

# IP102: 用于害虫识别的大规模基准数据集\*

伍小平<sup>1</sup>, 展翅<sup>1</sup>, 来煜昆<sup>2</sup>, 程明明<sup>1</sup>, 杨巨峰<sup>1†</sup>

<sup>1</sup>南开大学      <sup>2</sup>卡迪夫大学

{xpwu95, chizhan\_nt}@163.com, LaiY4@cardiff.ac.uk, {cmm, yangjufeng}@nankai.edu.cn

## Abstract

害虫是影响农产品产量的主要因素之一。准确识别害虫有利于及时采取预防措施, 避免经济损失。然而, 现有的用于视觉分类任务的数据集主要包括常见物体, 如鲜花和狗等。这限制了深度学习技术在农业等特定领域的应用。本文中, 我们收集了一个大规模数据集用于害虫识别, 命名为IP102。具体来说, 该数据集包括超过75,000张图像, 分属102个类别, 数据呈自然的长尾分布。此外, 我们为约19,000张图像标注了边界框用于目标检测。IP102采用了分级分类法, 主要影响同一种特定农产品的害虫归属于同一父类别。此外, 我们在IP102数据集上进行了多项基准实验, 包括人为设计特征和基于深度学习特征的分类方法。实验结果表明, 该数据集存在类间和类内方差和数据不平衡的挑战。我们相信IP102将促进未来对实际害虫控制、细粒度视觉分类和不平衡学习领域的研究。我们已开源数据集和预训练模型, 详见 <https://github.com/xpwu95/IP102>。

## 1. 引言

众所周知, 害虫是对具有重要商业价值的农作物造成损害的主要原因 [8]。病虫害分类在农业病虫害预报中起着至关重要的作用, 对粮食安全和农业经济稳定至关重要 [10]。由于害虫种类繁多, 种间存在细微差异, 害虫识别严重依赖农业专家的专业知识 [1], 这意味着它即昂贵又耗时。随着机器学习和计算机视觉技术的发展, 自动化害虫识别吸引了越来越多的关注。

\*本文是CVPR 2019[43]论文的中文翻译版。

†Corresponding author

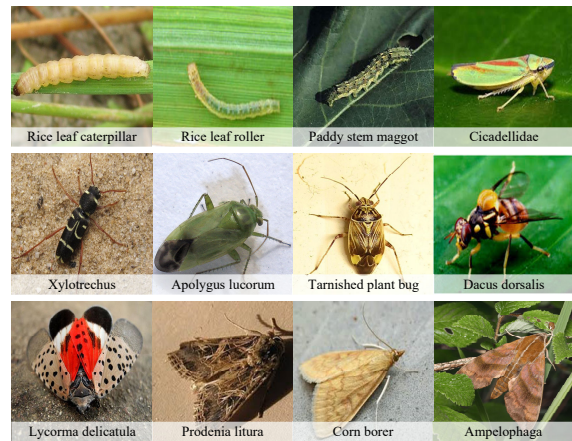


图 1. IP102数据集中的示例图像。每个图像属于不同种类的害虫。

此前关于虫害识别的大部分工作都可归属为传统的机器学习分类框架, 主要由两个模块组成: (1) 害虫图像的特征表示: 一系列人为设计的特征, 包括GIST [30]、SIFT [25]和SURF [3]等, 被用来表示整个图像。(2) 包括支持向量机 [4]和k-最近邻(KNN)分类器在内的机器学习分类器。这些基于特征的方法依赖于仔细选择的特征。如果从害虫图像中提取的特征不完整或错误, 后续分类器可能无法区分相似的害虫种类。

最近, 深度学习实现了鲁棒的特征学习, 并且在各种图像分类任务上取得了最先进的性能。众所周知, ImageNet大规模视觉识别挑战赛 (ILSVRC) 标志着深度学习快速发展的开始, 表明大规模图像数据集在推动深度学习进步方面发挥着关键作用。然而, 迄今为止用于害虫识别的深度学习方法还受限于小规模数据集, 其中仅包含很少的样本或害虫种类。同时, 现有公共数据集中的害虫图像大多是在受控的实验室环境

中收集的，不能很好地满足真实野外环境中害虫识别的要求。而且，虫害识别有不同于现有的目标分类或动物分类工作的特点 [41, 16, 27]。具体来说，不同害虫种类可能有高度相似的外观而相同种类害虫可能有不同存在形式，包括卵、幼虫、蛹和成虫，也即该任务存在显著的类内差异和种间相似度。

为了推进计算机视觉任务中害虫识别的研究，我们在这项工作中提出了IP102，这是一个新的大型害虫数据集。首先，我们使用常见的图像搜索引擎收集了超过300,000张图像，这些图像仅使用查询关键词进行弱标记。接下来，志愿者会检查每张图片，以确保其与害虫相关。然后，农业专家进一步检查图像并使用类别标签或边界框来标注图像。详细的数据集构建过程在后续章节中介绍。最后，我们的IP102数据集涵盖了102种常见农作物害虫，包括超过75,000张图像。与文献中目前可用的害虫数据集相比，IP102的规模要大得多，这有利于基于深度学习的方法。我们的数据集还涉及其他几个特征。首先，属于同一类别的图像可能会捕捉到同一类型害虫的不同生长形式。这种多样性是害虫数据集所独有的，但被以前的数据集忽略了。此外，类别不平衡是害虫的一种特性，因为某些物种更容易被观察到。我们的数据集满足数据分布不平衡的特征，就像在现实世界中一样。如图1所示为数据集中部分昆虫示例图像。

