专题 农业信息高质量图谱数据

葡萄多模态目标检测和语义分割数据集

陈文骏[1,2,3]，饶元[1,2,3*]，王丰仪[1,2,3]，张雨[1,2,3]，杨雨梦[1,2,3]，罗庆[1,2,3]，张通[1,2,3]，万天与[1,2,3]，刘心宇[1,2,3]，张梦宇[1,2,3]，张蕊[1,2,3]

1. 安徽农业大学，信息与计算机学院，合肥 230036
2. 农业农村部农业传感器重点实验室，合肥 230036
3. 智慧农业技术与装备安徽省重点实验室，合肥 230036

摘要： 葡萄果实的采摘点定位准确率依赖于目标检测和语义分割网络的性能。然而，在实际应用场景中，基于可见光图像的葡萄果实目标识别准确率和分割精度易受光照变化、复杂环境影响，往往表现不佳，且葡萄果实成串生长，现有苹果、梨子等多模态数据集难以满足串形葡萄果实的识别需求。构建基于可见光、深度、近红外的葡萄多模态目标检测和语义分割数据集，对于探索更高识别率和更强泛化能力的葡萄果实目标检测和语义分割模型至关重要。本数据集约 39.08 GB，共收集了在不同光照和遮挡条件下青色、紫色两类 6 个品种的葡萄高质量多模态视频流数据，并从中提取 3954 张图像样本进行语义分割和目标检测标注。在使用旋转、缩放、错切、平移，以及高斯模糊等图像增强手段下，可满足主流深度学习模型训练需要。本数据集可为多模态视觉数据融合、葡萄果实语义分割和目标检测等领域提供宝贵的基础数据资源，对促进农机装备智能化领域研究具有重要的实际应用价值。

关键词： 自然环境；葡萄；多模态；语义分割；目标检测

数据库（集）基本信息简介

数据库（集）名称	葡萄多模态目标检测和语义分割数据集
数据通信作者	饶元（raoyuan@ahau.edu.cn）
数据作者	陈文骏、饶元、王丰仪、张雨、杨雨梦、罗庆、张通、万天与、刘心宇、张梦宇、张蕊
数据时间范围	2021 年 8 月、2023 年 7 月
地理区域	安徽合肥
数据量	39.08 GB
数据格式	*.png, *.xml, *.json, *.mkv
数据服务系统网址	http://doi.org/10.57760/sciencedb.j00001.00883
基金项目	国家自然科学基金(32371993)；安徽省重点研究和开发计划面上攻关项目(202104a06020012, 202204c06020022)；安徽省高校自然科学重大项目(2022AH040125)。

www.sciengine.com/CSD
ISSN 2096-2223
CN 11-6035/N

文献 CSTR：
32001.14.11-6035.csd.2023.0132.zh

文献 DOI：
10.11922/11-6035.csd.2023.0132.zh

数据 DOI：
10.57760/sciencedb.j00001.00883

文献分类：生物学

收稿日期：2023-08-06
开放同评：2023-08-28
录用日期：2023-09-07
发表日期：2023-10-07

* 论文通信作者
饶元：raoyuan@ahau.edu.cn

数据集组成	本数据集包含原始多模态视频流文件，以 mkv 形式独立存放于 VID 文件夹中。本数据集所标注的图像数据采用特定帧间隔抽取于该视频流文件，按光照条件分别放置于正常光照（NORMAL）、微弱光照（POOR）两个文件夹中，各文件夹又包含青色（GREEN）、紫色（PURPLE）两子文件夹。每个多模态样本由可见光图像、深度图像、近红外图像、json 格式语义分割标注文件、xml 格式目标检测标注文件组成，并存储在对应的文件夹中： （1）RGB 文件夹存放葡萄的可见光图像，以 PNG 格式保存； （2）Depth 文件夹是对应可见光图像的深度图像，以 PNG 格式保存； （3）NIR 文件夹是对应可见光的近红外图像，以 PNG 格式保存； （4）Seg 文件夹是对应可见光图像的 json 格式标注文件，包含葡萄果实的像素标注信息； （5）VOC 文件夹是对应可见光图像的 VOC 格式目标检测标注文件，包含图像大小和图像中葡萄的位置信息。

引　言

近年来，我国葡萄产业蓬勃发展，葡萄种植面积增长率位居中国前五大果品之首，是世界葡萄生产大国[1-2]。然而，我国大多葡萄园基础设施装备简陋，田间作业以人力劳作为主，劳动强度大、生产效率低[3]。利用智能机器人采摘葡萄果实，可极大减轻人力工作量，实现降本增效。

葡萄采摘点检测是实现葡萄果实自动采摘的前提，葡萄采摘点的定位主要通过检测葡萄果实的外接轮廓生成感兴趣区域，再以图像处理或几何推理方法计算采摘点位置[4-5]。传统研究大多通过可见光图像提取葡萄果实颜色、纹理等轮廓特征，实现采摘点定位[5-8]。受树叶遮挡、果实重叠、背景复杂和光照多变等自然环境因素及葡萄颜色、形态差异等自身因素影响，同一模型对果实的识别准确率差异明显[7, 9-10]，在下游任务中易致使葡萄果实破损，造成经济损失[11]。引入深度、近红外图像等多模态图像数据可为葡萄果实的识别提供更丰富的特征信息，对提高果实识别模型的环境适应能力和泛化性有重要作用[12]。已有研究表明，引入多模态数据后，模型的检测与分割能力均有不同程度提升[13-15]。此外，阴天、夜间等采摘场景下，RGB 图像信息量受限导致模型检测效果表现不佳，而深度、近红外等不同模态视觉数据的引入有利于获取场景与目标的空间关系或特征差异，有效提高目标检测效果[16]。

现有果实数据集的研究对象大多以苹果、梨子等果实为主，果实体积较大、果实位置相对分散，且果实与枝叶颜色差别大。葡萄果实呈串形，果实体积小、果实位置相对密集，部分品种的葡萄果实与枝叶颜色相近，前人的研究内容及数据集难以同化至葡萄及类似形态果实的研究中，领域内尚缺乏适用于葡萄果实多模态目标检测与语义分割的训练样本。因此，构建基于可见光、深度、近红外的葡萄多模态目标检测和语义分割数据集，对于探索更高鲁棒性和更强泛化能力的葡萄果实目标检测和语义分割模型至关重要。

