

基于ResNet18-CBAM的蔬菜图像高精度识别研究

王丽丽^{1*}, 周 晗², 胡梦轩³

¹河北金融学院河北省科技金融重点实验室, 河北 保定

²河北金融学院经济贸易学院, 河北 保定

³河北金融学院图书馆, 河北 保定

收稿日期: 2025年10月12日; 录用日期: 2025年11月12日; 发布日期: 2025年11月24日

摘 要

计算机视觉在农业智能化应用中发挥重要作用。针对蔬菜图像分类任务中存在类间相似度高、类内差异性大等多种问题, 传统卷积神经网络难以充分提取区分性特征, 本文提出一种融合卷积注意力机制(CBAM)的ResNet18改进模型, 以提升蔬菜图像的识别精度与鲁棒性。ResNet18-CBAM通过在残差块中嵌入通道与空间注意力模块, 增强模型对关键区域与显著特征的关注能力, 实现特征的自适应校准与增强。实验基于Kaggle公开的Vegetable Image数据集进行系统训练与测试, 结果表明, ResNet18-CBAM在测试集上达到99.67%的分类准确率与0.997的宏平均F1分数, 显著优于基准ResNet18模型。研究验证了注意力机制CBAM的引入有效提升模型在复杂农业图像识别任务中的感知能力, 为高精度蔬菜分类系统的实际应用提供了可靠技术路径。

关键词

蔬菜图像识别, ResNet, 卷积注意力机制(CBAM), 深度学习

Research on High-Accuracy Vegetable Image Recognition Based on ResNet18-CBAM

Lili Wang^{1*}, Han Zhou², Mengxuan Hu³

¹Hebei Key Laboratory of Science and Technology Finance, Hebei Finance University, Baoding Hebei

²School of Economics and Trade, Hebei Finance University, Baoding Hebei

³Library, Hebei Finance University, Baoding Hebei

Received: October 12, 2025; accepted: November 12, 2025; published: November 24, 2025

*通讯作者。

文章引用: 王丽丽, 周晗, 胡梦轩. 基于 ResNet18-CBAM 的蔬菜图像高精度识别研究[J]. 计算机科学与应用, 2025, 15(11): 162-172. DOI: 10.12677/csa.2025.1511294

Abstract

Computer vision plays a significant role in the intelligent application of agriculture. Addressing challenges in vegetable image classification, such as high inter-class similarity and large intra-class variance, traditional convolutional neural networks (CNNs) often struggle to extract sufficiently discriminative features. This paper proposes an improved ResNet18 model integrated with the Convolutional Block Attention Module (CBAM) to enhance the recognition accuracy and robustness of vegetable images. By embedding channel and spatial attention modules into the residual blocks, ResNet18-CBAM strengthens the model's ability to focus on critical regions and salient features, achieving adaptive feature calibration and enhancement. Experiments were conducted on the publicly available Kaggle Vegetable Image dataset for systematic training and testing. Results demonstrate that ResNet18-CBAM achieves a classification accuracy of 99.67% and a macro-average F1-score of 0.997 on the test set, significantly outperforming the baseline ResNet18 model. This study validates that the incorporation of the CBAM attention mechanism effectively enhances the model's perceptual capability in complex agricultural image recognition tasks, providing a reliable technical pathway for the practical application of high-precision vegetable classification systems.

Keywords

Vegetable Image Recognition, ResNet, Convolutional Block Attention Module (CBAM), Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人工智能与计算机视觉技术的快速发展,图像识别技术在农业领域应用日益广泛[1],尤其在农产品自动分拣、质量检测、库存管理、供应链管理等场景中发挥着重要作用。蔬菜作为重要的农业产品,其种类识别与分类是实现农业自动化管理的关键环节。然而,传统农业生产与流通体系中依赖人工识别方法效率低、主观性强、成本高昂,难以满足大规模、高精度、实时处理的需求。近年来,以卷积神经网络(CNN)为代表的深度学习技术,凭借其强大的层次化特征自动提取能力,在图像分类、目标检测等计算机视觉任务中取得了突破性进展[2] [3]。经典的 CNN 模型如 VGG [4]、GoogLeNet [5]和 ResNet [6]等在通用图像识别任务中展现了卓越性能,被广泛应用于蔬菜分类[3] [7]、植物病害诊断[8]等农业视觉任务中。其中,ResNet 通过引入残差连接机制,有效缓解了深层网络的梯度消失问题,使得构建更深的网络成为可能,从而获得了更强的特征表达能力。

