



文献 CSTR:

32001.14.11-6035.csd.2024.0162.zh



文献 DOI:

10.11922/11-6035.csd.2024.0162.zh

数据 DOI:

10.57760/sciencedb.13135

文献分类: 农学

收稿日期: 2024-09-11

录用日期: 2025-10-23

发表日期: 2025-12-18

草莓多模态图像目标检测数据集

李科^{1*}, 徐启涛¹, 李玉杰¹, 刘昊天¹

1. 安徽农业大学, 信息与人工智能学院, 合肥 230036

摘要: 草莓在果实生长期内的精准检测一直是果园精准管理的关键技术难题。本数据集针对目前草莓目标检测领域缺乏公开的多模态图像数据集的问题, 通过实地采集和数据处理, 构建了草莓多模态图像目标检测数据集, 涵盖草莓生长期不同生长阶段、各种自然条件下的图像数据。本数据集包含可见光、深度和红外多模态图像, 总数据量为 31.36 GB。为了确保草莓多模态图像目标检测数据集的高质量和可靠性, 采用了严格的数据质量控制方法, 涵盖规范化采集、异常帧筛选、标注一致性、多模态对齐校准以及数据完整性和可读性的定期校验。本数据集为多模态图像数据融合与目标检测研究提供了宝贵的基础数据资源, 并可作为深度学习模型训练和测试的标准图库, 对推动草莓目标检测研究具有重要的应用价值。

关键词: 草莓检测; 多模态数据集; 图像处理; 深度学习; 农业自动化

数据库(集)基本信息简介

数据库(集)名称	草莓多模态图像目标检测数据集
数据通信作者	李科 (like@ahau.edu.cn)
数据作者	李科, 徐启涛, 李玉杰, 刘昊天
数据时间范围	每年11月至次年4月底
地理区域	安徽省合肥市市长丰县
数据量	31.36 GB
数据格式	*.png, *.txt
数据服务系统网址	http://doi.org/10.57760/sciencedb.13135
基金项目	安徽省高校协同创新计划(GXXT-2022-041); 安徽省高等学校质量工程项目(2022jyxm464); 安徽农业大学引进和稳定人才科研经费(yj2020-74)。
数据库(集)组成	数据集共包括4个数据文件夹, 其中: (1) rgb文件夹存放草莓的可见光图像, 以PNG格式保存, 共12840张, 数据量29.3 GB; (2) ir文件夹存放对应可见光图像的红外图像, 以PNG格式保存, 共12840, 数据量850 MB; (3) depth文件夹存放对应可见光图像的深度图像, 以PNG格式保存, 共12840, 数据量1.29 GB; (4) Annotations-YOLO文件夹存放对应可见光图像的YOLO格式标注文件, 包含图像中草莓的位置信息, 数据量3.41 MB。

* 论文通信作者

李科: like@ahau.edu.cn

引言

在现代农业生产中,精准农业和智能化技术的运用已成为提升作物品质和产量的重要手段^[1]。随着图像处理和机器学习领域的迅速发展,多模态图像目标检测技术已经在农业生产的自动化过程中展现出显著的价值。特别是在果实采摘、果蔬分类、病虫害识别以及成熟度评估等方面,该技术能够提供高效、无损的检测方法,从而极大地提高了农业生产的精细化管理水平^[2-6]。

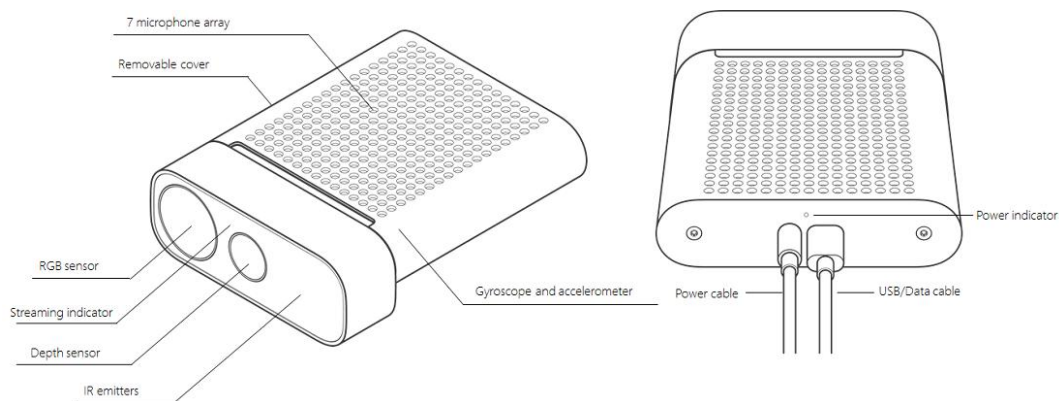
虽然通用的目标检测方法和数据集为计算机视觉领域提供了强大的基础,但它们在特定应用上仍存在限制。特别是在农业领域,由于环境的多变性和作物的特殊性,通用方法可能无法直接适用。此外,现有农业数据集往往只包含单一模态的图像,而现实世界中的条件可能需要结合多种信息源才能获得最佳结果^[7-11]。

因此,专门针对草莓并且融合了多模态信息的数据集能够更好地满足精准农业的需求。它不仅可以提高检测的准确性和鲁棒性,还可以作为研究更复杂问题的基础,如草莓成熟度预测、病虫害识别和采摘点定位等^[12-15]。尽管构建这样的数据集在资源和技术上有更大的挑战,但其潜在的学术价值和实际应用前景是非常高的。本研究旨在构建一个专门的草莓多模态图像目标检测数据集,该数据集结合了 RGB、红外和深度等多种模态的图像信息,以增强模型对草莓大小、形状、成熟度和采摘点的识别能力。通过融合不同模态的信息,可以更全面地捕捉到草莓的特征,提高检测的准确性和鲁棒性。此外,该数据集的建立将有助于推动多模态图像目标检测技术在农业领域的应用和发展,为研究人员提供一个标准化的测试和开发平台。这不仅是科技进步的体现,也是向智慧农业转型的重要步骤。通过本研究的深入开展,我们期望能够为农业领域带来切实的技术革新和经济效益。

1 数据采集和处理方法

1.1 数据采集方法

数据集中的草莓图像拍摄于安徽省合肥市长丰县数字草莓种植基地,数据采集时间是 2024 年 1 月,在大棚环境中使用 Kinect DK 深度相机拍摄高架草莓,图像获取设备与采集环境如图 1 所示,数据采集使用的 Azure Kinect 深度相机基本性能参数如表 1 所示。相机拍摄草莓的角度如图 2 所示,相机距离水果 30 - 60 厘米,分别从正前方、左前方、右前方和上方拍摄视频。为保证数据集样本的多样性,还需要在不同时段、不同光照条件下,对不同成熟度的大棚中的草莓进行拍摄,最终从视频帧中提取出 16625 张草莓图像。



(a)深度相机 Azure Kinect DK



(b)数字草莓种植区 Digital strawberry-growing area

图 1 图像获取设备与采集环境

Figure 1 Image acquisition equipment and data collection environment

表 1 Azure Kinect 深度相机基本性能参数

Table 1 Basic performance parameters of the Azure Kinect DK

指标项	性能参数
每秒帧数	30FPS
RGB 相机分辨率	RES_1080P
深度模式	NFOV_2X2BINNED
惯性测量单元	LSM6DSMUS/1.6kHz