为了验证我们提出的数据集的应用价值，我们还展示了最先进的分类和目标检测算法在该数据集上的性能。结果表明，该数据集具有挑战性，并为研究创造了新的机会。

我们的贡献总结如下：

- 据我们所知，我们建立了最大规模的害虫识别数据集，包括害虫分类和检测。整个数据集对科研界开源。
- 我们在数据集上开展了大量基于CNNs和人为设计特征的实验，并以此性能为基础建立了可供未来研究对比的基准。我们还在IP102数据集的目标检测分支上测试了数个最先进的目标检测模型。我们希望这可以推动害虫识别的研究。

## 2. 相关工作

在本节中，我们将介绍害虫识别方法相关的工作

并回顾现有数据集。

### 2.1. 害虫识别

病虫害早期识别有助于防治病虫害，提高农产品质量和产量 [35]。

近年来，计算机视觉社区中出现了许多计算机辅助的害虫识别系统 [32, 2]。我们将它们分为两种类型：人为设计特征的方法和基于深度学习特征的方法。

人为设计的特征如SIFT [25]和HOG [5]等在低级特征表示上表现良好(如颜色、边缘和纹理)。在早期，基于人为设计特征的方法是害虫识别的主要解决方案。Samanta等人[35]利用基于相关性的特征选择和人工神经网络，在包含609个样本的数据集上识别8种茶树上的害虫。在 [28, 32]种，SVM分类器被用来识别叶子图像中的粉虱、蚜虫和蓟马。这些方法倾向于提取几个典型的人为设计特征来表示害虫，然后在类别很少的小数据集上进行评估。然而，现实世界中害虫种类繁多。设计用于识别多种害虫的特征提取器是低效且耗时的。此外，人为设计的特征缺乏表示高级语义信息的能力。

近来，深度学习技术广泛吸引了研究人员的关注 [18, 34, 24]。深度卷积神经网络(CNNs)如GoogleNet [39]和ResNet [13]在图像分类任务中展现了优异的性能。现有几个工作 [23, 2]成功地应用CNNs来解决害虫识别问题。Liu等人 [23]通过训练深度CNN对稻田害虫进行分类，他们的数据集涵盖12个种类，包括大约5,000个训练样例。Alfarisy等人 [2]使用CaffeNet [14]进行稻田害虫识别。此外，[7]基于仿生方法取得了与深度CNN(如VGGNet [36])相当的结果。然而评估数据集仅包含563个样本。总的来说，这些基于深度特征的工作缺乏足够的样本来优化CNN的大量超参数。为促进进一步的科学研究和实际应用，我们希望解决种类和样本有限的问题。因此我们收集了大规模IP102数据集，涵盖102个种类共包括75,222个样本。

### 2.2. 相关数据集

现有一些与虫害识别相关的小数据集，如 [35, 42, 7]。大部分数据集仅包括不超过1,000个样本。例如 [40]仅包括20个种类的200张样例用于稻田害虫分类。随后，出现了几个更大的数据集。Xie等人 [45]提供了一个包含1,440个样本涵盖24类大田作物常见害虫的

数据集。然而每个类别仅包含60个样例，这难以训练一个CNN模型。为解决该问题，[23, 44, 2]提出了几个含有超过4,500个样本的数据集，每个类别有100个样例。然而，仅有[44]中的数据集已开源。此外，数据集[44]中害虫图像的背景、物体姿势都高度相似，导致难以应对真实场景中的复杂情景。表2说明了这些相关数据集的详细信息。相反，我们的IP102涵盖102种实际应用中常见害虫种类，并且是在野外环境中构建的。此外，IP102数据集包括75,222张图像，平均每个种类有737个样本。

### 3. 我们的害虫数据集

#### 3.1. 数据收集与数据标注

我们通过以下四个阶段收集和标注IP102数据集：1) 分类系统建立，2) 图像收集，3) 数据初筛和4) 专业数据标注。

##### 3.1.1 分类系统建立

我们为IP102数据集建立了分层分类系统。我们邀请了几位农业专家，讨论了常见的害虫种类。最终确定了102类，构建了如图4所示的层次结构。基于其主要侵害的农作物，每类害虫归属于一个上层类别(后文使用父类表示)。换言之，每种害虫都是某个特定父类的从属类(后文使用子类表示)。例如，稻茎蛆虫主要危害水稻，水稻属于大田作物。因此，在IP102的分类系统中，子类稻茎蛆从属于父类水稻和大田作物。IP102数据集的详细结构将在后续数据集结构小节3.3介绍。

