

基于属性特征知识图谱的细粒度葡萄园害虫识别

郑增威^{1,2}, 李彦臻^{1,2,3}, 刘 益^{1,2,3}, 林中琦^{1,2}, 向正哲^{1,2}, 何梦竹^{1,2}, 孙 霖^{1,2*}

(1. 浙大城市学院计算机与计算科学学院, 杭州 310015; 2. 智能植物工厂浙江省工程实验室, 杭州 310015;

3. 浙江大学计算机科学与技术学院, 杭州 310027)

摘 要: 针对葡萄园害虫识别过程中害虫种类繁多、形态属性复杂、实体间深层次关联关系差等因素导致的识别准确率不够高的问题, 该研究提出一种基于属性特征知识图谱的细粒度葡萄园害虫识别模型。该模型以视觉编码器作为提取图像高层表征的骨干网络, 并结合知识图谱具有在描述害虫实体属性特征和害虫实体间关联方面的优势, 将知识图谱所提供的细粒度属性特征和害虫实体关联特征信息用于葡萄园害虫分类研究。该方法在 2 个数据集上进行了测试: 一是 GP21 数据集, 该数据集由公开大规模害虫数据集 IP102 中 21 类葡萄园相关害虫类别图像构成; 二是 GP8 数据集, 该数据集包含从农业生产基地实地采集并手工标注的 8 种葡萄园害虫。试验结果表明, 该文所提模型性能明显优于普遍通用预训练网络模型, 分类准确率在 GP21 数据集和 GP8 数据集上分别达到了 91.21% 和 95.03%, 相比于仅使用视觉编码器分别增加 1.64 和 1.17 个百分点。这证明属性特征知识图谱的引入能够辅助视觉编码器捕获细粒度更高的葡萄园害虫特征信息, 有效解决了葡萄园害虫识别中的精度不够高的问题。

关键词: 知识图谱; 图像识别; 深度学习; 葡萄; 害虫; 关联特征; 图像分类

doi: 10.11975/j.issn.1002-6819.202302122

中图分类号: TP391; S24

文献标志码: A

文章编号: 1002-6819(2023)-11-0146-09

郑增威, 李彦臻, 刘益, 等. 基于属性特征知识图谱的细粒度葡萄园害虫识别[J]. 农业工程学报, 2023, 39(11): 146-154. doi: 10.11975/j.issn.1002-6819.202302122 <http://www.tcsae.org>

ZHENG Zengwei, LI Yanzhen, LIU Yi, et al. Identifying fine-grained vineyard pest using attribute knowledge graph[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(11): 146-154. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202302122 <http://www.tcsae.org>

0 引 言

中国葡萄种植面积大、产量高, 是主要的水果之一^[1]。在葡萄园作物栽培管理过程中, 虫害是制约其品质改良和产量增长的主要因素, 正确诊断作物害虫类别是科学防治的必要前提。随着人工智能技术的发展, 利用计算机视觉技术对农作物害虫图像进行自动识别和诊断已经成为国内外一大研究热点。

深度学习作为其中一项关键技术, 弥补了传统图像分类方法特征提取能力差、效率低等缺陷, 被广泛应用于农作物害虫的识别与诊断中。AYAN 等^[2]集成 Inception-V3^[3]、Xception^[4] 和 MobileNet^[5], 提出了 GAEnsemble 模型, 通过遗传算法确定预训练模型权重, 该模型具备良好的稳定性和作物害虫识别准确性。苏仕芳等^[6]利用 ImageNet-21k 预训练 VGG-16^[7-8], 并在葡萄叶片病害数据集上微调, 通过数据增强技术结合迁移学习的训练方式, 该模型对褐斑病、黑腐病等葡萄叶常见病害的识别准确率均达到 90% 以上。孙钰等^[9]将 VGG-16 用于无人机上对森林害虫进行检测。梁万杰等^[10]提出一种卷积神经网络模型的水稻害虫识别方法。TURKOGLU 等^[11]提

出了一种基于 LSTM (long short-term memory)^[12] 的卷积神经网络模型 MLP-CNNs, 实现了苹果害虫的精确识别。王林惠等^[13]优选 MoblieNet 作为害虫图像特征提取网络对柑橘害虫进行准确识别。

虽然上述研究在一定程度上解决了农作物害虫的识别问题, 但是目标物种仅仅局限于小麦、苹果、柑橘等, 对于葡萄园中的害虫识别仍然存在识别精度不足、针对性不强等问题。针对此, 找到一种能够更加精准识别葡萄园害虫种类的模型和方法已成为葡萄园提高产量和品质的迫切需求。

知识图谱 (knowledge graph, KG) 作为一种能够精确描述领域内复杂知识的数据模型, 被广泛应用于智能搜索、个性化推荐等领域。在农业领域知识图谱方面, 于何龙等^[14]构建了包含害虫种类、病害类别、病斑颜色、发病阶段等信息的水稻病虫害本体, 并开发了知识图谱与确定性因子模型相结合的水稻病虫害知识推理和智能诊断系统, 该系统支持领域知识检索, 能够为水稻作物生产实践提供指导。戈为溪等^[15]提出了一种基于知识图谱和案例推理的水稻精准施肥推荐模型, 该施肥推荐模型能够输出详细的施肥方案和精确的施肥量。吴赛赛等^[16]提出了一种基于深度学习的实体-关系抽取模型, 成功实现多源数据中有用知识的准确抽取, 所构建的病虫害知识图谱能够为其他农作物相关下游任务提供高质量的知识基础。郑泳智等^[17]对荔枝和龙眼害虫知识进行

收稿日期: 2023-02-21 修订日期: 2023-05-20

基金项目: 浙江省自然科学基金 (LGN21F020002, LGN22F020002); 浙江省重点研发计划 (2022C03037)

*通信作者: 孙霖, 副教授, 研究方向为深度学习。Email: sunl@hzcu.edu.cn

了研究,并基于知识图谱开发了智能问答系统,为害虫诊断和防治提供指导。知识图谱技术的引入实现了数据信息的有效利用,同时提高了专业知识在农业领域的应用能力。

然而,现有研究中知识图谱主要用于农业领域的知识检索和智能诊断,很少涉及将知识图谱与计算机视觉技术相结合以提高害虫图像识别精度的研究。同时对于覆盖果蔬作物害虫知识图谱的深入研究较少,尤其是针对葡萄园害虫这一垂直领域的系统仍有待开发。为解决上述问题,本研究提出了一种基于属性特征知识图谱的细粒度葡萄园害虫识别方法 ACKGViT (attribute characteristics knowledge graph enhanced vision transformer),利用从农业科学等网站收集的大量葡萄害虫相关知识语料,由领域专家指导构建了害虫属性特征知识图谱,用于增强视觉编码器对害虫图像细粒度特征信息的感知能力,从而实现葡萄园害虫精确识别。该方法可作为葡萄园害虫信息检索、智能推荐等下游应用的知识库基础,可以有效应用于作物品种选择、害虫防控等农业生产方面。

