



文献 CSTR:

32001.14.11-6035.csd.2023.0161.zh



文献 DOI:

10.11922/11-6035.csd.2023.0161.zh

数据 DOI:

10.57760/sciencedb.j00001.00947

文献分类: 生物学

收稿日期: 2023-10-07

开放同评: 2023-11-16

录用日期: 2024-08-24

发表日期: 2025-02-17

* 论文通信作者

邱荣洲: 49497479@qq.com

柑橘黄龙病田间症状识别图像数据集

池美香¹, 陈韶萍¹, 黄婷¹, 陈世雄², 梁勇², 邱荣洲^{1*}

1. 福建省农业科学院植物保护研究所, 福建省作物有害生物监测与治理重点实验室, 福州 350003

2. 福建省农业科学院数字农业研究所, 福州 350003

摘要: 柑橘黄龙病的快速识别有助于生产者尽早发现并清除柑橘黄龙病植株。基于深度学习的植物病害图像自动识别依赖于庞大的数据集, 然而开放的可用于柑橘黄龙病识别的数据集非常有限。本研究通过手机或相机拍摄, 整理了一组包含自然生态、白色背景板和黑色背景板 3 种采集环境的照片数据集, 收录了包含黄龙病及相关病害症状共 14 种类别的图片 9896 张。其中, 训练集样本 7192 张, 验证集样本 898 张, 测试集样本 1806 张 (其中 602 个样本具有 PCR 检测结果)。样本的症状类别均由植保专家进行分类和标注以确定数据的质量和可靠性。本数据集提供了 yolo 格式 (txt) 的标注文件和基于 YOLOv5l 训练获得的模型文件, 以期为基于 YOLO 卷积神经网络的植物病害分类、目标检测和自动识别领域的机器学习建模提供训练和测试样本。

关键词: 黄龙病; 柑橘病害; 症状识别; 图像识别; YOLO; 数据集

数据库 (集) 基本信息简介

数据库 (集) 名称	柑橘黄龙病田间症状识别图像数据集
数据通信作者	邱荣洲 (49497479@qq.com)
数据作者	池美香, 陈韶萍, 黄婷, 陈世雄, 梁勇, 邱荣洲
数据时间范围	2018–2021年
地理区域	福建, 地理范围为24°41'52"–27°4'56"N, 117°6'53"–119°31'35"E。
数据量	1.18 GB
数据格式	*.jpg, *.txt, *.json
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00947
基金项目	国家重点研发项目 (2021YFD1400800); “5511”协同创新工程 (XTCXGC2021011、XTCXGC2021017); 福建省农业科学院植物保护创新团队项目 (CXTD2021027)。
数据库 (集) 组成	数据集共包括4个数据文件: (1) trains.zip为训练集和验证集数据, 数据量541.4 MB; (2) tests.zip为测试集数据, 数据量156.1 MB; (3) models.zip为模型文件, 数据量464.3 MB; (4) PCRs.zip为PCR检测结果数据, 数据量48.1 MB。

引言

柑橘黄龙病是由韧皮部杆菌 (*Candidatus Liberibacter* spp., CLas) 引起的柑橘生产上的毁灭性病害, 可通过柑橘木虱在果园传播, 尽早发现并清除病株是阻止病害传播的有效手段。然而, 感染黄龙病的植株常出现叶片斑驳、均化黄化、缺锌状花叶、红鼻子果等多种症状^[1], 极大地增加了黄龙病诊断的难度。对黄龙病进行快速准确识别有助于生产者尽早发现黄龙病植株。

近几年来深度学习 (Deep Learning, DL) 被广泛应用于图像识别研究中, 国内外学者开展了利用图像识别算法来自动检测黄龙病的研究。例如, Choi 等^[2]基于卷积神经网络 (Convolutional Neural Networks, CNN) 构建了针对黄龙病、健康、锈蜘蛛为害状、风害的柑橘果实自动鉴别系统。Mohanty 等^[3]基于 CNN 构建了针对包括柑橘黄龙病在内的 26 种植物病害图像的识别器。戴泽翰等^[4]基于深度卷积神经网络构建了柑橘黄龙病症状识别模型; 陆健强等^[5]基于 Mixup 算法和卷积神经网络实现了柑橘黄龙病果实的智能识别。林少丹^[6]等构建了柑橘叶片的主叶脉韧皮部的显微图像数据集, 为在显微尺度下快速诊断柑橘黄龙病提供支撑。

