基于深度学习的多模态病虫害检测模型训练与验证研究

李　冰[1],王瀛龙[2]

(1.河南职业技术学院现代信息技术学院,郑州 450046;
2.河南省特种设备检验技术研究院,郑州 450046)

摘　要: 多模态病虫害数据集具有多样性,导致模型平均精度较低。基于此,文章设计了一种基于深度学习的多模态病虫害检测模型。首先,采集高清图像、音频、视频等多模态数据来整合现有资源构成数据集;其次,进行数据增强以提高模型的泛化能力,减少了过拟合风险;最后,选用深度学习中的MobileNet作为主干网络来捕捉关键特征,模型训练融合了多模态数据,并通过动态权重分配和注意力机制显著增强了检测效能。实验结果表明,在迭代60次时,设计模型的mAP达到0.87,远超文献[1]和文献[2]模型的0.72与0.79,能够快速收敛并准确检测多种病虫害。

关键词: 深度学习;MobileNet;多模态;病虫害检测;模型训练

中图分类号: TP391　　**文献标识码:** A

Training and validation of a multimodal pest detection model based on deep learning

LI Bing[1],WANG Yinglong[2]

(1.College of Modern Information Technology,Henan Polytechnic,Zhengzhou 450046,China;
2.Special Equipment Safety Inspection and Research Institute of Henan Province,Zhengzhou 450046,China)

Abstract: The diversity of multi modal pest and disease data sets leads to low average accuracy of the model. Based on this paper, a multi modal pest and disease detection model based on deep learning is designed. Firstly, multi modal data such as HD image, audio and video are collected, and existing resources are integrated to form a data set. Secondly, data enhancement is carried out to improve the generalization ability of the model and reduce the risk of overfitting. Finally, MobileNet in deep learning is selected as the backbone network to capture key features, and the model training incorporates multi modal data and significantly enhances detection performance through dynamic weight allocation and attention mechanisms. The experimental results show that after 60 iterations, the mAP of the design model reaches 0.87, which is far higher than the 0.72 and 0.79 of the model in literature [1] and [2]. It is proved that the design model can rapidly converge and accurately detect a variety of pests and diseases.

Key words: deep learning,MobileNet,multimodal,pest detection,model training

1　引言

病虫害的有效防控是保障农作物产量和质量的关键环节,传统的人工巡检和分类方法不仅效率低、成本高,且难以应对大规模农田和复杂多变的病虫害情况。因此,开发高效、智能的病虫害检测技术成为农业领域的研究热点。在文献[1]中,通过引入轻量级卷积神经网络ShuffleNet V2作为主干网络,并增加小目标检测层和注意力机制SEnet,模型不仅显著提高了检测精度和速度,还保持了轻量化。同时,文献[2]通过搭建地面数据采集设备和无人机遥感系统,从空、地2个角度全面获取猕猴桃冠层叶片的病虫害信息,结合图像处理技术统计被害叶片数量,从而计算出单株果树的病虫害率。然而,上述方法在多模态病虫害检测中仍有一定局限性,因此本文研究了基于深度学习的多模态病虫害检测模型。深度学习能够

基金项目: 河南省科技厅科技攻关(242102111190);河南省高等学校重点科研项目(24B520017)

模拟人脑神经网络的复杂结构，通过多层次的数据抽象与特征学习来自动挖掘并理解数据中的潜在规律与模式[3]。在图像识别、自然语言处理、语音识别等领域，它展现了卓越的性能与广阔的应用前景，通过融合不同来源的数据信息、优化模型结构和算法参数，可以进一步提升病虫害检测的效率和准确性。

2 采集多模态病虫害数据

在构建基于深度学习的多模态病虫害检测模型时，核心在于构建一个多元且均衡的数据集。其应囊括丰富的病虫害种类，可跨越不同农作物及其生长阶段，以确保模型能广泛适应。一方面，数据收集可利用现代设备捕捉病虫害的多模态特征，如视频、音频及高清图像；另一方面，可通过积极整合现有资源（包括国内外权威机构发布的病虫害图像库及多源信息[4]），以拓展数据集的深度与广度，数据来源如表1所列。

表1 数据来源

数据来源	病虫害种类数	图像数量	农作物种类	包含模态
国内农科院	300+	50 000+	20+	图像、视频
国际病虫害数据库	450+	80 000+	35+	图像、音频
农业科技公司合作	200+	30 000+	15+	图像、视频、文本描述
公开竞赛数据集	100+	20 000+	10+	图像、视频、标注信息
科研机构共享	150+	45 000+	25+	图像、音频、环境数据