##### 3.1.2 图像收集

我们利用互联网作为收集图像的主要来源，该方法广泛用于构建ImageNet [6]和Microsoft COCO [21]等数据集。图像收集第一个步骤依赖于常见图像搜索引擎，包括Google、Flickr和Bing等。我们使用每个子类的英文名称和对应的同义词作为查询关键字。每个关键词仅保留前2,000个结果。然后我们从几个专业的农业和昆虫科学网站进行搜索。除了图像形式，我们还收集包含害虫内容的视频片段。在视频片段中，我们以每秒5帧的速度捕获图像。因此，我们为IP102数据集收集了超过300,000幅候选图像。

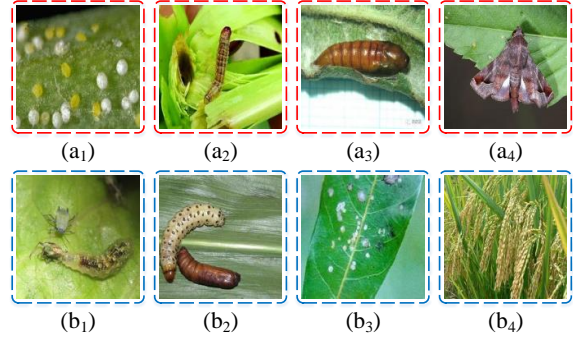


图2. 不同形式的害虫图像。红色虚线框表示不同形式的害虫，包括(a<sub>1</sub>)卵、(a<sub>2</sub>)幼虫(a<sub>3</sub>)蛹和(a<sub>4</sub>)成虫，以上都属于同一子类。被蓝色虚线框包围的图像被丢弃，因为其中没有害虫或有一种以上害虫。

##### 3.1.3 图像初筛

我们组织了6名志愿者来手动筛选候选图像。进行数据筛选前，他们主要接受三方面的训练内容，即1) 来自农业专家的虫害常识知识，2) IP102的分类系统和3) 害虫的不同形式。例如，图2展示了害虫的四种形式，包括卵、幼虫、蛹和成虫。即使处于生命周期的不同阶段，它们仍会对农产品造成不同程度的损害。在图像初筛阶段，志愿者删除图像中没有害虫或有超过一种害虫的图像，如图2所示。然后，我们将筛选后的图像转化为JPEG格式，并删除重复的和损坏的图像。最后，我们得到约120,000幅使用检索关键字弱标记的图像。每个父类标签根据IP102数据集的分类系统得到。

##### 3.1.4 专家数据标注

农业专家的数据标注是最重要的过程。在IP102的分类系统中，共有8种受虫害侵扰的农作物。对每种农作物，我们都会邀请主要研究相应作物的农业专家。因此，我们共邀请8名农业专家对初筛后的图像进行标注。为了便于标记，我们建立了一个问答(Q/A)系统。对Q/A系统交互页面上出现的图像，专家需回答图像属于哪种类别。专业数据标注由独立标注和协同标注组成。在独立标注阶段。每位农业专家只负责标注一种作物的父类。例如，对主要研究水稻的专家，他需要为父类是水稻的图像进行标注。在这种情况下，专家的Q/A系统种有15个类别选项。这些选项包括14种主

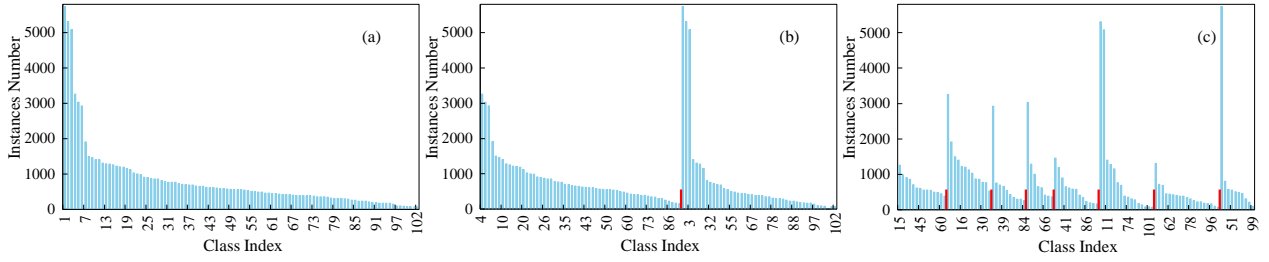


图 3. IP102数据集不同层级的样本数分布。红色校准尾部分别在子图(b)和子图(c)中拆分了2个父类和8个父类。



图 4. IP102数据集的分类系统。‘FC’和‘EC’分别代表大田作物和经济作物。在子类层级，仅展示35类图像。每个子类完整的列表可见于开源的IP102数据集。

要损害水稻作物的害虫类别和一个“其他”选项。“其他”代表该图像不属于这14种害虫类别、不包括害虫或包括超过一种害虫。下一阶段是协同标注。每位专家的Q/A系统中有103个类别选项(即，102个害虫类别和1个“其他”选项)。然后，这8名专家协同标注独立标注阶段标签为“其他”的图像。对每张图像，每位专家都需要进行标注，即在103个选项选择一个。最终的标注结果遵循一个严格的标准：一张图片只有在超过5位专家一致标注后才归属某个类别，否则将被删除。