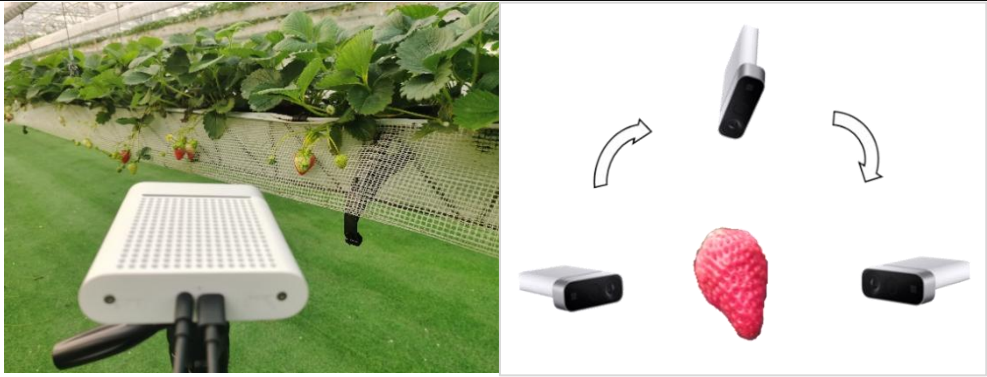


图 2 拍摄角度示意图

Figure 2 Illustration of photographing angles

1.2 数据处理与标注

1.2.1 数据预处理

数据预处理是将采集到的原始视频数据转化为适合后续图像处理和步骤。本数据集的预处理过程包括从视频中提取图像、图像对齐、格式转换和图像质量优化等步骤。从原始视频中提取 RGB 图像、深度图像和红外图像，使用 Pyk4a 工具将深度图像和红外图像与 RGB 图像进行对齐；由于深度图像和红外图像的位深度为 16 bit，提取后的图像数据需要进行适当的处理，将对齐后的深度图像和红外图像从 16 bit 转换为 8 bit，并将所有图像保存为 PNG 格式。为了确保数据的质量和一致性，对提取后的图像进行质量优化处理，主要步骤包括去除无效和污染数据、图像增强等。检查提取的图像，去除模糊或者无法识别的图像，对图像进行适当的增强处理，如调整亮度、对比度等，以确保图像质量的一致性，最后将处理后的图像按照一定的编号规则保存，以便后续的标注和使用。

1.2.2 数据标注

数据标注是将处理后的图像进行目标标注，生成用于训练和测试的标注文件。本研究采用方框标注的方法，对图像中的草莓进行标注。拍摄的草莓，大致能划分为三个生长状态阶段，分别为绿果（green strawberry）、白果（white ripening）以及红果（ripe strawberry）。准确判定草莓的成熟度极为重要，依据中国农业草莓行业标准，当草莓红色着色率超过 70%时，便达到可采摘并食用的标准，此类草莓被标注为红果；若红色着色率低于 70%，且呈现红白相间外观的，标注为白果；而通体皆为绿色的，就标注为绿果。如图 3、图 4 所示，首先使用 LabelImg 工具，通过矩形框对图像中的草莓进行手动标注，生成 YOLO 格式的标注文件。然后将部分手动标注的数据用于训练 YOLOv8n 模型，训练后的模型预测效果如图 5 所示，利用训练好的模型进行初步的目标检测，并生成标注文件。对生成的标注文件进行人工校正，确保标注的准确性。最后将处理和标注后的图像数据和标注文件按照一定的组织结构进行存储，方便后续的使用和管理。

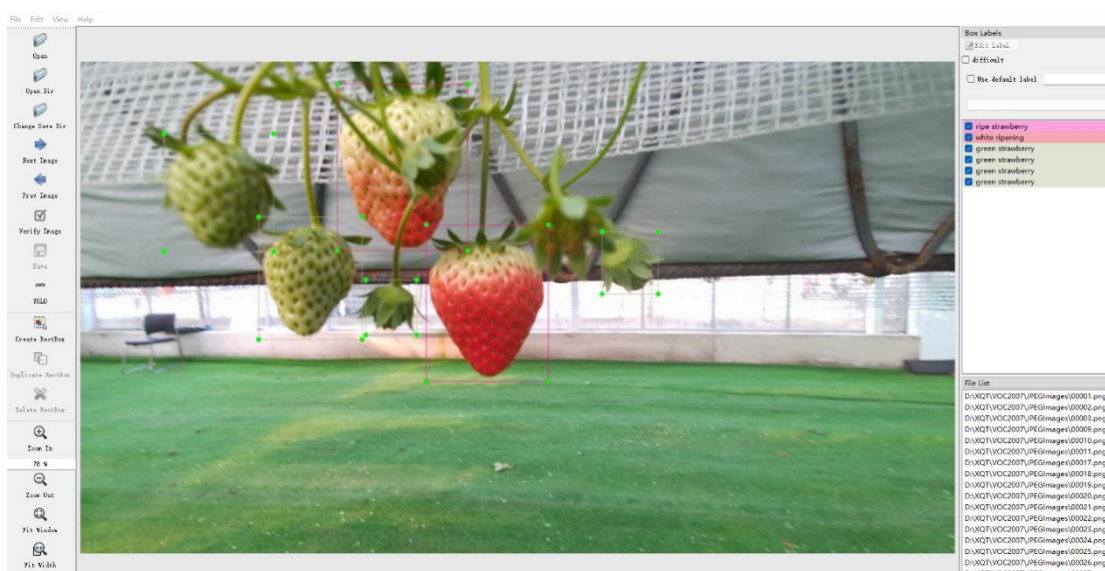


图 3 LabelImg 工具标记草莓标签

Figure 3 Strawberry labeling using the LabelImg tool

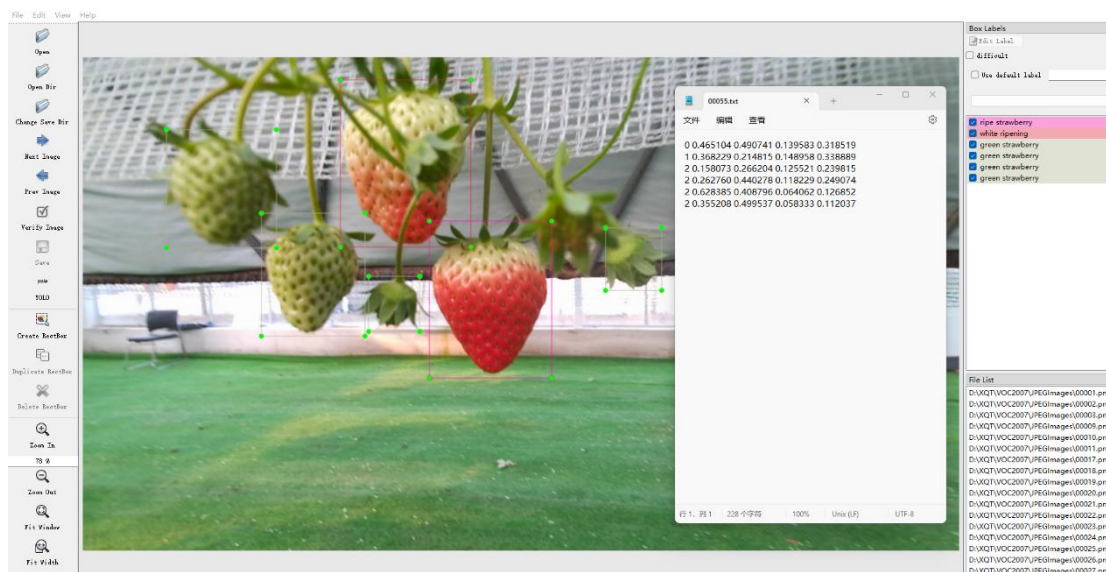


图 4 草莓数据集标注文件

Figure 4 Strawberry dataset labeling file



图 5 模型预测效果

Figure 5 Model prediction results

2 数据样本描述

本数据集通过视频录制采集真实场景，并采用等间隔抽帧生成静态图像样本。这种采集方式导致相邻样本间存在一定的时间连续性，表现为部分图像在视角、光照或目标姿态上具有渐进式变化，如图 6 所示。需要强调的是：抽帧间隔确保了样本间不存在完全相同的图像，且场景动态变化使得样本具有实质差异性；此类渐进式变化的样本模拟了真实场景的连续性，可增强模型对目标外观微小变化的鲁棒性，提升泛化能力；YOLO 等目标检测模型通过数据增强显式打破局部相似性，且训练早停策略基于独立验证集性能，因此时间连续性不会导致模型过拟合。