在基于目标检测和语义分割的葡萄对象检测中，果实的颜色及其与背景区分显著度是影响不同品种葡萄检测效果的主要因素。因此在相关项目支持下，我们选取青色和紫色两个颜色的葡萄，采集了白玫瑰、阳光玫瑰、醉金香、巨峰、甬优、夏黑等 6 个市面销售的主要葡萄品种在果园真实环境下的一组可见光、深度、近红外多模态数据，其内容涵盖顺光、逆光、运动模糊等主要采摘场景，采集于正常光照、微弱光照的两类典型光照条件，可作为实际生产环境的代表性数据。该数据可用于基于可见光、深度、近红外多模态融合的葡萄果实语义分割、目标检测及下游任务，探索阴天、

夜间场景的葡萄检测方法，为成串生长果实机械化采摘的算法研究奠定数据基础，对促进农机装备智能化领域研究具有重要实际应用价值。

1 数据采集和处理方法

1.1 数据采集

本数据集葡萄多模态数据分别于 2021 年 8 月和 2023 年 7 月采集自安徽省合肥市大圩生态区。采集设备型号为 Azure Kinect 深度相机，图像分辨率大小均为 1280×720。Azure Kinect 深度相机的相关参数如表 1 所示。

表 1　Azure Kinect 深度相机基本性能参数

Table 1　Basic performance parameters of Azure Kinect DK

指标项	性能参数
每秒帧数	30FPS
RGB 相机分辨率	RES_720P
深度模式	NFOV_2X2BINNED
惯性测量单元	LSM6DSMUS/1.6kHZ
感兴趣视场角	90°×59°/75°×65°

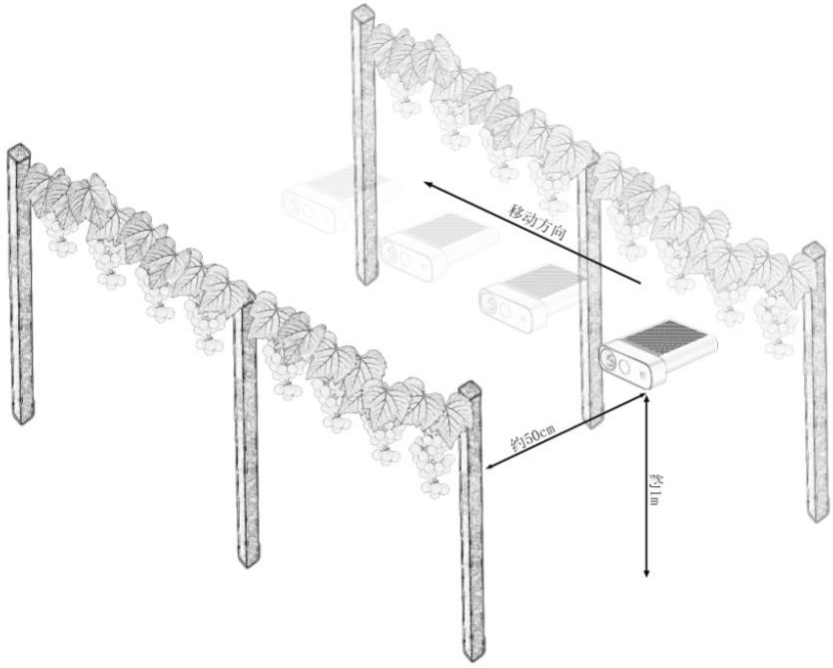


图 1　图像采集位置示意图

Figure 1　Diagram of image collection location

如图 1 所示，视频流数据采集时，相机视距约 50 cm，近地高度约 1 m，机位角度近似垂直于果实表面所在平面。采集视频流数据内容依光照条件分为顺光、逆光，依光照条件分为正常光照、微弱光照，依拍摄画面分为稳定、运动模糊，依葡萄颜色分为青色、紫色。为避免拍摄时相机移动缓慢导致相近时间拍摄大量相似数据，拍摄结束后按照 10 帧的间隔从原始视频流数据中均匀地抽取图像数据用于后续标注。

1.2 数据处理

因 Azure Kinect 深度相机可见光、深度、近红外镜头成像原理及拍摄位置不同，需先对齐并配准多模态图像数据。首先计算原始可见光图像数据中各二维像素点对应的三维空间位置，并将三维空间位置映射至原始深度数据与近红外数据，计算以可见光图像像素位置为基准的深度图像像素位置与近红外图像像素位置，生成对齐后的可见光、深度、近红外图像数据。接着采用编写的数字图像处理程序对所提取的图像数据依次进行筛选，删除相似、失真与无效图像。最后人工二次核查图像数据，保障数据集质量。数据处理流程如图 2 所示。

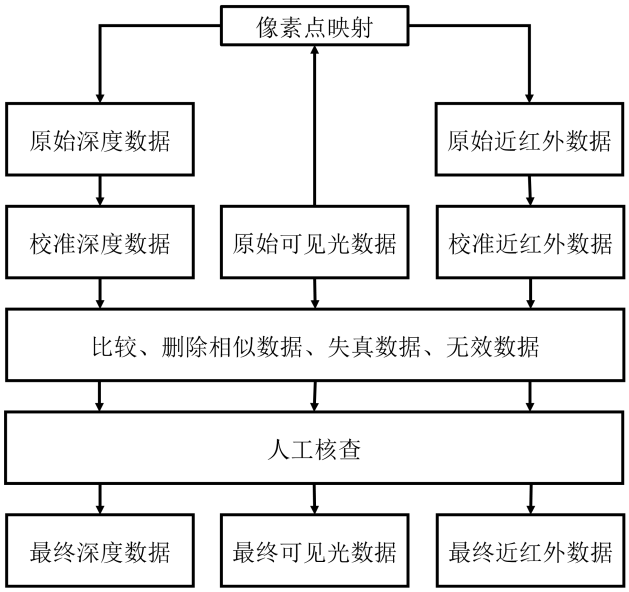


图 2 数据处理流程

Figure 2 Data processing flowchart

2 数据样本描述

2.1 多模态数据示例

本数据集包含相互对齐的可见光、深度和近红外视觉数据。可见光视觉数据以 RGB 图像存储，深度和近红外视觉数据以 16 bit 灰度图像存储。随机选取数据样本如图 3 所示。因 16 bit 深度、近红外图像特征肉眼辨识度低，示例采用 8 bit 下 Plasma、Gray 的色彩空间映射，对深度与近红外原始数据进行可视化处理。其中，深度图像的紫、黄色渐变代表相对距离由远到近。