1 葡萄园害虫知识图谱构建

1.1 葡萄园害虫知识获取与预处理

由于葡萄园害虫领域暂时没有公开可用的资料库和数据库能够直接作为试验材料,本研究涉及的葡萄害虫相关知识和数据通过专业农业网站、昆虫科学网站、维基百科、百度百科等知识库获取。利用 Scrapy 框架,共爬取包括绿盲蝽、大青叶蝉、葡萄二星叶蝉等 21 种葡萄园常见害虫在内的数据 1 264 条。通过正则表达式等数据清洗方式,将爬取的数据转化为规范化的葡萄害虫语料。清洗后的数据包括半结构化和非结构化两种类型:

1) 半结构化数据。数据包括例如“形态特征”“生

活习性”等目录或标题在内的半结构化信息,通过构造相应规则直接进行实体抽取。

2) 非结构化数据。将整段、整篇文本作为属性的数据,采用深度学习模型 Bi-LSTM-CRF^[18]对该类型的数据进行实体抽取,以将实体-属性抽取转换为序列标注任务。

1.2 葡萄园害虫属性特征知识图谱建模

知识图谱构建包括“自底向上”和“自顶向下”两种方式。自底向上是一种数据驱动方式,适用于开放领域的知识图谱构建;而自顶向下的构建方式是指在构建知识图谱之前,预先定义本体和模式。根据本研究是垂直于“葡萄园”种植行业的研究这一特性,采用自顶向下的方式构建葡萄园害虫属性特征知识图谱。将实例集合定义为<害虫类别,关系,属性特征>的三元组,并选用图数据库 Neo4j 作为知识存储方式。

为了更加精准地描述葡萄园害虫实体、属性特征以及实体关联信息,本文确定了常见葡萄园害虫的种类、生命周期、分布区域,以及它们对葡萄产量和品质的影响程度,从而更全面地理解葡萄园害虫问题。此外,还研究了针对不同害虫种类的防治策略和方法,在这项研究中,对害虫的属性特征进行了细致的分析,包括颜色特征、纹理特征和轮廓特征。这些特征有助于区分不同种类的害虫,为识别模型提供了丰富的信息,使得模型识别准确率得以上升。除此之外,还深入研究了害虫的生物学特性,如生活习性、繁殖方式、天敌关系等。通过对害虫的分类体系的了解,可以为知识图谱构建提供更为精细的分类依据。在收集了大量葡萄害虫相关知识语料后,本研究构建出了一个能够精确反映葡萄园害虫实体属性和关联信息知识图谱,最终知识图谱部分搭建效果展示如图 1 所示。

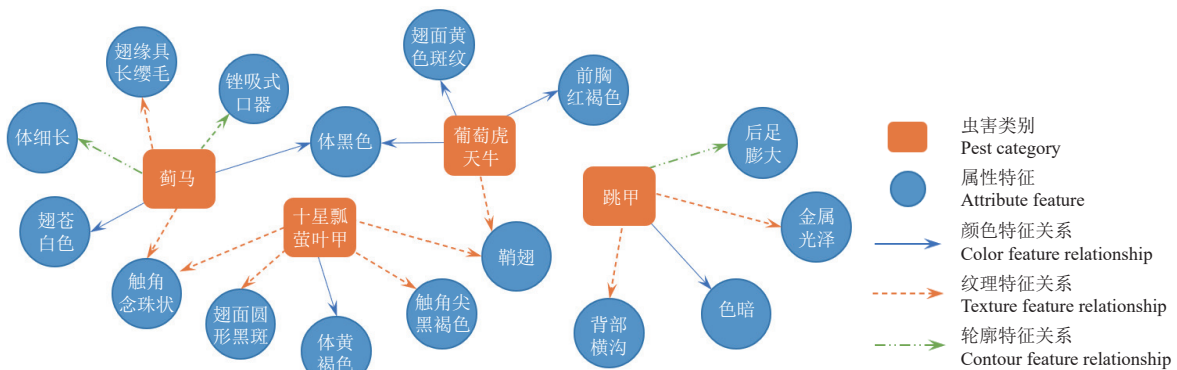


图 1 知识存储示例

Fig.1 Example of storing knowledge

2 ACKGViT 模型介绍

如图 2 所示,本文所提出的 ACKGViT 模型采用双分支结构,构建包括:基于属性特征知识图谱(attribute characteristics knowledge graph, ACKG)的害虫属性特征及关联特征提取分支,和基于深度学习网络 ViT^[19](vision transformer, ViT)的害虫图像高层语义表征提取分支,

结合两个分支获取的特征向量,用于葡萄害虫图像分类研究。ACKG 通过图卷积网络实现映射。其中值得说明的是,在仅使用单一知识图谱进行训练时,网络初始化参数可能会对最终优化效果造成干扰,因此,为了让知识图谱学习到的关系更符合实际领域知识,本文还引入了传统特征信息,作为知识图谱特征学习的目标,用于优化训练。

