOLO 数据格式并保存,形成可直接用于训练的数据样本。

1.3.1 柚树组件提取

使用 Grounded SAM 提取柚树组件过程如图 2 所示:首先,构建文本提示词组 [pomelo, branch, leaf, bag], 将其与图像一同输入 Grounding DINO 模

型进行检测及匹配,得到可能的目标类别与边界框信息。随后通过非极大值抑制去除冗余检测结果。接着,将检测框输入 Segment Anything 模型,实现对每个目标的精细掩码分割,提取出像素级的目标区域并保存。最终,通过人工筛选出其中较优的实例用于后续使用。筛选后,柚子组件库包含树叶图像 55 幅,柚子图像 18 幅。在本文的柚子目标检测数据集上对 Grounding DINO 模型进行了定量性能评估,使用默认参数时,其精确率为 92.1%,召回率为 14.6%,mAP50 为 14.0%,mAP50 - 95 为 8.2%。此外,Grounded SAM 提取的检测掩码存在一定误差,某些树叶目标的检测掩码覆盖了整片背景树叶区域 (如图 2 右下角红色区域)。尽管如此,Grounded SAM 相比传统人工抠图具有显著效率优势,尤其适用于快速构建图像组件库等任务,展现出较强的实用性和便利性。为评估模型的资源消耗,前期试验还对比了参数量与推理效率:Grounding DINO (SwinT - OGC 版本) 参数量为 1.72×10^8 ,在 RTX8000 上单幅图像平均推理时间为 0.21 s;而 YOLO v10 Nano 版本参数量仅为 2.71×10^6 ,推理时间为 0.08 s。在资源有限或需处理大规模图像的场景下,Grounded SAM 因计算开销较大,更适用于图像组件库构建,而非实时目标检测任务。

1.3.2 随机背景生成

使用 Stable Diffusion 模型根据提示词生成复杂

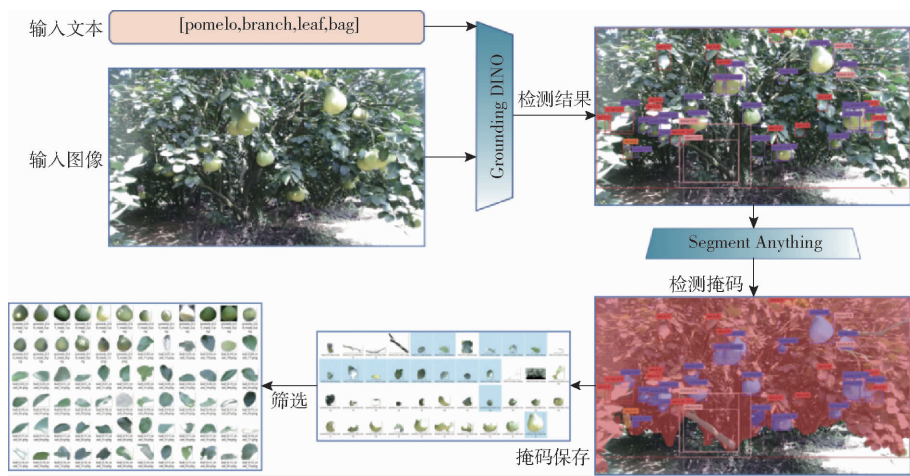


图 2 基于 Grounded SAM 的柚树组件提取流程图

Fig. 2 Component extraction process of pomelo trees based on Grounded SAM

多样的自然场景(如天空、草地或森林边缘)作为树木图像的背景。候选提示词包括“blue sky, grassland”, “lush greenery, rolling hills”, “morning mist, open fields”, “dappled sunlight, forest edge”, “gentle breeze, countryside”, “clear sky, meadow”。这些提示词大部分来源于采集的柚子图像中背景的典型特征,包含绿植覆盖的丘陵坡地、林缘空地、草地以及天空等自然环境元素,契合岭南地区丘陵地带柚子果园的真实生态环境特征。部分图像背景呈现晨雾、斑驳光影等光照变化,提示词涵盖了这类多样化的自然光照条件。此外,提示词的设计也考虑了背景样式的多样性,从而提升数据集在训练时的鲁棒性和泛化能力。生成时,从上述的候选提示词中抽取一个作为背景生成的候选提示词,模型将自动生成符合对应主题风格的图像。