基于深度学习的植物病害图像自动识别研究依赖于庞大的数据集, 这些数据集涵盖的种类、数量差异较大, 且没有开放共享; 同时测试集数据采集环境也较为单一, 无法对训练的模型进行多场景应用评估^[7-8]。鉴于此, 本数据集通过野外拍摄, 整理了一组包含自然生态、白色背景板和黑色背景板 3 种采集环境下的黄龙病及相关病害症状照片, 并提供了 yolo 格式 (txt) 记录照片中柑橘黄龙病症状种类和发病叶片、果实位置信息的标注文件以及基于 YOLOv5 训练获得的模型文件。本次发布的数据集收录了包括健康、典型黄龙病特征、疑似黄龙病特征、黄龙病相近特征和黄龙病不相关特征 5 个大类别, 包含训练集照片 8090 张, 测试集照片 1806 张。本数据集的建立为植物病害分类、目标检测和自动识别领域的机器学习建模提供训练和测试样本。

1 数据采集和处理方法

1.1 数据采集方法

本数据集样本图片采自福建省闽侯县、古田县、福安市、顺昌县、三明梅列区、永安市、尤溪县、漳州长泰区等地的 12 个柑橘种植园 (图 1)。所采集的柑橘品种包括芦柑 (*Citrus reticulata* Blanc Ponkan)、蕉柑 (*C. reticulata* var. *tankan*)、蜜橘 (*C. unshiu* Marc.)、沃柑 (Orah)、红美人 (*Hongmeiren citrus* hybrid)、砂糖橘 (*C. reticulata* cv *Shatangju*)、脐橙 (*C. sinensis* Osb. var. *brasiliensis* Tanaka)、沙田柚 (*C. grandis* var. *shatinyu* Hort)。在 2018–2021 年的 9 至 12 月间, 采集夏梢、秋梢上斑驳黄化状 (blotchy mottling)、红鼻子果 (red-nose fruit)、缺锌状 (zinc-deficiency)、叶脉黄化状 (vein-yellowing)、均匀黄化状 (uniform yellowing)、缺镁状 (magnesium-deficiency)、缺硼状 (boron-deficiency)、炭疽病 (anthracnose)、脂点黄斑病 (citrus greasy spot)、青苔 (citrus moss)、煤烟病 (sooty mould)、溃疡病 (canker)、健康叶片 (healthy leaf) 和健康果实 (healthy fruit) 的样本图像。用于图像采集的叶片要求叶型完整, 可自然展开, 果实要求完好无损, 果蒂清晰可见。

拍摄手机采用小米 MI9、华为、苹果等智能手机, 相机为 Sony-RX100 数码相机。摄像头与样本间的拍照距离 50–150 mm。采样时的天气有晴天、多云或雨天, 时间在 8:30–18:00 之间。为了提高样本的多样性, 将田间采集的部分样本带回实验室, 以白色板或黑色板为背景进行拍摄。