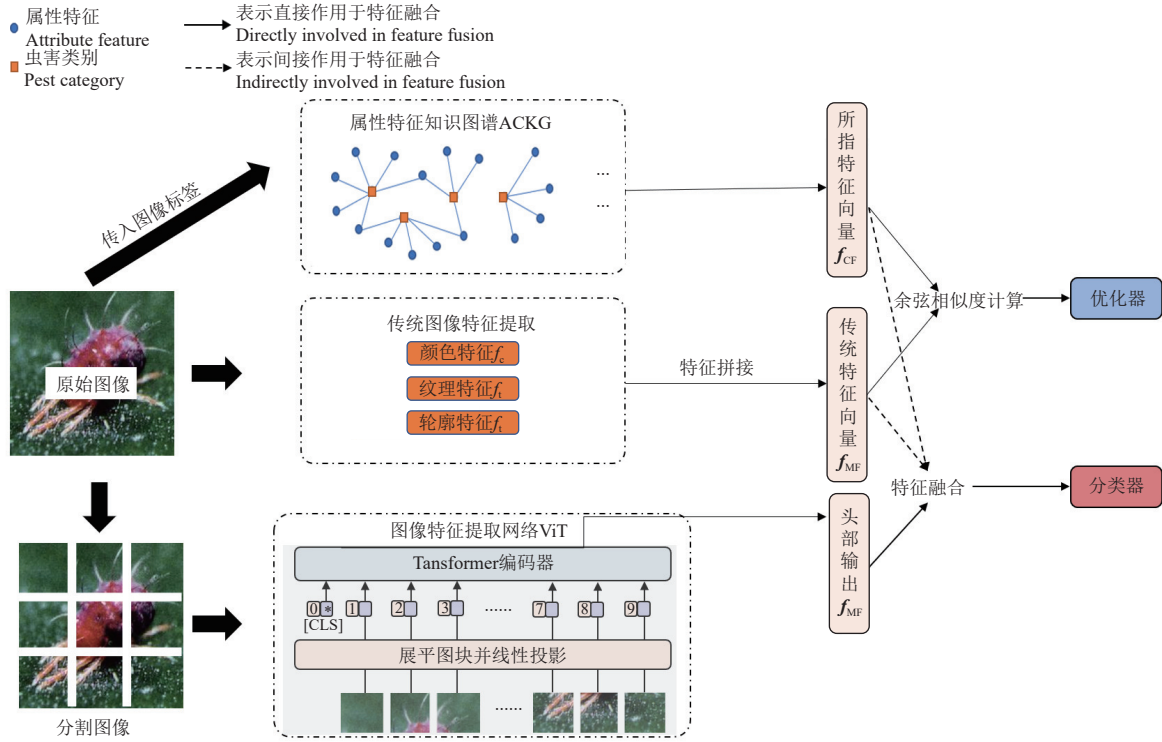


图2 ACKGViT 模型示意图

Fig.2 Illustration of the ACKGViT(attribute characteristics knowledge graph enhanced vision transformer) model

2.1 细粒度害虫属性特征提取网络

2.1.1 传统特征提取模块

ACKGViT 组合颜色特征、纹理特征和轮廓特征作为传统手工特征向量。对于输入的每一张图像，将颜色矩作为图像颜色表征，定义为 f_c ；使用局部二值模式（local binary patterns）^[20] 和灰度共生矩阵（gray-level co-occurrence matrix）^[21] 分别提取图像纹理特征并进行拼接（Concatenate），所得向量作为全局纹理特征 f_t ；轮廓特征 f_o 基于 Canny 边缘检测算法进行提取。最终传统特征 f_{MF} 由上述 3 种特征通过拼接操作得到：

$$f_{MF} = \text{Concat}(f_c, f_t, f_o) \quad (1)$$

2.1.2 属性特征知识图谱 ACKG

本研究利用 GAT（graph attention network, GAT）^[22] 网络将葡萄园病害知识图谱 ACKG 映射为可以进行训练的神经网络模块。GAT 是由 VELICKOVIC P 等^[22] 在 2018 年提出的一种图卷积网络模型，由堆叠图注意力（Attention）层构成，利用自注意力机制（Self-attention）聚合邻居节点信息，通过在训练过程中自适应学习邻居权值的方式，使得模型具备良好的可解释性和准确性。知识图谱中的点包括害虫类别 N_l 和害虫特征 N_f 两种类型：

$$N_l = \{N_{l_0}, N_{l_1}, \dots, N_{l_n}\} \quad (2)$$

$$N_f = \{N_{f_0}, N_{f_1}, \dots, N_{f_m}\} \quad (3)$$

式中 n 和 m 分别代表害虫类别总数和图谱中所有属性节点的数量。 l_0 代表害虫类别的第一个索引。式（2）表示知识图谱中害虫类别节点的集合，其中 N_{l_0} 是第一个害虫类别节点， N_{l_1} 是第二个害虫类别节点，依次类推，直到

N_{l_n} 为第 n 个害虫类别节点。同样，式（3）表示知识图谱中害虫特征节点的 N_f 集合，包括 m 个特征节点。

ACKG 的训练包括两个步骤：首先，基于输入害虫图像的标签（Label）在知识图谱中进行索引，得到该类害虫在知识图谱中对应节点的属性特征向量，记为 f_{CF} ；然后，与手工特征向量 f_{MF} 进行余弦相似度计算，得到相似度损失 \mathcal{L}_s ：

$$\mathcal{L}_s = \frac{\sum_{i=1}^k f_{MF_i} \cdot f_{CF_i}}{\sqrt{\sum_{i=1}^k (f_{MF_i})^2} \cdot \sqrt{\sum_{i=1}^k (f_{CF_i})^2}} \quad (4)$$

式中， k 代表特征向量的维度，与害虫类别总数相等。 f_{MF_i} 代表手工特征向量 f_{MF} 的第 i 个分量， f_{CF_i} 代表属性特征向量 f_{CF} 的第 i 个分量。式（4）计算的是相似度损失 \mathcal{L}_s ，其目的是衡量手工特征向量 f_{MF} 与知识图谱中属性特征向量 f_{CF} 之间的相似度。这里使用了余弦相似度作为相似度的度量方法，因为它能够很好地捕捉两个向量之间的角度关系，而不受长度影响。式（4）的分子部分计算了 f_{MF} 和 f_{CF} 之间的点积，而分母部分计算了各自的模长乘积。通过将点积除以模长乘积，可以得到两个向量之间的余弦相似度。这个相似度值可以用于评估手工特征与知识图谱中属性特征之间的一致性。在训练过程中，希望最小化相似度损失 \mathcal{L}_s ，以便使得手工特征与知识图谱中的属性特征更加一致。

测试时，利用每张图像的手工特征向量，与 ACKG 中所有表示害虫类别的节点所对应的特征向量进行余弦相似度计算，组合得到属性相似性特征向量 f_{CL} 。用 l_k 表示害虫类别节点索引，则 f_{CL} 表示为

$$f_{CL} = [\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_k] \quad (5)$$

2.2 害虫图像高层语义表征提取网络 ViT

由于视觉编码器 ViT^[19] 在多种预训练网络模型中表现最为优异, 本研究将 ViT 作为提取图像高层语义表征信息的骨干网络。

ViT 是由 Google 团队在 2020 年提出的一种图像分类模型, 通过在视觉任务中引入 Transformer^[23] 机制, ViT 表现出了良好的性能与可扩展性, 现被广泛应用于各类视觉任务。

对于输入图像 $X \in \mathbb{R}^{H \times W \times C}$, 即高、宽、通道数分别为 H 、 W 、 C , ViT 会首先将其切分为多个子图块 (Patch), 并展平为一维输入序列。令图块大小为 $P \times P$, 则该子图块序列表示为

$$X_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$$

式中, N 表示子图块数目。

接下来, 将每个子块投影为固定长度 D 的向量再输入到 Transformer 编码器, 即有:

$$X_p' \in \mathbb{R}^{N \times D}$$

在序列头部嵌入特殊字符 CLS, 从而将视觉问题转化为 seq2seq 问题。经过位置编码、层归一化以及多层感知机进行维度变换之后, 得到最终图像表征输出向量, 记为 f_{SF} 。为了获得更好的迁移效果, 本研究所使用的是经过 ImageNet-21k 预训练的 ViT 网络。