数据集中的草莓主要包含“丰香”“红颜”“女峰”这几个品种。这些品种一般从 11 月前后开始，陆续分批次挂果并逐渐成熟，这种过程会一直延续到次年 4 月末，因此数据适用时间范围为每年 11 月至次年 4 月底。草莓多模态图像目标检测数据集包含三种主要模态的图像数据：RGB 图像、深度图像和红外图像。数据集涵盖了草莓不同生长阶段（幼小草莓、红白半熟草莓和红熟期草莓）

在不同自然条件下（顺光、逆光等）的图像样本。每个生长阶段的数据均包含上述三种模态的图像，并附有 YOLO 格式的标注文件。

数据集的文件组织形式如表 2 所示。

表 2 数据集文件组织形式

Table 2 Organization of dataset files

文件夹	子文件夹名称/文件类型	文件内容/说明
Strawberries	RGB/*.png	草莓彩色图像
	Depth/*.png	草莓深度图像
	Infrared/*.png	草莓红外图像
	Annotations-YOLO/*.txt	YOLO 格式标注文件

2.1 图像数据样本

为了更好地理解草莓多模态图像目标检测数据集的构成和特点，本节将详细介绍数据集中各类典型样本的图像数据结构和特征。首先，为了使数据集能够适应不同的应用场景，我们对草莓采用了不同的拍摄角度，包括左平视、正平视、右平视；左俯视、正俯视、右俯视；左仰视、正仰视、右仰视，如图 6 所示。处于不同成熟状态的草莓，在图 7 中有所展示。此外，数据集涵盖了不同自然光照条件下（顺光、逆光等）的图像样本，如图 8 所示。

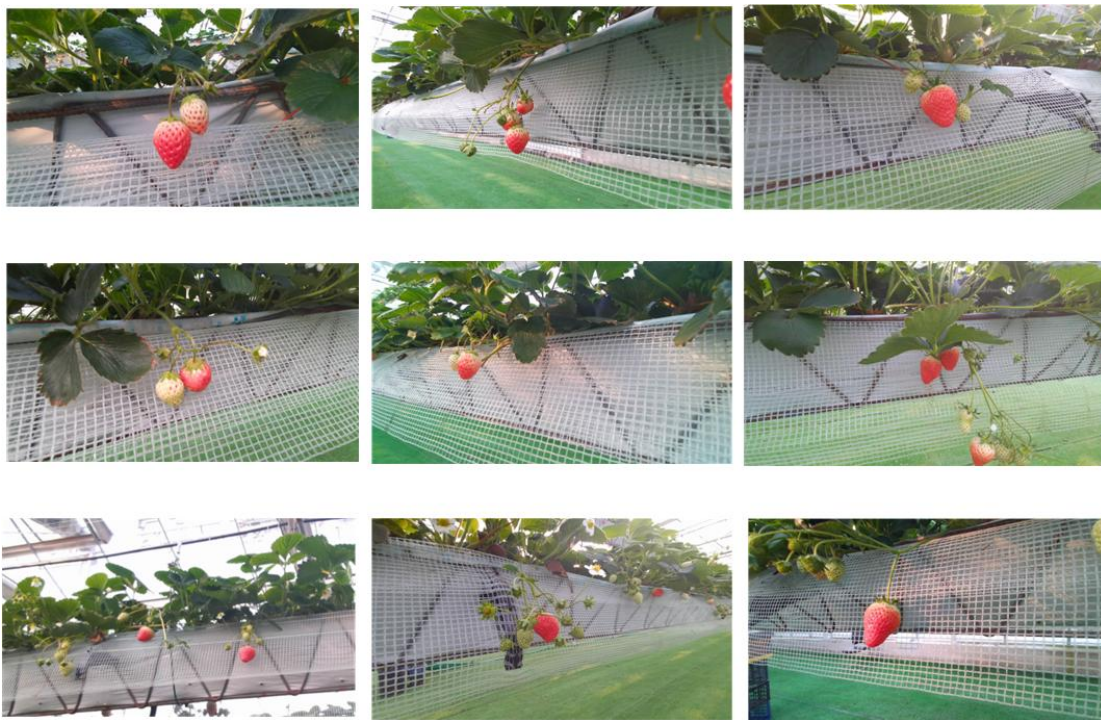


图 6 不同拍摄角度下的草莓

Figure 6 Strawberries captured from different viewing angles



(a)红果和绿果

(b)红果和白果

(c)白果和绿果

图 7 不同生长阶段的草莓

Figure 7 Strawberries at different growth stages



(a)顺光

(b)逆光

图 8 不同自然光照条件下的图像样本

Figure 8 Sample images under different natural lighting conditions

数据集包含 RGB、深度和红外三种模态的图像数据，涵盖了草莓在不同生长阶段和光照条件下的多样化样本。这些样本数据不仅为模型训练提供了丰富的视觉信息，还为多模态数据融合研究奠定了基础。本节选取包含不同成熟度草莓的典型图像作为样本展示，如下图所示。

(1) RGB 图像样本如图 9 所示。文件类型：/*.png；分辨率：1920×1080 像素；文件大小：3.59 MB；描述：该样本拍摄于晴天顺光条件下，图中展示了三种不同成熟度草莓果实，果实与绿叶之间对比明显，便于视觉识别和检测。



图 9 RGB 图像样本

Figure 9 RGB image sample

(2) 深度图像样本如图 10 所示。文件类型：/*.png；分辨率：640×576 像素；文件大小：117 KB；

描述：对应于 RGB 图像的深度图像，展示了果实的空间位置和相对距离信息，能够帮助模型理解目标的三维结构，下图示例为 Azure Kinect SDK 软件工具包中的 Azure Kinect Viewer 程序读取获得。