1.3.3 柚树生成

柚树生成由传统的分形树算法改良而成。图像

尺寸为 1 920 像素 × 1 080 像素,具体步骤如图 3 所示,为了提高生成柚树的自然真实性与多样性,主要从以下方面进行改进:

- (1)划定根部初始区域:树根的起始位置不再固定,在画布底部预设区域内随机选取,横坐标 $x \in [860, 1\,060]$ 像素,纵坐标 $y \in [1\,000, 1\,150]$ 像素。
- (2)随机初始宽度及枝宽衰减:主干初始宽度在 $[25, 45]$ 像素之间随机生成,并以 0.7 的缩放系数递减。主干初始长度在 $[250, 300]$ 像素之间,下层分支缩放系数在 $(0.7, 0.9)$ 之间。
- (3)随机及动态变化的生长方向:每层分支被拆分为 10 段,每段引入随机微小角度 $\pm 5.6^\circ$ 之间的偏移,同时长度随机扰动 $(0.85 \sim 1.15)$ 倍),模拟自然弯曲和生长走势。
- (4)左右子分支的非对称深度衰减:初始深度为 10 层。子分支深度衰减采用概率控制的随机方式:70% 概率减少 1 层,20% 减少 2 层,9% 减少 3

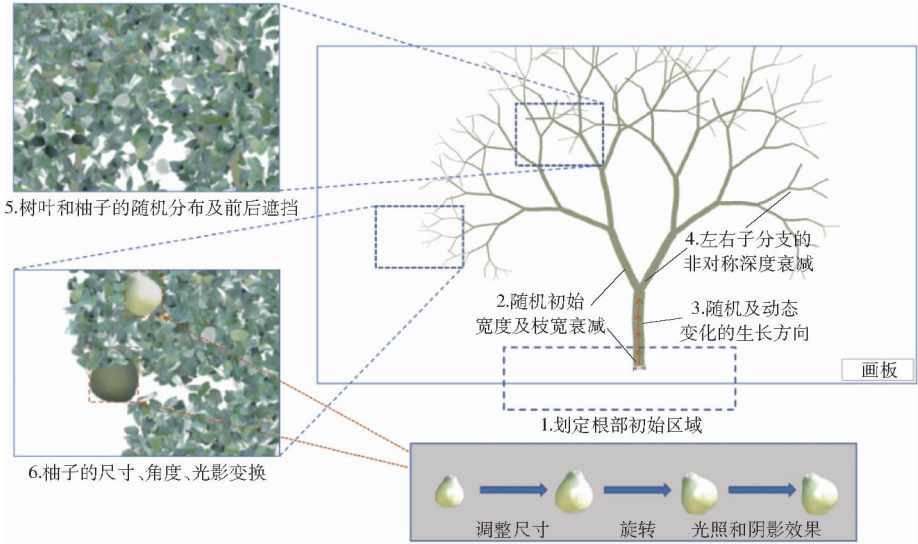


图 3 改进的柚树分形生成算法示意图

Fig. 3 Improved pomelo fractal tree generation diagram

层,1% 减少 4 层,形成不对称的树结构。

(5)树叶和柚子的随机分布及前后遮挡:树叶仅在深度小于 10 的分支生成,数量为深度的 3 倍,位置在 ± 13 倍深度范围内扰动,角度 $0^{\circ} \sim 360^{\circ}$,缩放后最大边长约 20 像素。柚子 10% 概率在深度不大于 2 的末端生成,偏移范围为 $x: [-10, 10]$ 像素、 $y: [-30, -10]$ 像素,宽高缩放至 40 ~ 60 像素,后生成的叶和果自动遮挡前者。

(6)柚子的尺寸、角度、光影变换:柚子生成时随机旋转 $0^{\circ} \sim 20^{\circ}$,尺寸为 40 ~ 60 像素,并施加亮度 $[0.7, 1.3]$ 、对比度 $[0.7, 1.3]$ 的光影扰动,以增强图像在多种光照条件下的自然感。

