

新疆棉田主要昆虫图像数据集 CottonInsect

www.sciengine.com/CSD

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14.11-6035.csd.2023.0130.zh



文献 DOI:

10.11922/11-6035.csd.2023.0130.zh

数据 DOI:

10.57760/sciencedb.j00001.00866

文献分类: 生物学

收稿日期: 2023-07-19

开放同评: 2023-07-20

录用日期: 2023-11-10

发表日期: 2024-04-07

杨满仙^{1,2,3}, 陈燕红^{1,2,3*}, 李雨晴^{1,2,3*}, 李永可^{1,2,3}

1. 新疆农业大学, 乌鲁木齐 830052
2. 智能农业教育部工程研究中心, 乌鲁木齐 830052
3. 新疆农业信息化工程技术研究中心, 乌鲁木齐 830052

摘要: 棉花是新疆重要经济作物, 由于种植结构调整、农药不合理使用与气候变化等原因, 棉花易受多种虫害侵袭, 影响其产量与品质。因此, 构建复杂场景下棉田昆虫图像数据集对棉田害虫综合治理具有基础性意义。本文构建了复杂场景下新疆棉田主要昆虫图像数据集, 用于真实环境下棉田昆虫的识别与检测。本数据集包含 13 种 (类) 常见的棉田昆虫, 共 3225 张图像, 原始图像共 24 GB, 均通过严格筛选, 确保图像质量。对每张图像的棉田昆虫进行人工标注, 构建图像分类数据集和目标检测数据集, 数据集大小共 29.7 GB。使用常见的深度学习模型对数据集进行评估, 确保数据集的可用性和可靠性。本数据为棉田昆虫图像分类、目标检测等提供基础数据, 对促进农业领域的发展、棉花虫害防治研究及提高棉花产量具有重要的实际应用价值。

关键词: 棉田昆虫; CottonInsect 数据集; 图像分类; 目标检测

数据库 (集) 基本信息简介

数据库 (集) 名称	新疆棉田主要昆虫图像数据集 CottonInsect
数据通信作者	陈燕红 (cyh@xjau.edu.cn); 李雨晴 (943655041@qq.com)
数据作者	杨满仙、陈燕红、李雨晴、李永可
数据时间范围	2022年
地理区域	新疆沙雅县
数据量	29.7 GB
数据格式	*.JPG、*.XML、*.JSON
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00866
基金项目	科技部科技创新2030——“新一代人工智能”重大项目(2022ZD0115805); 新疆维吾尔自治区重大科技专项课题(2022A02011-3)。
数据库 (集) 组成	数据集共包括3个数据文件, 共计29.7 GB。其中:(1) CottonInsect_image.zip是原始图像数据, 包含3225张JPG图像, 为棉铃虫、多异瓢虫、苜蓿盲蝽等13种 (类) 常见棉田昆虫不同时期图像, 数据量24 GB; (2) CottonInsect_classification.zip是图像分类数据, 包含txt标注文件、JPG图像文件和class_indices.json、classes.txt两种格式的分类文件, 数据量2.85 GB; (3) CottonInsect_detection.zip是目标检测数据, 包含xml标注文件、JPG图像文件和classes.txt类别文件, 数据量2.85 GB。

* 论文通信作者

陈燕红: cyh@xjau.edu.cn

李雨晴: 943655041@qq.com

引言

由于独特的地理位置和气候条件，新疆盛产优质棉花。根据国家统计局关于 2022 年棉花产量的公告^[1]，新疆维吾尔自治区棉花种植面积为 2496.9 万亩、总产量为 539.1 万吨，分别占同年中国棉花总种植面积和棉花总产量的 83.22%、90.2%。在棉花在种植生产过程中，病虫害是严重制约棉花产量及品质的关键因素之一。通过调查，棉田病虫害分布范围不均且害虫种类较多，无法通过单一的化学生物手段解决虫害。传统的棉花虫害防治是直接向棉田棉花植株全覆盖喷洒农药，这样不仅不利于棉花的生长还会造成严重的农药污染。因此，需要使用科学有效的方法检测棉花害虫。例如利用人工智能技术中的目标检测判断棉花虫害严重程度，及时给予精准防治，以避免大面积虫害、农药滥用。同时，在棉花害虫调查和管理中，对害虫天敌（益虫）的调查是重要的，以便全面了解生态系统中的相互关系，并采取可持续的害虫管理措施。因此结合深度学习方法精准检测棉花害虫和益虫，对于提高棉花产量和品质，推动地区经济发展至关重要。