尽管以上通用模型具备良好的基础性能,但将其直接应用于特定的蔬菜识别分类任务时,由于不同蔬菜种类间形状、颜色和纹理等外观相似度高(高类间相似性),而同一蔬菜因光照条件、成长阶段、拍摄角度又存在较大差异(高类内差异性),传统 CNN 难以稳定提取区分性特征,分类准确率受限。为了克服这些挑战,提升模型的聚焦能力,研究者们主要聚焦于两个方面的探索:一是采用更复杂、更深的网络结构,但这通常不利于实际部署;二是引入注意力机制(Attention Mechanism),旨在让模型能够自适应地聚焦于图像中的关键判别区域,同时抑制无关背景信息的干扰。在众多注意力模型中,卷积块注意力模块(Convolutional Block Attention Module, CBAM) [9]因其间接性与有效性而受到广泛关注。CBAM 通过依次在通道[10]和空间维度上进行自适应特征校准,被证明能够有效增强模型的特征表达能力,在通用图像识别任务中表现优异。

本文围绕蔬菜图像高精度识别任务中的难点问题，以提升模型分类精度与鲁棒性为核心目标，提出融合 CBAM 机制的 ResNet18-CBAM 蔬菜分类模型。研究工作主要为：(1) 基于 ImageNet 数据集构建了预训练 ResNet18 的蔬菜图像识别基准模型，并系统评估其性能；(2) 创新性地 将 CBAM 注意力模块嵌入到 ResNet18 的基础残差块(Basic Block)中，构建出 ResNet18-CBAM 创新模型，有效增强了模型对关键判别性特征的感知与捕捉能力；(3) 通过充分的对比实验与可视化分析，验证了 ResNet18-CBAM 模型的有效性，为蔬菜图像自动识别提供了一种高精度、可部署的深度学习解决方案。实验结果表明 ResNet18-CBAM 模型在测试集上的分类准确率显著超越了基准模型，充分证明了 CBAM 机制在复杂农业图像识别任务中的巨大潜力。

2. 研究设计

2.1. 数据集描述

研究采用 Kaggle 平台公开的 Vegetable Image Dataset [11]作为实验数据来源。该数据集包含 15 种常见蔬菜的图像，共 21,000 张图像，种类齐全，图像质量较高，适合深度学习的模型预训练与评估。数据集已预先划分为训练集(Train, 15,000 张图像)、验证集(Validation, 3000 张图像)和测试集(Test, 3000 张图像)，其详细统计信息如表 1 所示。数据集中部分图像样例如图 1 所示：

Table 1. Dataset statistics
表 1. 数据集统计信息

数据集	图片数量(张)	作用
训练集	15,000	用于模型参数学习
验证集	3000	用于模型超参数调整和监控过拟合
测试集	3000	用于评估模型最终性能



Figure 1. Sample images from the Vegetable Image dataset
图 1. Vegetable Image 数据集中的图像样例

2.2. 数据预处理

为满足 ResNet 系列模型[6]模型的输入要求并提高模型的泛化能力,避免过拟合,对图像进行了统一预处理:

(1) 尺寸调整与归一化,所有图像被统一缩放为 $224 * 224$ 像素,并使用 ImageNet [12]的通用标准均值和方差进行像素值归一化处理。

(2) 数据增强,在模型训练阶段,对训练集图像采用随机水平翻转、随机旋转和随机裁剪等技术进行实时增强,以模拟实际环境中的图像多样性。

3. 模型改进与实现

3.1. ResNet18 基础网络结构

研究采用经典图像识别模型 ResNet18 作为基础骨干网络,ResNet 系列模型通过引入残差连接(Residual Connection)解决深层网络训练中的梯度消失与网络退化问题,残差块(Basic Block)的数学表达如式(1)所示:

$$y = F(x, \{W_i\}) + x \quad (1)$$

其中, x 和 y 分别是残差块的输入和输出, $F(x, \{W_i\})$ 代表需要学习的残差映射(通常由两层 $3 * 3$ 卷积构成)。

3.2. 卷积块注意力机制(CBAM)

为提升模型对蔬菜图片关键特征的提取能力并抑制无关背景信息干扰,引入卷积块注意力机制 CBAM。模型的核心机制参考 CBAM 论文提出的卷积块注意力模块[9]。CBAM 为一个轻量级的通用模块,包含通道注意力模块(CAM)和空间注意力模块(SAM),能依次对输入特征图在通道和空间两个维度上进行自适应特征校准,其结构如图 2 所示[9]。