多模态病虫害数据集包括高清图像、声音记录及病虫害发生时的环境视频，它们共同构成了丰富的数据生态。考虑到病虫害在不同作物、生长周期及环境条件下的表现差异，数据采集需覆盖不同的地域、季节与作物种类，以确保模型的泛化能力。

在数据收集后，除了调整常规的亮度、对比度，还需对图像数据中的病虫害特有颜色、纹理特征进行优化，以减少背景干扰（如阴影、反光等）对识别效果的影响；需对音频数据进行噪声过滤，以凸显病虫害特有的声音模式；需对视频数据进行帧率调整、关键帧提取等，以提高处理效率与准确性。将图像、声音等信息有效整合，不仅能提升检测精度，还能为病虫害的精准识别与早期预警提供更为全面的信息支持。

3 进行数据增强

通过数据增强来提升模型的泛化能力，以减少过拟合风险。对于图像模态数据，将病虫害图像统一缩放至384×384分辨率，随后采用随机角度旋转及缩放操作，再裁剪至299×299像素，以模拟不同视角和距离下的观测情况。同时，引入色彩抖动和亮度调整，以模拟光照变化对图像的影响。对于音频模态数据，通过调整播放速度、加入环境噪声及回声效果，以增加病虫害声音的多样性。对于文本模态（如病虫害描述数据），实施同义词替换、句子重组及回译策略，生成语义相近但表述各异的文本描述，从而增强模型的语义理解能力。最后，所有模态的数据均经过标准化处理，图像数据通常被转换为归一化张量，将像素值缩放到[0,1][5]，假定初始的图像数据集合为I，为了进行统一化处理，归一化后处理后的图像数据集合可以通过以下公式计算：

$$I'=\frac{I-\min(I)}{\max(I)-\min(I)} \tag{1}$$

式中，$\min(I)$和$\max(I)$分别表示图像I在所有像素上的最小值和最大值。

音频数据转换为梅尔频谱图张量，而文本数据则通过词嵌入转化为数值向量。若文本中有一个词嵌入矩阵E，则单词w可以通过查找E中的相应行来获得嵌入向量v_w：

$$v_w=E\left[d_w\right] \tag{2}$$

式中，d_w表示单词w在词汇表中的索引。

依托多模态数据来增强流程，可确保数据的一致性和模型输入的有效性，丰富了数据的多样性。

4 基于深度学习选择模型的主干网络

在多模态病虫害检测模型中，针对增强后的多模态病虫害数据选择适宜的深度学习模型主干网络至关重要。鉴于图像中丰富的空间层次信息与细节，MobileNet通过深度可分离卷积有效平衡了精度与计算效率，成为优选。

作为深度学习中的轻量级模型代表，MobileNet的独特之处在于深度可分离卷积设计。首先，该设计将输入数据利用1×1卷积核实施通道数的减少策略，显著减少了计算过程中的资源消耗，有效降低了计算复杂度。其次，对缩减后的特征图执行空间卷积，以提取局部空间特征[6]，不仅保留了关键信息，还显著减少了参数数量。最终，通过通道堆叠恢复特征图深度，确保模型在保持高效的同时具备强大的特征表达能力，从而成为移动端及边缘计算场景下优选的主干网络。MobileNet的结构如图1所示。

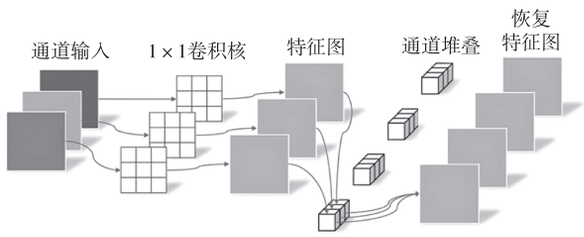

图1 MobileNet结构

在 MobileNet 等轻量级主干网络中，其深度可分离卷积的参数量 P 可简化为：

$$P=K\times M+(1\times 1)\times M\times N \tag{3}$$

式中，K 表示空间卷积核大小；M 表示输入通道数；N 表示输出通道数。

基于 MobileNet 的参数量，多模态病虫害检测模型参数量 P_M 可概括为：

$$P_M=\sum_n K_n^2(M+1)PN \tag{4}$$

式中，n 表示卷积核数量。

若将 MobileNet 作为替代传统卷积模型的主干网络，则参数压缩率 R 可定义为：

$$R=\frac{P_t}{P_M} \tag{5}$$

式中，P_t 表示传统卷积的参数量。在 MobileNet 轻量级主干网络中，针对深层特征图通道数增多[7]的情况，若忽略 1×1 卷积的通道数变化，则参数压缩率 R 可近似为：$R\approx\frac{K^2}{2}$。