应用以上机制后,每棵生成的柚树在结构和形态上都与其他柚树存在一定差异,有效提升了生成样本的多样性与自然感。在贴图柚子时,记录对象的位置坐标信息,并保存相应的标注格式。完成图像生成后,系统即可获得带有多样背景的柚树图像及其对应的标注信息,降低了人工标注成本。

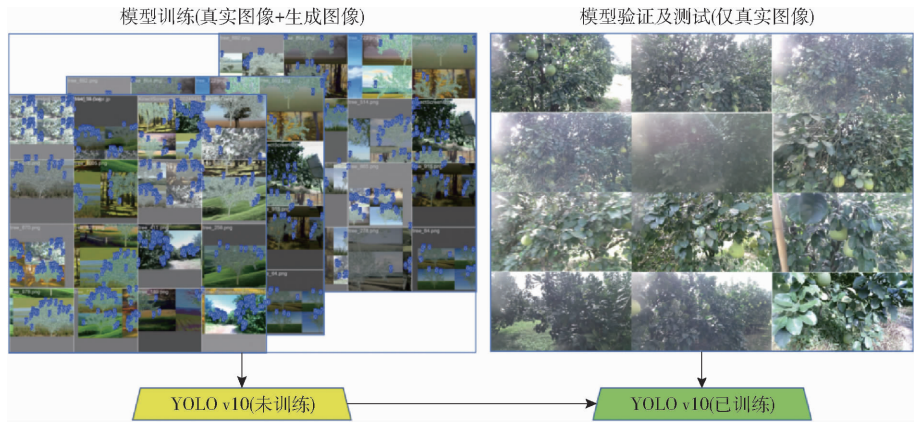


图 4 YOLO v10 的训练、验证及测试过程

Fig. 4 YOLO v10 training, verification and testing process

而当模型进行验证及测试时,仅使用真实图像。

主要使用 mAP50-95 作为模型性能的核心评估指标,它是一个较为严格的指标,通常比 mAP50 低,计算涉及多个概念,包括交并比 (Intersection over union, IoU)、真正例 (True positive, TP)、假正例 (False positive, FP)、假负例 (False negative, FN)、精确率 (Precision)、召回率 (Recall)、mAP、F1 分数等。训练迭代次数设定为 200 个 epochs。

2.2 柚子目标检测试验结果

2.2.1 不同训练配置下柚子目标检测结果示例

图 5 展示了在不同训练集配置下, YOLO v10 模型对两幅柚子图像的目标检测结果,其中绿色框为真正例,红色框为假正例,蓝色框为假负例。图 5a ~ 5d 对应第 1 幅图像在不同训练集配置下的检测表现。可以看到,当训练集中仅有 8 幅真实图像时,加入

2 试验结果分析

2.1 试验设置

试验使用的计算服务器 CPU 为 Intel(R) Xeon (R) Gold 6226R, GPU 为 RTX8000, 系统为 Ubuntu 20.04.2 LTS。Python 版本为 3.8.2, CUDA 版本 11.3, PyTorch 版本为 1.12.1。目标检测的模型为 YOLO v10 轻量化版本,验证平台为 Ultralytics。数据集包括柚子目标检测数据集以及橙子目标检测数据集,称之为真实数据集,真实数据集被划分为训练集、验证集、测试集,比例为 6:1:3。为了确保最终评估的可靠性和代表性,提高了测试集比例。在验证生成数据效果的试验中,训练集中根据相应的试验配置会包含部分真实数据和生成数据,验证集、测试集中只包含真实数据,以保证验证的有效性。

图 4 展示了本文验证试验中 YOLO v10 的训练、验证及测试过程。在训练过程中,真实数据与生成数据一起送入模型中进行训练,当开启图像拼接时,图像将由 4 幅图像拼接成 1 幅图像进行训练。

64 幅合成图像后, F1 分数从 54.5% 提升至 78.6%, 提升率达到 44.2%。而在训练集包含 221 幅真实图像时,加入 1024 幅合成图像后, F1 分数提升率为 5.4%。图 5e ~ 5h 展示了第 2 幅图像的检测结果。与第 1 幅图像的结果趋势一致,当训练集中仅有 8 幅真实图像时,添加 64 幅合成图像后, F1 分数提升了 177.4%,而在真实图像数量为 221 幅时,提升率为 22.1%。在图 5g 和图 5h 中,部分较小、较远的目标检测难度较大,且树叶遮挡的柚子也使得模型难以准确预测其边界,导致 IoU 小于 0.5,被判定为假正例。