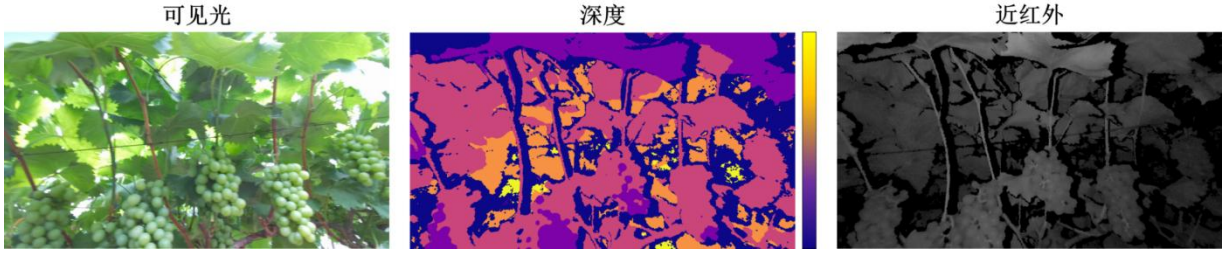


图 3 多模态样本示例

Figure 3 Example images of multimodal data

2.2 数据多样性示例

数据集由青色、紫色两个不同颜色的葡萄组成，选取白玫瑰、阳光玫瑰、醉金香 3 类品种作为青色葡萄的代表性品种，选取巨峰、甬优、夏黑 3 类品种作为紫色葡萄的代表性品种。随机选取两类颜色品种葡萄样本如图 4 所示，相同颜色品种的葡萄形态相似，大小、纹理等特征具有细微差异。



图 4 两类不同颜色葡萄样本示例

Figure 4 Example images of two different colors of grapes

数据集涵盖顺光、逆光、运动模糊等主要场景，模拟机械臂采摘角度，采集正视、俯视、仰视 3 类视角，因微弱光照下图像难以直观展示，图 5 列出了随机选取的部分正常光照条件下代表性样本。



图 5 代表性采摘场景样本示例

Figure 5　Example images in representative picking scenarios

在微弱光照等复杂环境下，RGB 图像噪声显著，难以提取特征信息，影响模型检测效果，多模态数据在此类环境中具有优势。数据集的微弱光照数据涵盖了自然光、人工光的两类典型场景，可用于研究此类环境下的检测和分割问题，随机选取部分样本如图 6 所示。

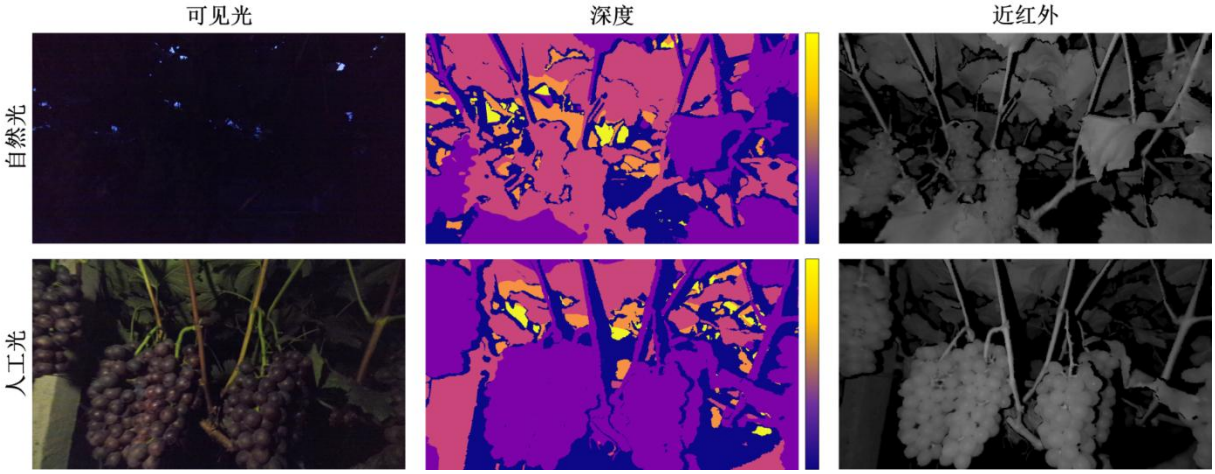


图 6 两类典型微弱光照场景样本示例

Figure 6　Example images with two representative poor illuminants

2.3　标注样本示例

为提高葡萄果实标注样本的可靠性和多样性，对不同遮挡情况下的葡萄果实分别采用不同的标注策略。对于葡萄果实无遮挡的样本，单串葡萄果实作为一个独立的分割对象。部分样本的葡萄果实分割示例以及据此生成的葡萄果实目标检测框可视化结果示例如图 7 所示。

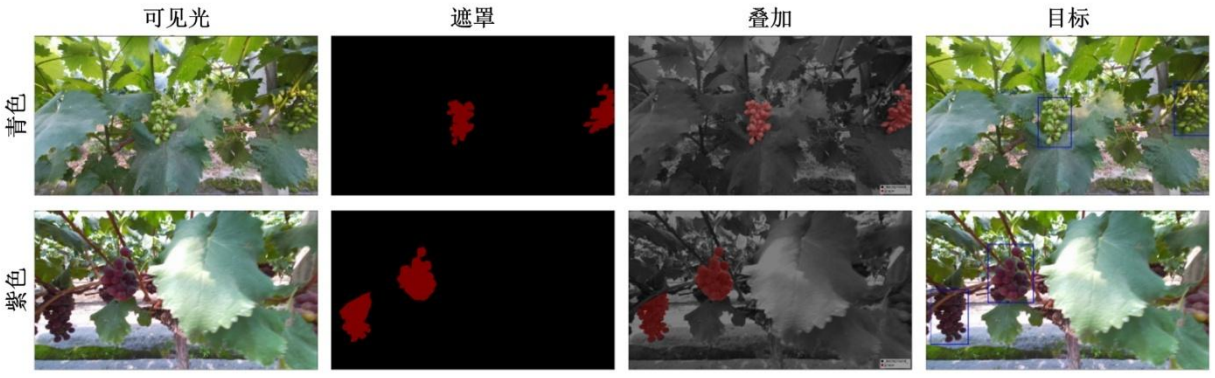


图 7 无遮挡葡萄串标注示例

Figure 7　Labeled examples of unobstructed grape bunches

对于树叶遮挡、果实间存在重叠现象的葡萄串，视野可见的葡萄果实每一部分均作为一个独立的分割对象。部分样本的葡萄果实分割示例以及据此生成的葡萄果实目标检测框可视化结果示例如图 8 所示。图 9 给出了图 8 的最左列可见光图像中蓝色遮罩区域标注细节的放大结果。