本文以 MobileNet 为核心主干网络，旨在以高效轻量的方式捕捉关键特征，同时优化模型颈部架构，实现更深层次的轻量化，加快病虫害检测流程，从而提升检测效率与精度。

5　训练多模态病虫害检测模型

针对农作物病虫害检测对多模态数据融合的迫切需求，特别是考虑到病虫害在不同生长阶段的多变形态及环境因素的复杂影响，本文对多模态病虫害检测模型进行了训练。该模型融合了图像、光谱及气象等数据，以全面捕捉病虫害的细微特征与动态变化。模型训练流程如图 2 所示。

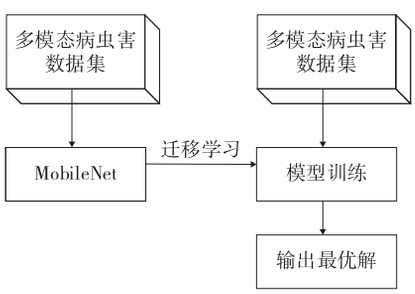

图 2　多模态病虫害检测模型训练流程

首先，采用多模态数据融合，将 640×640 像素的 RGB 图像、1 024 维的高光谱数据及时间序列的气象数据通过特定编码层转换为统一维度的特征向量。为提升模型在复杂场景下的识别能力，引入动态权重分配机制，通过式(6)可自动调整各模态数据的贡献度：

$$A=\exp\left(\alpha S_j\right) \tag{6}$$

式中，S_j 表示第 j 模态的特征质量评分；α 表示可学习参数，可调节评分的影响力。

随后，融合后的特征向量通过注意力模块提升多尺度跨模态注意力。该模块不仅考虑了通道间的相关性，还融入了跨模态的空间对齐[8]，通过式(7)可计算得到增强后的特征表示：

$$D=A\left(W_{lj}+\sigma F_l+F_j\right) \tag{7}$$

式中，W_{lj} 表示模态 l 与 j 之间的注意力权重；σ 表示激活函数；F_l 和 F_j 分别表示模态 l 和 j 的特征图。

使用增强后的特征对模型进行训练，选择 Adam 自动调整学习率，从而提高收敛速度与训练效果。在训练过程中，可以通过验证集来不断调整这些参数，以找到最优的组合。通过迁移学习来充分利用已有模型的知识和经验，新模型能够更快地收敛到最优解并提高性能。

6　多模态病虫害检测模型验证实验

6.1　实验环境

本次多模态病虫害检测模型验证实验在某农科院实验田展开，依托高性能计算平台，实验环境基本配置信息如表 2 所列，旨在最大化模型的精度与效率，为模型的训练与验证提供高效、稳定的平台。

表 2　实验环境配置

序号	配置	参数
1	CPU	AMD Ryzen Threadripper 3990X
2	CPU 主频	3.1 GHz
3	内存	256 GB DDR4
4	显卡	NVIDIA RTX A6000 * 2
5	显存	48 GB GDDR6 * 2
6	操作系统	Ubuntu 20.04 LTS
7	开发界面	Jupyter Notebook 6.4.0
8	编程语言	Python 3.9.5
9	CUDA 版本	11.3
10	CuDNN 版本	8.2.1
11	深度学习框架	TensorFlow 2.5.0

6.2　实验准备

将华为 Mate 40 Pro、OPPO Find X3、三星 Galaxy S21、小米 Mix 4 及 iPhone 13 Pro Max 等智能手机作为数据采集设备。实验时间为 3 月至 6 月，每日选取光照最稳定的上午 10 时至下午 3 时时段，以确保图像质量。病虫害种类涵盖小麦条锈病、白粉病及蚜虫、麦蜘蛛 4 种典型病虫害，如图 3 所示。

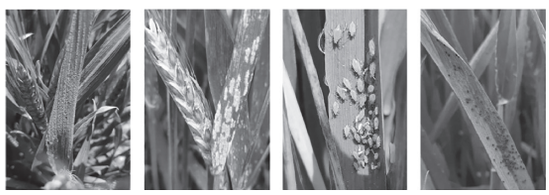

图 3　4 种典型病虫害

图像分辨率从4 800×6 400像素到3 264×2 448像素不等。最终，成功收集到高质量、多模态的病虫害图像数据共4 500张，利用LabelImg图像标注工具对图像中的目标进行了详尽标注。为了优化模型的训练效果及确保评估的精确度，依据7∶1.5∶1.5的比例进行数据集划分。数据集的具体构建细节如表3所列。