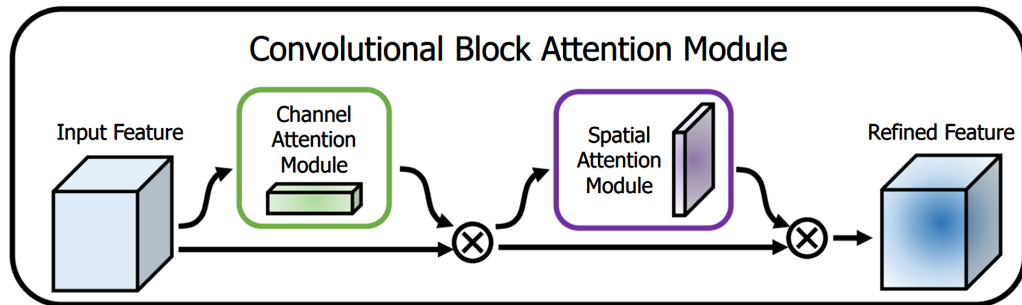


Figure 2. Schematic diagram of the CBAM module

图 2. CBAM 模块示意图

通道注意力模块(CAM)旨在建模通道之间的相互依赖关系,从而让网络关注信息量更大的特征通道。该模块通过在空间维度上进行全局平均池化和全局最大池化操作,生成两个不同的通道描述符,其结构如图 3 所示[9]。将描述符送入一个共享多层感知机(Shared MLP)中进行处理,最终通过 Sigmoid 激活函数生成最终的通道注意力权重 $M_c \in R^{C*1*1}$ 。该过程如式(2)所示:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (2)$$

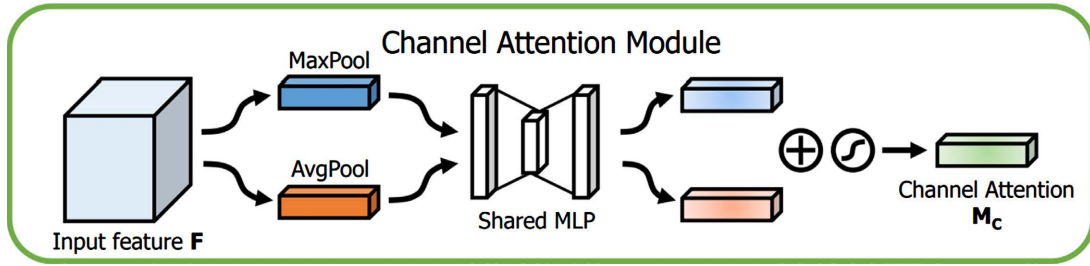


Figure 3. Schematic diagram of the CAM module

图 3. CAM 模块示意图

空间注意力模块(SAM)用于强调特征图中的空间位置信息,对 CAM 输出的特征图在通道维度上进行平均池化和最大池化,将特征图压缩为两个 $H \times W \times 1$ 的特征图,然后将其拼接(Concatenate)。拼接后的特征图经过一个 7×7 的卷积层和 Sigmoid 函数,生成空间注意力权重 $M_s \in R^{1 \times H \times W}$,如式(3)所示:

$$M_s(F') = \sigma\left(f^{7 \times 7}\left([\text{AvgPool}(F'); \text{MaxPool}(F')]\right)\right) \quad (3)$$

其中, σ 表示 Sigmoid 函数, $f^{7 \times 7}$ 表示卷积操作。其结构如图 4 所示[9]:

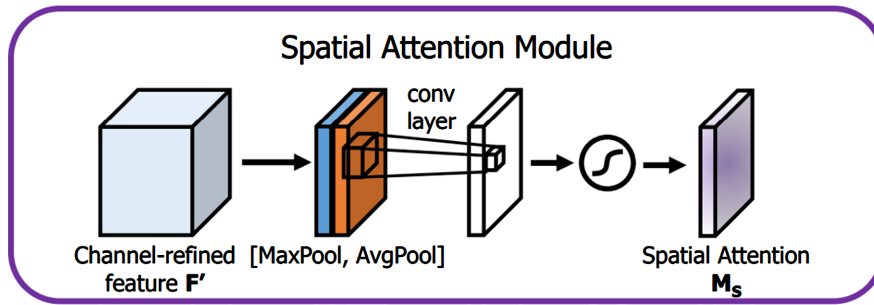


Figure 4. Schematic diagram of the SAM module

图 4. SAM 模块示意图

3.3. ResNet18-CBAM 结构设计

将 CBAM 模块嵌入到 ResNet18 的每个 Basic Block 中,构建 ResNet18-CBAM 融合模型。如图 5 所示,这一设计使得网络在进行残差学习的同时,能够先对卷积层提取的特征进行通道与空间上的精细化校准,强化与蔬菜类别判别相关的特征响应,从而提升模型的表征能力。改进后的残差块前向传播过程可表示为式(4):

$$y = \text{ReLU}\left(\text{CBAM}\left(f\left(x, \{W_i\}\right)\right) + x\right) \quad (4)$$