现阶段除 IP102 数据集^[2]以外，大多数害虫数据集的场景复杂度不够高，且数据集的规模较小，并且没有针对棉田昆虫的公开数据集可以使用。为了关注复杂背景下的棉田昆虫分类及检测的问题，为自然场景下棉田昆虫分类提供可行的数据标准，本研究根据课题组病虫害研究团队进行的棉花群落调查研究中涉及的 19 类昆虫，挑选出具有适合图像识别、迁移性和隐蔽性强等特点的 13 类昆虫进行研究，分别为苜蓿盲蝽、牧草盲蝽、绿盲蝽、黑食蚜盲蝽、茶翅蝽、中黑盲蝽、棉铃虫、草蛉、多异瓢虫、菱斑巧瓢虫、横斑瓢虫、食蚜蝇和蜂，构建了复杂背景下的小目标棉田昆虫数据集，并命名为 CottonInsect。

1 数据采集和处理方法

1.1 采集类别

新疆阿克苏地区棉花种植占新疆地区的 80%，是新疆种植棉花面积最大的地区。基于科技部科技创新 2030——“新一代人工智能”重大项目（2022ZD0115805）以及新疆维吾尔自治区重大科技专项课题（2022A02011-3），2022 年课题组前往新疆维吾尔自治区阿克苏地区沙雅县各乡镇棉田通过人工拍照采集构建棉田昆虫数据集。通过实地调查，结合《中国农业百科全书昆虫卷》^[3]和《棉花病虫害及其防治原色图册》^[4]等文献书籍，以及数据信息规范化整理，挑选出 13 种（类）沙雅地区常见的棉田昆虫类别进行研究，最终得到 CottonInsect 棉田昆虫识别研究图像数据集。13 种（类）棉田昆虫中文名、拉丁学名以及构建数据集使用的标注名称如表 1 所示，13 种（类）棉田昆虫图像如图 1 所示。

表 1 棉田昆虫类别

Table 1 Insect categories in cotton fields

序号	中文名	拉丁学名	标注名称
1	草蛉	<i>Chrysopa perla</i>	lacewing
2	多异瓢虫	<i>Hippodamia variegata</i>	ladybug_dy
3	菱斑巧瓢虫	<i>Oenopia conglobata</i>	ladybug_lbq
4	横斑瓢虫	<i>Coccinella transversoguttata</i>	ladybug_hb

序号	中文名	拉丁学名	标注名称
5	苜蓿盲蝽	<i>Adelphocoris lineolatus</i>	stinkbug_mx
6	牧草盲蝽	<i>Lygus pratensis</i>	stinkbug_mc
7	绿盲蝽	<i>Lygocoris lucorum</i>	stinkbug_lv
8	黑食蚜盲蝽	<i>Deraeocoris punctulatus</i>	stinkbug_hsy
9	茶翅蝽	<i>Halyomorpha halys</i>	stinkbug_ccc
10	中黑盲蝽	<i>Adelphocoris suturalis</i>	stinkbug_zh
11	食蚜蝇	<i>Syrphidae</i>	syrphid
12	棉铃虫	<i>Helicoverpa armigera</i>	bollworm
13	蜂	<i>Apoidea</i>	bee



图1 棉田昆虫图像

Figure 1 Insect images in cotton fields

1.2 采集设备

数据采集设备使用 vivo x27 后置原相机，图像大小设置为 2000 像素×1500 像素进行拍摄，图像总像素为 300 万像素，高分辨率的图像可以捕获更丰富的细节，该像素尺寸能适应大多数计算机视觉任务，为后续的数据处理和分析提供更高质量的原始数据。使用者可以根据具体需求和设备限制，将图像压缩为不同大小。在使用 vivo x27 后置原相机进行拍摄时，确保相机稳定，对焦准确，获取最清晰的图片。

1.3 数据处理

1.3.1 图像筛选

数据的质量是影响模型性能的重要因素之一，为确保数据集中图像的质量，并且可用于进行机

器学习或深度学习任务的训练，通过人工筛选原始图像，清除像素过低、拍摄角度不好以及拍摄目标不完整的图像，最终得到 3225 张高质量棉田昆虫图像，并将所有图像按流水号命名。