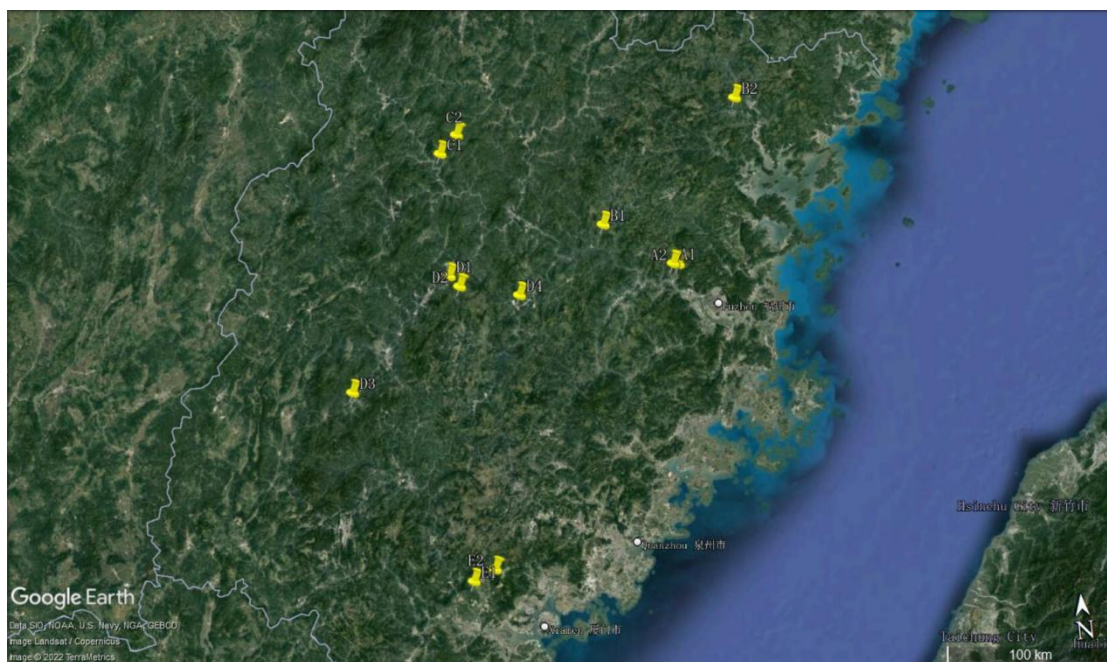


图 1 采集图片的柑橘果园地理位置

Figure 1 Geographical locations of citrus orchards, where plant images were acquired

A1、A2: 福州; B1、B2: 宁德; C1、C2: 南平; D1-4: 三明; E1, E2: 漳州

A1, A2: Fuzhou; B1, B2: Ningde; C1, C2: Nanping; D1-4: Sanming; E1, E2: Zhangzhou

1.2 数据处理方法

将原始图片用 Photoshop 统一处理为 JPG 格式, 图像分辨率为 72 像素/英寸, 按长边(宽度或高度)为 640 像素进行同比例缩小。用 HyperSpider LabelTool (自主开发标注软件) 标注叶片、果实目标边界框。标注包括边界框坐标和类别两类信息。标注时将柑橘叶片或果实的最小外接矩形框作为真实框, 以减少真实框内的背景像素。标注文件以图片同名的 yolo 格式(txt) 进行存储。

因采集的各症状的样本数量存在不平衡, 为了提高检测模型的泛化能力和鲁棒性, 同时避免模型过拟合, 另通过网络搜索和数据增强获取样本。数据增强采用 Photoshop 将样本按亮度值+50 进行统一处理, 且仅对采集样本不足 400 张的数据类别进行数据增强。

2 数据样本描述

根据叶片和果实症状差异, 将图像数据集划分为健康 (Dataset A)、典型黄龙病特征 (Dataset B)、疑似黄龙病特征 (Dataset C)、黄龙病相近特征 (Dataset D) 和不相关特征 (Dataset E) 5 个大类别, 并细分 S1-S14 14 个小类别 (表 1)。详细信息如下:

Dataset A: 在健康植株上采集的样本图片, 包括健康果实 (S1) 和叶片 (S2) 2 种 (图 2A)。

Dataset B: 在具有典型黄龙病特征的植株上采样, 包括斑驳黄化状 (S3)、红鼻子果 (S4) 2 种 (图 2B)。

Dataset C: 在具有疑似黄龙病特征的植株上采样, 这些特征可能是因为黄龙病引起的症状, 包括缺锌状 (S5)、叶脉黄化状 (S6)、均匀黄化状 (S7) 3 种 (图 2C)。

Dataset D: 主要包括一些与黄龙病特征相似的非黄龙病病害, 本文以缺镁状 (S8)、缺硼状 (S9)、

炭疽病（S10）、脂点黄斑病（S11）、青苔（S12）5种症状为例（图 2D）。

Dataset E: 主要包括非黄龙病引起的不具有黄龙病相近特征的危害, 这里以煤烟病（S13）、溃疡病（S14）（图 2E）。

表 1 在不同采集环境下的图片数据集的症状分类、采集品种和样本量

Table 1 Symptom categories, varieties, and sample sizes of the image database acquired under different conditions