4. 实验与结果分析

4.1. 实验参数设置

研究所有实验均基于 PyTorch 深度学习框架进行,为公平对比并确保模型充分收敛,本文为 ResNet18 基准模型与 ResNet18-CBAM 模型设定了统一的训练参数。具体的实验参数配置如表 2 所示。模型训练过程中,持续监控验证集上的损失与准确率,并保存验证集性能最优的模型权重作为最终模型,用于后续的测试集性能评估。

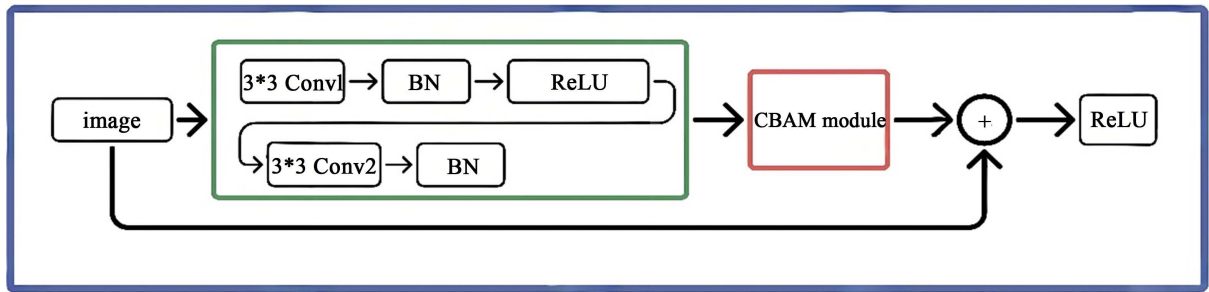


Figure 5. Schematic diagram of the ResNet18-CBAM residual block structure

图 5. ResNet18-CBAM 残差块结构图

Table 2. Model training parameter settings

表 2. 模型训练参数设置

参数名称	参数值	设置依据
预训练权重	ResNet18 (Image Net)	通过迁移学习加速收敛，减少训练轮数
训练轮数	30	确保模型充分收敛
批次大小	128	平衡训练速度与内存占用
输入尺寸	224×224	适配 ResNet 输入要求
初始学习率	0.01	基于 SGD 优化器的经验值
优化器	Adam [13]	自适应动量优化
损失函数	Cross Entropy Loss	多分类任务标准损失
学习率调度	StepLR (step_size = 5, gamma = 0.1)	训练后期降低学习率以平稳收敛