2.2.2 不同图像生成方式性能对比

试验中训练集中生成图像数量设置为 64、128、256、512 和 1 024 共 5 种配置,并对其结果取平均以进行统计。在随机粘贴增强中,生成图像数量的增

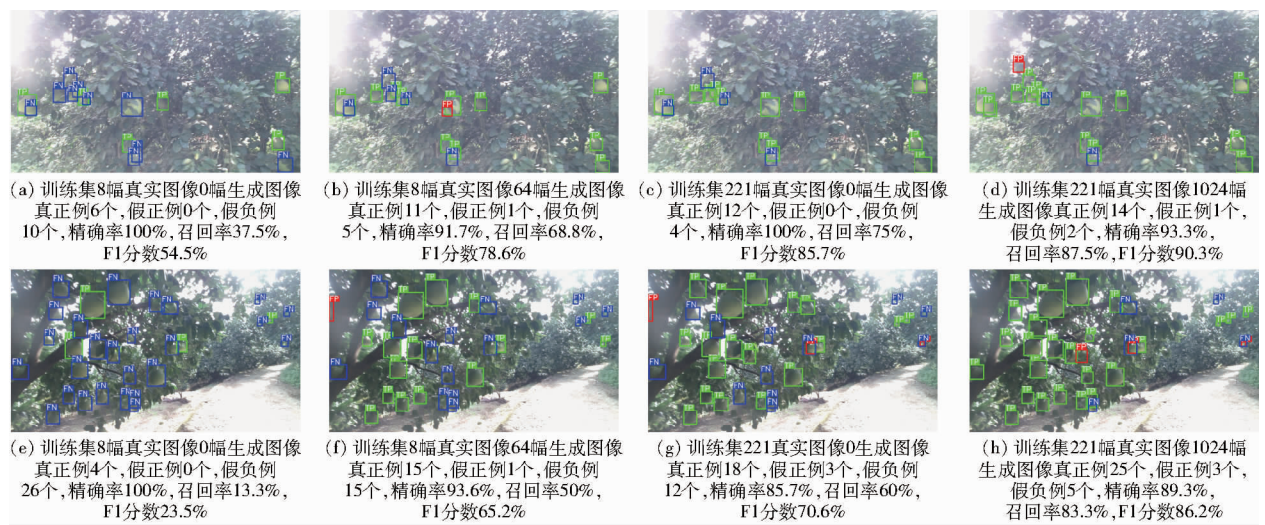


图5 不同训练集配置下的柚子检测结果示例

Fig. 5 Examples of pomelo detection under different training configs

加反而导致检测性能下降;而在其余 2 种增强方式中,各配置间性能差异较小,整体表现较为一致。当训练集中真实图像数量为 221 幅,生成图像数量为 512 幅时,达成全局的最优性能, mAP50 - 95 为 38.4%,其对应的精确率为 76.9%,召回率为 62.7%, mAP50 为 70.3%。图 6 对比了 3 种图像生成方式在目标检测任务中的性能差异。结果显示,随机粘贴方式难以有效提升模型性能,这在一定程度上是因为引入了与原始数据集特征差异较大的图像数据,进而导致模型预测准确率下降。相比之下,本文方法在训练集中不包含任何真实图像的情况下,较大提升了模型性能,提升率高达 662.3%。随着真实图像数量的增加,性能提升率逐渐减小,最终趋近于仅使用真实数据的性能水平。而采用空白背景生成柚树的方式,仅在真实图像数量为零的情况下对模型性能有一定提升,在其他配置下与仅使用真实图像的性能基本持平。