数据集 Dataset	症状类别 Symptom type	品种 variety	采集环境 A Acquisition condition A			采集环境 B Acquisition condition B	采集环境 C Acquisition condition C
			训练集 Training set	验证集 Validation set	测试集 A Test set A	测试集 B Test set B	测试集 C Test set C
健康 Health (Dataset A)	健康果实 Healthy fruit (S1)	a, b	483+54*	60	57	43	43
	健康叶片 Healthy leaf (S2)	a, b	471	60	56	43	43
典型黄龙病特征 typical HLB symptoms (Dataset B)	斑驳黄化状 Blotchy mottling (S3)	a, c, d	586	70	67	43	43
	红鼻果 “Red-nose” fruit (S4)	a, c, e	185+23*+208 [#]	16+10*+26 [#]	26	43	43
疑似黄龙病特征 suspected HLB symptoms (Dataset C)	缺锌状 Zinc-deficiency (S5)	a, c, f	622	80	76	43	43
	叶脉黄化状 Vein-yellowing (S6)	a, c, e	391	47	44	43	43
	均匀黄化状 Uniform yellowing (S7)	a, b, c	275+275 [#]	35+35 [#]	34	43	43
黄龙病相近特	缺镁状	a, g	262+262 [#]	33+33 [#]	32	43	43

数据集 Dataset	症状类别 Symptom type	品种 variety	采集环境 A Acquisition condition A			采集环境 B Acquisition condition B	采集环境 C Acquisition condition C
			训练集 Training set	验证集 Validation set	测试集 A Test set A	测试集 B Test set B	测试集 C Test set C
征 HLB-like symptoms (Dataset D)	Magnesium-deficiency (S8)						
	缺硼状 Boron-deficiency (S9)	e	254+254 [#]	32+32 [#]	32	43	43
	炭疽病 Anthracnose (S10)	a, h	286+286 [#]	36+36 [#]	36	43	43
	脂点黄斑病 Citrus greasy spot (S11)	a, g	147+55*+ 202 [#]	20+7*+27 [#]	27	43	43
	青苔 Citrus moss (S12)	a, g	361	45	44	43	43
黄龙病不相关 特征 HLB-irrelevant symptoms (Dataset E)	煤烟病 Sooty mould (S13)	a, g	327+327 [#]	42+42 [#]	36	43	43
	溃疡病 Canker (S14)	b, e	298+298 [#]	37+37 [#]	35	43	43
Total			7192	898	602	602	602

注: a 表示柑橘品种芦柑 *Citrus reticulata* Blanc Ponkan; b 表示柑橘品种沃柑 Orah; c 表示柑橘品种蕉柑 *C. reticulata* var. *tankan*; d 表示柑橘品种沙田柚 *C. grandis* var. *shatinyu* Hort; e 表示柑橘品种脐橙 *C. sinensis* Osb. var. *brasiliensis* Tanaka; f 表示柑橘品种红美人 *Hongmeiren citrus* hybrid; g 表示柑橘品种蜜橘 *C. unshiu* Marc.; h 表示柑橘品种砂糖橘 *C. reticulata* cv *Shatangju*; *表示图片来源网络搜索; #表示图片来源于数据增强。

根据采集设备和拍照环境的不同，将样本划分为 Acquisition condition A、B 和 C（表 1），详细信息如下：Acquisition condition A 使用华为、小米手机在田间自然生境或在室内纯色背景板（白色或黑色）上的采集的照片，共 8692 张；Acquisition condition B 使用相机和苹果手机拍照，拍照背景为田间自然生境的数据，共采集 602 张照片；Acquisition condition C 将 Acquisition condition B 采集

的叶片或果实样本摘取下来放在白色背景板上用相机或手机进行二次拍照，共采集 602 张照片。其中，Acquisition condition B 和 C 的图像采集的植株、采集时间和采集人均与 Acquisition condition A 不同。