图像中害虫位置的检测也非常重要。它将帮助农业专家或用户更好地找到害虫具体位置(尤其是某些并不明显的图像)。此外，现实世界中识别害虫相对更复杂。当目标害虫不显着时，杂乱的背景会误导分类器，并且图像中可能存在多个害虫样本需要分别进行

表 1. IP102数据集的训练/验证/测试(表示为Train/Val/Test)在不同类别层级上进行划分，并有对应的不平衡率(Imbalance Ratio, IR)。`'class'`代表父类下属的子类的数量。`'FC'`和`'EC'`分别代表大田作物和经济作物。

	<i>Super-Class</i>	Class	Train	Val	Test	IR
FC	Rice	14	5,043	843	2,531	6.4
	Corn	13	8,404	1,399	4,212	27.9
	Wheat	9	2,048	340	1,030	5.2
	Beet	8	2,649	441	1,330	15.4
	Alfalfa	13	6,230	1,037	3,123	10.7
EC	Vitis	16	10,525	1,752	5,274	74.8
	Citrus	19	4,356	725	2,192	17.6
	Mango	10	5,840	971	2,927	61.7
IP102	FC	57	24,602	4,098	12,341	39.4
	EC	45	20,721	3,448	10,393	80.8
	IP102	102	45,095	7,508	22,619	80.8

识别。场景中的病虫害防治措施需要准确的病虫害位置和每一种病虫害的类别。因此，通过空间信息的样本感知识别来有效地检测害虫可以减轻现实场景的复杂性。它还可以通过去除不相关的背景特征来提高分类性能。考虑到标注边界框的难度和成本，我们从每个类中随机选择部分图像构建目标检测任务的子集。专家按照Pascal VOC [9]的格式标注害虫的边界框。

### 3.2. 数据集划分

IP102数据集包括75,222幅图像，涵盖102种害虫，而其中最小的类别仅有71幅样本。为了在IP102上获得

表 2. 与现有害虫相关数据集的比较。'class'代表类别数量。'Avail'代表数据集是否开源。'Y'和'N'分别代表'yes'和'no'。'Avg'代表每个类的平均样本数。

Dataset	Year	Class	Avail	Sample	Avg
Samanta 等人 [35]	2012	8	N	609	76
Wang 等人 [42]	2012	9	Y	225	25
Venugoban 等人 [40]	2014	20	N	200	10
Xie 等人 [45]	2015	24	Y	1,440	60
Liu 等人 [23]	2016	12	N	5,136	428
Xie 等人 [44]	2018	40	Y	4,500	113
Deng 等人 [7]	2018	10	Y	563	56
Alfarisy 等人 [2]	2018	13	N	4,511	347
IP102	2019	102	Y	75,222	737

更可靠的测试结果，测试集上的每个类别都应该有足够的样本。因此我们大致按照6 : 1 : 3的比例进行划分。训练集、验证集和测试集均在子类级别进行划分。具体地说，IP102对分类任务划分出45,095幅图像用于训练、7,508幅图像用于验证和22,619幅图像用于测试。不同级别的详细划分可见于表 1。每个集合对应的图像列表可见于IP102数据集。对于目标检测任务，共有18,983幅标注的图像。我们将带有边界框标注的图像划分出15,178幅和3,798幅图像分别用于训练和测试。

### 3.3. 数据集结构

IP102数据集有分层结构，图 4展示了详细的分层系统。每个子类根据该害虫主要损害的作物划分至对应的父类。例如，子类朱砂叶螨(*tetranychus cinnbarinus*, TC)的父类为柑橘。8种农作物(例如大米、玉米和小麦)进一步归类为两个父类(即大田作物和经济作物)。举例来说，柑橘属于父类经济作物。此外，表 1展示了不同父类下子类的数据分布。

### 3.4. 与其他数据集的对比

在表 2中，我们将 IP102 与几个现有的与害虫识别任务相关的数据集进行了比较。相比于最大的数据集 [23, 44, 2]，我们的数据集包含14倍的样本

量。就种类多样性而言，最大的和第二的数据集仅包括40和8种类别。然而，现实世界中有大量的害虫，而我们的IP102包括102种类别。就每个类别的平均样本数而言，IP102对比数据集至少多309个样本。除了统计上的区别，当前仅有一半数据集是开源的，且只有[44]的规模较大。受限于此，大多数现有的与害虫相关的数据集(例如 [40, 45, 7])很难应用到实际应用中。