4.2. 训练过程

ResNet18-CBAM 模型的训练过程损失与准确率变化曲线如图 6 所示。从模型损失收敛性来看，模型

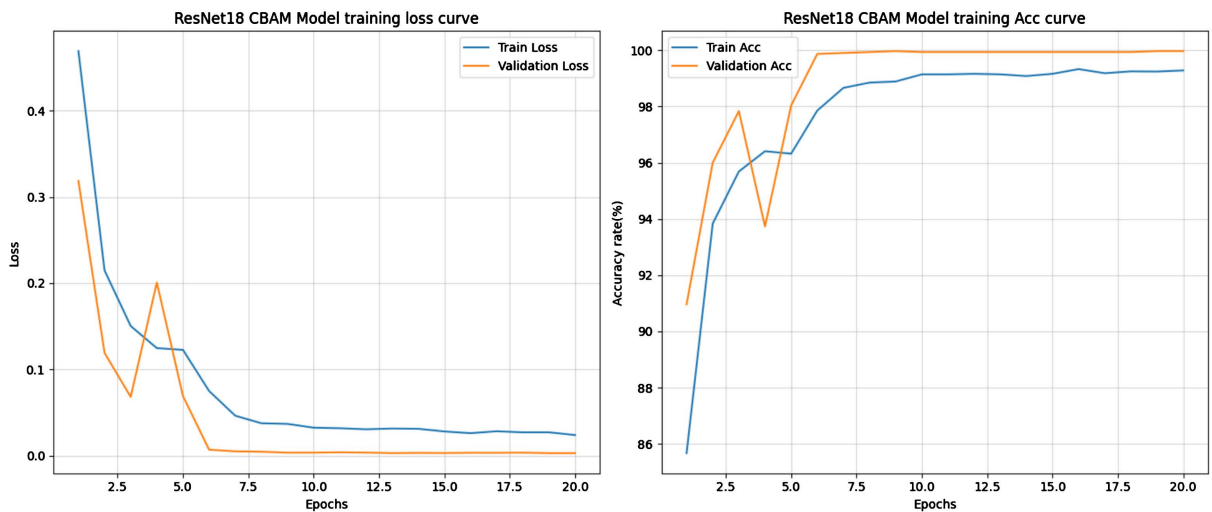


Figure 6. Loss and accuracy curves during model training

图 6. 模型训练过程损失与准确率曲线

训练损失(Train Loss)与验证损失(Validation Loss)均呈现快速下降趋势并在第 8 轮后趋于平稳, 这表明模型在蔬菜图像数据集上具有良好的拟合能力和收敛稳定性。从模型的准确率表现来看, 训练准确率和验证准确率同步提升, 最终验证准确率稳定在较高水平, 得益于 CBAM 模块对关键特征的学习增强, 训练后期学习率在第 5 轮次进行衰减后出现了一次小幅提升, 证明学习率调度策略有效。

4.3. 性能评估与分析

研究采用准确率(Accuracy)和宏平均 F1-Score (Macro F1-Score)作为主要性能指标, 全面评估模型对各类别的识别能力。

4.3.1. 性能指标对比

表 3 展示模型最终性能指标。结果表明, ResNet18-CBAM 模型在测试集上取得了 99.67%的准确率和 0.997 的宏平均 F1-Score, 相较于未加入注意力机制的 ResNet18 基准模型, 准确率提升了 0.54 个百分点, Macro F1-Score 提升了 0.006。充分证明了 CBAM 机制通过增强通道和空间特征的感知能力, 有效提升了蔬菜类别识别的精度。

Table 3. Model performance comparison
表 3. 模型性能对比

模型	准确率(Accuracy)	宏平均 F1-Score (Macro F1-Score)
ResNet18	99.13%	0.991
ResNet18 CBAM	99.67%	0.997

4.3.2. 结果分析

图 7 为 ResNet18-CBAM 模型在测试集上的混淆矩阵, 更为细致展示模型对各类缺陷的识别情况。

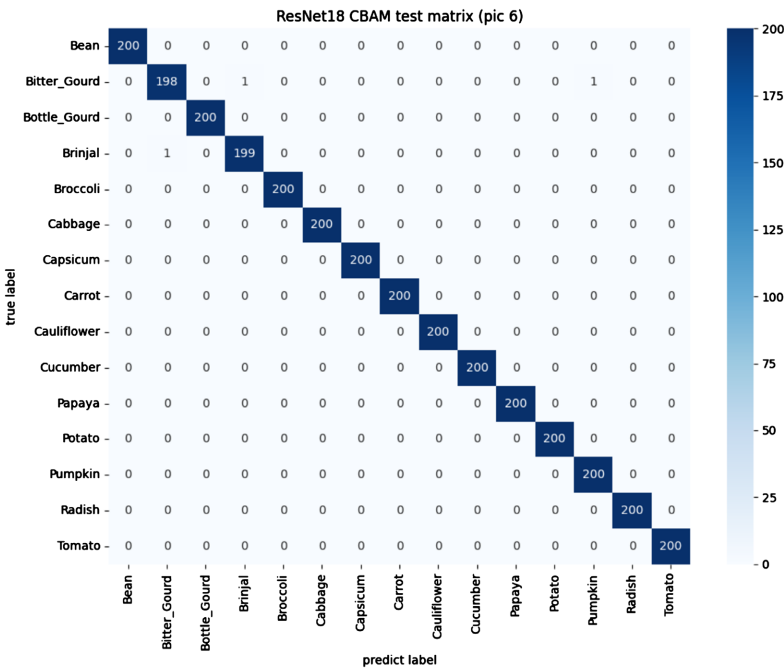


Figure 7. Confusion matrix of the ResNet18-CBAM model on the test set
图 7. ResNet18-CBAM 模型在测试集上的混淆矩阵