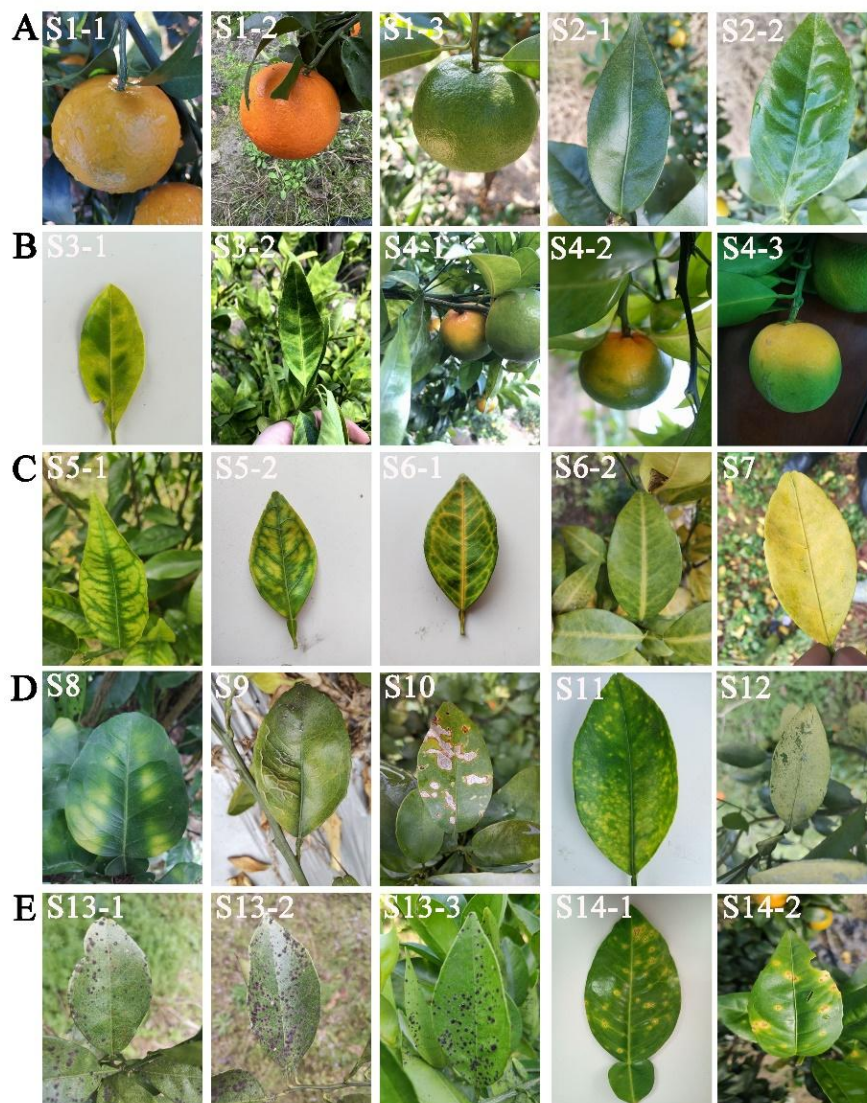


图 2 数据集中的不同症状的叶子和果实图像实例

Figure 2 Example of leaf and fruit images with different symptoms in the dataset

A, dataset A (健康); B, dataset B (典型黄龙病特征); C, dataset C (疑似黄龙病特征); D, dataset D (黄龙病相近特征); E, dataset E (黄龙病不相关特征)。S1, 健康果实; S2, 健康叶片; S3, 斑驳黄化状; S4, 红鼻子果; S5, 缺锌状; S6, 叶脉黄化状; S7, 均匀黄化状; S8, 缺镁状; S9, 缺硼状; S10, 炭疽病; S11, 脂点黄斑病; S12, 青苔; S13, 煤烟病; S14, 溃疡病。

A, dataset A (healthy); B, dataset B (typical HLB symptoms); C, dataset C (suspected HLB symptoms); D, dataset D (HLB-like symptoms); E, dataset E (HLB-irrelevant symptoms). S1, healthy fruit; S2, healthy leaf; S3, blotchy mottling; S4, “red-nose” fruit; S5, zinc-deficiency; S6, vein-yellowing; S7, uniform yellowing; S8, magnesium-deficiency; S9, boron-deficiency; S10, anthracnose; S11, citrus greasy spot; S12, citrus moss; S13, sooty mould; S14, canker.