### 3.5. 多样性和难度

生命周期不同阶段的害虫对农产品的危害程度不同。因此，我们在数据收集和标注期间保留包含所有生命周期的图像。图 2(a<sub>1</sub>-a<sub>4</sub>)展示了IP102中害虫不同形式的示例图像，包括卵、幼虫、蛹和成虫。对于分类模型，很难将它们归入同一类别，因为很难提取判别特征。除了生物多样性，数据分布的不平衡也不容忽视。如图 3所示，三个子图显示了数据集在不同级别的不平衡分布，其中(a)、(b)和(c)分别展示了102个子类、2个父类和8个父类的实例数量分布。具体地说，基于IP102数据集的分层标签系统，102个子类根据害虫主要侵害的农作物(如水稻和玉米)划分为8个父类，进而被划分为2个父类(即大田作物和经济作物)。不同层级的不平衡分布给不平衡学习领域和分层标签的使用带来了挑战。表 1也展示了IP102在大部分父类层级有较高的不平衡率(IR)(即，高于9IR [12])。不平衡的数据会导致分类模型学习到具有相对较多训练样本的类别的偏差。

## 4. 实验评估

特征的选择通常在图像识别中起着重要的作用。为了全面评估IP102数据集，我们首先分别利用人为设计的特征和深度特学习征评估分类性能。随后，我们在IP102的子集上评估了几个目标检测框架。

### 4.1. 实验设置

SVM分类器根据LIBLINEAR [11]采用一对多的策略进行训练。KNN分类器的最近邻个数设为5。训练深度学习网络时，我们采用64的mini-batch使用随机梯度下降优化器微调所有层。学习率设为0.01，每40轮使用因子0.1对学习率进行衰减。权重衰减和动量参数分别设为0.0005和0.9。为避免过拟合，我们使用了dropou [37]，设为0.3。我们保持这些深度学习模型的基本架

表 3. SVM和KNN分类器在IP102数据集上不同评估指标下的分类性能。特征表示分别使用人为设计的特征和深度学习特征。

#	Methods	SVM						KNN					
		Pre	Rec	F1	GM	M <sub>AUC</sub>	Acc	Pre	Rec	F1	GM	M <sub>AUC</sub>	Acc
Handcrafted Feature	CH	9.7	3.2	2.5	0.3	12.0	12.9	18.2	14.2	15.0	8.3	16.8	15.8
	Gabor [29]	8.5	3.9	3.6	0.5	12.1	14.2	22.0	14.9	16.5	9.1	20.0	19.2
	GIST [30]	12.2	3.8	3.8	0.6	12.1	13.1	19.1	15.1	15.4	9.2	19.2	18.2
	SIFT [25]	25.1	6.3	6.8	1.0	19.9	18.1	19.4	10.3	12.1	5.6	15.9	13.1
	SURF [3]	28.2	7.3	8.3	1.5	21.2	19.5	21.3	11.5	13.4	7.1	17.5	14.7
	LCH [38]	7.2	5.0	4.7	0.9	11.1	13.1	21.6	14.7	16.1	8.3	19.0	16.8
Deep Feature	Alexnet [17]	41.5	16.4	21.0	9.3	32.5	28.3	36.7	32.4	33.5	23.9	41.0	40.7
	GoogleNet [39]	45.8	25.8	30.4	16.0	41.9	40.5	36.8	31.7	33.0	23.3	41.6	40.7
	VGGNet [36]	43.4	37.6	39.1	28.3	48.1	48.7	41.9	37.8	39.0	29.8	47.6	47.1
	ResNet [13]	43.6	39.1	40.6	31.0	48.7	49.5	43.7	39.1	40.5	30.7	48.2	49.4

构不变，仅将最后的全连接层由1,000改为我们目标的类别数。输入图像尺寸固定为224 × 224。基于深度学习特征的实验均使用PyTorch [31]实现，在一块12GB显存的NVIDIA Titan X GPU上进行。

#### 4.2. 评价标准

IP102有不平衡的类别分布。我们采用了几个分类任务常用的综合指标，包括准确率、召回率、F-measure、G-mean和M<sub>AUC</sub>。准确率(表示为Pre)描述分类器不将负样本标记为正样本的能力。召回率(表示为Rec)表示找到一个特定类别的所有正样本的能力。F1权衡了准确率和召回率。G-mean(表示为GM)评估类别敏感性并指出在多数类和少数类上的平衡分类性能。micro average scheme M<sub>AUC</sub> [15]定义为曲线下度量下的面积。对于目标检测任务，我们使用平均准确率(Average Precision, AP)(IoU=[.50:.05:.95])、AP<sup>.50</sup> (IoU=.50)和AP<sup>.75</sup> (IoU=.75)作为评价标准。IoU被定义为检测到的框和真实数据之间的交集。IoU的阈值越大，检测难度越大。

#### 4.3. 基于人为设计特征的分类结果

我们从IP102数据集中提取几个人为设计的纹理和颜色特征，包括颜色直方图(CH)、LCH [38]、Gabor

表 4. 不同深度学习模型的性能。‘st’代表从头开始训练。

Method	F1	GM	Acc	F1 <sup>st</sup>	GM <sup>st</sup>	Acc <sup>st</sup>
AlexNet [17]	34.1	27.0	41.8	29.1	22.2	35.3
GoogleNet [39]	32.7	21.3	43.5	27.0	11.3	40.2
VGGNet [36]	38.7	30.9	48.2	33.3	25.5	41.4
ResNet [13]	40.1	31.5	49.4	29.6	22.2	35.7