1.3.2 图像标注

在图像分类的过程中，正确的类别标注是非常关键的。为保证标注的正确性，课题组邀请了病虫害研究团队的 3 名专家来协助进行图像类别的标注，构建用于图像分类的数据集。在图像分类数据集的文件中规定，0 代表“lacewing”，1 代表“ladybug_dy”，2 代表“ladybug_lbq”，3 代表“ladybug_hb”，4 代表“stinkbug_mx”，5 代表“stinkbug_mc”，6 代表“stinkbug_lv”，7 代表“stinkbug_hsy”，8 代表“stinkbug_ccc”，9 代表“stinkbug_zh”，10 代表“syrphid”，11 代表“bollworm”，12 代表“bee”，并保存为 json 和 txt 两种格式以供使用。接着将每张图像的名称和所属类别编号记录在 txt 文件中，例如图片名为 00001 的图片对应的类别是“lacewing”，在 txt 文件中记录为“00001.jpg 0”。完成分类后，txt 文件中记录所有图像的名称和所属类别。

在虫害防治中，面对复杂的棉田场景时需要识别昆虫的准确位置和类别。因此构建用于棉田昆虫目标检测的数据集，使用 LabelImg 工具^[5]对数据图像中昆虫目标的位置和类别进行人工标注。选择标注模式为 VOC 模式，新建标注框并拖动到适当的位置和大小，使标注框完全贴合待标注对象。选择对象对应的类别标签并保存，在保存文件夹下可以看到图片标注后生成 PASCAL VOC 2007 格式的 xml 标注文件。xml 标注文件中记录有图像名、图像大小、昆虫种类和昆虫的位置信息等。标注图片示例如图 2 所示，标注框上方显示昆虫类别名，使用不同颜色的标注框区分不同的昆虫类别。



图 2 图像标注示例

Figure 2 Image annotation examples

1.3.3 数据增强

数据增强是一种扩展训练数据集的方法，通过先验知识产生相似的生成数据，常用于深度学习任务中。数据增强可以增加训练的数据量，提高模型的泛化能力，解决样本不均衡等问题。数据增强可以通过几何变换和像素变换的方式，包括有监督和无监督的数据增强。有监督数据增强是基于已有数据进行扩充，包括单样本和多样本的数据增强，单样本数据增强方法包括旋转、剪裁、缩放、抖动，以及添加噪声、模糊，调节亮度、饱和度、直方图均衡化等。多样本数据增强方法包括 SMOTE^[6]、Mosaic^[7]、Mixup^[8]等。无监督的数据增强包括 GAN^[9]、Autoaugmentation^[10]、Diffusion^[11]等。本文数据集中存在样本不均衡问题，可以采用图像增强方法对昆虫图像进行扩充，获得更好的训练效果。图 3 展示了使用旋转方法进行数据增强的效果，左图为原始昆虫图像，右图为旋转之后效果。通过旋转图像，可以模拟昆虫在自然环境中可能出现的不同角度，从而帮助模型更加有效的学习和识别昆虫。

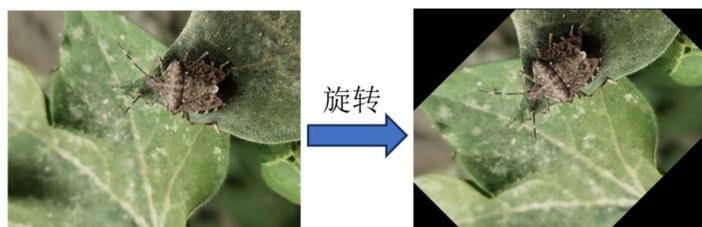


图3 图像增强示例

Figure 3 Image enhancement example

1.3.4 数据集划分

CottonInsect 数据集包含棉田害虫与益虫共 13 种(类), 共计 3225 张图像。图像分类任务按 8:1:1 的比例随机划分为训练集、验证集和测试集。其中 2587 张图像用于训练, 323 张图像用于验证, 315 张图像用于测试。表 2 中展示了该图像分类数据集中每类昆虫图像划分情况统计。目标检测任务共有 3225 张标注图像, 3485 个标注框, 按 9:1 的比例对带有标注框的图像随机划分, 2451 张图像用于训练, 1034 张图像用于测试, 表 3 中展示了该目标检测数据集中每类昆虫图像划分情况统计。

表 2 图像分类数据集划分数量统计

Table 2 Partitioning count statistics of the image classification dataset

类别	名称	训练集	验证集	测试集	总计
害虫	苜蓿盲蝽	97	12	11	120
	棉铃虫	430	54	53	537
	牧草盲蝽	369	46	46	461
	绿盲蝽	11	1	1	13
	黑食蚜盲蝽	233	29	28	290
	茶翅蝽	118	15	14	147
	中黑盲蝽	14	2	1	17
益虫	草蛉	312	39	38	389
	多异瓢虫	574	72	71	717
	菱斑巧瓢虫	132	16	16	164
	横斑瓢虫	129	16	16	161
	食蚜蝇	39	5	4	48
	蜂	129	16	16	161
CottonInsect		2587	323	315	3225