2.3 分类器

ACKGViT 结合知识图谱提取的图像属性特征和 ViT 提取的图像高层语义表征特征用于训练分类器。训练、测试时融合后的特征 f_{train} 、 f_{test} 分别表示为

$$f_{train} = f_{SF} + f_{CF} \quad (6)$$

$$f_{test} = f_{SF} + f_{CL} \quad (7)$$

分类器由全连接网络和 Softmax 函数构成。输入是害虫图像特征向量, 即 f_{train} 或 f_{test} , 输出是害虫类别。

模型损失 \mathcal{L} 用交叉熵损失函数 \mathcal{L}_c 和余弦损失函数 \mathcal{L}_s 表示:

$$\mathcal{L}_c = - \sum_i y_i \lg p(\hat{y}_i) \quad (8)$$

$$L = \mathcal{L}_c + \mathcal{L}_s \quad (9)$$

式中, y_i 和 \hat{y}_i 分别表示输入害虫图像的真实标签和预测标签, $p(\hat{y})$ 表示 \hat{y} 的预测概率。

训练分类器方法依据与原理如下: 1) 图像属性特征提取: 利用知识图谱中的概念节点和它们之间的关系来捕捉图像中的属性信息。知识图谱通过对领域知识的结构化表示, 为模型提供了关于病虫害的高级信息。利用图注意网络 (graph attention networks, GAT) 对知识图谱进行编码, 从而为每个概念节点生成一个特征向量。这些特征向量可以视为害虫类别的属性特征。2) 高层语义表征特征提取: 使用 ViT 从图像中提取高层语义表征特征。ViT 通过将图像分割为固定大小的 patches, 并将它们线性嵌入到特征空间中, 然后应用 Transformer 结构

进行特征抽取。这样, ViT 可以捕获图像中的全局上下文信息, 从而生成具有高层语义的特征表示。3) 训练阶段: 将知识图谱提取的图像属性特征与 ViT 提取的高层语义表征特征进行融合。融合通过特征加法操作进行融合。融合后的特征向量包含了病虫害的属性信息和图像的高级语义信息, 可以更好地表征害虫类别, 提高分类性能。这样做可以进一步引导模型关注与目标类别相关的特征, 减小背景噪声的影响。具体过程如图 3 所示。

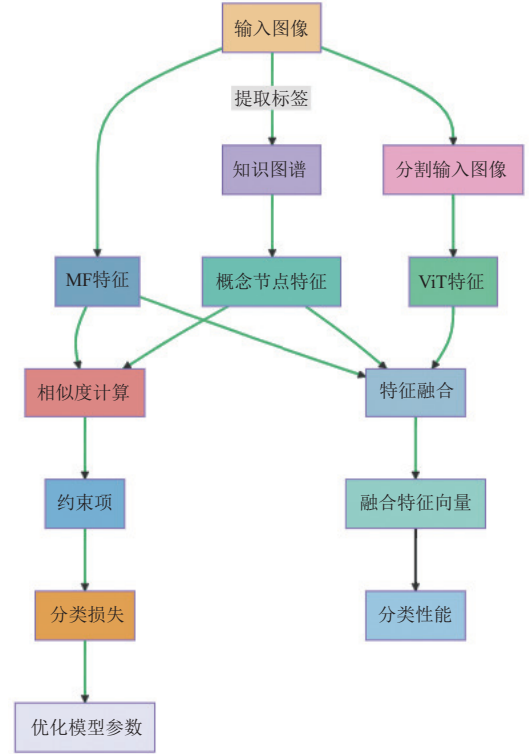


图 3 ACKGViT 模型具体运作机理

Fig.3 Specific operating mechanism of ACKGViT(attribute characteristics knowledge graph enhanced vision transformer) model

2.4 数据集

本研究测试所用数据集包括从大规模公开数据集 IP102^[24] 中挑选的 GP21 数据集和从农业生产基地实地采集的 GP8 数据集。

GP21 数据集。GP21 数据集来源于 IP102^[24] 数据集, 该数据集包括从专业农业网站和昆虫科学网站上收集的共计 75 222 个样本。由于本任务聚焦于葡萄园害虫的细粒度识别问题, 所以从中挑选了 21 种葡萄园常见害虫图像, 构成 GP21 数据集, 用于分类研究, 其中, 训练样本和测试样本总数分别为 10 303 和 1 714。

GP8 数据集。该数据集采集于杭州浙大城市学院植物工厂基地。如表 1 所示, 在 3 位农业专家指导下, 利用远程可视化自动害虫监测系统 iMETOS iSCOUT, 采集了包含绿盲蝽、大青叶蝉和麦二叉蚜等在内的 8 种当季葡萄园害虫图像, 通过人工筛选和标注, 最终获得 GP8 数据集, 如图 4 所示, 其样本总数为 1 365。将 GP8 数据集按照 3:1 的比例进行划分, 得到训练集和测试集。其中, 训练样本总数为 1 023, 测试样本总数为 342。

表 1 GP8 数据集包含害虫类别及相应样本量

Table 1 Taxonomy and its corresponding sample size of the GP8 dataset

害虫名称 Pest name	样本量 Sample size
绿盲蝽 <i>Apolygus lucorum</i>	173
大青叶蝉 <i>Cicadella viridis</i>	165
麦二叉蚜 <i>Schizaphisgraminum</i>	168
斜纹夜蛾 <i>Spodoptera litura fabricius</i>	171
十星瓢蚧叶甲 <i>Oides decempunctata billberg</i>	162
康式粉蚧 <i>Pseudococcus comstocki kuwana</i>	165
斑衣蜡蝉 <i>Lycorma delicatula</i>	182
葡萄虎天牛 <i>Xylotrechus pyrrhoderus bates</i>	179

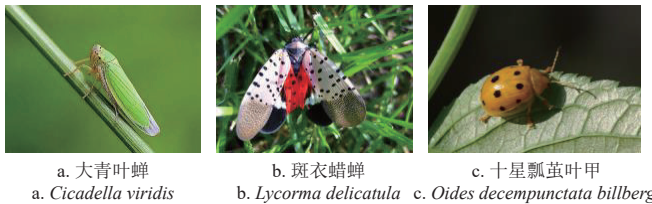


图 4 GP8 数据集中的样本示例

Fig. 4 Different example images of the GP8 dataset

2.5 试验设置

本文所提出的 ACKGViT 模型采用两层的 GAT 网络, 所含 MLP (multi-layer perceptron, MLP) 隐藏层维度为 16, 注意力头 Z 设置为 4, 输出结点数目与数据集中害虫类别总数保持一致; ViT 使用默认参数配置。在将图像输入到网络进行特征提取之前, 通过缩放操作 (Resize) 保证其空间尺度为 224×224 。训练时观察到模型在 50 个轮次时已完全收敛, 本试验将训练轮次设置为 50, 学习率设置为 0.001。