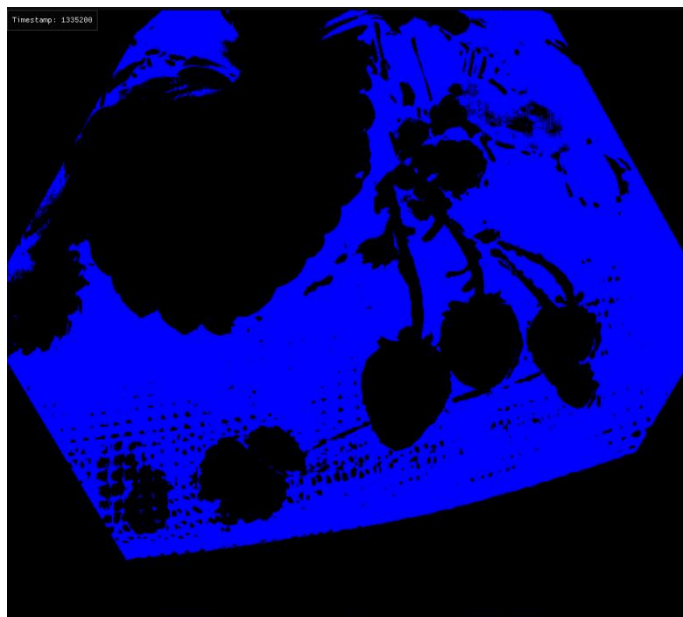


图 10 深度图像样本

Figure 10 Depth image sample

(3) 红外图像样本如图 11 所示。文件类型：/*.png; 分辨率：640×576 像素; 文件大小：355 KB; 描述：红外图像提供了关于植株健康状态的额外信息，果实与背景的热辐射差异明显，可用于增强目标检测的鲁棒性。

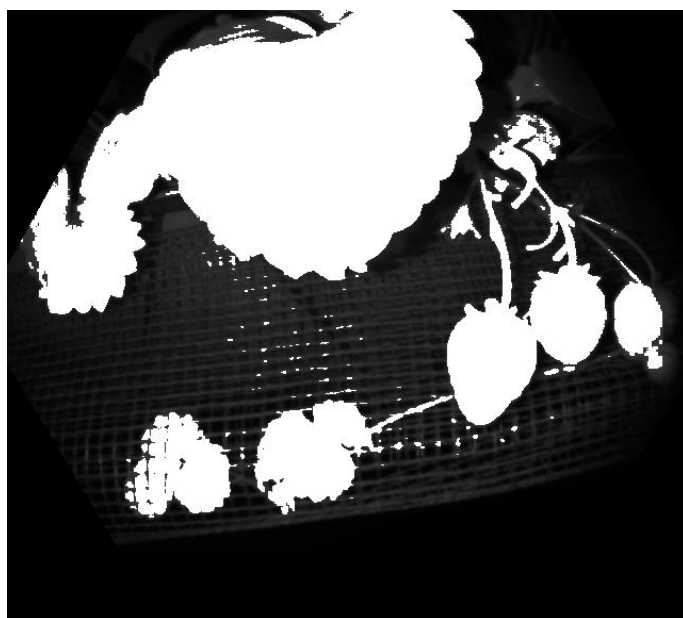


图 11 红外图像样本

Figure 11 Infrared image sample

以上举例的数据样本涵盖了草莓不同生长阶段和不同自然条件下的典型图像样本，具有代表性和精简性。通过这些样本，读者可以充分了解数据集的结构和内容，便于理解和使用数据集。

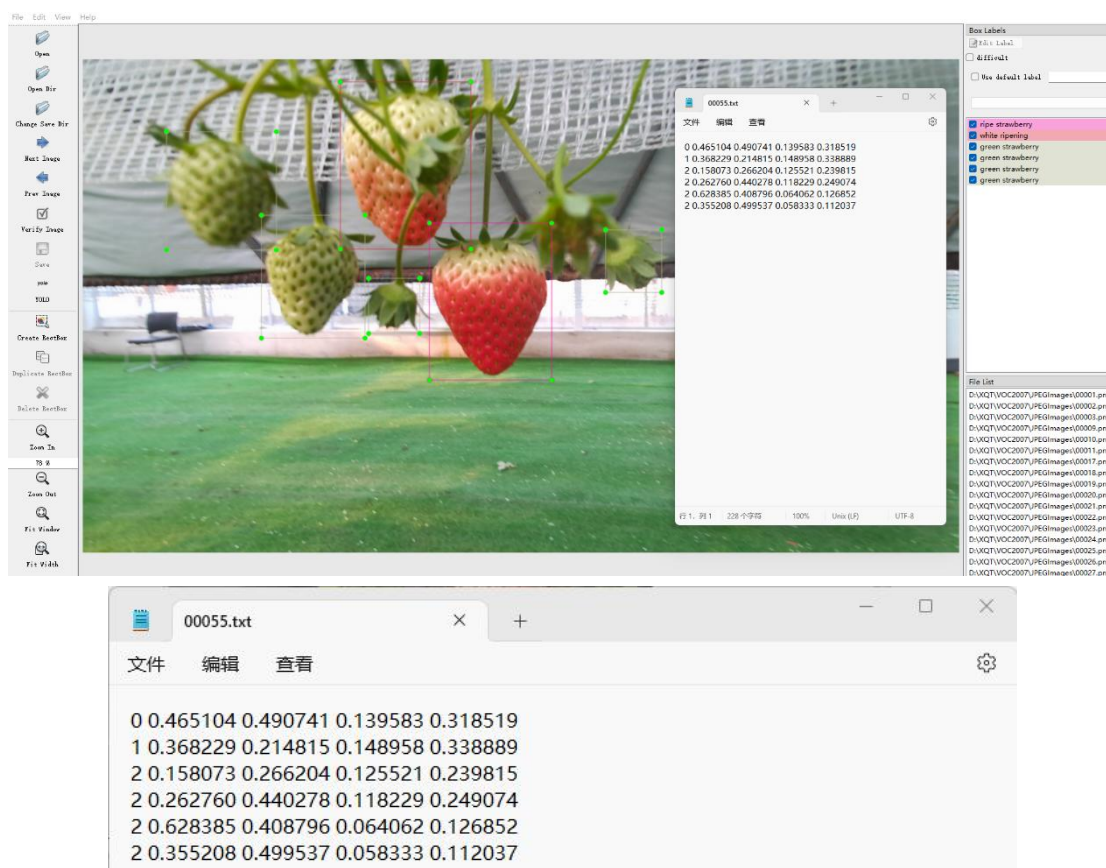
2.2 标注数据样本

为了满足草莓在自然环境下各类农业操作的应用需求,本数据集对草莓果实进行了详细的标注,采用了矩形框标注方法。标注过程分为单类标注和多类标注两种方法,以支持不同层次的目标检测需求。单类标注主要用于标注图像中草莓果实的大小和位置。通过在图像中绘制矩形框,准确标记每个草莓果实的边界,以支持基础的目标检测算法研究。这种标注方法适用于简单场景下的草莓检测,能够帮助模型识别果实的基本特征。多类标注则进一步对草莓果实的成熟情况进行分类,包括以下三类:绿果(green strawberry)、白果(white ripening)以及红果(ripe strawberry)。本数据集的标注结果以 YOLO 格式保存,以适应不同目标检测模型的需求,标注结果以.txt 格式保存,包含每个目标的详细位置和类别信息,支持复杂的目标检测和实例分割任务。如图 12 所示,每个目标对象的标注信息包括其类别和位置,每个标签文件包含多行,每行代表图像中的一个目标。每行的格式如下:

`<object-class> <x_center> <y_center> <width> <height>`

其中,<object-class>是目标的类别索引,从 0 开始。<x_center>和<y_center>是目标中心的坐标,相对于图像宽度和高度的比例值,范围在 0 到 1 之间。<width>和<height>是目标的宽度和高度,也是相对于图像宽度和高度的比例值。