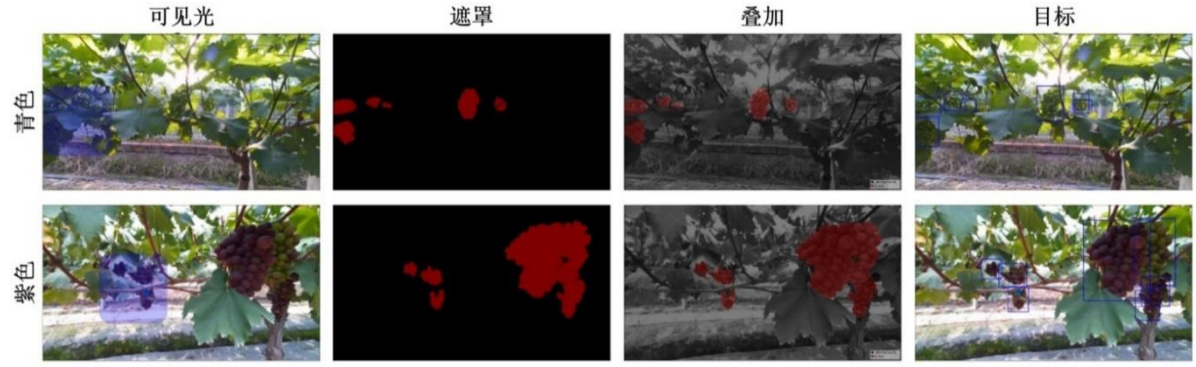


图 8 被遮挡葡萄串标注示例

Figure 8 Labeled examples of occluded grape bunches

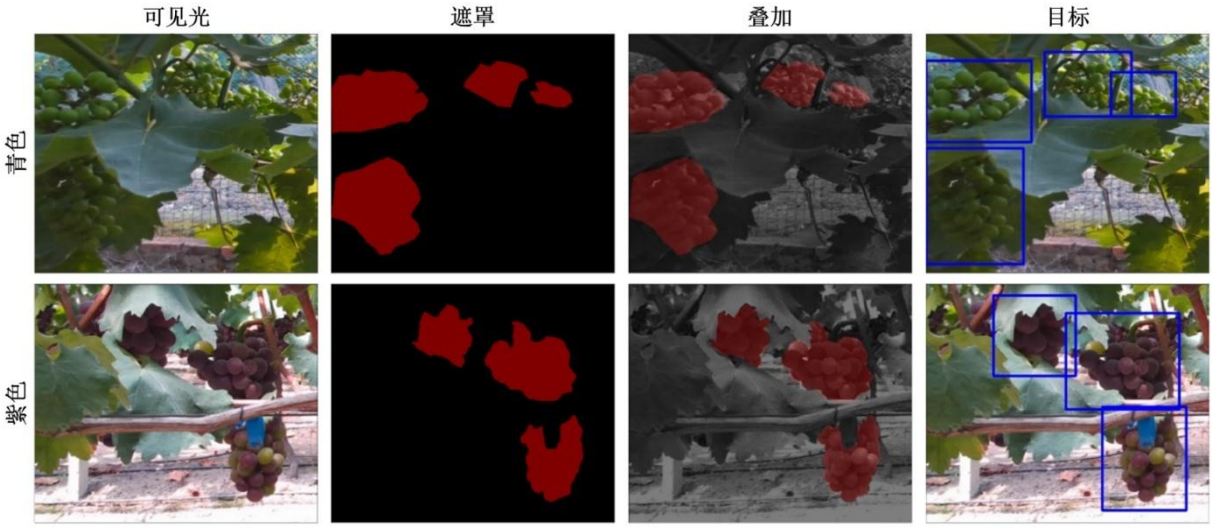


图 9 被遮挡葡萄串局部区域标注细节

Figure 9 Annotation detail of partial area in occluded grape bunches

对于微弱光源、人工光源的样本，采用人工在近红外图像上进行标注的方式，对于无遮挡和被遮挡葡萄串，标注方式同上。图 10 给出了在近红外图像上进行标注的过程预览图，以及与可见光图像的对比。



图 10 近红外图像中葡萄串标注过程

Figure 10 Labeling process of grape bunches in near-infrared images

不同品种的葡萄果实颜色和形态不同，相对于背景等周边环境的显著度明显有别，故葡萄的检测和分割难度往往也存在差异。为提高葡萄多模态数据集的多样性，图像数据含青、紫两种颜色的葡萄。每张图像中含有多个葡萄串目标，青、紫色葡萄在正常、微弱光照条件下葡萄样本及标记量

如表 2 所示。数据集中图像样本可通过旋转、放缩、错切、平移[17]、高斯模糊[18]、提高亮度、提高对比度、提高饱和度等数据增强手段进一步增加数据量。

表 2 葡萄图像样本与标记量

Table 2 Labeling number for grape image samples and bunches

光照条件	颜色	样本数量（张）	标记数量（个）
正常光照	绿色	953	3645
	紫色	1000	4822
微弱光照	绿色	1015	4091
	紫色	986	4259

视频流中存在丰富的时间-空间信息，其前后帧数据在基于帧差法的深度学习模型中具有参考价值。对于无监督、自监督、半监督模型，可通过未标注数据扩充数据量，提高模型对采摘场景的理解能力。此外，还可用于葡萄长势监测、产量估计等实际应用中。因此，提供了部分具有参考价值的原始多模态视频流数据，以五通道 mkv 方式存储。视频分辨率为 1280×720，帧率为 30 帧每秒。

2.4 标注文件描述

为满足现有基于深度学习的葡萄串目标检测、分割模型训练的需要，对葡萄多模态数据集进行语义分割和检测框标注。数据集内的果实分割标注文件以 json 为后缀，使用主流数据标注工具 LabelMe[19]生成，含 LabelMe 版本、多边形遮罩顶点位置、图片名称、图片内容、图片宽高等信息，其具体内容及释义见表 3。数据集内葡萄果实目标检测标注文件以 xml 为后缀，格式参考已公布的 VOC2007、VOC2012 数据集[20]，依据多边形遮罩各顶点坐标最大、最小值生成，含图像名称、文件路径、图像宽高、目标检测框矩形位置等信息，其具体内容及释义见表 4。