相关试验在装有 NVIDIA RTX 3090 GPU 和 Intel Core i9 10900K CPU 的机器上进行。训练过程使用 SGD 优化器, GP21 数据集和 GP8 数据集完成 50 个轮次的训练分别需要大约 4 和 0.5 h。与其他数据集相关文献一致, 本试验将准确率 (accuracy, A) 作为评估模型性能的指标, 同时也列举了 F_1 分数 (F_1), 精确率 (precision, P), 召回率 (recall, R) 指标的计算结果。

$$A = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (10)$$

$$P = \frac{T_p}{T_p + F_p} \quad (11)$$

$$R = \frac{T_p}{T_p + F_n} \quad (12)$$

$$F_1 = \frac{2PR}{P+R} \quad (13)$$

式中, T_p 是正确预测的正样本数量, T_n 是正确预测的负样本数量, F_p 是错误预测的正样本数量, F_n 是错误预测的负样本数量。

2.6 模型算法的复杂度分析

在评估所提出的葡萄园害虫识别模型的性能之外, 还需要关注算法复杂度和计算量等方面的分析。这些分析有助于了解模型在实际应用中的效率和可扩展性。

首先, 考虑模型的时间复杂度。模型的主要组成部分包括视觉编码器 (ViT) 和图卷积网络 (graph convolutional network, GCN)。对于 ViT, 其时间复杂

度主要取决于图像的分辨率和 ViT 的层数。假设输入图像的大小为 $H \times W$, ViT 的层数为 L , 那么 ViT 的时间复杂度大致为 $O(L \times H \times W)$ 。对于 GCN, 其时间复杂度主要取决于图的节点数和 GCN 的层数。假设知识图谱有 N 个节点, GCN 的层数为 K , 那么 GCN 的时间复杂度大致为 $O((K \times N)^2)$ 。因此, 整个模型的时间复杂度大约为 $O(L \times H \times W + (K \times N)^2)$ 。

其次, 关注模型的空间复杂度。同样地, 模型的空间复杂度主要取决于 ViT 和 GCN。对于 ViT, 其空间复杂度主要与图像分辨率和 ViT 的参数数量有关。设 ViT 的参数数量为 P_v , 那么 ViT 的空间复杂度大约为 $O(P_v)$ 。对于 GCN, 其空间复杂度主要与知识图谱的节点数和 GCN 的参数数量有关。设 GCN 的参数数量为 P_g , 那么 GCN 的空间复杂度大约为 $O(P_g + N)$ 。因此, 整个模型的空间复杂度大约为 $O(P_v + P_g + N)$ 。

最后, 讨论模型的计算量。计算量主要受输入图像大小、模型参数数量以及训练迭代次数等因素的影响。在训练阶段, 模型需要对大量图像进行前向传播和反向传播, 以更新模型参数。设训练迭代次数为 T , 那么整个模型的计算量大约为 $O(T(L \times H \times W + (K \times N)^2))$ 。

综上所述, 本研究所提出的模型在时间复杂度、空间复杂度和计算量方面的分析表明, 虽然引入知识图谱和图卷积网络增加了一定的计算负担, 但总体上仍在可接受范围内。此外, 随着硬件性能的提升和算法优化技术的发展, 该模型在实际应用中将具有较高的效率和可扩展性。

3 结果与分析

3.1 不同模型在数据集上的性能比较

表 2 分别列出了预训练网络 VGG-16、ResNet-152^[21]、Inception-V3、Xception、MobileNet、SqueezeNet^[22] 和 ViT 在 GP21 和 GP8 测试集上的性能。

表 2 不同模型在 GP21 和 GP8 数据集上的性能对比

Table 2 Performance comparison of the different models on the GP21 and GP8 datasets.

数据集 Dataset	模型 Model	准确率 Accuracy	F_1 分数 F_1 score	精确率 Precision	召回率 Recall
GP21	VGG-16	85.13	77.08	79.34	75.45
	ResNet-152	87.49	79.59	81.07	79.03
	Inception-V3	87.18	79.11	80.71	78.48
	Xception	84.49	75.78	77.95	74.58
	MobileNet	85.19	76.84	78.70	75.59
	SqueezeNet	76.79	67.28	70.89	65.23
	ViT	89.57	83.05	84.98	81.70
	ACKGViT	91.21	85.95	87.52	84.99
GP8	VGG-16	88.30	86.75	87.21	86.29
	ResNet-152	91.23	89.33	90.11	88.56
	Inception-V3	90.64	88.71	89.92	87.53
	Xception	87.72	85.34	87.66	83.14
	MobileNet	88.89	87.01	88.96	85.14
	SqueezeNet	80.12	78.26	83.39	73.72
	ViT	93.86	92.05	94.28	89.92
	ACKGViT	95.03	93.98	95.17	92.82

从表 2 中可以看出, ViT 模型在 Accuracy 和 F_1 指标上都明显优于其他模型。相比于目前视觉任务中最高频使用的模型之一 ResNet-152, ViT 的 Accuracy 和 F_1 值在 GP21 数据集上分别提高了 2.08 和 3.46 个百分点, 在

GP8 数据集上分别提高了 2.63 和 2.72 个百分点。这是因为相比于卷积神经网络模型，ViT 的注意力机制使得它具备低层的全局特征学习能力，所以利用 ViT 提取的高层表征能够更精细地整合害虫图像全局和局部信息，因此，本研究将 ViT 作为骨干网络构建 ACKGViT 模型。

ACKGViT 的性能在表 3 最后一行展示，相比于 ViT，ACKGViT 的 Accuracy 和 F_1 值在 GP21 数据集上分别提高了 1.64 和 2.90 个百分点，在 GP8 数据集上分别提高了 1.17 和 1.93 个百分点，这是因为 ViT 在识别形状相似的物体时能力不足^[27]，而知识图谱能够提供不同类别害虫之间的细节信息，从而辅助 ViT 区分害虫类型。

3.2 消融试验

为了进一步分析知识图谱的引入对于葡萄园害虫分类性能的提升作用，本研究设计了 3 组消融试验：1) 将 ACKGViT 模型移除手工特征 (w/o MF)，使提取的手工特征不参与特征融合；2) 将 ACKGViT 模型移除知识图谱 (w/o KG)，使知识图谱提取的属性特征不参与特征融合；3) 同时移除手工特征和知识图谱 (w/oMF \cup KG)，使得手工特征和知识图谱提取的属性特征均不参与特征融合。