例如,如果一个图像中有一个红果(ripe strawberry)目标,其类别索引为 0,中心坐标为图像宽度的 46.5104%和高度的 49.0741%,标注框宽度为图像宽度的 13.9583%,标注框高度为图像高度的 31.8519%,则对应的标签文件内容为:0 0.465104 0.490741 0.139583 0.318519。



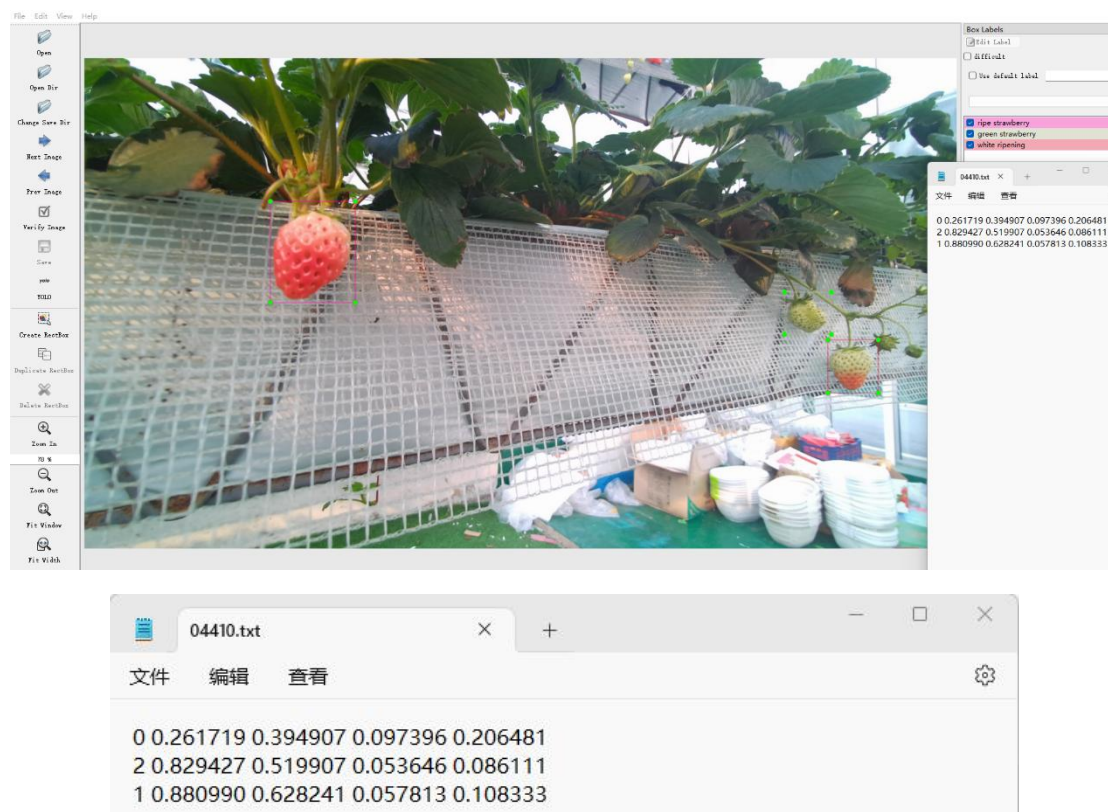


图 12 草莓数据标注样本

Figure 12 Samples of strawberry data labeling

3 数据质量控制和评估

为了确保草莓多模态图像目标检测数据集的高质量 and 可靠性, 采用了严格的数据质量控制方法。使用高精度的 Azure Kinect 深度相机进行数据采集, 确保图像分辨率为 1920×1080 像素。在实际果园环境下进行数据采集, 覆盖不同光照条件 (顺光、逆光) 和不同成熟度, 以保证数据的多样性和代表性。确保摄像头与草莓果实表面所在平面尽量垂直, 保证图像清晰, 减少因拍摄角度引起的变形和误差。使用 Pyk4a 工具对深度图像和红外图像进行与 RGB 图像的对齐, 确保不同模态图像之间的空间一致性。在数据处理过程中, 剔除模糊、污染严重或无法识别的图像, 以保证数据的有效性。将 16 bit 的深度和红外图像转换为 8 bit, 确保图像在后续处理中的兼容性和易用性。对每一张图像进行人工标注或半自动标注下的人工确认, 并生成 YOLO 格式的标注文件。初始标注由经验丰富的标注员进行, 确保标注的准确性。

为了验证数据集的可靠性, 采用了定性评估和定量评估方法。在定性评估中, 我们首先进行了样本检查, 随机抽取数据集中的图像样本, 进行视觉检查, 确保图像清晰、标注准确。然后进行标注验证, 对标注文件进行抽样检查, 确认标注框位置和大小准确性。

在完成数据准备与标注后, 本研究需要客观、全面地评估所提出的模型在草莓果实检测和采摘点定位上的性能表现。选择合适的评价指标不仅有助于比较不同模型或改进版本的优劣, 还能为后续的参数调整与结构优化提供参考依据。基于研究目标与相关文献惯例, 本研究主要采用的评价指标包括精确率 (Precision)、召回率 (Recall)、平均精度均值 (mAP)。

(1) 精确率 (Precision)

精确率用于衡量检测结果中真正的正样本 (True Positive, TP) 在模型判定为正样本 (预测检测到的目标) 中的比例, 常用公式如式(1)所示。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

其中, TP 表示真实为正且被模型正确检测为正的数量, FP 表示真实为负但被模型错误检测为正的样本数量。精确率越高意味着模型预测为目标的草莓中, 真正属于草莓的比例越大, 也即减少了“误检”情况。对草莓果实检测来说, 如果精确率过低, 则在实际使用中会导致机器人频繁对不存在的果实进行抓取尝试, 影响作业效率。

(2) 召回率 (Recall)

召回率衡量了所有正样本中被正确检测到的比例, 常用公式参见式 (2)。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

其中, FN 表示真实为正但被模型错误检测为负 (漏检) 的数量。召回率越高意味着在全部真实草莓目标中, 被正确识别的比例越大, 从而降低了对草莓果实的漏检概率。在实际应用中, 如果召回率过低, 部分成熟可采摘的草莓将被遗漏, 无法实现对果实的全面检测。

(3) 平均精度均值 (mAP, mean Average Precision)

mAP 是目标检测领域常用的综合指标, 表示对每一个类别计算平均精度 (AP, Average Precision), 再对所有类别取平均值。对于单类别 (如只检测草莓) 的情况, mAP 可以退化为其 AP 值。先根据模型输出的预测框与真实标注框之间的 IoU (Intersection over Union) 确定是否属于正确检测 (True Positive), 并将检测结果按照置信度或预测分数排序; 对每一个阈值 threshold 下计算 Precision 和 Recall 得出 PR 曲线; 再对 PR 曲线进行积分, 得到 AP, 最后对所有类别的 AP 进行平均得到 mAP。mAP 能够反映模型在不同置信度阈值下的检测性能, 能够更好地衡量模型在多场景、多类别下的综合表现。本研究中虽主要关注草莓果实, 但对于不同成熟度或品种的草莓, 也可将其视为不同类别进行评估, mAP 能呈现模型在多类别检测时的全局性能。

