



任治洲,梁琨,王泽宇,等. 结合改进 CBAM 和 MobileNetV2 算法对小麦病斑粒分类[J]. 南京农业大学学报,2024,47(3):583-591.
REN Zhizhou,LIANG Kun,WANG Zeyu, et al. Combining improved CBAM and MobileNetV2 algorithms for classifying diseased wheat kernels[J].
Journal of Nanjing Agricultural University,2024,47(3):583-591.

结合改进 CBAM 和 MobileNetV2 算法对小麦病斑粒分类

任治洲,梁琨*,王泽宇,张群,郭雅欣,郭嘉琦

(南京农业大学人工智能学院,江苏 南京 210031)

摘要:[目的]小麦病斑粒智能检测对于高效、快速、准确评估麦粒的品级有重要意义。现有小麦病斑粒分类的神经网络存在参数数量大、运算复杂等缺点,不便于在移动端部署模型,影响了小麦病斑粒现场分类的效率。本文提出了一种用于小麦病斑粒分类的轻量级神经网络算法。[方法]本研究在轻量化网络 MobileNetV2 的基础上进行改进,加入改进的 CBAM(convolutional block attention module)注意力机制,对改进的模型全整型量化后在边缘计算设备上部署。利用该模型对 4 种小麦籽粒(赤霉病粒、腥黑穗病粒、破损粒和正常粒)分类。[结果]相比于改进前的 MobileNetV2 网络,结合改进的注意力机制和 MobileNetV2 网络在准确率、精准率、召回率上分别提升 3.15%、3%和 3%;全整型量化后的改进模型对小麦赤霉病粒、腥黑穗病粒、破损粒和正常粒的识别准确率分别达到 99%、94%、99%和 96%。该模型大小仅有 2.36 MB,在边缘计算设备的单次推理时间仅为 96.95 ms。[结论]本文改进后的算法模型的准确率提升、大小减少、推理速度加快,可为小麦病斑粒分类模型去冗余提供指导。

关键词:小麦病斑粒;注意力机制;轻量级神经网络;全整型量化

中图分类号:TP18;S858.31

文献标志码:A

文章编号:1000-2030(2024)03-0583-09

Combining improved CBAM and MobileNetV2 algorithms for classifying diseased wheat kernels

REN Zhizhou, LIANG Kun*, WANG Zeyu, ZHANG Qun, GUO Yaxin, GUO Jiaqi

(College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China)

Abstract:[Objectives] Intelligent detection of diseased wheat kernels is important for the efficient, rapid, and accurate evaluation of wheat kernel quality. Existing deep neural network models for the classification of diseased wheat kernels have disadvantages such as large numbers of parameters and complexity of operations, which make it unsuitable for the deployment of the model on edge computing devices, thus affecting the efficiency of on-site classification of diseased wheat kernels. In this paper, a lightweight neural network algorithm for diseased wheat kernel classification was proposed. [Methods] In this study, the model was developed based on the lightweight network MobileNetV2 and added an improved CBAM (convolutional block attention module) attention mechanism. The improved model was fully integer quantized and deployed to mobile devices. Moreover, the proposed model was applied to classify four types of wheat kernels (fusarium-damaged kernels, common bunt of wheat kernels, broken kernels, and normal kernels). [Results] Compared to the previous MobileNetV2 network, the model combining the improved attention mechanism and MobileNetV2 network was improved, and the accuracy, precision, and recall rates for the model were improved by 3.15%, 3% and 3%, respectively. The improved model after full integer quantization achieved 99%, 94%, 99% and 96% recognition accuracy for fusarium-damaged kernels, common bunt of wheat kernels, broken kernels, and normal kernels, respectively. In addition, the size of the model was 2.36 MB, and the single inference time of this model at the edge computing device was 96.95 ms. [Conclusions] The improved algorithm of this paper has increased the model accuracy, reduced the size of the model, and accelerated the model inference speed. This study can provide guidance for the de-redundancy of fusarium-damaged kernel classification models.

Keywords: diseased wheat kernels; attention mechanism; lightweight neural networks; full integer quantization

小麦是我国的三大储备粮食之一,麦粒质量的精准检测对我国经济发展和食品安全具有重要意义^[1]。麦粒品质是影响小麦价格的重要因素之一,小麦病斑粒数量是衡量麦粒品质的限制指标,小麦病斑粒包括黑胚粒与赤霉病粒。目前,病斑粒的检测方式主要采用人工检测和图像检测。其中人工麦粒检

收稿日期:2023-08-02

基金项目:江苏省自然科学基金项目(BK20221518);江苏省农业科技自主创新资金项目[CX(23)1002]

*通信作者:梁琨,副教授,研究方向为农产品无损检测,E-mail:kliang@njau.edu.cn。

测效率低、耗时长、检测成本高;图像检测方面,研究人员采用计算机视觉与深度学习相结合的技术在麦粒品质分级^[1]、病斑粒检测^[2]方面进行了研究,对后续实现小麦病斑粒移动端检测有着重要意义。

神经网络算法应用于小麦病斑粒识别成为当下研究的热点。祝诗平等^[2]利用深度神经网络 AlexNet 实现了小麦破损籽粒的筛选,识别准确率达 96.67%。冯继克等^[3]通过特征选择实现 6 个品种小麦籽粒的多分类,识别准确率达 94.26%。但基于深度神经网络的麦粒识别方法存在很多不完善的地方,网络层数深、参数量大、模型冗余高等缺点导致其难以在移动端设备部署。基于此,轻量级神经网络算法被应用于农业无损检测。张旭等^[4]改进了轻量化模型 ShuffleNet 用于苹果叶部病害诊断,该模型平均推理时间缩减 87.94%。李云霞等^[5]利用可见光谱与改进 MobileNetV2 算法估算冬小麦分蘖数,该方法加快了模型的推理速度。Zhang 等^[6]提出了将注意力机制加入 ResNet 网络用于小麦不完善粒识别,该模型识别准确率高达 97.5%。这些研究表明利用轻量级神经网络、结合注意力机制方法可以有效提高模型识别准确率、降低网络参数量、加快模型推理时间,但尚未提出将 2 种方法结合应用于小麦病斑粒移动端检测的相关研究。