3 组消融试验结果如表 3 所示。从表 3 中可以看出，移除知识图谱和手工特征 (w/oMF \cup KG) 使得模型性能 accuracy 和 F_1 在 GP21 数据集上分别下降 1.64 和 2.9 个百分点，在 GP8 数据集上分别下降 2.34 和 2.96 个百分点。移除手工特征 (w/o MF) 和移除知识图谱 (w/o KG) 使得模型性能 accuracy 在 GP21 数据集上分别下降 1.35 和 1.55 个百分点，在 GP8 数据集上分别下降 1.46 和 1.76 个百分点，同时， F_1 在 GP21 数据集上分别下降 2.32 和 2.36 个百分点，在 GP8 数据集上分别下降 2.10 和 2.60 个百分点。上述结果表明：1) 通过引入知识图谱的方式辅助 ViT 获取更加精确的害虫信息是有效的；2) 仅使用传统特征和知识图谱在提升模型性能上作用不大，主要原因是：传统特征提取方法在表达图像高层语义信息方面存在缺陷，而仅使用知识图谱无法有效训练图卷

积网络，从而使得结点特征向量表征不足；3) 在 ACKGViT 模型学习中，知识图谱作用大于传统特征，这是因为除了图像浅层表征外，知识图谱还能够提供不同类别之间的关联特征等额外信息作为辅助分类的线索。例如，如图 1 所示，知识图谱能够通过“足”的长短、“后翅”颜色等属性特征来区分都含有“体暗褐色”特征的“斜纹夜蛾”和“葡萄短须蛾”。另一方面，知识图谱基于卷积神经网络实现，能够从原始图像中提取到复杂抽象的深层特征，这些特征可以增强模型对于分类任务的健壮性。

表 3 知识图谱和手工特征消融试验结果

Table 3 Ablation study results of knowledge graph and

数据集 Dataset	消融试验操作 Ablation experiment operation	handcrafted features				%
		准确率 Accuracy	F_1 分数 F_1 score	精确率 Precision	召回率 Recall	
GP21	无操作 (No operation)	91.21	85.95	87.52	84.99	
	移除手工特征 (w/o MF)	89.86	83.63	86.24	81.84	
	移除知识图谱 (w/o KG)	89.66	83.59	86.00	81.85	
	同时移除手工 特征和知识图谱 w/oMF \cup KG	89.57	83.05	84.98	81.70	
	无操作 (No operation)	95.03	93.98	95.17	92.82	
GP8	移除手工特征 (w/o MF)	93.57	91.84	93.62	90.13	
	移除知识图谱 (w/o KG)	93.27	91.33	93.28	89.46	
	同时移除手工 特征和知识图谱 w/oMF \cup KG	92.69	91.02	92.50	89.59	

注：表中“w/o”表示移除操作。MF 表示手工特征，KG 表示知识图谱。

Note: In the table, "w/o" represents the removal operation. "MF" stands for manual features, and "KG" refers to the knowledge graph.

3.3 可视化案例子

为了验证知识图谱的引入在葡萄园害虫特征挖掘方面的可用性，本研究进一步比较了 ACKGViT 模型与骨干网络 ViT 模型学习到的害虫特征向量，并利用可视化技术^[28]直观展示对比结果。

如图 5 所示，在不同复杂程度的背景条件下，对于不同类别的害虫，ACKGViT 方法都可以准确关注到害虫区域。

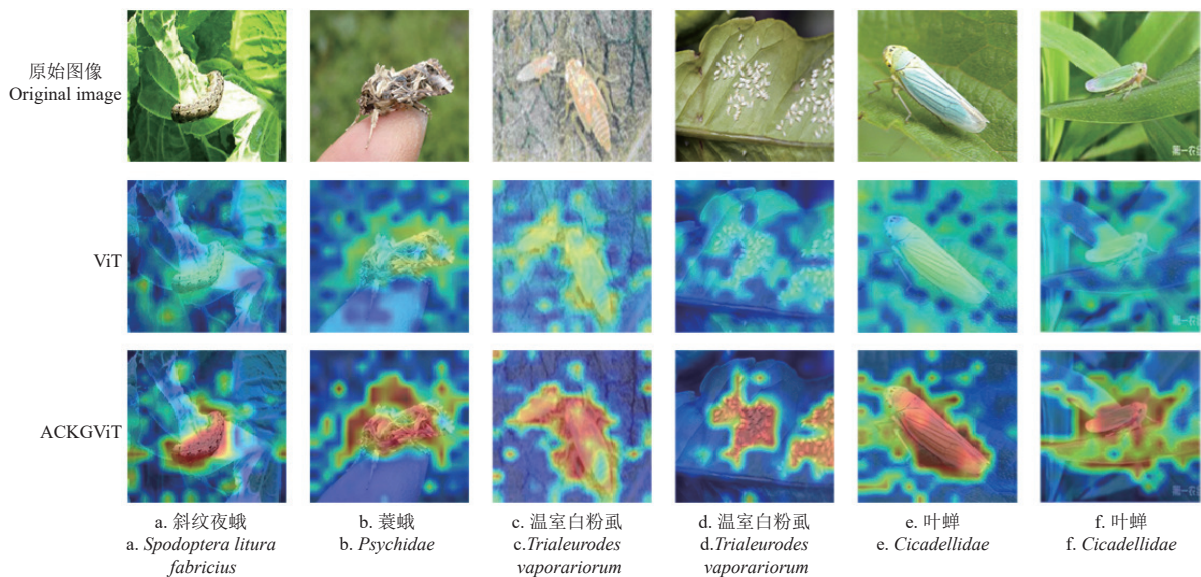


图 5 ACKGViT 和 ViT 可视化结果对比

Fig.5 Visualization of ACKGViT method and ViT method

此外,相比于 ViT 方法,ACKGViT 能够更加准确地将害虫区域和背景环境区分,并且描绘出更加清晰的轮廓曲线,证明 ACKGViT 方法能够捕捉到更加细节、更加完整的害虫特征信息并用于分类研究。

3.4 不同相似度计算方法的对比试验

在本章节中将对不同相似度计算方法进行对比试验,以评估它们在葡萄园害虫识别任务中的性能。相似度计算方法是评估两个实体之间相似程度的关键因素,对于基于知识图谱的葡萄园害虫识别方法来说尤为重要。通过比较不同相似度计算方法,可以找到适合该任务的最佳方法,从而提高识别精度和效率。为了达到上述目的,本研究选取了以下几种常见的相似度计算方法进行对比试验:余弦相似度、欧几里得距离、曼哈顿距离、皮尔森相关系数、Jaccard 系数。为了确保试验结果的可靠性,将在相同的数据集上对各种相似度计算方法进行评估。试验结果如表 4 所示。