2.2.3 不同数据增强方式性能对比

所有与目标检测相关的数据增强方法的定义及参数设置均采用 Ultralytics 默认配置。在试验中,首先关闭所有数据增强选项,测试在不同真实图像数量的训练集上模型的 mAP50 - 95 性能,此设定标记“无增强”。随后,分别单独启用图像混合、图像拼接等共计 13 种增强方式进行训练与评估,以观察各自的独立效果。接着,测试启用所有增强方式的整体性能,以及默认增强配置和默认配置配合本文方法的组合效果。使用本文方法时训练集中生成图像的数量分别设置为 256、256、512。表 1 展示了各类增强策略下的性能比较结果(省略了效果相对一般的增强方法)。从表 1 可以看出,本文方法在少样本场景下有效提升了模型性能,表现优于其他增强

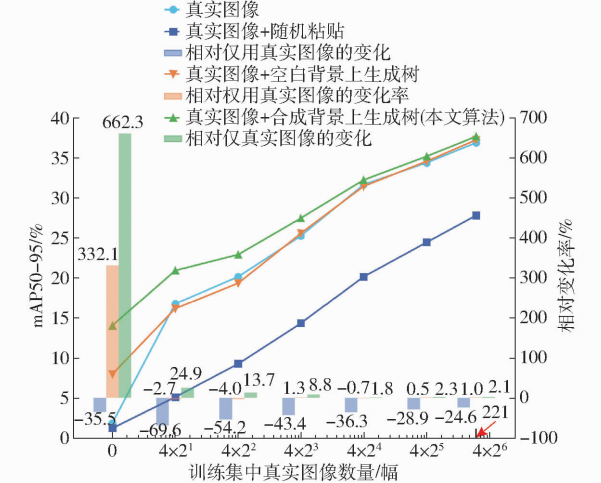


图6 不同图像生成方式对柚子目标检测性能的影响趋势图

Fig. 6 Trend of performance impact from different image generation methods on pomelo object detection

表 1 柚子目标检测不同数据增强方式对比

Tab.1 Comparison of data augmentation methods for pomelo detection							%
训练集中真实图像数量							
增强方法	8		64		221		
	mAP50 - 95	相对变化率	mAP50 - 95	相对变化率	mAP50 - 95	相对变化率	
增强	12.21	0	20.96	0	27.31	0	
图像拼接	13.61	11.47	26.56	26.72	32.62	19.44	
透视变换	13.79	12.94	24.36	16.22	32.45	18.82	
缩放变换	14.73	20.64	27.57	31.54	33.80	23.76	
平移变换	12.13	-0.66	26.27	25.33	31.65	15.89	
旋转角度	10.44	-14.50	26.51	26.48	32.64	19.52	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
本文方法	14.68	20.23	23.77	13.41	29.46	7.87	
全部应用	12.84	5.16	21.72	3.63	25.99	-4.83	
默认	16.77	37.35	31.68	51.15	36.92	35.19	
默认 + 本文方法	22.06	80.67	33.13	58.06	38.35	40.42	

策略。而如缩放变换和图像拼接等增强方式,则在真实图像数量较多时展现出更优的性能表现。此外,垂直翻转和剪切变换等增强方式在本数据集上效果不佳,导致启用全部增强方式后的整体性能未达到预期。默认配置在大多数情况下优于单一增强方式,而将本文方法与默认配置结合后,进一步提升了模型性能,取得了最优结果。

2.3 橙子目标检测试验结果

在橙子目标检测数据集上,采用与前述试验相同的训练配置与评估流程。图 7 是生成的橙树图像示例。

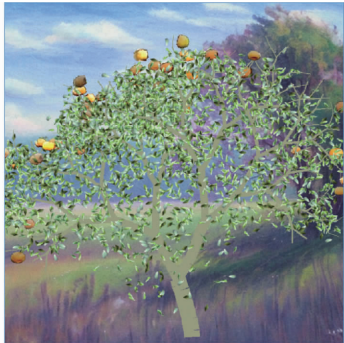


图 7 橙树生成示例

Fig.7 Example of orange tree generation

图 8 展示了在不同真实图像数量条件下,采用本文方法进行数据增强对模型性能的提升效果。在训练集不包含任何真实图像(数量为 0)时,本文方法带来的性能提升最为显著,提升率高达 212.9%。当真实图像数量分别为 9、16、32 和 64 时,性能提升率依次为 16.5%、14.0%、5.2% 和 4.1%。当训练集中真实图像数量为 1 302 幅,生成图像数量为 512 幅时,达成全局最优性能,mAP50 - 95 为 54.0%,对应的精确率为 90.3%、召回率为 87.8%、mAP50 为 94.0%。随着训练集中真实图像数量的持续增加,性能提升率逐渐减小,但在测试集规模内表现出相对优势,进一步验证了本文方法在多种样本配置下的有效性。