MobileNet 网络是目前农业无损检测领域使用最多的轻量级网络^[7],该网络采用深度可分离卷积的方法,降低模型参数量、加快模型推理速度。MobileNetV2 模型继承了深度可分离卷积方法的优势,在此基础上又引入倒残差与线性瓶颈结构。目前将 MobileNetV2 算法用于小麦病斑粒多分类的研究较少,由于病斑粒种类繁多,且特征不易提取,利用轻量级神经网络会忽略细节特征产生识别误差。故本研究在轻量化网络 MobileNetV2 的基础上,针对网络结构进行改进,加入改进的 CBAM 注意力机制,并对比改进前后网络的检测效果,以提高小麦病斑粒移动端检测能力。同时设计消融试验与 Grad-CAM 神经网络可视化试验验证算法的有效性。将改进的 MobileNetV2 模型通过全整型量化并部署在边缘计算设备上可以实现籽粒样本的现场实时检测,为实现小麦品质快速、精确检测提供有效保障。

1 材料与方法

1.1 试验数据采集与分析

1.1.1 数据采集 选用由江苏省粮油质量检测中心提供的白硬麦作为试验材料,其收获时间为 2023 年 6 月。按照国家标准《小麦:GB 1351—2023》将小麦人工分为正常粒、赤霉病粒、腥黑穗病粒、破损粒 4 类,分类标准见表 1。在自主搭建的试验箱内,以黑色绒布板作为背景,将每一批 100 粒小麦按照 10×10 的分布进行随机摆放。本试验图像采集的设备为迈德威视 MV-GE1600-T-CL 型千兆网工业相机,相机像素为 1 600 万,焦距为 6 mm。将相机垂直固定在距离待测籽粒 25 cm 处。从上述 4 种小麦籽粒中各选取 1 700 粒,共采集 6 800 幅麦粒图像作为试验数据集。

表 1 小麦籽粒分类标准

Table 1 Classification standards of wheat kernel

类型 Types	特征 Feature
正常粒 Normal kernels	形态呈椭圆形或长圆形,略扁平,籽粒颜色通常为淡黄色或浅棕色,且籽粒表面光滑圆润 The shape of the kernels is oval or oblong, slightly flat, the color of the kernels is usually light yellow or light brown, and the surface of the kernels is smooth and rounded
赤霉病粒 Fusarium-damaged kernels	籽粒皱缩、呆白,粒面呈紫色,或有明显的粉红色霉状物 The kernels are wrinkled, and dull-white, the surface of the kernels is purple, or has an obvious pink mold
腥黑穗病粒 Common bunt of wheat kernels	籽粒比正常粒短,体型更粗,外表存在灰白色外膜,碾碎后内部充满黑色粉末,且散发出浓烈的鱼腥臭味 The kernels are shorter and thicker than normal, with a grayish-white outer membrane, and when crushed, the interior is filled with black powder and emits a strong fishy odor
破损粒 Broken kernels	压扁、破损,伤及胚与胚乳的籽粒 The kernels are crushed, and broken, and the embryo and endosperm of the kernels are injured

1.1.2 数据集制作与划分 对工业相机拍摄的原始数据集进行灰度化、二值化,对二值化后的单个麦粒图像求其最小外接矩形坐标,利用坐标从原始数据集中分割出单个小麦籽粒图像并保存,将单个小麦籽粒图像归一化成 224×224 像素的 RGB 图像用作模型训练。对归一化后的 6 800 幅麦粒图像,按照 8:1:1 比例划分^[8],其中训练集 5 440 个样本,验证集 680 个样本,测试集 680 个样本。图像分割后的数据集样本如图 1 所示。

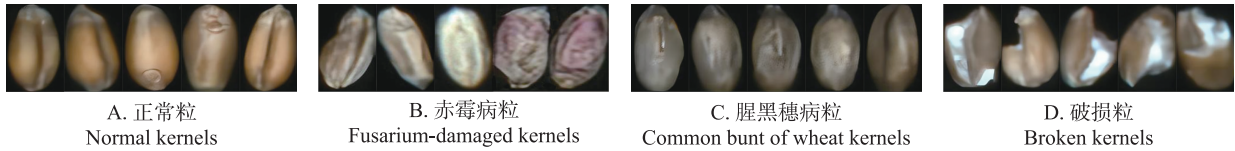


图 1 图像分割数据集样本

Fig. 1 Image segmentation of dataset sample

1.2 MobileNetV2 模型结构

神经网络模型复杂度高、推理时间长,而轻量级神经网络 MobileNetV2 模型参数量小、复杂度低、推理速度快。MobileNetV2 模型^[7]采用深度可分离卷积、倒残差与线性瓶颈结构。倒残差结构与 ResNet 中的残差结构顺序相反,将普通的 3×3 卷积替换为 1×1 逐点卷积(pointwise convolution, PW)作为扩展层, 3×3 的深度卷积(depthwise convolution, DW)作为特征提取层和 1×1 的逐点卷积作为降维层,并使用瓶颈结构控制其通道数,通过短路连接将输入的低级特征与卷积后的高级特征结合,促进了特征信息的融合。除此之外利用线性激活代替 ReLU 激活,以减少其信息的损失。倒残差与线性瓶颈结构如图 2 所示。

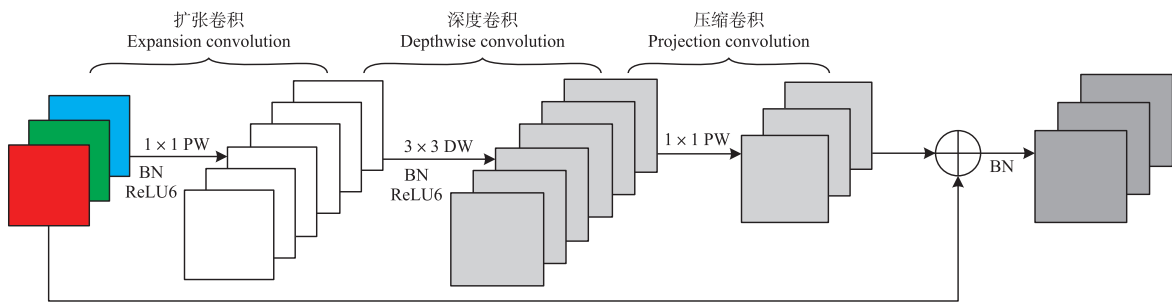


图 2 MobileNetV2 倒残差与线性瓶颈结构图

Fig. 2 MobileNetV2 inverted residual and linear bottleneck structure diagram

PW:逐点卷积 Pointwise convolution;DW:深度卷积 Depthwise convolution;BN:批量归一化 Batch normalization.