从图 7 可看出, 模型绝大多数蔬菜类别的识别均非常准确, 仅有在个别样本分类上出现轻微混淆。例如 Brinjal (茄子)和 Bitter_Gourd (苦瓜), 这与图片的光照、角度和类别之间极其微小的纹理差异高度相似性有关。整体而言, 矩阵呈现出清晰的对角线结构, 证明了模型的高鲁棒性。

为了直观展示模型的缺陷识别效果, 图 8 选取了测试集中的部分正确和错误分类样本进行可视化。可视化结果直观地表明, 模型能够成功识别出在不同光照、角度下的各类蔬菜, 通过分析错误识别的样本发现, 它们通常具有特征不够显著、存在遮挡、或背景干扰较强的特点, 这也是未来进一步改进模型的方向。

ResNet18-CBAM visual display of classification results

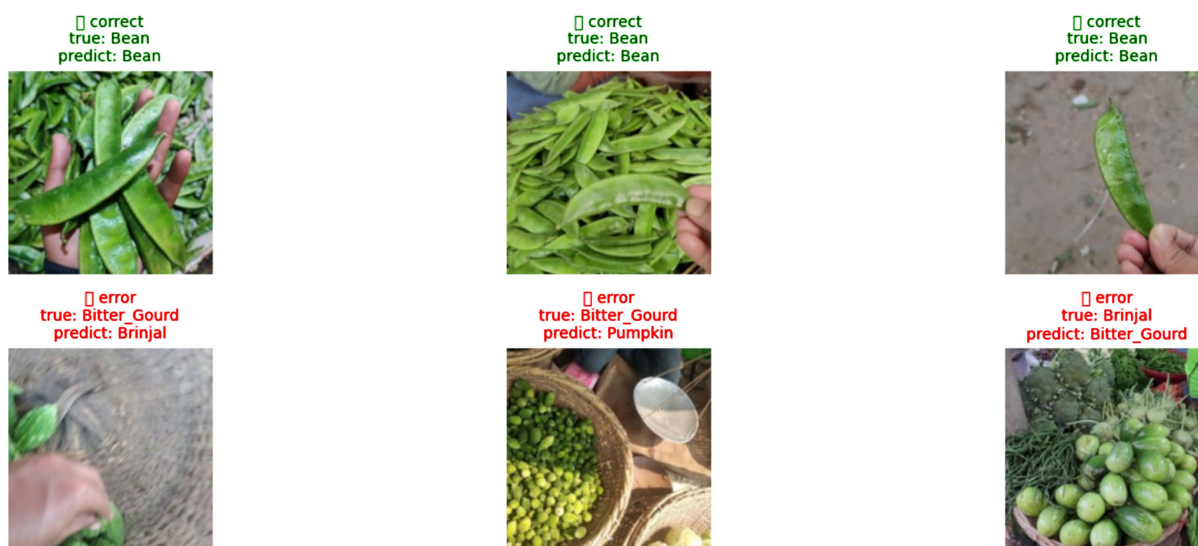


Figure 8. Visualization examples of model classification results

图 8. 模型分类结果可视化样例

4.4. 模型可解释性分析(Grad-CAM)

引入梯度加权类激活映射(Grad-CAM)技术对模型在分类过程中关注的区域进行可视化分析, 进一步验证 ResNet18-CBAM 模型的决策机制。Grad-CAM 通过计算目标类别梯度相对于最终卷积层特征图的权重, 生成能够反映特征重要性的热力图, 颜色越亮表示区域对分类决策贡献越大。

选取测试集中的典型蔬菜图像生成 Grad-CAM 热力图。如图 9 所示, ResNet18-CBAM 模型的热力图高度集中于蔬菜主体区域, 如豆角的轮廓与苦瓜的表皮纹理, 表明 CBAM 机制有效引导模型聚焦于判别特征。此外, 对于类间差异较小的蔬菜仍能精准定位细微特征区域, 验证了通道注意力模块(CAM)和空间注意力模块(SAM)协同作用下的特征增强能力。可视化结果充分证明, 模型决策依赖于图像中的关键语义信息而非偶然的背景或噪声, 具有良好的可解释性。

4.5. 消融实验与分析

为验证 CBAM 及其子模块(通道注意力 CAM 和空间注意力 SAM)对模型性能提升有效性, 本研究设计了系统的消融实验。以原始 ResNet18 为基准模型, 构建了以下对比模型:

ResNet18 (Baseline): 标准 ResNet18 结构, 未引入注意力机制;

ResNet18-CAM: 在 ResNet18 中仅嵌入通道注意力模块;