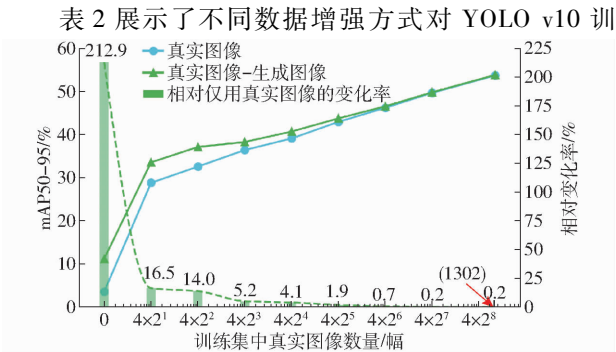


图 8 本文方法对橙子目标检测性能的提升效果

Fig.8 Improvement of orange object detection performance by proposed method

练性能的影响,其中训练集中生成图像的数量设置为 256。从表 2 可以看出,本文方法在多个设置下均能够有效提升模型的检测精度。然而,与缩放变换和图像拼接等经典增强方式相比,其性能提升率相对较小。同时,默认增强配置整体优于仅使用单一增强方式的设置。进一步地,将本文方法与 YOLO 默认配置相结合后,模型取得了全局最优的检测性能。

表 2 橙子目标检测不同数据增强方式对比

Tab.2 Comparison of data augmentation methods for orange detection

增强方法	真实图像数量					
	8		128		1 302	
	mAP50 – 95	相对变化率	mAP50 – 95	相对变化率	mAP50 – 95	相对变化率
无	25.48	0	35.28	0	47.37	0
图像拼接	28.08	10.20	40.40	14.51	52.55	10.94
水平翻转	24.52	–3.77	34.93	–0.99	47.29	–0.17
透视变换	27.86	9.34	41.82	18.54	52.30	10.41
缩放变换	29.91	17.39	42.58	20.69	53.36	12.65
平移变换	28.06	10.13	40.76	15.53	52.86	11.59
旋转角度	28.51	11.89	40.48	14.74	46.54	–1.75
⋮	⋮	⋮	⋮	⋮	⋮	⋮
本文方法	26.56	4.24	37.90	7.43	47.61	0.51
全部应用	21.52	–15.54	30.77	–12.78	41.17	–13.09
默认	28.83	13.15	42.96	21.77	53.73	13.43
默认 + 本文方法	35.04	37.52	44.15	25.14	53.91	13.81

3 结论

(1)在柚子目标检测任务中,当训练集真实图像数量分别为 0、8、16、32、64 幅时,使用本文方法后 YOLO v10 模型 mAP50 - 95 性能提升率依次达到 662.3%、24.9%、13.7%、8.8%、1.8%,当训练集中真实图像数量为 221 幅,生成图像数量为 512 幅时,达成全局最优性能,精确率 76.9%、召回率 62.7%、mAP50 为 70.3%、mAP50 - 95 为 38.4%;迁移到橙子目标检测任务,相同数据规模下的性能提升率分别为 212.9%、16.5%、14.0%、5.2%、4.1%,当训练集中真实图像数量为 1 302 幅,生成图像数量为 512 幅时,达成全局最优性能,mAP50 - 95 为 54.0%,对应的精确率 90.3%、召回率 87.8%、mAP50 为 94.0%,展现出良好的泛化能力。本文方法在零样本和少样本学习条件下对 YOLO v10 轻量化版本模型性能具有明显提升效果,随着训练集真实样本数量的增加,生成数据对模型性能的边际增益逐渐降低,但仍能有效拟合真实数据分布,保持一定的性能提升。

(2)与基于空白背景生成树的方法相比,本文方法在所有数据配置下均取得更优性能,而前者仅在零样本场景中表现出优势;相较于图像拼接等传统基于变换的数据增强方式,本文方法在柚子检测任务中的少样本场景下表现更优,在橙子检测任务中虽未取得单项最佳,但在所有配置下均能有效提升模型性能,且在默认训练配置下引入本文方法,可获得全局最优表现。

参 考 文 献

[1] 齐海军,周星星,罗毅智,等. 数字化技术促进广东柚果优质优价的路径研究[J]. 中国果树, 2023(12):124 – 129.