1.3 CBAM 注意力机制模块

注意力机制^[9]可以重构神经网络模型权重,弱化次要特征,强调主要特征,用于加强网络特征提取能力。CBAM^[10]是一种轻量化注意力机制,参数量小,可移植性高。CBAM 融合了通道注意力机制(channel attention)与空间注意力机制(spatial attention),可以使模型从通道和空间 2 个维度分别关注对分类起决定作用的像素区域。CBAM 结构如图 3 所示。CBAM 模块对输入的特征图,先分别平均池化和最大池化;再

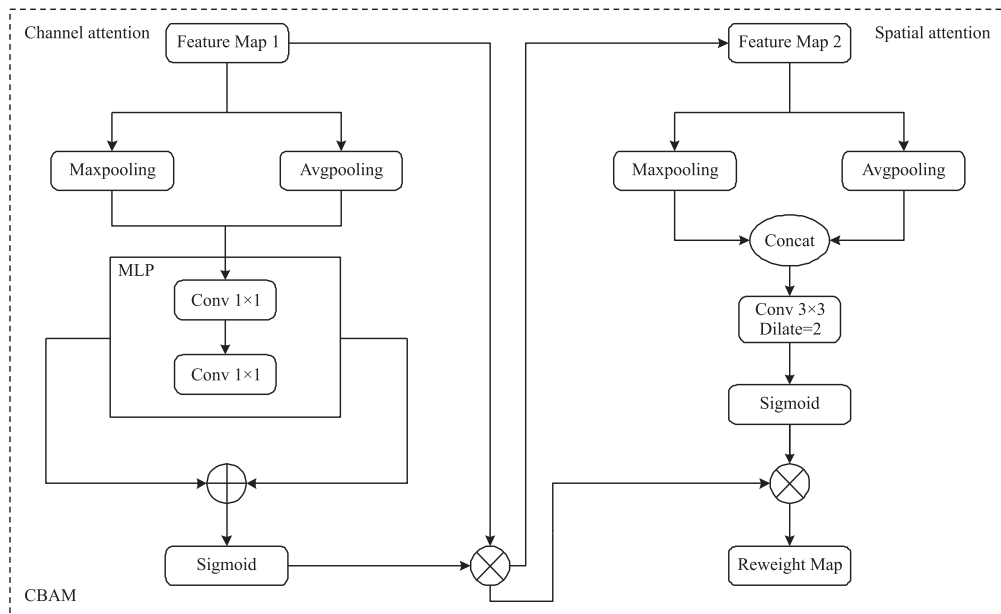


图 3 CBAM 结构图

Fig. 3 CBAM structure diagram

通过同一个多层感知机(multilayer perceptron, MLP)和 Sigmoid 激活函数,学习通道权重;最后将其分别平均池化和最大池化后通过一个 3×3 的空洞卷积和 Sigmoid 激活函数,学习空间各点权重,最终生成权重优化的特征图。

1.4 小麦病斑粒识别模型

针对神经网络推理速度慢、硬件算力需求较高的问题,本研究采用轻量化神经网络模型替代深度神经网络模型完成小麦病斑粒的分类检测,减少模型参数量且充分提取小麦病斑粒图像特征。改进前的 MobileNetV2 模型参数量小,模型识别准确率较低,因此本研究对 MobileNetV2 进行改进,得到改进的模型(以下称为 S-MobileNetV2 模型),其网络结构如图 4 所示。S-MobileNetV2 模型^[11]在网络结构上可以划分为输入端、特征提取网络、分类网络和输出端 4 部分。其特征提取网络修改了 MobileNetV2 末尾三层的 Bottleneck,将其替换成 1×1 的常规卷积,减少模型参数量,同时减少深度网络的计算损失。加入改进的 CBAM,该算法对于输入特征先通过一个共享的线性多层感知机与 Hard-Sigmoid 激活函数,学习通道权重;再通过 3×3 的卷积核与 Hard-Sigmoid 函数,获得空间各点权重,通过 Hard-Sigmoid 函数后不仅能获得良好的跨通道信息,而且该方法无幂计算,可以有效避免梯度消失现象。超参数的设置影响模型整体的性能,经过多次训练试验,本模型优化器采用 Adam 优化器,训练批量大小为 32,训练轮数为 50 次,初始学习率设置为 0.000 1,动量因子设置为 0.9。