表 4 不同相似度计算方法的性能对比

Table 4 Performance comparison of the different similarity calculation methods %

数据集 Dataset	相似度计算方法 Similarity calculation method	准确率 Accuracy	F_1 分数 F_1 score	精确率 Precision	召回率 Recall
GP21	欧几里得距离	75.25	65.74	69.35	63.61
	曼哈顿距离	83.86	74.62	76.18	73.29
	皮尔森相关系数	82.41	73.19	75.32	71.84
	Jaccard 系数	87.63	80.92	82.76	79.45
	余弦相似度	91.21	85.95	87.52	84.99
GP8	欧几里得距离	78.36	76.42	80.97	72.51
	曼哈顿距离	87.11	85.24	87.39	83.47
	皮尔森相关系数	86.24	83.56	85.69	81.78
	Jaccard 系数	91.23	89.15	90.87	87.68
	余弦相似度	95.03	93.98	95.17	92.82

根据试验结果分析,在 GP21 数据集上,使用余弦相似度相比于欧几里得距离、曼哈顿距离、皮尔森相关系数和 Jaccard 系数,准确率分别提高了 15.96 个百分点、7.35 个百分点、8.80 个百分点和 3.58 个百分点; F_1 分数分别提高了 20.21 个百分点、11.33 个百分点、12.76 个百分点和 5.03 个百分点。在 GP8 数据集上,使用余弦相似度相比于欧几里得距离、曼哈顿距离、皮尔森相关系数和 Jaccard 系数,准确率分别提高了 16.67 个百分点、7.92 个百分点、8.79 个百分点和 3.8 个百分点; F_1 分数分别提高了 17.56 个百分点、8.74 个百分点、10.42 个百分点和 4.83 个百分点。通过上述分析可以得出余弦相似度计算方法在葡萄园害虫识别任务中性能表现最优的结论,通过对比试验为该任务找到了最佳的相似度计算方法,从而为后续研究和实际应用提供了有益的启示。

4 结 论

本研究基于领域农业专家的丰富经验和害虫知识语料构建了一套详尽的葡萄园害虫属性特征知识图谱。在此基础上,本研究提出了一种基于知识图谱的细粒度害虫分类双分支模型 ACKGViT。该模型利用图卷积网络 GAT 将知识图谱转化为可供训练的神经网络模块,同时

结合传统特征对网络参数进行训练和优化。主要结论如下:

1) 与骨干网络 ViT 相比,ACKGViT 在葡萄园害虫分类任务中表现出了明显的性能提升。具体而言,ACKGViT 在 GP21 数据集上的准确率和 F_1 指标分别提高了 1.64 和 2.90 个百分点,而在 GP8 数据集上,这两个指标分别提高了 1.17 和 1.93 个百分点。

2) 知识图谱和手工特征消融试验结果表明:移除知识图谱所在分支使得模型性能准确率和 F_1 分数在 GP21 数据集上分别下降 1.64 和 2.90 个百分点,在 GP8 数据集上分别下降 2.34 和 2.96 个百分点。移除手工特征和移除知识图谱使得模型性能 Accuracy 在 GP21 数据集上分别下降 1.35 和 1.55 个百分点,在 GP8 数据集上分别下降 1.46 和 1.76 个百分点,同时, F_1 在 GP21 数据集上分别下降 2.32 和 2.36 个百分点,在 GP8 数据集上分别下降 2.10 和 2.60 个百分点。证明本文所提方法的有效性。

3) 不同相似度计算方法的性能对比试验结果表明:在 GP21 数据集上,使用余弦相似度相比于其他相似度计算方法,准确率和 F_1 分数分别提高了最多 15.96 个百分点和 20.21 个百分点。相应地,在 GP8 数据集上,准确率和 F_1 分数相比于其他方法最多提高了 16.67 个百分点和 17.56 个百分点。这些结果充分证明,余弦相似度在葡萄园害虫识别任务中的性能表现最优。

在未来的研究中将会继续深入探索知识图谱在害虫图像分类任务中的应用:1) 图谱权重优化:研究如何更有效地利用知识图谱中的权重信息,进一步提高害虫图像分类的性能;2) 动态知识图谱构建:实时更新和扩展知识图谱,以适应不断变化的农业环境和新出现的病虫害类型;3) 多模态数据融合:探讨将其他数据源(例如气象数据、土壤信息等)与图像数据融合,提供更丰富的上下文信息以提高分类准确性。通过以上研究方向的探索,期望为智慧农业的发展贡献力量,从实际问题出发,促进农业生产的可持续发展。

[参 考 文 献]

- [1] 乔虹,冯全,张芮,等.基于时序图像跟踪的葡萄叶片病害动态监测[J].农业工程学报,2018,34(17):167-175.
QIAO Hong, FENG Quan, ZHANG Rui, et al. Dynamic monitoring of grape leaf disease based on sequential images tracking[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(17): 167-175. (in Chinese with English abstract)
- [2] AYAN E, ERBAY H, VARÇIN F. Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks[J]. Computers and Electronics in Agriculture, 2020, 179: 105809.
- [3] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Santiago, Chile,