数据集包括四部分：第一部分为训练集和验证集，标注了症状类别编号的样本图片、标注文件及 JSON 格式的标注对象说明文件，根据症状类别差异组合成 Dataset A、Dataset B、Dataset C、Dataset

D、Dataset E 共 5 种图像数据集，分别对应 A-train.zip、B-train.zip、C-train.zip、D-train.zip、E-train.zip 5 个训练集压缩文件和 A-val.zip、B-val.zip、C-val.zip、D-val.zip、E-val.zip 5 个验证集压缩文件。各症状数据文件夹参照表 1 的症状类别代码 S1-S14 进行命名。第二部分为测试集，根据采集设备、拍照环境和症状类别组合的不同，将测试集划分为 testA、testB、testC 共 3 个测试集，其中 testA 为使用华为、小米手机在田间自然生境或在室内纯色背景板（白色或黑色）上的采集的照片，testB 为使用相机和苹果手机拍照，拍照背景为田间自然生境的数据，testC 为在白色背景板上用相机或手机进行拍照，分别对应 testA.zip、testB.zip、testC.zip 3 个测试集压缩文件。第三部分为模型文件，通过使用 Dataset A + B + C、Dataset A + B + C + D、Dataset A + B + C + D + E 3 种不同数据类别组合的数据集分别训练 Yolov5l 网络结构获得，模型存放在 models 压缩文件内。第四部分为 PCR 检测结果，将 testB 的 602 个样本分别进行 CLas PCR 检测，结果阳性的样本图片为 PCR+.zip 压缩文件，阴性的为 PCR-.zip 压缩文件。

3 数据质量控制和评估

数据来源方面，本数据集收录了 14 种类别的柑橘黄龙病田间症状识别图片 9896 张，其中利用手机或相机在田间采集拍照 7516 张；通过网络搜索获取样本 149 张，包括健康果实 54 张（训练集）、脂点黄斑病 62 张（训练集 55 张，验证集 7 张）、红鼻果 33 张（训练集 23 张，验证集 10 张）。为以示区别，来源网络的图片在文件命名时以 net 开头，如“net*.jpg”（*表示任意字符）；通过数据增强获取样本 2380 张，以“PS_L50_”开头进行文件命名，如“PS_L50_*.jpg”。

样本的症状类别由植保专家进行分类。黄龙病症状分类参照《柑桔黄梢（黄龙）病研究 I. 病情调查》^[9]和《柑橘黄龙病症状与“*Candidatus Liberibacter asiaticus*”PCR 检测结果的相关性分析》^[1]有关柑橘黄龙病症状的分型方法。数据标注由植保专家完成，可及时剔除过程中出现的不符合要求的文件，保证本数据集中图像标注的质量和可靠性。

4 数据价值

本数据集的采集环境丰富，收录了包含健康、典型黄龙病特征、疑似黄龙病特征、黄龙病相近特征和黄龙病不相关特征 5 个大类 14 种小类的样本图片，涉及的柑橘品种有 8 种，采集地点涵盖了福建省福州、宁德、三明、漳州等地 12 个柑橘果园。同时，本数据集还提供了基于不同采集环境的 3 种测试集，以及部分样本还提供了 PCR 检测结果，可为柑橘黄龙病、柑橘病害等植物病害 AI 识别相关领域的进一步研究提供可靠的测试样本数据。通过设计多环境的不同测试集，本数据集可应用于不同深度学习算法研究的比较分析，Qiu 等^[10]基于本数据集研发了柑橘黄龙病田间智能识别系统，实现了在田间利用手机拍照进行自动识别黄龙病症状的功能，为种植农户防控柑橘黄龙病提供技术支持。本数据集的开放共享将为今后进一步利用机器学习建模识别柑橘病害提供了丰富的样本库。

5 数据使用方法和建议

柑橘黄龙病的症状复杂多样，不同品种、不同部位上表现的症状也不同，在初、中、后不同时期其病斑的形状、大小、色泽存在差异，且是一个不断变化的过程，在数据分类过程中可能存在因

不同专家的分类标准不一导致出现分类偏差问题，因此使用本数据集的图像数据时可根据具体需求进行二次筛选。本数据集可直接应用于 yolo 系列的模型训练，标注对象类别说明存储在 label_object.json 文件，数据集样本按症状类别代码 S1-S14 进行文件夹命名，数据使用者可自由选择数据组合进行模型训练，以满足不同的应用需求。本数据集还提供基于 yolov5l 深度学习网络训练获得模型文件 HLB_ABC、HLB_ABCD、HLB_ABCDE，数据使用者可使用不同的深度学习网络进行模型训练，比较不同算法的识别精度。