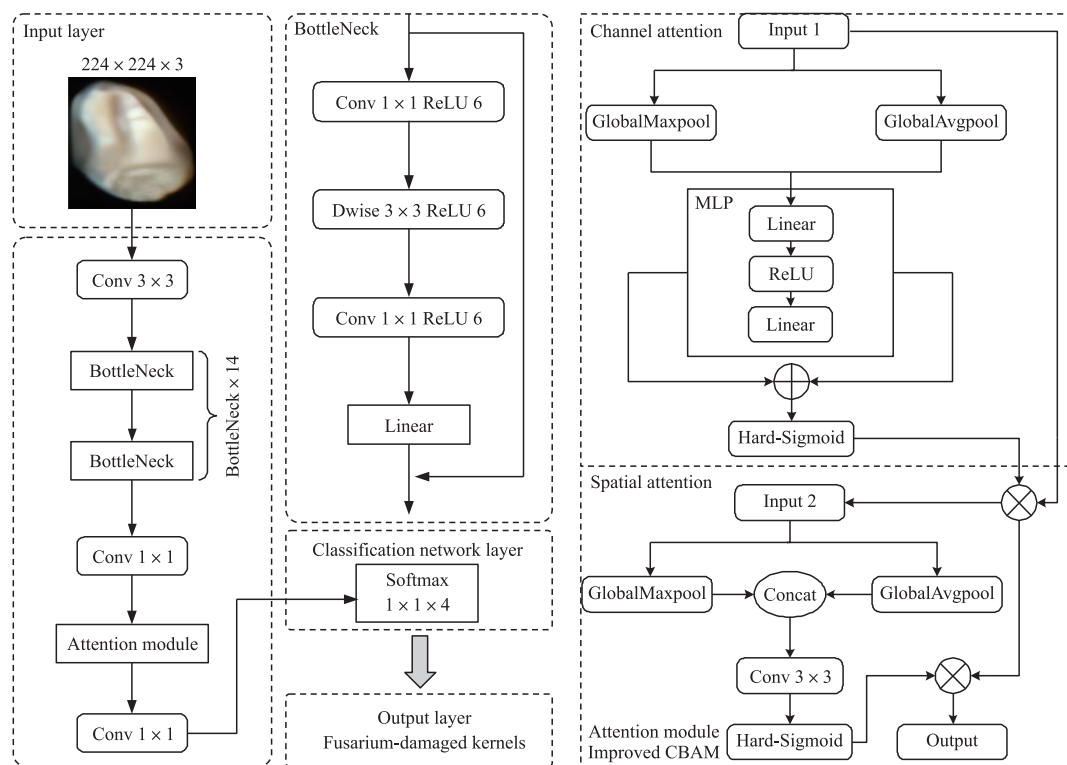


图 4 结合改进的 CBAM 和 MobileNetV2 网络结构图

Fig. 4 Combining improved CBAM and MobileNetV2 network structure diagram

1.5 模型量化与部署

本文针对训练后的模型进行量化^[12],主要采用全整型量化(INT8 型量化)算法,原理如图 5 所示。全整型量化方法可以将权重、激活值与输入值全部量化成 INT8 型,并将所有中间层运算过程置于 INT8 下执行,可以实现内存占用低、识别速度快、模型去冗余的效果。全整型量化将 32 位浮点型输入数据、权重参数量化为 8 位整型,进行卷积运算,运算结果生成 32 位整型数据进行激活,最后再将激活的数据反量化为 32 位浮点型输出。根据上述原理对实际模型进行量化,具体流程如图 6 所示。全整型量化采用代表性数据集进行训练,代表性数据集能够提供模型在典型情况下的统计特性,使模型更好适应实际推理任务。对代表性数据集训练完成后,冻结模型权重,确定量化参数,并按照图 5 所示原理进行权重量化与激活值量化,将权重参数乘以缩放因子并四舍五入至最近整数。量化工作完成后对模型进行重构,针对重构完的模型可以选择微调参数,也可选择直接进行模型评估。