[29]、GIST [30]、SIFT [25]和SURF [3]。然后，我们利用SVM和KNN分类器基于人为设计的特征构建基线模型。

表 3展示了人为设计的特征的性能。我们可以看到相比于纹理特征(Gabor [29])，颜色(CH)特征在大多数评价标准中表现较差。这表明在野外害虫检测中，纹理特征起着更重要的作用。如图 1所示，大面积单调的背景色使得模型难以通过颜色特征区分害虫。最好的人为设计的特征仅取得了大约19.5%的准确率，该模型使用SURF [3]特征和SVM分类器。主要原因是这些人设计的特征既不能捕捉到与害虫有关的综合信息，

也不能消除真实环境中害虫图像的噪声。此外，许多不同的害虫具有相似的外观，但传统的人为设计的特征不足以捕捉细微的差异。IP102和原有的小数据集 [19, 45] 上准确率的巨大差异也表明IP102有更高的识别难度。

#### 4.4. 基于深度学习特征的分类结果

深度学习特征被证明对图像分类是有效的。在本节中，我们将在IP102数据集上评价最先进的深度卷积神经网络的性能，包括AlexNet [17]、GoogleNet [39]、VGGNet-16 (VGGNet) [36]和ResNet-50 (ResNet) [13]。

所有的网络均在ImageNet [6]上预训练然后在IP102数据集上进行微调。我们通过移除模型架构中的最后一层来从CNN中提取深度学习特征。随后，我们利用这些深度学习特征来训练SVM和KNN分类器。表 3展示了深层特征的分类性能。与其他三个模型相比，ResNet在大部分指标上表现最好。因此ResNet可作为IP102最好的特征表示，即使它的特征维度(2,048)比VGGNet低(4,096)。此外，深层特征表示总体上超过人为设计的特征。这展示了深度学习模型的特征学习能力。然后，我们进一步看到KNN分类器总体上比SVM分类器性能更好。特别是对AlexNet的特征，KNN模型结果在大部分指标上超过SVM模型。基于AlexNet模型特征，KNN分类器取得了40.7%的准确率而SVM分类器仅有28.3%准确率。此外，SVM分类器仅取得了16.4%召回率和9.3%G-mean的较差性能表现。这说明AlexNet的深层特征具有低敏感性。

表 4显示了深度模型在不同评估指标上的softmax分类性能。请注意，ResNet在所有指标上都取得了最佳结果。而49.4%准确率和31.5%G-mean之间的巨大差异显示了我们IP102数据集的高度不平衡性。分类模型偏向于那些具有大量样本的类别。此外，最高的准确率仅为49.4%表明IP102数据集具有挑战性。我们还从头开始训练深度学习模型，即不在ImageNet数据集上进行预训练。与微调的模型相比结果差得多，这是由于这些深度学习模型具有大量超参数，并且很容易在训练样本较少的类上过拟合。

#### 4.5. 目标检测结果

我们在IP102数据集上评估了几种最先进的目标检测方法。两阶段方法包括Faster R-CNN (FRCN)

表 5. 具有不同层次标签的分类性能。每一行代表每个作物对应的子类上的结果。

<i>Super-Class</i>		Pre	Rec	F1	GM	M <sub>AUC</sub>	Acc
FC	Rice	31.5	30.0	30.4	28.3	32.3	32.1
	Corn	55.1	54.4	54.6	50.3	61.9	62.2
	Wheat	37.5	34.5	35.5	29.3	52.1	53.0
	Beet	51.6	49.5	50.4	45.3	62.0	62.2
	Alfalfa	42.1	41.2	41.4	38.1	46.2	46.4
EC	Vitis	78.2	76.3	77.1	74.9	86.8	86.7
	Citrus	69.6	68.5	68.8	65.2	76.6	76.6
	Mango	75.8	74.7	75.1	72.3	89.0	89.0

表 6. 不同IoU阈值下目标检测方法的平均精度性能。

Method	Backbone	AP	AP <sup>.50</sup>	AP <sup>.75</sup>
FRCNN [34]	VGG-16	21.05	47.87	15.23
FPN [20]	ResNet-50	28.10	54.93	23.30
SSD300 [22]	VGG-16	21.49	47.21	16.57
RefineDet [46]	VGG-16	22.84	49.01	16.82
YOLOv3 [33]	DarkNet-53	25.67	50.64	21.79

[34]和FPN [20](使用FRCN作为目标检测框架的主干网络)。他们通过首先在特征图上滑动窗口来扫描潜在目标，然后对它们进行分类并回归相应的框坐标来检测目标。一阶段的方法包括SSD300 [22]、RefineDet [46]和YOLOv3 [33]，这些方法直接回归每个目标的种类和位置。表 6的检测性能展示了基于区域候选方法的两阶段检测器(FPN)相对于一阶段统一方法(SSD300、RefineDet和YOLOv3)的优越性。我们观察到，在深度网络中结合来自多层的特征图(FPN和YOLOv3)对于目标尺寸的多尺度适应是有效的。