数据作者分工职责

池美香（1971—），女，福建省邵武市人，工程师，研究方向为数字植保技术。主要承担工作：数据集设计，照片拍摄，数据标注，论文撰写。

陈韶萍（1990—），女，福建省泉州市人，助理研究员，研究方向为分子生物学。主要承担工作：照片拍摄，数据标注，论文修改。

黄婷（1998—），女，福建省政和县人，科研助理，研究方向为数字植保技术。主要承担工作：照片拍摄，数据标注。

陈世雄（1978—），男，福建省三明市人，科研助理，研究方向为农业工程。主要承担工作：照片拍摄。

梁勇（1986—），男，福建省福州市人，农艺师，研究方向为农业工程。主要承担工作：照片拍摄。

邱荣洲（1979—），男，福建省尤溪县人，副研究员，研究方向为数字植保技术。主要承担数据集设计，照片拍摄，数据标注，论文修改。

参考文献

- [1] 许美容, 陈燕玲, 邓晓玲, 等. 柑橘黄龙病症状与"*Candidatus Liberibacter asiaticus*" PCR 检测结果的相关性分析[J]. 植物病理学报, 2016, 46(3): 367-373. DOI:10.13926/j.cnki.apps.2016.03.010. [XU M R, CHEN Y L, DENG X L, et al. Correlations between symptoms of citrus huanglongbing and PCR detection of '*Candidatus Liberibacter asiaticus*'[J]. Acta Phytopathologica Sinica, 2016, 46(3): 367-373. DOI:10.13926/j.cnki.apps.2016.03.010.]
- [2] CHOI D, LEE W, SCHUELLER J K, et al. A precise fruit inspection system for Huanglongbing and other common citrus defects using GPU and deep learning technologies[C]//ISPA. Proceedings of 13th International Conference on Precision Agriculture, St Louis, Missouri: ISPA. 2016: 1-6.
- [3] MOHANTY S P, HUGHES D P, SALATHE M. Using deep learning for image-based plant disease detection[J]. Front Plant Sci, 2016, 7: 1-7. DOI:10.3389/fpls.2016.01419.
- [4] 戴泽翰, 郑正, 黄莉舒, 等. 基于深度卷积神经网络的柑橘黄龙病症状识别[J]. 华南农业大学学报, 2020, 41(4): 111-119. DOI:10.7671/j.issn.1001-411X.201909031. [DAI Z H, ZHENG Z, HUANG L S, et al. Recognition of Huanglongbing symptom based on deep convolutional neural network[J]. Journal of South China Agricultural University, 2020, 41(4): 111-119. DOI:10.7671/j.issn.1001-411X.201909031.]
- [5] 陆健强, 林佳翰, 黄仲强, 等. 基于 Mixup 算法和卷积神经网络的柑橘黄龙病果实识别研究[J]. 华南农业大学学报, 2021, 42(3): 94-101. DOI:10.7671/j.issn.1001-411X.202008041. [LU J Q, LIN J H,

- HUANG Z Q, et al. Identification of citrus fruit infected with Huanglongbing based on Mixup algorithm and convolutional neural network[J]. Journal of South China Agricultural University, 2021, 42(3): 94-101. DOI:10.7671/j.issn.1001-411X.202008041.]
- [6] 林少丹, 李效彬, 杨碧云, 等. 适用于小样本显微图像数据集的柑橘黄龙病快速诊断模型[J]. 农业工程学报, 2022, 38(12): 216-223. DOI: 10.11975/j.issn.1002-6819.2022.12.025. [LIN S, LI X, YANG B, et al. Detecting citrus Huanglongbing from few-shot microscopic images using an improved DETR[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(12): 216-223. DOI: 10.11975/j.issn.1002-6819.2022.12.025.]
- [7] BARBEDO J G A. Factors influencing the use of deep learning for plant disease recognition[J]. Biosystems engineering, 2018, 172: 84-91. DOI:10.1016/j.biosystemseng.2018.05.013.
- [8] BARBEDO J G A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification[J]. Computers and electronics in agriculture, 2018, 153: 46-53. DOI: 10.1016/j.compag.2018.08.013.
- [9] 林孔湘. 柑桔黄梢(黄龙)病研究I.病情调查[J]. 植物病理学报, 1956. 2(1): 1-11. DOI:10.13926/j.cnki .apps .1956.01.001. [LIN K X. Observations on Yellow Shoot of Citrus[J]. Acta Phytopathologica Sinica, 1956. 2(1): 1-11. DOI:10.13926/j.cnki .apps .1956.01.001.]
- [10] QIU R Z, CHEN S P, CHI M X, et al. An automatic identification system for citrus greening disease (Huanglongbing) using a YOLO convolutional neural network[J]. Frontiers in Plant Science, 2022, 13: 1002606. DOI:10.3389/fpls.2022.1002606.