随后我们采用模型验证的方式进行了定量评估, 将处理后的草莓数据集随机划分为训练集、验证集和测试集。具体划分比例为: 8 : 1 : 1。划分过程确保了各类别的样本在三个子集中分布均衡。训练集用于模型参数的更新, 验证集用于超参数调整和早停以避免过拟合, 测试集用于最终评估模型的泛化性能。本研究选取了 Faster R-CNN、YOLOv11n、YOLOv10n、YOLOv9n、YOLOv8n 五个基准模型, 并与自己改进的 SD-YOLOv8n 模型进行系统对比。对比的指标包括检测精度 (mAP)、精确率 (Precision)、召回率 (Recall), 以全方位证明本数据集可以衡量各模型的综合性能。检测模型实验结果对比如图 13 所示, 可视化效果如图 14 所示。

如表 3 所示, 将准确率 (Precision)、召回率 (Recall)、平均精度 (mAP) 作为评估指标对数据集的目标检测模型训练与验证性能进行定量评估, 通过模型验证过程, 得到了以下实验结果:

从图 13 和表 3 中可以看出, SD-YOLOv8 以最高的精确率 (0.904)、召回率 (0.875) 及最高的 mAP50 (0.932)、mAP50-95 (0.82) 夺得整体“最佳”, 同时保持了最短的推理时间 (4 ms)。该模型最适合对精度与速度都有较高需求的场景。YOLOv8 以 0.882 的精确率和 0.922 的 mAP50 排名第二, 速度 4.7 ms, 也非常接近 SD-YOLOv8, 适用面较广。YOLOv9 在召回率上相对突出 (0.873),

如果应用场景更注重漏检率,就可优先考虑 YOLOv9; 不过其精确率略低于 YOLOv8、SD-YOLOv8。YOLOv10 和 YOLOv11 在 mAP50 上与 YOLOv8 处于同一水平,但在 mAP50-95 等指标上微有波动; 速度分别为 5.2 ms 和 7 ms,也可根据计算资源和检测需求进行选用。尽管在精确率、召回率上仍具一定竞争力, FasterR-CNN 也常被作为更传统的、可解释性较强的双阶段检测器示例,但其推理速度(59 ms)难以与单阶段 YOLO 模型相比。在需要高实时性的场景下,其并非最优选择。

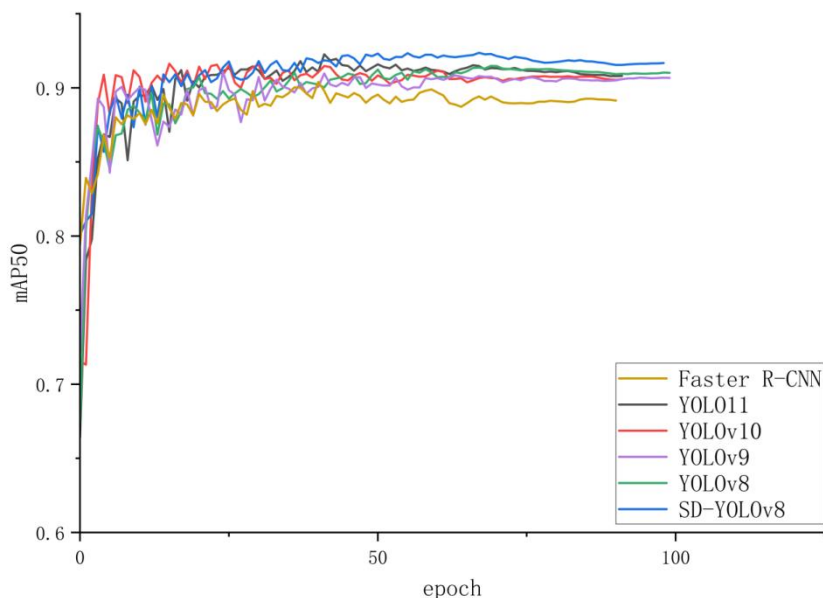


图 13 检测模型平均精度对比

Figure 13 Comparison of average accuracy of test models



图 14 草莓检测可视化结果

Figure 14 Visualization results of strawberry test results

表 3 数据集的目标检测模型训练与验证性能定量评估

Table 3 Quantitative evaluation of target test model training and validation performance on the dataset

检测模型	Precision	Recall	mAP50	mAP50-95	检测速度/ms
Faster R-CNN	0.858	0.837	0.904	0.772	59
YOLOv11	0.864	0.862	0.915	0.78	7
YOLOv10	0.862	0.848	0.915	0.781	5.2
YOLOv9	0.849	0.873	0.907	0.777	4.9
YOLOv8	0.882	0.844	0.922	0.785	4.7
SD-YOLOv8	0.904	0.875	0.932	0.82	4

对于成熟草莓这一颜色显著、目标特征相对易于识别的类别，如表 4 和图 15 所示，各模型均取得了较高的 Precision(≥ 0.94)和 Recall(≥ 0.918)。整体排名上,SD-YOLOv8 最优(P=0.973,R=0.933),随后是 YOLOv8 (0.966,0.924)与 YOLOv9 (0.964,0.929)。其他模型虽略低,但仍保持在 0.94 以上的高水准。所有模型在 mAP50 上均超过 0.96,足以反映成熟草莓在本数据集中的识别难度较小;在更严格的 mAP50-95 上也均高于 0.85,差异主要体现在精细定位与边界重叠严格度方面。SD-YOLOv8 再次以 0.979 (mAP50)和 0.887 (mAP50-95)领跑, YOLOv8 (mAP50-95=0.868)与 YOLOv9 (0.864)紧随其后,显示出对红色草莓在多种尺度、不同视角下的强泛化能力。

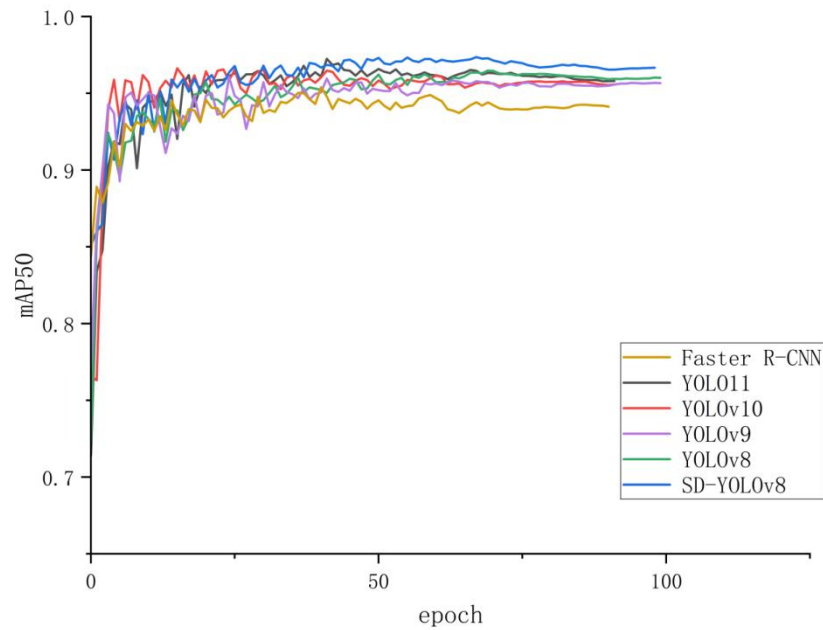


图 15 成熟草莓检测结果对比

Figure 15 Comparison of test results for ripe strawberries

表 4 成熟草莓检测结果对比

Table 4 Comparison of test results for ripe strawberries

检测模型	Precision	Recall	mAP50	mAP50-95
Faster R-CNN	0.941	0.918	0.966	0.859

检测模型	Precision	Recall	mAP50	mAP50-95
YOLOv11	0.95	0.924	0.972	0.857
YOLOv10	0.956	0.931	0.973	0.862
YOLOv9	0.964	0.929	0.971	0.864
YOLOv8	0.966	0.924	0.972	0.868
SD-YOLOv8	0.973	0.933	0.979	0.887