#### 4.6. 进一步分析

在表 5中，我们进一步评价深度模型在每个父类上的性能。在我们的IP102数据集的分层结构上，每个子类都从属于一个父类。每个父类是IP102的一个子集，涵盖102种害虫的一部分。例如，对于父类水稻，我们的目标是将IP102子集中的某个示例归类到14个种

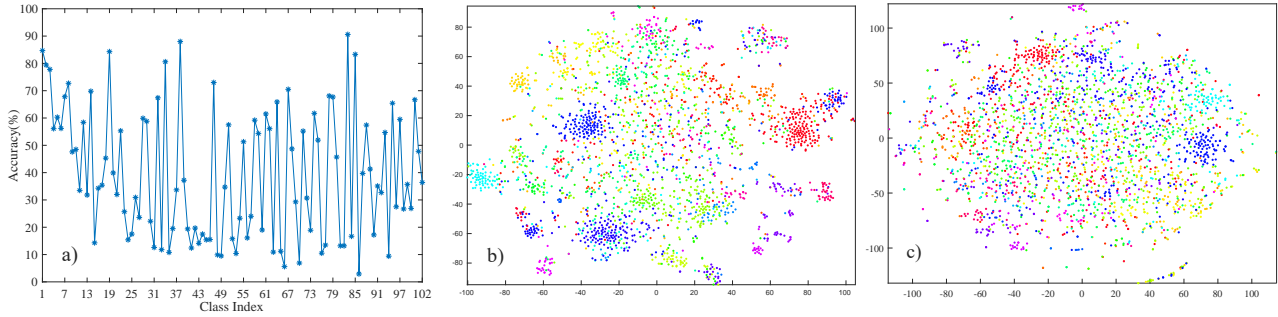


图 5. (a) ResNet在IP102中每个子类的top-1精度。(b)和(c)2D t-SNE [26]在IP102上特征嵌入的可视化。(b) ResNet在ImageNet预训练后在IP102上微调的结果。(c) ResNet在IP102上从头训起。

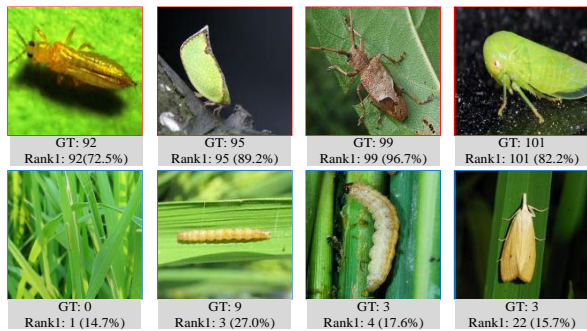


图 6. ResNet分类器在“芒果”(上方)和“水稻”(下方)子类上的结果示例。上行的图像是正确分类的，下行的图像是错误分类的。

类中的一个。父类的详细的类比分布可见于表 1。我们选择ResNet [13]为基础CNN模型，该模型上一小节在IP102上性能表现最佳。我们还展示基于不平衡学习评价指标的分类结果，因为IP102的样本数在父类层级仍是不平衡的，如图 3所示。从表 5可见，模型性能在8个不同父类上是不同的。此外，性能最好的“芒果”和最差的“水稻”之间精度的差距是56.9%。如图 6所示为在两个父类上的分类结果。我们可以看到“芒果”上的害虫在形状、颜色和背景等方面具有易判别的特征。至于“水稻”，图像由于三方面原因容易被误分类。首先，物体和背景之间的颜色相似。害虫在大量背景信息下难以区分。其次，类内差异较大，如图 2所示。这些害虫通常会在作物的整个生命周期中不同程度地损害作物，并且很难对它们进行正确分类，尤其是在幼虫时期。第三，不同类别害虫往往相似，如亚洲水稻螟和黄螟。因此，如图 7所示虫害识别的困难也给检测任务带来了挑战。即使目标被准确地检测到，

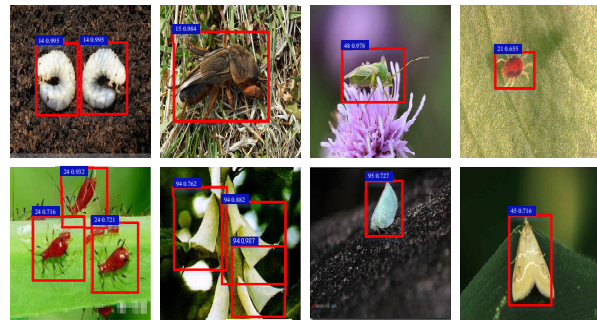


图 7. IP102数据集上的目标检测的示例结果。顶行显示正确检测到的图像。底行显示了一些失败案例，例如右边两张正确检测到但错误分类的图像。

但它也可能被错误分类。

此外，如图 5(a)所示，我们展示了IP102每个子类上ResNet [13]的分类准确率结果。图 5(b)和图 5(c)展示了IP102基于t-SNE [26]的特征嵌入的可视化结果。可见，基于在ImageNet [17]上预训练的模型，ResNet在特征空间中能更好地提取特征来区分不同的害虫。