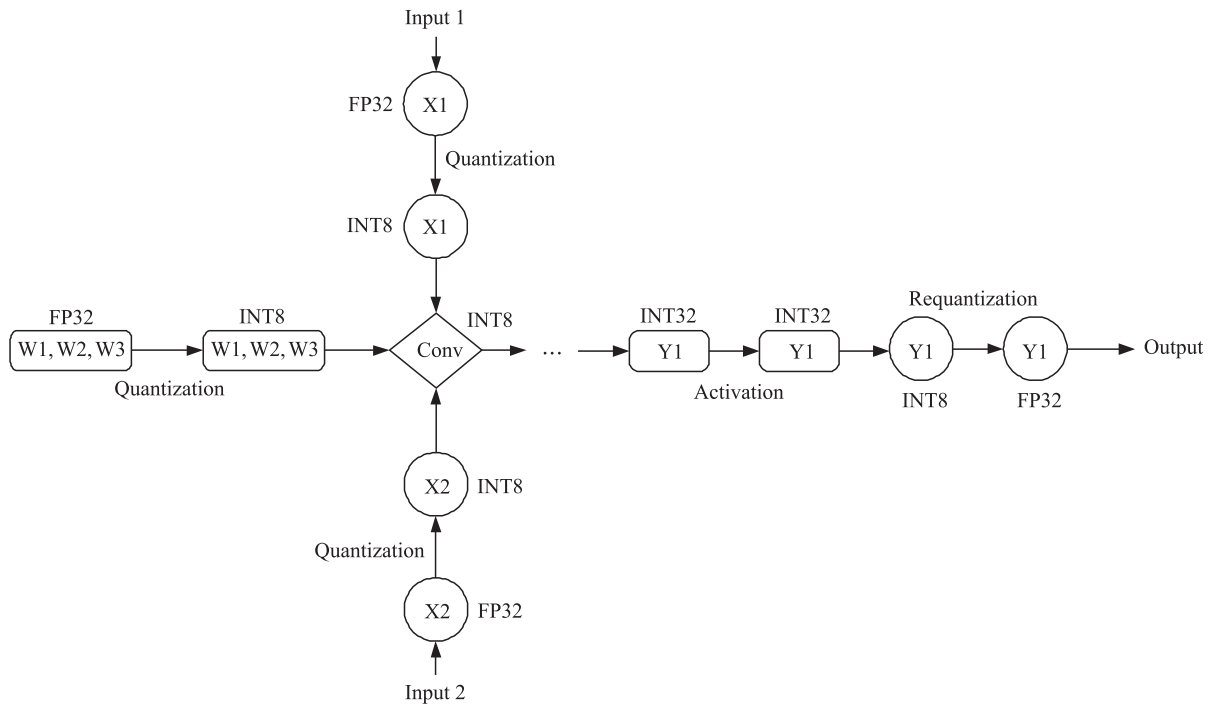


图 5 全整型量化原理图

Fig. 5 Schematic diagram of full integer quantization

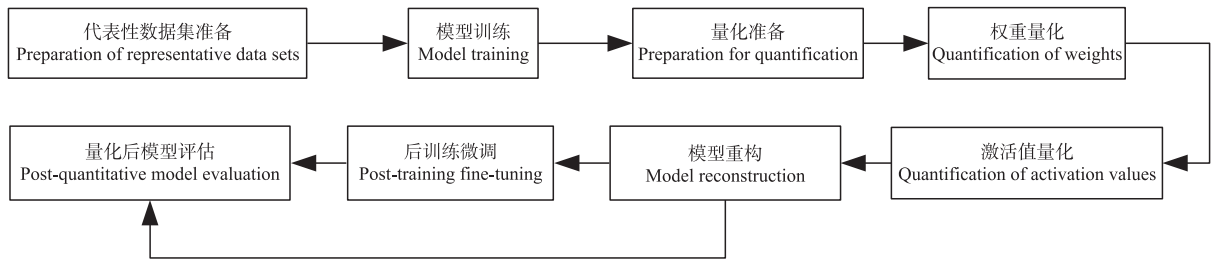


图 6 全整型量化流程图

Fig. 6 Full integer quantization flowchart

1.6 试验平台与环境配置

本研究使用 Windows 10(64 位)操作系统,硬件配置: Intel(R) Xeon(R) Platinum 8338C CPU® 2.60 GHz 处理器, 42 GB 内存, NVIDIA RTX3090, 24 GB 显卡。软件环境为 Pycharm 2020.1、Python 3.8、Tensorflow 2.5.0、CUDA11.2。本研究部署在 Ubuntu18.04 操作系统上,边缘计算设备采用 NVIDIA Jetson Nano,该设备硬件配置: CPU 采用四核 ARM Cortex-A57 MPCore 处理器, 4 GB 内存, GPU 加速采用 128 核 NVIDIA Maxwell 架构,软件环境为 Python3.8、Tensorflow2.5.0。

1.7 模型评价指标

为了验证本试验改进模型的有效性^[13-15],利用混淆矩阵评价模型识别精度效果,混淆矩阵行标签表示预测的真实类别,列标签表示模型推理类别,主对角线上数值越高,其余部分数值越低,则模型识别效果越好。本研究使用模型参数量(parameters)、浮点运算总量(floating point operations per second, FLOPs)、模型推理速度作为衡量模型复杂度的指标;采用准确率(accuracy, A)、精准率(precision, P)、召回率(recall, R)作为评价模型的指标;使用交叉熵损失(cross entropy)评估模型预测输出与真实值的差距。

$$A = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$P = \frac{TP}{TP+FP} \tag{2}$$

$$R = \frac{TP}{TP+FN} \tag{3}$$

式中: TP 表示模型预测为正的正样本数; TN 表示模型预测为负的正样本数; FP 表示模型预测为负的正样本数; FN 表示模型预测为正的负样本数。

2 结果与分析

2.1 改进前后模型结果对比

为对比轻量化注意力机制 CBAM 与 MobileNetV2 模型改进前后效果,进行消融试验对比模型大小、参数量、浮点运算量和准确率^[16-18],结果如表 2 所示。由表 2 可知,加入改进后 CBAM 的 MobileNetV2 网络,模型大小从 26.48 MB 提升至 28.32 MB,准确率从 94.37% 提升至 97.35%,可见改进的 CBAM 虽会增加模型大小,但可以提高模型识别准确率。改进前、后的 MobileNetV2 算法准确率从 94.37% 提升至 96.47%,且模型大小从 26.48 MB 减少至 14.15 MB,可见改进的 MobileNetV2 算法可以提高识别精度且减少模型大小。加入改进后的 CBAM 注意力机制的 MobileNetV2 模型大小从 26.48 MB 减少至 14.80 MB,浮点运算量从 312.38 M 减少至 261.01 M,模型准确率从 94.37% 提升至 98.44%,可见改进后的模型性能明显提升,模型变小,冗余度降低,准确率提升。模型改进前后的准确率变化曲线如图 7 所示,3 种模型在训练迭代 50 次后,逐步达到收敛,可见改进 CBAM 和 MobileNetV2 模型准确率高于改进前的 MobileNetV2 网络,且准确率曲线拟合波动较小,模型收敛效果更好。

表 2 消融试验结果

Table 2 Results of ablation experiment

试验编号 Experiment number	影响因素 Influencing factors		模型大小/MB Model size	参数量/M Parameters	浮点运算量/M FLOPs	准确率/% Accuracy
	网络 Network	注意力机制 Attention mechanism				
1	MobileNetV2		26.48	2.26×10^6	312.38	94.37
2	MobileNetV2	CBAM	28.32	2.41×10^6	317.48	95.74
3	MobileNetV2	改进的 CBAM Improved CBAM	28.32	2.41×10^6	317.48	97.35
4	S-MobileNetV2		14.15	1.19×10^6	260.93	96.47
5	S-MobileNetV2	CBAM	14.80	1.24×10^6	261.01	97.68
6	S-MobileNetV2	改进的 CBAM Improved CBAM	14.80	1.24×10^6	261.01	98.44

对比原 MobileNetV2 模型和改进后的模型对 4 种小麦籽粒的预测准确率^[19]。由表 3 可以看出:改进前模型对腥黑穗病粒、赤霉病粒、破损粒、正常粒的识别率分别为 100%、86%、93%、91%,而改进后的模型对腥黑穗病粒、赤霉病粒、破损粒、正常粒的识别率分别为 99%、94%、96%、99%。本研究中,观察到部分赤霉病籽粒的特征并不显著,这导致原始的 MobileNetV2 模型在此类籽粒的识别上存在较高的误判率。通过对模型进行改进后,赤霉病籽粒的识别准确率从原先的 86% 提升至 94%。然而,对腥黑穗病粒的识别准确率则从 100% 略微下降至 99%,说明改进模型对某些特征较为模糊的腥黑穗病粒仍然存在一定的识别误差。尽管如此,改进后的模型总体展现出更强的预测能力。

表 3 改进前后模型预测准确率对比

Table 3 Comparison of model prediction accuracy before and after improvement

真实样本 Real samples	改进前模型 Pre-improved model				改进后模型 Improved model				%
	腥黑穗病粒 CK	赤霉病粒 FK	破损粒 BK	正常粒 NK	腥黑穗病粒 CK	赤霉病粒 FK	破损粒 BK	正常粒 NK	
腥黑穗病粒 CK	100	0	0	0	99	0	0	1	
赤霉病粒 FK	1	86	0	13	0	94	0	6	
破损粒 BK	9	0	91	0	1	0	99	0	
正常粒 NK	2	4	1	93	0	4	0	96	

Note: CK: Common bunt of wheat kernels; FK: Fusarium-damaged kernels; BK: Broken kernels; NK: Normal kernels.