本研究采用模型验证的方式进行了定量评估，通过实验结果可以看出不同模型使用该数据集训练之后，检测精度都很高，说明数据集质量可靠。

4 数据价值

与国内外现有的草莓图像目标检测数据集相比，本数据集在数据来源、数据整理加工过程和质量控制等方面具有以下创新性。数据来源多样性：本数据集覆盖了不同自然条件（如晴天、阴天、顺光、逆光等）下的草莓图像，确保了数据的多样性和代表性；本数据集涵盖了草莓的幼小期、成熟期和采摘期等不同生长阶段，提供了丰富的样本数据。数据整理加工过程：除了传统的 RGB 图像，本数据集还包括深度图像和红外图像，增强了目标检测的准确性和鲁棒性；采用了先进的图像处理技术，如使用 Pyk4a 工具对深度图像和红外图像进行对齐，确保了多模态图像数据的一致性和高质量。质量控制严格：结合人工标注和半自动标注，确保了标注数据的准确性和一致性；对无效和污染数据进行了严格筛选和剔除，保证了数据集的高质量。

本数据集在多个潜在应用领域具有重要价值。农业智能化管理：可用于草莓果园的智能化管理系统，提升果园管理效率和生产效益。智能机器人采摘：为开发智能机器人采摘系统提供训练数据，提高采摘机器人在复杂环境下的识别和操作能力。深度学习研究：作为深度学习目标检测模型的标准数据集，为研究人员提供高质量的训练和测试数据。多模态数据融合研究：促进多模态数据融合技术的发展，提升目标检测的精度和鲁棒性。

通过与现有草莓图像目标检测数据集的比较，可以进一步体现本数据集的优势：现有数据集大多仅包含单一模态的 RGB 图像，而本数据集则涵盖了 RGB、深度和红外三种模态图像，提高了数据的丰富性和检测精度。现有数据集的标注通常较为粗略，而本数据集通过严格的人工和半自动标注校验，确保了标注数据的高精度。现有数据集通常针对单一应用场景，本数据集则涵盖了农业智能化管理、智能机器人采摘、深度学习研究等多个应用领域。

综上所述，草莓多模态图像目标检测数据集在数据来源、数据整理加工过程、质量控制等方面具有显著的创新性和优势，为相关研究和应用提供了宝贵的数据资源。

5 数据使用方法和建议

使用本数据集时，读者可以使用 Python、OpenCV、PIL (Python Imaging Library)、YOLO (You Only Look Once)、TensorFlow/PyTorch 读取保存的数据集中的图像文件，以下对这些推荐的软件和工具进行简要说明：

Python：作为主要的编程语言，广泛应用于数据处理和深度学习模型的开发。

OpenCV: 开源计算机视觉库, 用于图像处理和计算机视觉任务。

PIL (Python Imaging Library): 用于图像加载、处理和增强。

YOLO (You Only Look Once): 实时目标检测系统, 适用于目标检测模型的训练和验证。

TensorFlow/PyTorch: 深度学习框架, 用于构建和训练神经网络模型。

通过以上方法和建议, 研究人员可以高效地利用本数据集进行草莓目标检测的研究和应用, 推动相关领域的发展。

致 谢

感谢安徽省合肥市长丰县数字草莓种植基地提供拍摄基地。

数据作者分工职责

李科(1984—), 男, 安徽省合肥市人, 博士, 讲师, 研究方向是信号处理。主要承担工作: 提出数据采集与处理方法、图像拍摄。

徐启涛(1999—), 男, 安徽省阜阳市人, 硕士, 研究方向是农业信息化技术, 主要承担工作: 图像拍摄、数据标注、数据集结构设计。

李玉杰(2000—), 男, 安徽省淮南市人, 硕士, 研究方向是农业信息化, 主要承担工作: 图像数据的标注。

刘昊天(2002—), 男, 安徽省宿州市人, 硕士, 研究方向是农业信息化技术, 主要承担工作: 图像拍摄与校对。

参考文献

- [1] 赵瑞雪, 杨晨雪, 郑建华, 等. 农业智能知识服务研究现状及展望[J]. 智慧农业(中英文), 2022, 4(4): 105–125. DOI:10.12133/j.smartag.SA202207009. [ZHAO R X, YANG C X, ZHENG J H, et al. Agricultural intelligent knowledge service: overview and future perspectives[J]. Smart Agriculture, 2022, 4(4): 105–125. DOI:10.12133/j.smartag.SA202207009.]
- [2] 张翔鹤, 刘婷婷, 王晓丽, 等. 辽宁富士苹果品质及病害近红外光谱数据集[J]. 农业大数据学报, 2022, 4(2): 30–35. DOI:10.19788/j.issn.2096-6369.220205. [ZHANG X H, LIU T T, WANG X L, et al. Near-infrared spectral data set of apple quality and disease of fuji apple in Liaoning Province[J]. Journal of Agricultural Big Data, 2022, 4(2): 30–35. DOI:10.19788/j.issn.2096-6369.220205.]
- [3] 王丰仪, 饶元, 罗庆, 等. 毛桃多模态图像目标检测数据集[J/OL]. 中国科学数据, 2022, 7(4). (2022-12-02). DOI: 10.11922/11-6035.csd.2022.0055.zh. [WANG F Y, RAO Y, LUO Q, et al. A dataset of multi-modal peach images for object detection[J/OL]. China Scientific Data, 2022, 7(4). (2022-12-02). DOI: 10.11922/11-6035.csd.2022.0055.zh.]
- [4] 张雨, 饶元, 陈文骏, 等. 不同成熟度番茄果实多模态图像数据集[J/OL]. 中国科学数据, 2023. (2023-09-29). DOI: 10.11922/11-6035.csd.2023.0154.zh. [ZHANG Y, RAO Y, CHEN W J, et al. A multi-modal image dataset of tomatoes at different ripening stages[J/OL]. China Scientific Data, 2023. (2023-09-29). DOI: 10.11922/11-6035.csd.2023.0154.zh.]