表 3 json 文件有效信息及释义

Table 3 Valid information and corresponding interpretation for the json file

键名	解释
shapes	记录各个分割信息，以列表形式存储。其子项中，“label”项记录类别名称，在本数据集中青、紫葡萄分别为“green_grape”、“purple_grape”；“points”项记录闭合多边形每个顶点的 x、y 坐标信息。
imagePath	记录对应的图片名称
imageData	记录 base64 加密后的以二进制方式读取的 png 格式图片内容
imageHeight	记录图片高度信息，均为 720（像素）
imageWidth	记录图片宽度信息，均为 1280（像素）

表 4 xml 文件有效信息及释义

Table 4 Valid information and corresponding interpretation in the xml file

键名	解释
filename	记录图像名称
path	记录文件路径

键名	解释
size	记录图像尺寸。其子项中，“width”“height”“depth”项分别记录图像宽度（像素）、高度（像素）、通道数，本数据集中各项数值分别为“1280”“720”“3”。
object	记录各目标检测框信息。其子项中，“name”项记录类别信息，在本数据集中青、紫葡萄分别为“green_grape”、“purple_grape”；“bndbox”项记录目标检测框位置信息，由“xmin”“ymin”“xmax”“ymax”4 组数据组成，分别记录目标检测框左上角、右下角的横、纵坐标。

3 数据质量控制和评估

3.1 数据质量控制

为保证青色、紫色葡萄多模态目标检测和语义分割数据集的数据质量，数据处理环节严格按照上述流程进行，未经后期加工处理。葡萄果实分割信息均为人工依葡萄果实轮廓进行标注。通过分割多边形遮罩获取外接矩形，并适当扩展矩形范围，再进行人工校正，得到 VOC 格式的葡萄果实检测框标注。所有图像数据经过多次核查校验，以确保数据质量。

3.2 数据评估

主流的采摘点检测、产量估计和机械采摘等下游任务实现方法多样，但均以目标检测或语义分割作为先要条件[21-24]。因此，采用目标检测和语义分割任务的有效性评估，能够作为各类下游任务的典型指标。YOLOv7[25]是当下热门的目标检测模型，其检测速度和精度表现突出。使用标准规格 YOLOv7 模型对本数据集青色、紫色葡萄进行目标检测测试，测试所得精确率（Precision）、召回率（Recall）、$mAP_{0.5}$、$mAP_{0.5:0.95}$ 数据及结果可视化如表 5、图 11 所示。

表 5 目标检测数据评估

Table 5 Object detection data evaluation

类别	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5:0.95}$
青色	0.880	0.852	0.887	0.621
紫色	0.873	0.866	0.885	

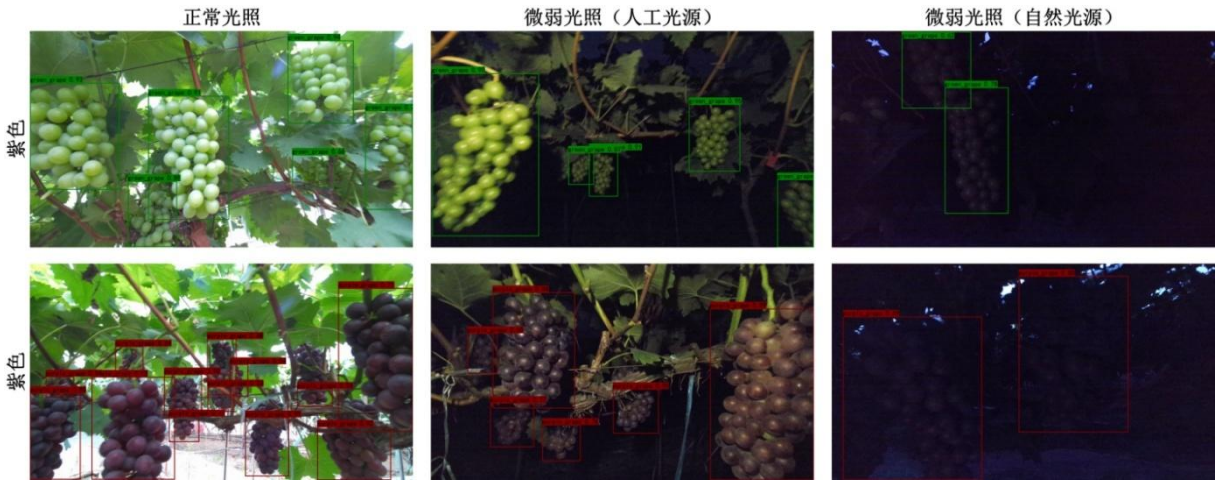


图 11 目标检测结果可视化

Figure 11 Visualization of object detection results

DeepLabV3+[26]是经典的语义分割模型，在工业和研究中被广泛应用。使用 DeepLabV3+分别对本数据集青色、紫色葡萄进行语义分割测试，测试所得精确率（Precision）、召回率（Recall）、mIoU 数据及可视化结果如表 6、图 12 所示。