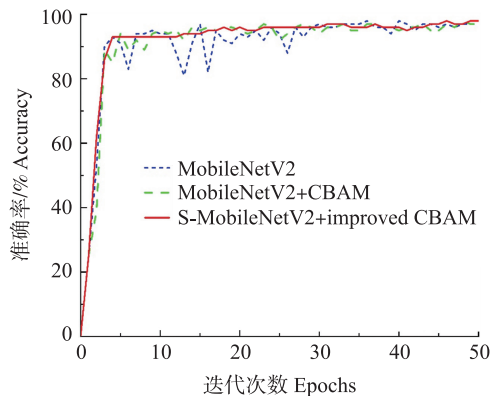


图 7 改进前后 MobileNetV2 的准确率变化曲线
Fig. 7 Accuracy variation curve of MobileNetV2 before and after improvement

2.2 Grad-CAM 网络可视化试验

为了进一步探究 CBAM 注意力机制加入改进模型后的优化效果^[20-21],分析网络特征权重的集中区域,选取部分小麦籽粒数据集图片,采用 Grad-CAM(class activation mapping)方法对改进前后模型进行可视化分析,将特征提取网络最后一层卷积层的特征图可视化,可以直观看出特征权重集中区域。图 8 中红色区域表示该区域对模型分类结果影响最大,表示其权重集中于该像素区域。图 8-A 显示待测籽粒原图像;图 8-B 为改进前模型的激活热力图,改进前的热力图没有将权重集中到图像内部的细微特征,仅关注图像的轮廓特征;图 8-C 为改进后模型的激活热力图,改进后的模型加入注意力机制后,网络更加注意样本籽粒本体的细微特征。图 8-C 中,左 3 列可见红色热区为腥黑穗病粒黑色褶皱凹陷区及黑色空心区,中间 3 列可见红色热区为赤霉病粒的褶皱区及粉红色霉头区,右 3 列可见红色热区为破损粒断层区。对比图 8-B 与图 8-C,改进后模型对小麦籽粒细微特征更加关注,红色高亮区更加集中。表明,改进后的模型在一定程度上提高了模型细微特征提取的能力。

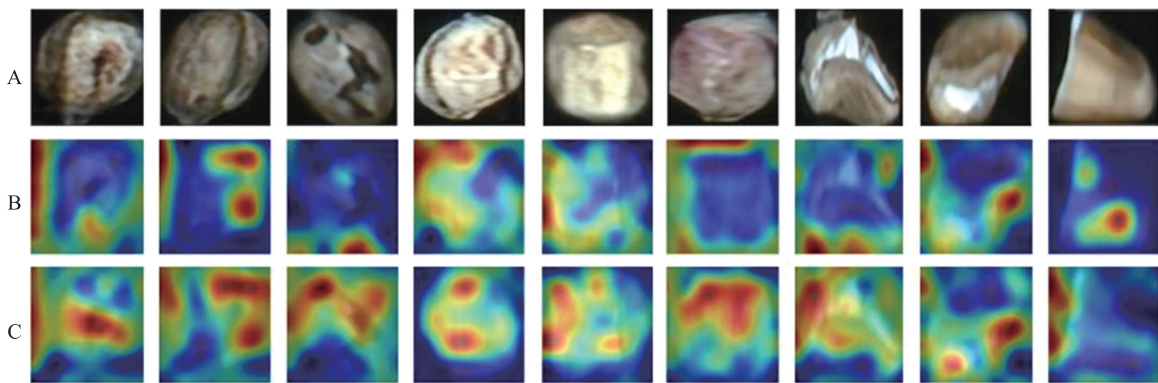


图 8 Grad-CAM 可视化示例

Fig. 8 Visualization example of Grad-CAM

A、B、C 图表示籽粒原图像、改进前模型的激活热力图、改进后模型的激活热力图。图左 3 列为腥黑穗病粒,图中间 3 列为赤霉病粒,图右 3 列为破损粒。

Figures A,B and C represent the original image of the grain, the activation heatmap of the pre-improved model, and the activation heatmap of the post-improved model. The three columns on the left of the figure show the diseased grains with black spikes, the three columns in the middle show the diseased grains with *Fusarium oxysporum*, and the three columns on the right show the damaged grains.

2.3 与其他模型的对比试验

为了进一步验证改进后模型的识别效果,选取近年来表现良好的轻量级模型 MobileNetV3^[22]、ShuffleNetV1^[23]和传统残差网络 ResNet50^[24]与本文改进的模型算法对比,数据集均采用相同的籽粒数据集,硬件配置相同,各类超参数设置相同,迭代 50 次,试验对比结果如表 4 所示。由表 4 可知,ShuffleNetV1 模型最小仅有 9.97 MB;本文改进后的模型其次,模型大小为 14.35 MB;ResNet50 最大,模型大小为 225.38 MB。改进的 MobileNetV2 模型^[25]对于小麦不完善籽粒数据集的识别准确率达 98.44%,相较于 ShuffleNetV1、MobileNetV3、ResNet50 分别提高了 3.15%、3.23%和 1.38%。

表 4 不同模型对比结果

Table 4 Comparison results of different models

模型 Model	模型大小/MB Model size	参数量/M Parameters	浮点运算量/M FLOPs	准确率/% Accuracy	精准率/% Precision	召回率/% Recall	推理时间/ms Inference time
ShuffleNetV1	9.97	1.21×10 ⁶	124.89	95.29	95	95	57.4
MobileNetV2	26.48	2.26×10 ⁶	317.48	94.37	94	94	59.8
MobileNetV3	35.58	3.05×10 ⁶	299.61	95.21	96	95	66.3
ResNet50	225.38	2.53×10 ⁸	1 404.69	97.06	96	97	165.5
S-MobileNetV2-Improved CBAM	14.35	1.24×10 ⁶	261.01	98.44	98	98	84.2