2 数据样本描述

本数据集图像文件夹中包含 13 种(类)不同昆虫, 这些图像数据都是在不同地点、不同环境、不同时间收集到的, 包含棉田昆虫在真实环境中不同姿态以及生长周期不同阶段的图像, 共有高质量 jpg 图像数据 3225 张。本数据集部分棉田昆虫示意图如图 4 所示。

表 3 目标检测数据集划分数量统计

Table 3 Partitioning count statistics of the target detection dataset

类别	名称	训练集	测试集	总计
害虫	苜蓿盲蝽	79	41	120
	牧草盲蝽	328	146	474
	绿盲蝽	11	2	13
	黑食蚜盲蝽	237	96	333
	茶翅蝽	164	53	217
	中黑盲蝽	13	4	17
	棉铃虫	377	160	537
益虫	草蛉	265	125	390
	多异瓢虫	602	247	849
	菱斑巧瓢虫	118	46	164
	横斑瓢虫	115	46	161
	食蚜蝇	32	17	49
	蜂	110	51	161
CottonInsect		2451	1034	3485



图 4 数据集特性展示

Figure 4 Demonstration of the dataset characteristics

与 IDADP^[12]等数据集相比, CottonInsect 数据集包含了真实环境中棉田昆虫不同生命周期的生长状态。与 IP102^[2]数据集相比, CottonInsect 数据同样具有不同类别棉田昆虫相似性非常高, 同一类别昆虫不同生长周期阶段差异大等特点, 例如图 4 中, 多异瓢虫、横斑瓢虫和菱斑巧瓢虫不属于

同一类别但形态、大小和颜色都非常相近，并且多异瓢虫和横斑瓢虫仅有头部黑斑形状有差别；棉铃虫的幼虫和成虫，多异瓢虫从卵、蛹、幼虫到成虫，均属于同一类别，但特征差异非常大。因此，利用符合真实环境中昆虫分布特点的昆虫数据集进行实验，能够很好地评估昆虫图像识别算法的效果和性能。

3 数据质量控制和评估

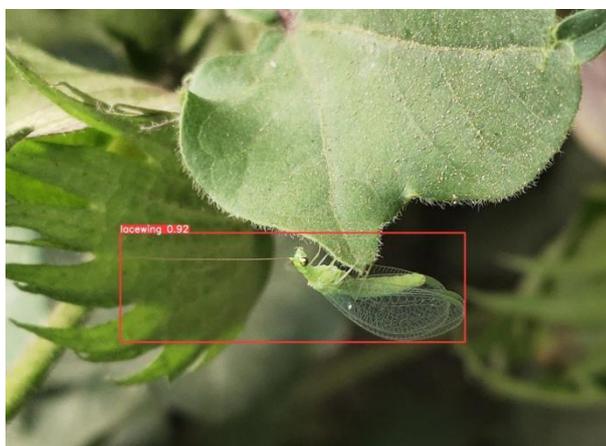
本研究中的数据集在专业技术人员的指导下，进行了数据的筛选和清洗。为提高数据的质量要求做了以下工作：（1）从采集设备、采集标准以及采集种类等几方面进行规范化约束；（2）在不同角度、光照强度和不同的实验背景下拍摄，确保数据的多样性；（3）对数据集进行清洗，通过人工筛选将拍摄不清晰、像素质量低的图像进行剔除，以提高数据集整体质量；（4）在高质量的图像数据基础上，按照现有公开数据集通用标注方法进行人工标注，确保文本数据与其配对图像的准确性和可靠性。

为进一步验证数据的可靠性，使用目前在深度学习领域较为流行的 Resnet34 及 Resnet50 网络模型进行图像分类，加载模型在公开数据集上训练过的权重进行迁移学习，通过调整参数，在 GPU 云平台训练得到图像分类结果如表 4 所示，可以看出 Resnet50 模型要比 Resnet34 模型分类效果好。使用 YOLOv5x 目标检测模型进行棉田昆虫目标检测，通过训练 mAP0.5 值为 0.93。通过图 5a 可以看出识别种类为成虫时期 lacewing（草蛉），目标框能完全框出草蛉的位置包括细小的触角；通过图 5b 可以看出尽管多异瓢虫图像有部分遮挡仍能准确识别出成虫时期的 ladybug_dy（多异瓢虫），识别置信度依然高达 0.94；图 5c 表明模型可以准确识别出图像中出现的多个目标，幼虫时期 ladybug_dy（多异瓢虫）和 stinkbug_hsy（黑食蚜盲蝽），该模型能较好识别复杂背景下棉田昆虫目标。以上结果均表明本数据是高质量可用于训练的。