表3 多模态的病虫害图像数据集

病虫害种类	训练集	验证集	测试集	合计
小麦条锈病	1 050	225	225	1 500
白粉病	560	120	120	800
蚜虫	798	171	171	1 140
麦蜘蛛	742	159	159	1 060
合计	3 150	675	675	4 500

训练迭代次数设定为60轮，初始学习率调整为3e-4，并在第45轮时以0.2的衰减率调整学习率，以促进模型收敛。采用Adam优化器，设置运行平均值和平方系数分别为0.85与0.99，以增强优化过程的稳定性，同时设置权重衰减为5e-5来减少过拟合。特别地，针对模型中处理查询向量与采样偏移量的部分，将其学习率独立调整为原值的0.2倍。在损失函数中，将针对病虫害区域分类的权重提升至3，以强化分类性能。可变形注意力模块与多模态融合模块均配置12个注意力头，并设置采样点数至50，以捕捉更精细的空间信息。

6.3 实验指标

为验证本研究提出的基于深度学习的多模态病虫害检测模型的有效性，将丰富的多模态病虫害图像数据集载入内存，随后应用一系列图像预处理技术，并随机打乱数据集中图像的排列顺序，以确保实验结果的公正性。在保持深度学习框架参数配置一致的前提下，将本文模型与文献[1]中的基于改进YOLOv8的小麦叶片病虫害检测轻量化模型、文献[2]中的基于空地多源信息的猕猴桃果园病虫害检测模型进行对比实验。为了全面评估各模型在给定数据集上的表现能力，将平均精度(mAP)作为核心评价指标，mAP是衡量所有病虫害种类目标检测平均准确度的关键指标，计算公式为：

$$mAP=\frac{\sum_{i=1}^{C}AP_i}{C} \tag{8}$$

式中，AP_i表示第i种病虫害的平均检测精度；C表示病虫害种类数。通过应用不同的训练迭代次数，以获取各模型在数据集上的性能表现，平均精度mAP对比结果如表4所列。

表4 平均精度mAP对比结果

迭代次数	文献[1]模型	文献[2]模型	本文模型
10	0.55	0.60	0.65
20	0.60	0.68	0.72
30	0.65	0.72	0.78
40	0.68	0.75	0.82
50	0.70	0.77	0.85
60	0.72	0.79	0.87

6.4 实验结果分析

由表4可知，随着训练迭代次数的增加，所有模型的平均精度mAP均有所提升，这表明更长的训练时间有助于模型学习更多的特征并改进其性能。然而，本文模型在不同迭代次数下的表现均优于文献[1]和文献[2]模型的性能。在迭代次数为10时，本文模型的mAP超过了另外2个模型，显示出较快的收敛速度和较强的初始性能。随着迭代次数的增加，本文模型的mAP持续提升，并在迭代次数为60时达到0.87，显著高于文献[1]模型的0.72和文献[2]模型的0.79。这一结果表明，本文提出的基于深度学习的多模态病虫害检测模型在处理复杂病虫害图像数据集时具有更高的检测精度和更强的泛化能力。

7 结束语

通过对基于深度学习的多模态病虫害检测模型的训练与验证，本文不仅提升了病虫害检测的准确性与效率，还展示了深度学习在农业智能化领域的巨大潜力。未来，将继续优化模型结构，探索更多模态信息的融合方法，以期实现更精准、高效的病虫害检测，为农业生产的可持续发展贡献智慧力量。

参考文献：

[1] 杨锋，姚晓通.基于改进YOLOv8的小麦叶片病虫害检测轻量化模型[J].智慧农业(中英文)，2024，6(1)：147-157.

[2] 闫云才，郝硕亨，高亚玲，等.基于空地多源信息的猕猴桃果园病虫害检测方法[J].农业机械学报，2023，54(S2)：294-300.

[3] 张友为，王鑫鑫，范晓飞.基于深度学习的玉米和番茄病虫害检测技术研究进展[J].江苏农业科学，2024，52(10)：10-20.

[4] 慕君林，马博，王云飞，等.基于深度学习的农作物病虫害检测算法综述[J].农业机械学报，2023，54(S2)：301-313.

[5] 王铭慧，张怀清，樊江川，等.基于深度学习网络实现番茄病虫害检测与识别[J].中国农业大学学报，2023，28(11)：165-181.

[6] 蔡易南，肖小玲.基于改进YOLOv5n的葡萄叶病虫害检测

(下转第96页)