2.4 模型量化效果验证试验

为了验证全整型量化的效果^[26],对改进后的模型进行量化试验,对比不同的量化方法对模型性能的影响,其评估结果如表 5 所示。由表 5 可见:在模型大小方面,就改进后的模型而言,Float16 型量化可使模型从 14.80 MB 减小至 4.56 MB,全整型量化可使模型从 14.80 MB 减小至 2.36 MB;在推理时间方面,全

整型量化模型推理用时最短, X86 环境下推理时间约为 52.01 ms, ARM 环境下推理时间约为 96.95 ms, 全整型量化后模型空间占用小、推理时间短。为得到全整型量化模型的识别效果, 利用相同的测试集针对改进的 MobileNet V2 原模型和量化后模型进行测试评估, 上述 2 种模型对测试集 4 种小麦籽粒识别准确率完全相同, 由此可见全整型量化有助于网络去冗余, 便于将模型部署在边缘计算设备之上。

表 5 改进后模型量化评估表

Table 5 Quantitative evaluation table of model after improvement

量化算法 Quantization algorithm	模型大小/MB Model size	X86 CPU 推理时间/ms Inference time on X86 CPU	ARM CPU 推理时间/ms Inference time on ARM CPU
无量化 Without quantification	14.80	84.21	970.16
Float16 型量化 Float16 quantization	4.65	55.53	97.72
全整型量化 Full integer quantization	2.36	52.01	96.95

3 结论

1) 本文将改进后的 MobileNetV2 模型引入改进的轻量化注意力机制 CBAM, 能使网络更加集中于小麦病斑籽粒的特征信息, 降低其特征冗余度, 减少模型参数量。本文改进后的 CBAM 注意力机制, 增加了网络的稀疏性, 优化了模型的权重参数, 提升了模型的推理准确率。

2) 相较于神经网络与常规轻量级网络, 改进后的模型在准确率、精准率、召回率等评估值上分别提升 3.15%、3%、3%, 模型参数量仅为 1.24×10^6 , 模型大小仅为 14.35 MB, 减少了模型的计算复杂度, 对小麦病斑籽粒识别具有更加出色的性能, 可用于移动端边缘计算平台的部署。

3) 针对本文改进的网络进行全整型量化并在边缘计算设备 NVIDIA Jetson Nano 上部署测试, 全整型量化后的改进模型大小仅有 2.36 MB, 该模型在 Jetson Nano 上的单次推理时间仅为 96.95 ms, 模型推理准确率均与量化前模型相同, 对小麦赤霉病粒、腥黑穗病粒、破损粒和完善粒的识别准确率分别达 99%、94%、99% 和 96%。

4) 本文改进的模型对小麦病斑粒(即赤霉病粒与腥黑穗病粒)识别的平均准确率为 96.5%, 平均准确率小幅度提升。模型大小为 14.35 MB, 相较模型 ResNet 的 225.38 MB, 模型大小大幅度下降。

参考文献 References:

- [1] 张玉荣, 王强强, 吴琼, 等. 基于 Python-OpenCV 图像处理技术的小麦不完善粒识别研究[J]. 河南工业大学学报(自然科学版), 2021, 42(6): 105-112.
Zhang Y R, Wang Q Q, Wu Q, et al. Research on imperfect wheat grain recognition based on Python-OpenCV image processing technology[J]. Journal of Henan University of Technology(Natural Science Edition), 2021, 42(6): 105-112(in Chinese with English abstract).
- [2] 祝诗平, 卓佳鑫, 黄华, 等. 基于 CNN 的小麦籽粒完整性图像检测系统[J]. 农业机械学报, 2020, 51(5): 36-42.
Zhu S P, Zhuo J X, Huang H, et al. Wheat grain integrity image detection system based on CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(5): 36-42(in Chinese with English abstract).
- [3] 冯继克, 郑颖, 李平, 等. 基于特征选择的小麦籽粒品种识别研究[J]. 中国农机化学报, 2022, 43(7): 116-123.
Feng J K, Zheng Y, Li P, et al. Research on the identification of wheat grain varieties based on feature selection[J]. Journal of Chinese Agricultural Mechanization, 2022, 43(7): 116-123(in Chinese with English abstract).
- [4] 张旭, 周云成, 刘忠颖, 等. 基于改进 ShuffleNetV2 模型的苹果叶部病害识别及应用[J]. 沈阳农业大学学报, 2022, 53(1): 110-118.
Zhang X, Zhou Y C, Liu Z Y, et al. Identification and application of apple leaf diseases based on improved ShuffleNetV2 model[J]. Journal of Shenyang Agricultural University, 2022, 53(1): 110-118(in Chinese with English abstract).
- [5] 李云霞, 马浚诚, 刘红杰, 等. 基于可见光谱与轻量级卷积神经网络的冬小麦分蘖数估算[J]. 光谱学与光谱分析, 2023, 43(1): 273-279.
Li Y X, Ma J C, Liu H J, et al. Tillering number estimation of winter wheat based on visible spectrogram and lightweight convolutional neural network[J]. Spectroscopy and Spectral Analysis, 2023, 43(1): 273-279(in Chinese with English abstract).
- [6] Zhang W W, Ma H M, Li X H, et al. Imperfect wheat grain recognition combined with an attention mechanism and residual network[J]. Applied Sciences, 2021, 11(11): 5139.
- [7] 张鹏程. 基于 MobileNetV2 的轻量化柑橘害虫识别方法研究[D]. 武汉: 华中农业大学, 2023.
Zhang P C. Research on lightweight citrus pest identification method based on MobileNetV2[D]. Wuhan: Huazhong Agricultural University, 2023(in Chinese with English abstract).