表 6 语义分割数据评估

Table 6 Data evaluation on semantic segmentation

类别	Precision	Recall	IoU	mIoU
青色	0.915	0.921	0.848	0.854
紫色	0.916	0.933	0.859	

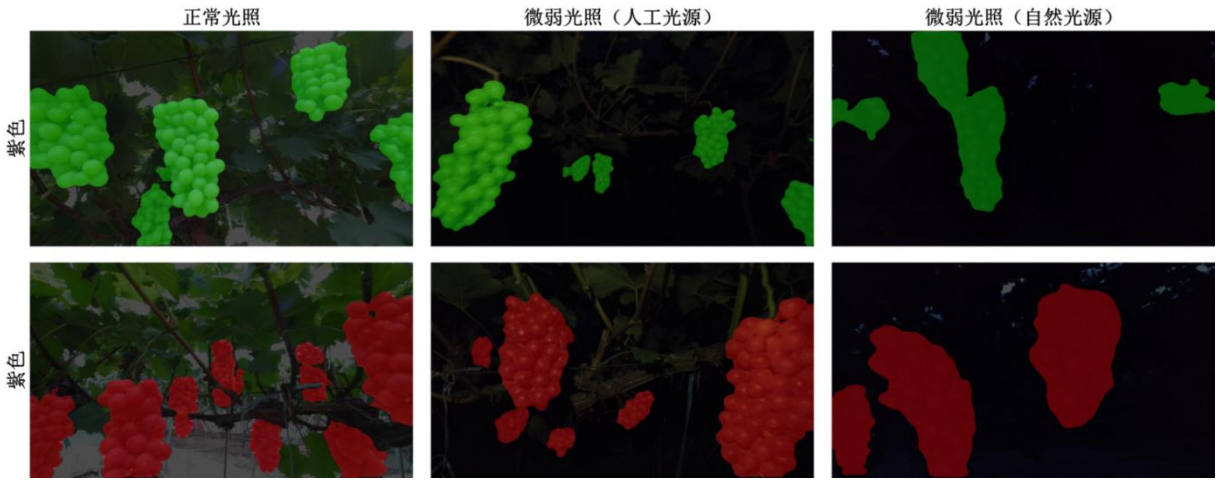


图 12 语义分割结果可视化

Figure 12 Visualization of semantic segmentation results

表 5、表 6 中，精确率（Precision）、召回率（Recall）、mAP、mIoU 的计算公式如下式（1、2、4、6）所示。此外，$mAP_{0.5}$ 代表交并比（IoU）为 0.5 下的 mAP 值，$mAP_{0.5:0.95}$ 代表交并比（IoU）在 0.5 至 0.95 之间，增幅为 0.05 的各 mAP 平均值。式中，TP 表示真实正样本数，FP 表示虚假正样本数，FN 表示虚假负样本数，C 为类别数，在 $mAP_{0.5}$ 项中取值为 1，在 $mAP_{0.5:0.95}$ 和 mIoU 项中取值为 2。

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

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

$$AP = \int_0^1 Presicion\ d(Recall) \#(3)$$

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C} \#(4)$$

$$IoU = \frac{TP}{TP + FP + FN} \#(5)$$

$$mIoU = \frac{\sum_{i=1}^{C} IoU_i}{C} \#(6)$$

如表 5、表 6 所示，青色、紫色葡萄目标检测的 $mAP_{0.5}$ 约为 0.88，语义分割的 IoU 约为 0.85，训练数据表现理想，说明本数据集数据量和标注质量可满足模型训练与验证需求。在目标检测评估

测试中，在样本数量相近的训练数据下，紫色葡萄检测精度低于青色葡萄，模型从紫色葡萄样本中提取的特征不够突出。在语义分割评估测试中，青色葡萄分割精度低于紫色葡萄，模型从果实边缘提取的特征不够突出。引入深度、近红外多模态数据后，模型分类、分割性能均有一定提升，可满足多模态目标检测与语义分割算法的研究需求。

表 7 主流模型基线数据

Table 7 Baseline data for mainstream models

模型	类别	Precision	Recall	$mAP_{0.5}$ / IoU	$mAP_{0.5:0.95}$ / mIoU
YOLOv5-s	青色	0.914	0.823	0.917	0.681
	紫色	0.897	0.844	0.927	
YOLOX-s	青色	0.894	0.878	0.938	0.726
	紫色	0.896	0.920	0.945	
YOLOv8-s	青色	0.910	0.856	0.931	0.750
	紫色	0.918	0.902	0.935	
PSPNet	青色	0.869	0.890	0.784	0.844
	紫色	0.889	0.882	0.794	

表 7 给出了主流的目标检测模型 YOLOv5-s、YOLOX-s 和 YOLOv8-s，以及语义分割模型 PSPNet 在本数据集上的训练基线数据。在检测精度方面，紫色葡萄的数据普遍高于青色葡萄；在召回率方面，青色葡萄的数据普遍高于紫色葡萄。此外，无锚框的 YOLOX 和 YOLOv8 模型的检测性能高于有锚框的 YOLOv5 和 YOLOv7，说明数据集包含丰富的多尺度目标。葡萄果实的分割精度普遍低于果实检测精度，准确高效的提取果实的边缘和轮廓特征，是当前葡萄果实检测和分割及其下游任务亟待解决的关键问题。

4 数据价值

本数据集采集了青色、紫色两个颜色的葡萄在顺光、逆光、运动模糊等真实自然条件下，以及正常光照、微弱光照等不同光照环境下的高质量多模态图像数据，旨在扩充当前多模态葡萄果实乃至串形水果果实在目标检测与语义分割领域的训练样本。本数据集可为实际环境下葡萄果实目标检测、语义分割、采摘点检测的相关研究提供宝贵的基础数据资源，对促进农业机械装备智能化领域研究具有重要的实际应用价值。

5 数据使用方法和建议

对于原始视频流数据，使用 ffmpeg 工具进行多模态图像提取，使用 Azure Kinect SDK 进行多模态图像对齐后即可用于训练集扩充以及目标检测、语义分割模型的测试验证。对于已标注数据，使用前，先使用批量替换工具修改 xml 文件内的文件路径为使用者电脑中的绝对路径。本数据集所包含的全部数据，可通过 Python 中的 PIL、OpenCV、json、ElementTree 等模块，或主流深度学习框架自带模块进行读取，再输入到深度学习模型中进行训练。

致　谢

感谢安徽省合肥市大圩生态区相关农场提供拍摄基地。

数据作者分工职责

陈文骏（2002—），男，安徽省合肥市人，本科在读，研究方向是人工智能。主要承担工作：数据采集、数据清洗、数据整理、数据标注、论文撰写。

饶元（1982—），男，安徽省合肥市人，博士，教授，研究方向是计算机应用技术。主要承担工作：提出数据采集与处理方法。

王丰仪（1998—），女，安徽省宿州市人，硕士，研究方向是机器视觉。主要承担工作：数据采集、论文修改。

张雨（2001—），女，安徽省马鞍山市人，本科在读，研究方向是机器视觉。主要承担工作：数据采集、数据标注。

杨雨梦（2001—），女，安徽省宿州市人，学士，研究方向是人工智能。主要承担工作：数据采集。

罗庆（1997—），男，安徽省六安市人，学士，研究方向是人工智能。主要承担工作：数据采集。

张通（1998—），男，安徽省宿州市人，学士，研究方向是人工智能。主要承担工作：数据采集。

万天与（1999—），男，江西省南昌市人，学士，研究方向是机器视觉。主要承担工作：数据采集。

刘心宇（2004—），女，安徽省亳州市人，本科在读，研究方向是机器视觉。主要承担工作：数据标注。

张梦宇（2004—），女，安徽省亳州市人，本科在读，研究方向是机器视觉。主要承担工作：数据标注。

张蕊（2004—），女，安徽省宿州市人，本科在读，研究方向是机器视觉。主要承担工作：数据标注。

参考文献

[1] 王静, 周广胜. 中国欧亚种酿酒葡萄种植分布的主要气候影响因子与气候适宜性[J]. 生态学报, 2021, 41(6): 2418–2427. DOI: 10.5846/stxb201903180507. [WANG J, ZHOU G S. The climatic suitability and climatic impact factors affecting the wine grapes (*Vitis vinifera* L.) planting distribution in China[J]. Acta Ecologica Sinica, 2021, 41(6): 2418–2427. DOI: 10.5846/stxb201903180507.]