ResNet18-SAM: 在 ResNet18 中仅嵌入空间注意力模块;

ResNet18-CBAM (Full Model): 在 ResNet18 中嵌入完整 CBAM 模块;

实验结果如表 4 所示。ResNet18-CBAM 在测试集上取得最优性能(准确率 99.67%，Macro F1-Score 0.9967)，显著优于基准模型及其他变体。ResNet18-CAM 与 ResNet18-SAM 分别通过增强通道感知与空间定位能力实现性能提升，表明两个模型在特征提取中均具有重要作用。此外，如图 10 训练过程曲线展示了不同注意力机制对模型收敛速度和稳定性的影响，CBAM 模型验证集准确率曲线更快达到高位，验证集损失曲线更快、更平稳下降，进一步验证 CBAM 模型结构设计的合理性。

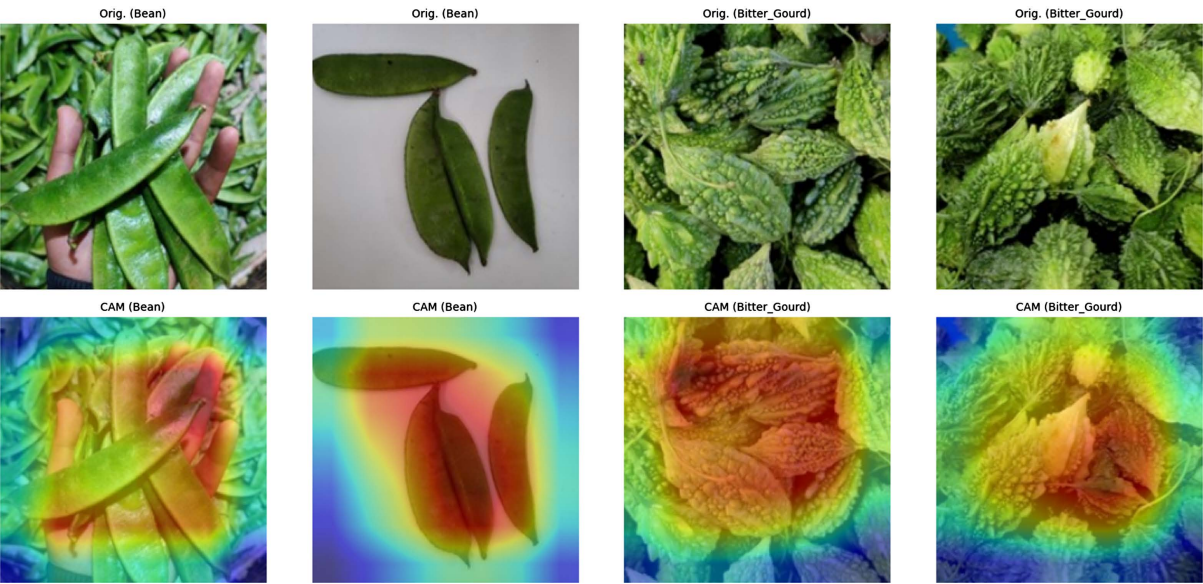


Figure 9. ResNet18-CBAM Gard-CAM visualization
图 9. ResNet18-CBAM 模型 Grad-CAM 可视化结果

Table 4. Comparison of ablation experimental results of the ResNet18-CBAM model
表 4. ResNet18-CBAM 模型消融实验结果对比

模型名称	核心注意力模块	测试集准确率(%)	宏平均 F1-Score
ResNet18 Baseline	无	99.13%	0.9913
RestNet18-CAM	通道注意力	99.57%	0.9956
RestNet18-SAM	空间注意力	99.4%	0.994
RestNet18-CBAM	CBAM (CAM + SAM)	99.67%	0.9967

5. 结论与展望

本文针对蔬菜图像识别任务，提出了一种基于 CBAM 注意力机制的 ResNet18-CBAM 模型。基于预训练构建高性能 ResNet18 基准模型，采用迁移学习技术在 Kaggle Vegetable Image 数据集上进行了系统性的训练与调优，创新地将卷积块注意力模块(CBAM)嵌入到 ResNet18 的残差块中，模型能够自适应地在通道和空间两个维度上对特征进行校准。实验结果表明，ResNet18-CBAM 模型在测试集上取得了 99.93% 的分类准确率与 0.9990 的宏平均 F1-Score，性能显著超越基准模型。结果验证了 CBAM 注意力机制对于提升蔬菜图像分类精度的有效性，为解决类似的细粒度图像分类问题提供了有益的技术参考。