表 4 图像分类实验结构

Table 4 Structure of the image classification experiment

类别	模型	Epoch	Accuracy	F1-score
图像分类	Resnet50	200	0.96	0.91
	Resnet34	200	0.85	0.82



(a)



(b)



(c)

图5 目标检测结果

Figure 5 Target detection results

4 数据价值

CottonInsect 数据集是通过实地考察和咨询专家获得的高清、真实的棉田昆虫数据资源。数据集具备类别数量分布不均、类别间差异小和类内差异大的特性，反映了真实环境中棉田昆虫的多样性和复杂性。另外，考虑到在真实环境下昆虫目标尺寸较小的情况，数据集中昆虫在图像中所占比例较小，使得本数据集适用于实际复杂背景下的昆虫监控，提高了图像分类和目标检测算法的准确性和实用性，为相关领域研究提供可靠的数据基础。

CottonInsect 数据集的建立为新疆棉花虫害防治提供了研究基础，有助于改善棉花虫害防治成效并推动新疆地区棉花产业生产过程信息化。在田间调查实验中，判断虫害的严重程度不仅需要考虑害虫数量，还需要考虑害虫的天敌益虫的数量。当害虫数量与益虫数量达到平衡时，不需要进行化学防治，依靠天敌益虫就可以抑制虫害的扩散，避免过度防治导致药害的发生。未来将通过发布新版本来更新和扩充数据集，致力于进一步构建和丰富棉田昆虫数据资源，以便为相关研究人员提供统一数据，便于不同算法在同一标准下进行比较。

致谢

感谢所有参与本数据集工作人员的辛勤付出，感谢各位专家的指导。

数据作者分工职责

杨满仙（1998—），女，云南德宏人，硕士，研究生，研究方向为计算机视觉。主要承担工作：数据采集、数据标注、汇总整理及论文撰写。

陈燕红（1979—），女，新疆乌鲁木齐人，双硕士，副教授，研究方向为视觉-语言智能表征学习。主要承担工作：总体方案设计与组织实施。

李雨晴（1998—），女，新疆乌鲁木齐人，硕士，研究生，研究方向为计算机视觉。主要承担工作：数据采集、数据标注、汇总整理。

李永可（1985—），男，河南许昌人，硕士，副教授，研究方向为智慧农业。主要承担工作：总体方案设计与组织实施。

参考文献

- [1] 国家统计局. 国家统计局关于 2022 年棉花产量的公告 [EB/OL]. (2022-12-26) [2023-6-25]. http://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901689.html. [National Bureau of Statistics. Announcement of the National Bureau of Statistics on cotton production in 2022 [J]. (2022-12-26) [2023-6-25]. http://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901689.html.]
- [2] WU X P, ZHAN C, LAI Y K, et al. IP102: a large-scale benchmark dataset for insect pest recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2020: 8779–8788. DOI: 10.1109/CVPR.2019.00899.
- [3] 中国农业百科全书总编辑委员会昆虫卷编辑委员会中国农业百科全书编辑部. 中国农业百科全书-昆虫卷[M]. 北京: 农业出版社, 1990. [General Editorial Committee of the Chinese Agricultural Encyclopedia Insect Volume Editorial Committee of the Chinese Agricultural Encyclopedia Editorial Department. Encyclopedia of Chinese Agriculture - Insect Volume [M]. Beijing: Agriculture Press, 1990.]
- [4] 崔金杰, 马奇祥. 棉花病虫害及防治原色图册[M]. 北京: 金盾出版社, 2007. [CUI J J, MA Q X. Cotton Pests and Diseases and control primary color Atlas [M]. Beijing: Golden Shield Publishing House, 2007.]
- [5] TZUTALIN. LabelImg, Git code (2015)[EB/OL].(2015) [2022.05-2022.06]. <https://github.com/tzutalin/labelImg>.
- [6] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J/OL]. arXiv: 2004.10934, 2020.
- [8] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond Empirical Risk Minimization[J]. 2017.
- [9] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.