## 5. 结论

在本工作中，我们收集了一个用于害虫识别的大规模数据集，命名为IP102，共包括超过75,000幅图像，涵盖102个物种。与现有数据集相比，IP102符合真实环境中害虫分布的几个特点(例如多样性和类别不平衡)。同时，我们还在我们的数据集上评估了一些最先进的识别方法。结果表明，目前的人为设计特征方法和深度学习特征方法还不能很好地处理害虫识别。我们希望这项工作能推进对几个基本问题(例如细粒度视觉分类和不平衡学习等)以及常见目标分类和检测任务的未

来研究。

## 致谢

This work was supported by the NSFC (No. 61876094, 61620106008, 61572264), Natural Science Foundation of Tianjin, China (No. 18JCYBJC15400, 18ZXZNGX00110, 17JCJQC43700), the National Youth Talent Support Program, and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

## 参考文献

- [1] H Al Hiary, S Bani Ahmad, M Reyalat, M Braik, and Z Al-rahmaneh. Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications*, 17(1):31–38, 2011.
- [2] Ahmad Arib Alfarisy, Quan Chen, and Minyi Guo. Deep learning based classification for paddy pests & diseases recognition. In *ICMAI*, 2018.
- [3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. In *ECCV*, 2006.
- [4] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- [5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.
- [7] Limiao Deng, Yanjiang Wang, Zhongzhi Han, and Renshi Yu. Research on insect pest image detection and recognition based on bio-inspired methods. *Biosystems Engineering*, 169:139–148, 2018.
- [8] Juan J Estruch, Nadine B Carozzi, Nalini Desai, Nicholas B Duck, Gregory W Warren, and Michael G Koziel. Transgenic plants: An emerging approach to pest control. *Nature Biotechnology*, 15(2):137, 1997.
- [9] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.
- [10] Fina Faithpraise, Philip Birch, Rupert Young, J Obu, Bassey Faithpraise, and Chris Chatwin. Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters. *International Journal of Advanced Biotechnology and Research*, 4(2):189–199, 2013.
- [11] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9(Aug):1871–1874, 2008.
- [12] Alberto Fernández, Salvador García, María José del Jesus, and Francisco Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [14] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM MM*, 2014.
- [15] Qi Kang, Lei Shi, MengChu Zhou, XueSong Wang, QiDi Wu, and Zhi Wei. A distance-based weighted undersampling scheme for support vector machines and its application to imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems*, 29(9):4152–4165, 2018.
- [16] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *ICCV Workshop*, 2013.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.
- [19] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [20] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.
- [21] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014.
- [22] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single shot multibox detector. In *ECCV*, 2016.

- [23] Ziyi Liu, Junfeng Gao, Guoguo Yang, Huan Zhang, and Yong He. Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific Reports*, 6:20410, 2016.
- [24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [25] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [27] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.
- [28] M Manoja and J Rajalakshmi. Early detection of pest on leaves using support vector machine. *International Journal of Electrical and Electronics Research*, 2(4):187–194, 2014.
- [29] Rajiv Mehrotra, Kameswara Rao Namuduri, and Nagarajan Ranganathan. Gabor filter-based edge detection. *Pattern Recognition*, 25(12):1479–1494, 1992.
- [30] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.
- [31] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS Workshop*, 2017.
- [32] R Uma Rani and P Amsini. Pest identification in leaf images using SVM classifier. *International Journal of Computational Intelligence and Informatics*, 6(1):30–41, 2016.
- [33] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.
- [35] RK Samanta and Indrajit Ghosh. Tea insect pests classification based on artificial neural networks. *International Journal of Computer Engineering Science*, 2(6):336, 2012.
- [36] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [37] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [38] Michael J Swain and Dana H Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
- [39] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, 2015.
- [40] Kanesh Venugoban and Amirthalingam Ramanan. Image classification of paddy field insect pests using gradient-based features. *International Journal of Machine Learning and Computing*, 4(1):1–5, 2014.
- [41] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.
- [42] Jiangning Wang, Congtian Lin, Liqiang Ji, and Aiping Liang. A new automatic identification system of insect images at the order level. *Knowledge-Based Systems*, 33:102–110, 2012.
- [43] Xiaoping Wu, Chi Zhan, Yukun Lai, Ming-Ming Cheng, and Jufeng Yang. Ip102: A large-scale benchmark dataset for insect pest recognition. In *IEEE CVPR*, 2019.
- [44] Chengjun Xie, Rujing Wang, Jie Zhang, Peng Chen, Wei Dong, Rui Li, Tianjiao Chen, and Hongbo Chen. Multi-level learning features for automatic classification of field crop pests. *Computers and Electronics in Agriculture*, 152:233–241, 2018.
- [45] Chengjun Xie, Jie Zhang, Rui Li, Jinyan Li, Peilin Hong, Junfeng Xia, and Peng Chen. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. *Computers and Electronics in Agriculture*, 119:123–132, 2015.
- [46] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, and Stan Z Li. Single-shot refinement neural network for object detection. In *CVPR*, 2018.