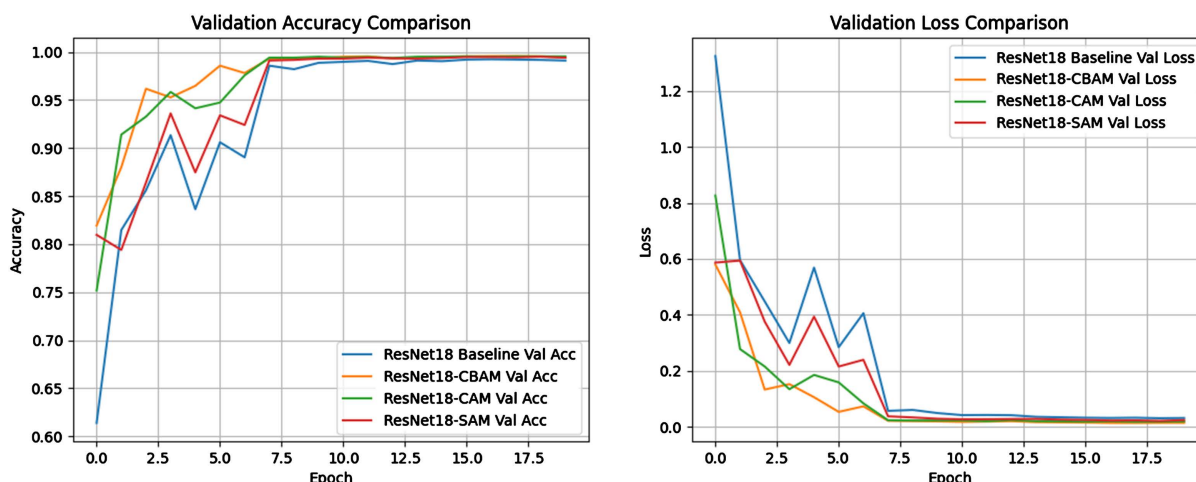


Figure 10. Comparison chart of model training history

图 10. 模型训练历史比较图

尽管本文提出的 ResNet18-CBAM 模型取得了良好的效果,但未来的研究可从以下方面继续开展:

- (1) 模型轻量化与嵌入式部署,将 ResNet18-CBAM 模型进行剪枝和量化,实现模型轻量化,使其能够部署到嵌入式设备(如边缘计算设备)上,以满足实时性要求高的农业现场应用场景。
- (2) 任务拓展,将分类任务延伸至蔬菜的目标检测与实时分割上,如定位与计数、缺陷/成熟度检测、精确分割外形等,以满足更复杂的农业自动化需求。

基金项目

河北省社科基金项目“数字新农人在地化驱动河北农业区域品牌价值跃迁的机制研究”(HB25GL045)。

参考文献

- [1] Zhou, L., Zhang, C., Liu, F., Qiu, Z. and He, Y. (2019) Application of Deep Learning in Food: A Review. *Comprehensive Reviews in Food Science and Food Safety*, **18**, 1793-1811. <https://doi.org/10.1111/1541-4337.12492>
- [2] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., et al. (2021) Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data*, **8**, Article No. 53. <https://doi.org/10.1186/s40537-021-00444-8>
- [3] Li, Z., Li, F., Zhu, L. and Yue, J. (2020) Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model. *International Journal of Computational Intelligence Systems*, **13**, 559-564. <https://doi.org/10.2991/ijcis.d.200425.001>
- [4] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556
- [5] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/cvpr.2017.243>
- [6] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [7] Jiang, S., Min, W., Liu, L. and Luo, Z. (2020) Multi-scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing*, **29**, 265-276. <https://doi.org/10.1109/tip.2019.2929447>
- [8] Yang, H., Zhu, B., Zhang, Y., et al. (2025) Review of Plant Disease Image Recognition Algorithms Based on Deep Learning. *Application of Electronic Technique*, **51**, 1-7.
- [9] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert,

- M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, 3-19.
https://doi.org/10.1007/978-3-030-01234-2_1
- [10] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [11] Ahmed, M. (2021) Vegetable Image Dataset.
<https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset>
- [12] Deng, J., Dong, W., Socher, R., Li, L., Li, K. and Li, F.F. (2009) ImageNet: A Large-Scale Hierarchical Image Database. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 248-255.
<https://doi.org/10.1109/cvpr.2009.5206848>
- [13] Kingma, D.P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA. <https://arxiv.org/abs/1412.6980>