- [10] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: Learning Augmentation Policies from Data.[J]. arXiv: Computer Vision and Pattern Recognition, 2018.
- [11] QIAO P, DOU Y, FENG W, et al. Learning non-local image diffusion for image denoising[C]//Proceedings of the 25th ACM international conference on Multimedia. 2017: 1847-1855.
- [12] 陈雷, 袁媛. 大田作物病害识别研究图像数据集[J/OL]. 中国科学数据, 2019, 4(4). (2019-06-11). DOI: 10.11922/csdata.2019.0008.zh. [CHEN L, YUAN Y. An image dataset for field crop disease identification[J/OL]. China Scientific Data, 2019, 4(4). (2019-06-11). DOI: 10.11922/csdata.2019.0008.zh.]

论文引用格式

杨满仙, 陈燕红, 李雨晴, 等. 新疆棉田主要昆虫图像数据集 CottonInsect [J/OL]. 中国科学数据, 2025, 10(1). (2024-04-03). DOI: 10.11922/11-6035.csd.2023.0130.zh.

数据引用格式

杨满仙, 陈燕红, 李雨晴, 等. 新疆棉田主要昆虫图像数据集 CottonInsect [DS/OL]. V2. Science Data Bank, 2023. (2023-11-03). DOI: 10.57760/sciencedb.j00001.00866.

A dataset of common insects in Xinjiang cotton fields:

CottonInsect

YANG Manxian^{1,2,3}, CHENG Yanhong^{1,2,3*}, LI Yuqing^{1,2,3*}, LI Yongke^{1,2,3}

1. Xinjiang Agricultural University, Urumqi 830052, P.R. China
2. Engineering Research Center of Intelligent Agriculture Ministry of Education, Urumqi 830052, P.R. China
3. Xinjiang Agricultural Informatization Engineering Technology Research Center, Urumqi 830052, P.R. China

*Email: cyh@xjau.edu.cn (CHENG Yanhong); 943655041@qq.com (LI Yuqing)

Abstract: Cotton is an important cash crop in Xinjiang. However, due to the adjustments in planting structure, inadvisable pesticide application, and climate change, cotton is vulnerable to various pests, which affect its yield and quality. Therefore, the construction of an insect image dataset in cotton fields under complex conditions holds fundamental significance for the integrated pest management strategies in cotton fields. In this paper, we compiled a dataset of common insects in Xinjiang cotton fields under complex conditions for the identification and detection of insects in cotton fields. This dataset covers 13 species (classes) of common cotton field insects, comprising a total of 3,225 images, with a data size of 24 GB, all of which have been rigorously screened to ensure image quality. Moreover, each image in the dataset has been manually labeled to identify the cotton field insects present. In this way, we obtained both an image classification dataset and a target detection dataset, with a data size of 29.7 GB. The dataset has undergone rigorous evaluation using standard deep learning models to ensure its accessibility and reliability. The dataset can provide basic data for insect image classification and target detection in cotton fields, and has

important practical application value for facilitating agricultural field development, cotton pest control research and improving cotton yield.

Keywords: Cotton field insects; CottonInsect dataset; image classification; target detection

Dataset Profile

Title	A dataset of common insects in Xinjiang cotton fields: CottonInsect
Data corresponding author	CHENG Yanhong (cyh@xjau.edu.cn), LI Yuqing (943655041@qq.com)
Data authors	YANG Manxian, CHEN Yanhong, LI Yuqing, LI Yongke
Time range	2022
Geographical scope	Shaya County, Xinjiang
Data volume	29.7 GB
Data format	*.JPG、*.XML、*.JSON
Data service system	< https://doi.org/10.57760/sciencedb.j00001.00866 >
Sources of funding	National Key R&D Program of China (2022ZD0115805); Provincial Key S&T Program of Xinjiang (2022A02011-3).
Dataset composition	The dataset consists of three data files with a data size of 29.7 GB. (1) CottonInsect_image.zip contains the original image data, including 3,225 JPG images of 13 common cotton field insects, such as cotton bollworm, ladybug and alfalfa blind bug, at different periods, with a data size of 24 GB; (2) CottonInsect_classification.zip contains the image classification data. (2) CottonInsect_classification.zip contains the image classification data, including txt annotation files, JPG image files, and category files in two formats, class_indices.json and classes.txt, with a data size of 2.85 GB; (3) CottonInsect_detection.zip contains the target detection data, including .xml annotation file, JPG image file and classes.txt category file, with data size of 2.85 GB.