[2] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning-method overview and review of use for fruit detection and yield estimation[J]. Computers and Electronics in Agriculture, 2019,162:219 – 234.

[3] TANG Y, QIU J, ZHANG Y, et al. Optimization strategies of fruit detection to overcome the challenge of unstructured background in field orchard environment: a review[J]. Precision Agriculture, 2023,24:1183 – 1219.

[4] XIAO F, WANG H, XU Y, et al. Fruit detection and recognition based on deep learning for automatic harvesting: an overview and review[J]. Agronomy, 2023,13:1625.

[5] UKWUOMA C C, ZHIGUANG Q, BIN HEYAT M B, et al. Recent advancements in fruit detection and classification using deep learning techniques[J]. Mathematical Problems in Engineering, 2022,2022:1 – 29.

[6] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘mangoyolo’[J]. Precision Agriculture, 2019,20(6):1107 – 1135.

[7] GOYAL K, KUMAR P, VERMA K. AI-based fruit identification and quality detection system[J]. Multimedia Tools and Applications, 2023,82(16):24573 – 24604.

[8] LATIF G, MOHAMMAD N, ALGHAZO J. Deepfruit: a dataset of fruit images for fruit classification and calories calculation[J]. Data in Brief, 2023,50:109524.

[9] TYAGI L. Fruit detection dataset[DB/OL]. Kaggle. 2023. <https://www.kaggle.com/dsv/4922010>.

[10] ZAROUIT Y, ZEKKOURI H, OUHDA M, et al. Date fruit detection dataset for automatic harvesting[J]. Data in Brief, 2024,52:109876.

[11] GENÉ-MOLA J, VILAPLANA V, ROSELL-POLO J R, et al. Multi-modal deep learning for fuji apple detection using RGB-D cameras and their radiometric capabilities[J]. Computers and Electronics in Agriculture, 2019,162:689 – 698.

[12] 李丽,梁继元,张云峰,等. 基于改进 YOLO v5 的复杂环境下柑橘目标精准检测与定位方法[J]. 农业机械学报, 2024, 55(8):280 – 290.
LI Li, LIANG Jiyuan, ZHANG Yunfeng, et al. Accurate detection and localization method of citrus targets in complex environments based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(8):280 – 290. (in Chinese)

[13] 赵博,柳苏纯,张巍朋,等. 面向樱桃番茄采摘识别的轻量化 Transformer 架构优化研究[J]. 农业机械学报, 2024, 55(10):62 – 71, 105.
ZHAO Bo, LIU Suchun, ZHANG Weipeng, et al. Performance optimization of lightweight Transformer architecture for cherry tomato picking[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(10):62 – 71, 105. (in Chinese)

[14] 赵文锋,黄袁爵,钟敏悦,等. 基于 OrchardYOLOP 的火龙果园多任务视觉感知方法[J]. 农业机械学报, 2024,55(11):160 – 170.
ZHAO Wenfeng, HUANG Yuanjue, ZHONG Minyue, et al. Multi-task visual perception method in dragon orchards based on OrchardYOLOP[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(11):160 – 170. (in Chinese)

[15] 马鹏伟,周杰. 基于改进 YOLOv7 的复杂环境下的葡萄成熟度检测[J]. 农业工程学报, 2025,41(3):171 – 178.
MA Pengwei, ZHOU Jie. Detecting grape ripeness in complex environments using improved YOLOv7[J]. Transactions of the CSAE, 2025,41(3):171 – 178. (in Chinese)

[16] 郑文轩,杨瑛. 基于频域数据增强与轻量化 YOLO v7 模型的成熟期香梨目标检测方法[J]. 农业机械学报, 2024, 55(5):244 – 253.
ZHENG Wenxuan, YANG Ying. Mature stage pear detection method based on frequency domain data augmentation and lightweight YOLO v7 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(5):244 – 253. (in Chinese)

[17] XU M, YOON S, FUENTES A, et al. A comprehensive survey of image augmentation techniques for deep learning[J]. Pattern Recognition, 2023,137:109347.