- [8] 田浩楠,华婧伊,张少帅,等. 基于红外热成像与线性回归拟合的母猪体温检测技术研究[J]. 智能化农业装备学报(中英文),2023,4(1):36-41.
Tian H N, Hua J Y, Zhang S S, et al. Research on the measurement of sow body temperature based on infrared thermography and linear regression fitting[J]. Journal of Intelligent Agricultural Mechanization,2023,4(1):36-41(in Chinese with English abstract).
- [9] 彭红星,徐慧明,刘华籍. 融合双分支特征和注意力机制的葡萄病虫害识别模型[J]. 农业工程学报,2022,38(10):156-165.
Peng H X, Xu H M, Liu H N. Model for identifying grape pests and diseases based on two-branch feature fusion and attention mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering,2022,38(10):156-165(in Chinese with English abstract).
- [10] 王欣,王美丽,边党伟. 融合 MobileNetV2 和注意力机制的轻量级人像分割算法[J]. 计算机工程与应用,2022,58(7):220-228.
Wang X, Wang M L, Bian D W. Algorithm for portrait segmentation combined with MobileNetV2 and attention mechanism[J]. Computer Engineering and Applications,2022,58(7):220-228(in Chinese with English abstract).
- [11] 张栋,姜媛媛. 基于改进 MobileNetV2 的钻杆计数方法[J]. 工矿自动化,2022,48(10):69-75.
Zhang D, Jiang Y Y. Drill pipe counting method based on improved MobileNetV2[J]. Journal of Mine Automation,2022,48(10):69-75(in Chinese with English abstract).
- [12] Shen Y, Yin Y X, Li B, et al. Detection of impurities in wheat using terahertz spectral imaging and convolutional neural networks[J]. Computers and Electronics in Agriculture,2021,181:105931.
- [13] 王志强,于雪莹,杨晓婧,等. 基于 WGAN 和 MCA-MobileNet 的番茄叶片病害识别[J]. 农业机械学报,2023,54(5):244-252.
Wang Z Q, Yu X Y, Yang X J, et al. Tomato leaf diseases recognition based on WGAN and MCA-MobileNet[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(5):244-252(in Chinese with English abstract).
- [14] Guo M, Ma Y T, Yang X J, et al. Detection of damaged wheat kernels using an impact acoustic signal processing technique based on Gaussian modelling and an improved extreme learning machine algorithm[J]. Biosystems Engineering,2019,184:37-44.
- [15] 宋泰年,秦伟伟,梁卓,等. 面向轻量化网络的改进双通道注意力机制图像分类方法[J]. 航空兵器,2021,28(5):81-85.
Song T N, Qin W W, Liang Z, et al. Improved dual-channel attention mechanism image classification method for lightweight network[J]. Aero Weaponry,2021,28(5):81-85(in Chinese with English abstract).
- [16] 杜晓冬,滕光辉,刘慕霖,等. 基于轻量级卷积神经网络的种鸡发声识别方法[J]. 农业机械学报,2022,53(10):271-276.
Du X D, Teng G H, Liu M L, et al. Recognition method of breeding birds' vocalization based on lightweight convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(10):271-276(in Chinese with English abstract).
- [17] 孙俊,朱伟栋,罗元秋,等. 基于改进 MobileNet-V2 的田间农作物叶片病害识别[J]. 农业工程学报,2021,37(22):161-169.
Sun J, Zhu W D, Luo Y Q, et al. Recognizing the diseases of crop leaves in fields using improved MobileNet-V2[J]. Transactions of the Chinese Society of Agricultural Engineering,2021,37(22):161-169(in Chinese with English abstract).
- [18] 孟亮,郭小燕,杜佳举,等. 一种轻量级 CNN 农作物病害图像识别模型[J]. 江苏农业学报,2021,37(5):1143-1150.
Meng L, Guo X Y, Du J J, et al. A lightweight CNN model for image recognition of crop disease[J]. Jiangsu Journal of Agricultural Sciences,2021,37(5):1143-1150(in Chinese with English abstract).
- [19] Javanmardi S, Miraei-Ashtiani S H, Verbeek F J, et al. Computer-vision classification of corn seed varieties using deep convolutional neural network[J]. Journal of Stored Products Research,2021,92:101800.
- [20] Yang X J, Guo M, Lyu Q S, et al. Detection and classification of damaged wheat kernels based on progressive neural architecture search[J]. Biosystems Engineering,2021,208:176-185.
- [21] Lingwal S, Bhatia K K, Tomer M S. Image-based wheat grain classification using convolutional neural network[J]. Multimedia Tools and Applications,2021,80(28):35441-35465.
- [22] Koonce B. MobileNetV3[M]//Convolutional Neural Networks with Swift for Tensorflow. Berkeley, CA: Apress,2021:125-144.
- [23] Angelova A, Zhu S H. Efficient object detection and segmentation for fine-grained recognition[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. June 23-28,2013, Portland, OR, USA. IEEE,2013:811-818.
- [24] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). June 27-30,2016, Las Vegas, NV, USA. IEEE,2016:770-778.
- [25] 黄乾峰,董琴,韦静. 改进 MobileNetV2 算法的番茄叶片病害种类识别[J]. 计算机系统应用,2023,32(1):385-391.
Huang Q F, Dong Q, Wei J. Identification of tomato leaf disease varieties by improved MobileNetV2 algorithm[J]. Computer Systems and Applications,2023,32(1):385-391(in Chinese with English abstract).
- [26] 杨国威,许志旺,房臣,等. 融合剪枝与量化的目标检测网络压缩方法[J]. 计算机工程与应用,2022,58(22):108-115.
Yang G W, Xu Z W, Fang C, et al. Object detection network compression method based on pruning and quantization[J]. Computer Engineering and Applications,2022,58(22):108-115(in Chinese with English abstract).
- [27] 张博. 基于深度学习的小麦外观品质机器视觉检测研究[D]. 杨凌:西北农林科技大学,2019.
Zhang B. Machine vision inspection of wheat appearance quality based on deep learning[D]. Yangling: Northwest A&F University,2019(in Chinese with English abstract).