- 2015: 1-9.
- [4] CHOLLET F. XCEPTION. Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Venice, Italy, 2017: 1251-1258.
- [5] HOWARD AG, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2023-02-21]<https://arxiv.org/abs/1704.04861>.
- [6] 苏仕芳, 乔焰, 饶元. 基于迁移学习的葡萄叶片病害识别及移动端应用[J]. 农业工程学报, 2021, 37(10): 127-134.
- SU Shifang, QIAO Yan, RAO Yuan, et al. Recognition of grape leaf diseases and mobile application based on transfer learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(10): 127-134. (in Chinese with English abstract)
- [7] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 248-255.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2023-02-21]<https://arxiv.org/abs/1409.1556>.
- [9] 孙钰, 周焱, 袁明帅, 等. 基于深度学习的森林虫害无人机实时监测方法[J]. 农业工程学报, 2018, 34(21): 74-81.
- SUN Yu, ZHOU Yan, YUAN Mingshuai, et al. UAV real-time monitoring for forest pest based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(21): 74-81. (in Chinese with English abstract)
- [10] 梁万杰, 曹宏鑫. 基于卷积神经网络的水稻害虫识别[J]. 江苏农业科学, 2017, 45(20): 241-243.
- [11] TURKOGLU M, HANBAY D, SENGUR A. Multi-model LSTM-based convolutional neural networks for detection of apple diseases and pests[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(7): 3335-3345.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [13] 王林惠, 兰玉彬, 刘志壮, 等. 便携式柑橘害虫实时检测系统的研制与试验[J]. 农业工程学报, 2021, 37(9): 282-288.
- WANG Linhui, LAN Yubin, LIU Zhizhuang, et al. Development and experiment of the portable real-time detection system for citrus pests[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 282-288. (in Chinese with English abstract)
- [14] 于合龙, 沈金梦, 毕春光, 等. 基于知识图谱的水稻病虫害智能诊断系统[J]. 华南农业大学学报, 2021, 42(5): 105-116.
- YU Helong, SHEN Jinmeng, BI Chunguang, et al. Intelligent diagnostic system for rice diseases and pests based on knowledge graph[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 42(5): 105-116. (in Chinese with English abstract)
- [15] 戈为溪, 周俊, 袁立存, 等. 基于知识图谱与案例推理的水稻精准施肥推荐模型[J]. 农业工程学报, 2023, 39(2): 126-133.
- GE Weixi, ZHOU Jun, YUAN Licun, et al. Recommendation model for rice precision fertilization using knowledge graph and case-based reasoning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(2): 126-133. (in Chinese with English abstract)
- [16] 吴赛赛, 周爱莲, 谢能付, 等. 基于深度学习的作物病虫害可视化知识图谱构建[J]. 农业工程学报, 2020, 36(24): 177-185.
- WU Saisai, ZHOU Ailian, XIE Nengfu, et al. Construction of visualization domain-specific knowledge graph of crop diseases and pests based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(24): 177-185. (in Chinese with English abstract)
- [17] 郑泳智, 吴惠舜, 朱定局, 等. 基于荔枝和龙眼病虫害知识图谱的问答系统[J]. 计算机与数字工程, 2021, 49(12): 2618-2622.
- ZHENG Yongzhi, WU Huilin, ZHU Dingju, et al. A question and answer system based on knowledge graph of litchi and longan disease pests[J]. Computer and Digital Engineering, 2021, 49(12): 2618-2622. (in Chinese with English abstract)
- [18] YAN L, LI S. Grape diseases and pests named entity recognition based on BiLSTM-CRF[C]//In 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). Chongqing, China, 2021: 2121-2125.
- [19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C] //International Conference on Learning Representations. Shanghai, China, 2021: 167-198.
- [20] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling[C]//International Conference on Computer Vision. Miami, USA, 2009: 32-39.
- [21] MOHANAIH P, SATHYANARAYANA P, GURUKUMAR L. Image texture feature extraction using GLCM approach[J]. International Journal of Scientific and Research Publications., 2013, 3(5): 1-5.
- [22] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2017-10-30)[2023-02-21] <https://arxiv.org/abs/1710.10903>.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing

- Systems, 2017: 30-36.
- [24] WU X, ZHAN C, LAI Y K, et al. IP102: A large-scale benchmark dataset for insect pest recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, California, USA, 2019: 8787-8796.
- [25] HE K, ZHANG X, REN S, SUN J. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA, 2016: 770-778.
- [26] IANDOLA F N, HAN S, MOSKEWICZ W, et al. Squeeze Net: AlexNet-level accuracy with 50 x fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24)[2023-02-21] <https://arxiv.org/abs/1602.07360>.
- [27] RAGHU M, UNTERTHINER T, KORNBLITH S, et al. Do vision transformers see like convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2021, 34: 12116-12128.
- [28] ABHISHEK D, RAMAKRISHNA V, MICHAEL C, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: 2017: 618-626.

Identifying fine-grained vineyard pest using attribute knowledge graph

ZHENG Zengwei^{1,2}, LI Yanzhen^{1,2,3}, LIU Yi^{1,2,3}, LIN Zhongqi^{1,2}, XIANG Zhengzhe^{1,2}, HE Mengzhu^{1,2}, SUN Lin^{1,2*}

(1. School of Computer & Computing Science, Zhejiang University City College, Hangzhou 310015, China; 2. Intelligent Plant Factory of Zhejiang Province Engineering Lab, Hangzhou 310015, China; 3. College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

Abstract: Pest infestation is one of the main reasons for the decrease in crop yield and quality in vineyards. Deep learning can be expected to identify the pest species for scientific prevention and control strategies, in order to improve the production level of vineyards. In this study, an identification model was proposed, termed as ACKGViT (attribute characteristics knowledge graph enhanced vision transformer), for the fine-grained vineyard pest using constructed attribute characteristics knowledge graph (KG). Particularly, the KG was fully utilized to treat the variety, complex morphology, and low correlation of vineyard pests. A specific description was also given for the entity attribute characteristics and inter-entity association of pests. The improved approach was then applied in various agricultural settings. Pest recognition challenges were clarified across diverse domains. The inherent structure of KGs incorporated valuable contextual complements with the visual features extracted by the visual encoder. The graph convolutional network facilitated the efficient learning of pest attributes and relationships from the KG. This information was then seamlessly integrated into the classification. ViT (vision transformer) was used as the backbone network to extract the high-level representation of images. The fine-grained attribute was combined with the associated features provided by KG for pest identification in vineyards. The KG recognition model of vineyard pests was also trained and optimized using traditional features of images. Traditional image features (such as color, texture, and shape) were extracted using various techniques, including histogram-based methods, wavelet transform, and local binary patterns. The essential information was obtained about the pest species and their characteristics, enabling the KG to accurately capture the associations between different pests. Two datasets were evaluated: one was the GP21 dataset, which was composed of 21 vine-related pest categories from the IP102 public large-scale pest dataset, and another was the GP8 dataset, including eight vine-related pest categories that were collected from agricultural production sites and hand-labelled. The experimental results show that the performance of ACKGViT was significantly better than the general pre-training networks. The accuracies of GP21 and GP8 datasets reached 91.21% and 95.03% respectively, which increased by 1.64 and 1.17 percentage points, compared with the original ViT. The attribute characteristic KG was introduced to assist the ViT with more pest information on the vineyard. In conclusion, accurate and rapid pest identification was offered for the fine-grained vineyard pest recognition model with attribute feature KGs. The improved model can substantially contribute to the targeted and effective pest prevention and control strategies, and further improve the grape yield, quality, and overall productivity in vineyards. Future work can expand the KG into additional attributes and relationships. Other sources of information can also be introduced, such as temporal data or expert knowledge. Furthermore, the proposed approach can be extended into other agricultural domains, indicating the versatility of pest recognition under various settings. Additionally, better performance can be expected to achieve in the high efficiency of the graph convolutional network and the full integration of KG information during classification.

Keywords: knowledge graph; image recognition; deep learning; grape; pest; associated features; image classification