[18] CHEN Y, YAN Z, ZHU Y. A comprehensive survey for generative data augmentation[J]. Neurocomputing, 2024,600:128167.

[19] RAHNEMOONFAR M, SHEPPARD C. Deep count: fruit counting based on deep simulated learning[J]. Sensors, 2017,17:905.

[20] 黄新林,张隆飞,唐小伟. 基于改进遗传算法的家电回收车辆路径规划方法[J]. 同济大学学报(自然科学版),2024,52(1):27-34.
HUANG Xinlin, ZHANG Longfei, TANG Xiaowei. Path planning method for household appliance recycling vehicle based on improved genetic algorithm[J]. Journal of Tongji University(Natural Science), 2024,52(1):27-34. (in Chinese)

[21] WALID E, HAIKAL E A, AJITH A, et al. A comparative study of the improvement of performance using a PSO modified by ACO applied to TSP[J]. Applied Soft Computing, 2014, 25:234-241.

[22] XUE J, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.

[23] 周江伟,邵洁,曹盛. 改进麻雀搜索的点云配准算法[J]. 计算机工程与应用,2025,61(11):295-305.
ZHOU Jiangwei, SHAO Jie, CAO Sheng. Point cloud registration algorithm based on improved sparrow search[J]. Computer Engineering and Applications, 2025,61(11):295-305. (in Chinese)

[24] 周建新,侯宏瑶,郑日成. 加入淘汰机制的改进麻雀搜索算法[J]. 火力与指挥控制,2024,49(3):65-72.
ZHOU Jianxin,HOU Hongyao,ZHENG Richeng. Improved sparrow search algorithm with elimination mechanism[J]. Fire Control & Command Control, 2024,49(3):65-72. (in Chinese)

[25] ZHANG Y, LI J, FEI L,et,al. Operational performance estimation of vehicle electric coolant pump based on the ISSA-BP neural network[J]. Energy, 2023, 268:126701.

[26] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, 1995: 1942-1948.

[27] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory [C] // MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. IEEE, 1995: 39-43.



(上接第 318 页)

[20] PAWARA P, BOSHCHENKO A, SCHOMAKER L R B, et al. Deep learning with data augmentation for fruit counting[C]//International Conference on Artificial Intelligence and Soft Computing. Springer, 2020:203-214.

[21] ZHANG W, CHEN K, WANG J, et al. Easy domain adaptation method for filling the species gap in deep learning-based fruit detection[J]. Horticulture Research, 2021,8:119.

[22] 张林鎡,巴音塔娜,曾庆松. 基于 StyleGAN2-ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. 农业机械学报, 2024,55(1):241-252.
ZHANG Linxuan, BA Yintana, ZENG Qingsong. Grape disease detection method based on StyleGAN2-ADA and improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(1):241-252. (in Chinese)

[23] ZHANG Z, TANIMOTO Y, IWATA M, et al. Data generation using pix2pix to improve YOLO v8 performance in UAV-based Yuzu detection[J]. Smart Agricultural Technology, 2025,10:100777.

[24] REN T, LIU S, ZENG A, et al. Grounded SAM: assembling open-world models for diverse visual tasks[J]. arXiv Preprint, arXiv:2401.14159, 2024.

[25] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[26] KULKARNI S. Orange dataset for object detection [DB/OL]. Roboflow, 2022. <https://www.kaggle.com/datasets/shreyas29k/orange-dataset-for-object-detection>.

[27] LIU S, ZENG Z, REN T, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection [M]. Springer Nature Switzerland, 2025:38-55.

[28] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023:4015-4026.

[29] NEWMAN W I, TURCOTTE D L, GABRIELOV A M. Fractal trees with side branching[J]. Fractals, 1997,5:603-614.

[30] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.

[31] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection[C]//Advances in Neural Information Processing Systems: Curran Associates, Inc., 2024:107984-108011.