论文引用格式

池美香, 陈韶萍, 黄婷, 等. 柑橘黄龙病田间症状识别图像数据集[J/OL]. 中国科学数据, 2025, 10(1). (2025-02-17). DOI: 10.11922/11-6035.csd.2023.0161.zh.

数据引用格式

池美香, 陈韶萍, 黄婷, 等. 柑橘黄龙病田间症状识别图像数据集[DS/OL]. V2. Science Data Bank, 2024. (2024-08-05). DOI: 10.57760/sciencedb.j00001.00947.

A dataset of field symptoms for identifying Citrus Huanglongbing

CHI Meixiang¹, CHEN Shaoping¹, HUANG Ting¹, CHEN Shixiong²,
LIANG Yong², QIU Rongzhou^{1*}

1. Fujian Key Laboratory for Monitoring and Integrated Management of Crop Pests, Institute of Plant Protection, Fujian Academy of Agricultural Sciences, Fuzhou 350013, P.R. China

2. Institute of Digital Agriculture, Fujian Academy of Agricultural Sciences, Fuzhou 350003, P.R. China

*Email: 49497479@qq.com (QIU Rongzhou)

Abstract: The rapid identification of Citrus Huanglongbing can help producers to detect and eliminate infected plants as early as possible. The automatic recognition of plant diseases based on deep learning

relies on large datasets, but there are few open datasets available for citrus huanglongbing recognition. In this study, a datasets of images was collected using a mobile phone or camera, involving three different collection environments: natural environment, white background, and black background. The dataset includes a total of 9,896 images including 14 categories of symptoms related to Huanglongbing and other associated diseases. Among these, 7,192 samples were used for the training set, 898 for the validation set, and 1,806 for the test set (of which 602 had PCR test results). The symptom categories of the samples were classified and annotated by plant detection experts to ensure the quality and reliability of the data. The datasets presents a YOLO format (txt) annotation file and model files trained using YOLOv5l, aiming to provide training and testing samples for developing models of machine learning based on the YOLO convolutional neural network in the fields of plant disease classification, object detection and automatic recognition.

Keywords: Huanglongbing; citrus disease; symptom recognition; image recognition; YOLO; dataset

Dataset Profile

Title	A dataset of field symptoms for identifying Citrus Huanglongbing
Data corresponding author	QIU Rongzhou (49497479@qq.com)
Data authors	CHI Meixiang, CHEN Shaoping, HUANG Ting, CHEN Shixiong, LIANG Yong, QIU Rongzhou
Time range	2018–2021
Geographical scope	Fujian Province (24°41'52"–27°4'56"N, 117°6'53"–119°31'35"E)
Data volume	1.18 GB
Data format	*.jpg, *.txt, *.json
Data service system	< https://doi.org/10.57760/sciencedb.j00001.00947 >
Sources of funding	The National Key R&D Program of China (2021YFD1400800); “5511” Collaborative Innovation Project of High-quality Agricultural Development and Surpassment in Fujian Province (XTCXGC2021011, XTCXGC2021017); the Innovation Team of Plant Protection and the Fujian Academy of Agricultural Sciences (CXTD2021027).
Dataset composition	The dataset consists of 4 data files as follows: (1) trains.zip is the training sets and validation sets, with a data volume of 541.4 MB; (2) tests.zip is testing sets, with a data volume of 156.1 MB; (3) models.zip is a model file, with a data volume of 464.3 MB; (4) PCRs.zip is PCR detection results data, with a data volume of 48.1 MB.