[2] 吕鹏超, 戚伟波, 胡怡林, 等. 我国果品产业发展趋势分析[J]. 中国果菜, 2022, 42(7): 75–79, 74. [LV P C, QI W B, HU Y L, et al. The development trend analysis of China's fruit industry[J]. China Fruit & Vegetable, 2022, 42(7): 75–79, 74.]

[3] 田淑芬. 中国葡萄产业与科技发展[J]. 农学学报, 2018, 8(1): 143–147. [TIAN S F. The development of grape industry and technology in China[J]. Journal of Agriculture, 2018, 8(1): 143–147.]

[4] 罗陆锋, 邹湘军, 熊俊涛, 等. 自然环境下葡萄采摘机器人采摘点的自动定位[J]. 农业工程学报, 2015, 31(2): 14–21. DOI: 10.3969/j.issn.1002-6819.2015.02.003. [LUO L F, ZOU X J, XIONG J T, et al. Automatic positioning for picking point of grape picking robot in natural environment[J]. Transactions of

the Chinese Society of Agricultural Engineering, 2015, 31(2): 14–21. DOI: 10.3969/j.issn.1002-6819.2015.02.003.]

[5] 宁政通, 罗陆锋, 廖嘉欣, 等. 基于深度学习的葡萄果梗识别与最优采摘定位[J]. 农业工程学报, 2021, 37(9): 222–229. DOI: 10.11975/j.issn.1002-6819.2021.09.025. [NING Z T, LUO L F, LIAO J X, et al. Recognition and the optimal picking point location of grape stems based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(9): 222–229. DOI: 10.11975/j.issn.1002-6819.2021.09.025.]

[6] 雷旺雄, 卢军. 葡萄采摘机器人采摘点的视觉定位[J]. 江苏农业学报, 2020, 36(4): 1015–1021. [LEI W X, LU J. Visual positioning method for picking point of grape picking robot[J]. Jiangsu Journal of Agricultural Sciences, 2020, 36(4): 1015–1021.]

[7] 田锐, 郭艳玲. 基于机器视觉的葡萄自动识别技术[J]. 东北林业大学学报, 2008, 36(11): 95–97. DOI: 10.3969/j.issn.1000-5382.2008.11.035. [TIAN R, GUO Y L. Automatic identification techniques of grape based on computer vision[J]. Journal of Northeast Forestry University, 2008, 36(11): 95–97. DOI: 10.3969/j.issn.1000-5382.2008.11.035.]

[8] 李元强, 高何璇, 何玉英, 等. 基于双目视觉的葡萄识别定位及跟踪方法研究[J]. 林业机械与木工设备, 2021, 49(4): 52–59, 64. [LI Y Q, GAO H X, HE Y Y, et al. Positioning and tracking method based on binocular vision[J]. Forestry Machinery ＆ Woodworking Equipment, 2021, 49(4): 52–59, 64.]

[9] 李会宾, 史云. 果园采摘机器人研究综述[J]. 中国农业信息, 2019, 31(6): 1–9. [LI H B, SHI Y. Review on orchard harvesting robots[J]. China Agricultural Informatics, 2019, 31(6): 1–9.]

[10] 郑小南, 杨凡, 李富忠. 葡萄图像检测技术研究进展[J]. 现代农业科技, 2020(12): 78–79, 82. DOI: 10.3969/j.issn.1007-5739.2020.12.040. [ZHENG X N, YANG F, LI F Z. Research progress on grape image detection technology[J]. Modern Agricultural Science and Technology, 2020(12): 78–79, 82. DOI: 10.3969/j.issn.1007-5739.2020.12.040.]

[11] 王思玉, 牛成强, 李钊, 等. 典型果蔬采摘机器人研究现状与趋势[J]. 新疆农机化, 2021(6): 14–17, 21. [WANG S Y, NIU C Q, LI Z, et al. Current research status and development trends of typical fruit and vegetable harvest robot[J]. Xinjiang Agricultural Mechanization, 2021(6): 14–17, 21.]

[12] 陈品岚, 朱立学, 张世昂. 面向采摘机器人的水果精准识别算法研究进展[J]. 现代农业装备, 2022, 43(2): 8–13, 42. [CHEN P L, ZHU L X, ZHANG S A. Research progress of fruit accurate recognition algorithm for picking robot[J]. Modern Agricultural Equipment, 2022, 43(2): 8–13, 42.]

[13] 李贤琴, 娄全福. 采摘机器人识别技术研究: 基于深度图像和计算机网络协同处理[J]. 农机化研究, 2023, 45(3): 241–244. [LI X Q, LOU Q F. Research on recognition technology of picking robot based on the collaborative processing of depth image and computer network[J]. Journal of Agricultural Mechanization Research, 2023, 45(3): 241–244.]

[14] 孙宝霞, 郑镇辉, 胡文馨, 等. 基于 RGB-D 的龙眼实时检测与定位方法[J]. 林业工程学报, 2022, 7(3): 150–157. DOI: 10.13360/j.issn.2096-1359.202107028. [SUN B X, ZHENG Z H, HU W X, et al. Real-time detection and localization method of Longan based on RGB-D[J]. Journal of Forestry Engineering, 2022, 7(3): 150–157. DOI: 10.13360/j.issn.2096-1359.202107028.]

[15] LEHNERT C, SA I, MCCOOL C, et al. Sweet pepper pose detection and grasping for automated crop harvesting[C]//2016 IEEE International Conference on Robotics and Automation. Stockholm, Sweden.

IEEE: 2428–2434. DOI:10.1109/ICRA.2016.7487394.

[16] FU L S, GAO F F, WU J Z, et al. Application of consumer RGB-D cameras for fruit detection and localization in field: a critical review[J]. Computers and Electronics in Agriculture, 2020, 177: 105687. DOI: 10.1016/j.compag.2020.105687.

[17] 梁琳, 何卫平, 雷蕾, 等. 光照不均图像增强方法综述[J]. 计算机应用研究, 2010, 27(5): 1625–1628. DOI: 10.3969/j.issn.1001-3695.2010.05.006. [LIANG L, HE W P, LEI L, et al. Survey on enhancement methods for non-uniform illumination image[J]. Application Research of Computers, 2010, 27(5): 1625–1628. DOI: 10.3969/j.issn.1001-3695.2010.05.006.]