- [5] 陈文骏, 饶元, 王丰仪, 等. 葡萄多模态目标检测和语义分割数据集[J/OL]. 中国科学数据, 2023. (2023-10-07). DOI: 10.11922/11-6035.csd.2023.0132.zh.[CHEN W J, RAO Y, WANG F Y, et al. A multi-modal dataset for grape object detection and semantic segmentation[J/OL]. China Scientific Data, 2023. (2023-10-07). DOI: 10.11922/11-6035.csd.2023.0132.zh.]
- [6] 罗庆, 饶元, 金秀, 等. 基于改进 YOLOv5s 和多模态图像的树上毛桃检测[J]. 智慧农业(中英文), 2022, 4(4): 84–104. DOI:10.12133/j.smartag.SA202210004. [LUO Q, RAO Y, JIN X, et al. Multi-class on-tree peach detection using improved YOLOv5s and multi-modal images[J]. Smart Agriculture, 2022, 4(4): 84–104. DOI:10.12133/j.smartag.SA202210004.]
- [7] 周宏平, 金寿祥, 周磊, 等. 基于多模态图像的自然环境下油茶果识别[J]. 农业工程学报, 2023, 39(10): 175–182. DOI:10.11975/j.issn.1002-6819.202303054. [ZHOU H P, JIN S X, ZHOU L, et al. Recognition of *Camellia oleifera* fruits in natural environment using multi-modal images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(10): 175–182. DOI:10.11975/j.issn.1002-6819.202303054.]
- [8] 杜亚茹, 黄媛, 杜鹏飞, 等. 基于多模态融合技术的番茄灰霉病智能协同诊断模型研究[J]. 中国农机化学报, 2023, 44(11): 115–122. DOI:10.13733/j.jcam.issn.2095-5553.2023.11.018. [DU Y R, HUANG Y, DU P F, et al. Study on intelligent collaborative diagnosis model of tomato *Botrytis* based on multi-mode fusion technology[J]. Journal of Chinese Agricultural Mechanization, 2023, 44(11): 115–122. DOI:10.13733/j.jcam.issn.2095-5553.2023.11.018.]
- [9] 高芳征, 汤文俊, 陈光明, 等. 基于改进 YOLOv3 的复杂环境下西红柿成熟果实快速识别[J]. 中国农机化学报, 2023, 44(8): 174–183. DOI:10.13733/j.jcam.issn.2095-5553.2023.08.024. [GAO F Z, TANG W J, CHEN G M, et al. Fast recognition of ripe tomato fruits in complex environment based on improved YOLOv3[J]. Journal of Chinese Agricultural Mechanization, 2023, 44(8): 174–183. DOI:10.13733/j.jcam.issn.2095-5553.2023.08.024.]
- [10] 孙俊, 钱磊, 朱伟栋, 等. 基于改进 RetinaNet 的果园复杂环境下苹果检测[J]. 农业工程学报, 2022, 38(15): 314–322. DOI:10.11975/j.issn.1002-6819.2022.15.034. [SUN J, QIAN L, ZHU W D, et al. Apple detection in complex orchard environment based on improved RetinaNet[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(15): 314–322. DOI:10.11975/j.issn.1002-6819.2022.15.034.]
- [11] 张羽丰, 杨景, 邓寒冰, 等. 基于 RGB 和深度双模态的温室番茄图像语义分割模型[J]. 农业工程学报, 2024, 40(2): 295–306. DOI:10.11975/j.issn.1002-6819.202309169. [ZHANG Y F, YANG J, DENG H B, et al. Semantic segmentation model for greenhouse tomato images using RGB and depth bimodal[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(2): 295–306. DOI:10.11975/j.issn.1002-6819.202309169.]
- [12] 张继成, 李德顺. 基于深度残差学习的成熟草莓识别方法[J]. 中国农机化学报, 2022, 43(2): 136–142. DOI:10.13733/j.jcam.issn.2095-5553.2022.02.019. [ZHANG J C, LI D S. Ripe strawberry recognition method based on deep residual learning[J]. Journal of Chinese Agricultural Mechanization, 2022, 43(2): 136–142. DOI:10.13733/j.jcam.issn.2095-5553.2022.02.019.]
- [13] 邱畅, 田光兆, 赵嘉威, 等. 基于改进 YOLOv5 的草莓病害识别[J]. 中国农机化学报, 2024, 45(3): 198–204. DOI:10.13733/j.jcam.issn.2095-5553.2024.03.027. [QIU C, TIAN G Z, ZHAO J W, et

- al. Strawberry disease identification based on improved YOLOv5[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(3): 198–204. DOI:10.13733/j.jcam.issn.2095-5553.2024.03.027.]
- [14] 杨珍, 龚惟新, 李凯, 等. 高架草莓的果实识别与果梗分割[J]. 农业工程学报, 2023, 39(17): 172–181. DOI:10.11975/j.issn.1002-6819.202305134. [YANG Z, GONG W X, LI K, et al. Fruit recognition and stem segmentation of the elevated planting of strawberries[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(17): 172–181. DOI:10.11975/j.issn.1002-6819.202305134.]
- [15] 高云茜, 邓三鹏. 基于 YOLOv5-en 算法的草莓采摘机器人目标检测技术[J]. 实验技术与管理, 2023, 40(10): 178–183, 216. DOI:10.16791/j.cnki.sjg.2023.10.027. [GAO Y X, DENG S P. Target detection technology of strawberry picking robot based on YOLOv5-en[J]. Experimental Technology and Management, 2023, 40(10): 178–183, 216. DOI:10.16791/j.cnki.sjg.2023.10.027.]

论文引用格式

李科, 徐启涛, 李玉杰, 等. 草莓多模态图像目标检测数据集[J/OL]. 中国科学数据, 2025, 10(4). (2025-12-18). DOI: 10.11922/11-6035.csd.2024.0162.zh.

数据引用格式

李科, 徐启涛, 李玉杰, 等. 草莓多模态图像目标检测数据集[DS/OL]. V2. Science Data Bank, 2025. (2025-04-01). DOI:10.57760/sciencedb.13135.

A dataset of strawberry multimodal image target detection

LI Ke^{1*}, XU Qitao¹, LI Yujie¹, LIU Haotian¹

1. Anhui Agricultural University, College of Information and Artificial Intelligence, Hefei 230036, P.R. China

*Email: like@ahau.edu.cn

Abstract: Accurate detection of strawberries during their growth period has always been a critical technical challenge for precision orchard management. To address the lack of publicly available multimodal image datasets in the field of strawberry object detection, this dataset was constructed through field data collection and standardized data processing, and includes multimodal image data of strawberries covering different growth throughout the fruit development period and a wide range of natural environmental conditions. The dataset contains visible light, depth, and infrared images, with a total data volume of 31.36 GB. To ensure the high quality and reliability of the strawberry multimodal image target detection dataset, a rigorous data quality control method was used, including standardized acquisition protocols, abnormal frame filtering, annotation consistency checks, multimodal alignment and calibration, as well as periodic verification of data integrity and readability. It provides valuable foundational data resources for multimodal image data fusion and target detection research and can serve as a benchmark dataset for training and testing deep learning models, demonstrating significant practical value in advancing strawberry target detection research.

Keywords: strawberry detection; multimodal dataset; image processing; deep learning; agricultural automation

Dataset Profile

Title	A dataset of strawberry multimodal image target detection
Data corresponding author	LI Ke (like@ahau.edu.cn)
Data authors	LI Ke, XU Qitao, LI Yujie, LIU Haotian
Time range	November to the end of April each year
Geographical scope	Changfeng County, Hefei City, Anhui Province
Data volume	31.36 GB
Data format	*.png, *.txt
Data service system	http://doi.org/10.57760/sciencedb.13135
Sources of funding	The University Synergy Innovation Program of Anhui Province (No. GXXT-2022-041); Anhui Provincial Quality Engineering Project of Higher Education Institutions (2022jyxm464); Anhui Agricultural University Introduction and Stabilization of Talents Research Funding (No. yj2020-74).
Dataset composition	The dataset consists of 4 subsets: (1) the rgb folder, which stores visible light images of strawberry, saved in PNG format, comprising a total of 12,840 images with a data volume of 29.3 GB; (2) the ir folder, which stores infrared images corresponding to the visible light images, saved in PNG format, with a total of 12,840 images with a data volume of 850 MB; (3) the depth folder, which stores depth images corresponding to the visible light images, saved in PNG format, comprising 12,840 images with a data volume of 1.29 GB; (4) the Annotations-YOLO folder, which stores YOLO-format annotation files corresponding to the visible light images and containing positional information of strawberries in each image, with a data volume of 3.41 MB.