[18] 邓泓, 刘志超, 彭莹琼, 等. 基于 Fibonacci 采样的数据预处理方法研究[J]. 江西师范大学学报(自然科学版), 2021, 45(1): 60–66. [DENG H, LIU Z C, PENG Y Q, et al. The study on data preprocessing method based on fibonacci sampling[J]. Journal of Jiangxi Normal University (Natural Sciences Edition), 2021, 45(1): 60–66.]

[19] RUSSELL B C, TORRALBA A, MURPHY K P, et al. LabelMe: a database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1): 157–173. DOI: 10.1007/s11263-007-0090-8.

[20] EVERINGHAM M. The PASCAL Visual Object Classes Homepage[EB/OL]. (2009–09–09) [2022–08–16]. http://host.robots.ox.ac.uk/pascal/VOC/index.html.

[21] ZHU Y J, LI S S, DU W S, et al. Identification of table grapes in the natural environment based on an improved Yolov5 and localization of picking points[J]. Precision Agriculture, 2023, 24(4): 1333–1354. DOI: 10.1007/s11119-023-09992-w.

[22] JIN Y C, LIU J Z, WANG J, et al. Far-near combined positioning of picking-point based on depth data features for horizontal-trellis cultivated grape[J]. Computers and Electronics in Agriculture, 2022, 194: 106791. DOI: 10.1016/j.compag.2022.106791.

[23] ZHAO R Z, ZHU Y C, LI Y H. An end-to-end lightweight model for grape and picking point simultaneous detection[J]. Biosystems Engineering, 2022, 223: 174–188. DOI: 10.1016/j.biosystemseng.2022.08.013.

[24] AMIT C, MANDEEP S. Computer vision and machine learning based grape fruit cluster detection and yield estimation robot[J]. Journal of Scientific and Industrial Research (JSIR), 2022: 866 – 872. DOI: 10.56042/jsir.v81i08.57971.

[25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada. IEEE, 2023: 7464–7475. DOI: 10.1109/cvpr52729.2023.00721.

[26] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 833–851. DOI: 10.1007/978-3-030-01234-2_49.

论文引用格式

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, 2025, 10(1). (2023-10-07). DOI: 10.11922/11-6035.csd.2023.0132.zh.

数据引用格式

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[DS/OL]. Science Data Bank, 2023. (2023-08-28). DOI: 10.57760/sciencedb.j00001.00883.

A dataset of grape multimodal object detection and semantic segmentation

CHEN Wenjun[1,2,3], RAO Yuan [1,2,3*], WANG Fengyi [1,2,3], ZHANG Yu [1,2,3], YANG Yumeng [1,2,3], LUO Qing [1,2,3], ZHANG Tong [1,2,3], WAN Tianyu [1,2,3], LIU Xinyu [1,2,3], ZHANG Mengyu [1,2,3], ZHANG Rui [1,2,3]

1. College of Information and Computer Science, Anhui Agricultural University, Hefei 230036, P. R. China

2. Key Laboratory of Agricultural Sensors, Ministry of Agriculture and Rural Affairs, Hefei 230036, P. R. China

3. Anhui Provincial Key Laboratory of Smart Agricultural Technology and Equipment, Hefei 230036, P. R. China

*Email: raoyuan@ahau.edu.cn

Abstract: The accuracy of grape picking point localization relies on the performance of grape detection and semantic segmentation network. However, in practical application scenarios, the accuracy and segmentation precision of grape targets based on visible light images are susceptible to light variations and complex environments, often resulting in poor performance. Moreover, grapes grow in bunches, and the existing multimodal datasets for apples and pears may not adequately fulfill the recognition needs of bunch-shaped grapes. The construction of a dataset of grape multimodal object detection and semantic segmentation based on visible light, depth, and near-infrared multimodal object detection and semantic segmentation is crucial to exploring better recognition rates and stronger generalization capabilities for grape detection and semantic segmentation models. This dataset, totaling about 39.08 GB, contains high-quality multimodal video stream data of green and purple grapes, including six varieties captured under different illumination and obscuration conditions. Additionally, the dataset offers 3,954 labeled image samples extracted from the aforementioned multimodal video stream. Using techniques such as rotation, deflation, mis-slicing, panning, and Gaussian blur, the dataset can be augmented for the training implementation of mainstream deep learning models. The dataset can provide valuable basic data resources for multimodal fusion, grape semantic segmentation, and object detection, and have important practical application value for promoting research in the field of agricultural machinery and equipment intelligence.

Keywords: natural environment; grapes; multimodal; semantic segmentation; object detection

Dataset Profile

Title	A dataset of grape multimodal object detection and semantic segmentation
Data corresponding author	RAO Yuan (raoyuan@ahau.edu.cn)
Data author(s)	CHEN Wenjun, RAO Yuan, WANG Fengyi, ZHANG Yu, YANG Yumeng, LUO Qing, ZHANG Tong, WAN Tianyu, LIU Xinyu, ZHANG Mengyu, ZHANG Rui
Time range	August, 2021 and July, 2023
Geographical scope	Hefei City, Anhui Province
Data volume	39.08 GB
Data format	*.png, *.xml, *.json, *.mkv
Data service system	<http://doi.org/10.57760/sciencedb.j00001.00883>
Source(s) of funding	The National Natural Science Foundation of China (32371993); Key Research and Development Projects in Anhui Province (202104a06020012, 202204c06020022); the Natural Science Major Project for Anhui Provincial University (2022AH040125).
Dataset composition	This dataset contains the original multimodal video file, which is stored independently in the VID folder in mkv format. The labeled image data are sampled from the video files at certain frame interval and grouped into two categories based on illumination conditions: normal illumination (NORMAL) and poor illumination (POOR); each of them is further divided into two sub-folders, namely green grapes (GREEN) and purple grapes (PURPLE). Each multimodal sample comprises of visible, depth, near-infrared images, annotation information files of segmentation and object detection, saved in the corresponding folders: (1) The RGB folder stores the visible images of the grapes, saved in PNG format. (2) The Depth folder stores the depth image corresponding to the visible images, saved in PNG format. (3) The NIR folder stores the near-infrared images corresponding to visible images, saved in PNG format. (4) The Seg folder stores the json format annotation file corresponding to the visible images, containing pixel annotation information for the grapes. (5) The VOC folder stores the VOC format object detection annotation file corresponding to the visible images, containing the size of each image and the position of the grapes in the image.