

面向边缘部署的苹果叶部病害轻量化诊断方法

蔡鑫 聂荟珊 杜甜甜 南新元

(新疆大学智能科学与技术学院, 乌鲁木齐 830017)

摘要: 针对高性能苹果叶部病害诊断模型难以在边缘设备高效部署的问题, 提出一种病灶感知的渐进式压缩框架 (L-EAFP)。首先适配一个高性能跨域模型 (EAFP-Med ST) 作为教师模型, 通过病灶加权剪枝和多目标病灶特征锚定蒸馏, 将知识高保真迁移至超轻量级学生模型 (ShuffleNetV2), 最后结合自适应量化完成压缩。在包含 9 263 幅图像的苹果叶部病害数据集上实验结果表明, 所提框架将模型参数量从 3.157×10^7 压缩至 3.500×10^5 (压缩率为 98.9%), 压缩后模型诊断准确率仍保持 98.58%。病灶感知压缩策略在模型轻量化与诊断性能保真度之间取得了良好平衡, 为高性能农业 AI 模型在边缘端的部署提供了有效方案。

关键词: 苹果叶部病害识别; 模型压缩; 病灶感知; 知识蒸馏; 跨域模型迁移

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-1298(2026)08-0235-11

OSID:



Lightweight Diagnostic Method for Apple Leaf Diseases Oriented to Edge Deployment

CAI Xin NIE Huishan DU Tiantian NAN Xinyuan

(School of Intelligent Science and Technology, Xinjiang University, Urumqi 830017, China)

Abstract: Aiming to address the critical bottleneck whereby high-performance convolutional neural networks for apple leaf disease diagnosis are difficult to deploy efficiently on resource-constrained edge devices due to high computational complexity, a novel lesion-aware progressive compression framework (L-EAFP) was proposed. The methodology began by adapting a robust high-performance cross-domain model, EAFP-Med ST, to function as a teacher network. To facilitate high-fidelity knowledge migration to an ultra-lightweight ShuffleNetV2 student model, the framework introduced a lesion-weighted pruning algorithm that prioritized the preservation of pathological feature channels while rigorously eliminating redundancy. Concurrently, a multi-objective lesion feature anchoring distillation technique was applied to strictly align the student model's attentional focus on disease regions. Finally, an adaptive quantization mechanism was incorporated to further compress the model bit-width. Experimental validations performed on a comprehensive dataset of 9 263 apple leaf images demonstrated the framework's effectiveness. The results revealed that L-EAFP reduced the model parameter count from 3.157×10^7 to 3.500×10^5 , achieving a remarkable compression rate of 98.9%. Crucially, despite this drastic reduction, the compressed model retained a diagnostic accuracy of 98.58%. These findings indicated that the proposed lesion-aware strategy successfully achieved an optimal balance between model lightweighting and diagnostic performance fidelity, providing a highly effective technical solution for the deployment of advanced agricultural AI models on mobile and embedded edge terminals.

Key words: apple leaf disease recognition; model compression; lesion-aware; knowledge distillation; cross-domain model adaptation

0 引言

近年来,在数字化浪潮的推动下,现代农业正加速朝着精准化、智能化与自动化方向蓬勃发展^[1]。

作为全球种植范围最广、经济价值最高的果品之一,苹果产业的健康稳定发展对于保障果农增收及农业经济增长具有重要意义^[2]。然而,在苹果漫长生长周期中,其叶片极易受到黑星病、锈病等多种真菌性

病害侵袭,这些病害若不能被及时发现与阻断,不仅会直接削弱果树树势,更会导致果实品质下降与产量锐减^[3]。长期以来,传统的病害诊断主要依赖农业专家或经验丰富的果农进行人工巡查,这种方式不仅劳动强度大、效率低,且受主观经验影响极大,难以覆盖大规模种植区域,常常导致因诊断滞后而错过最佳防治窗口期。因此,利用基于深度学习^[4]的先进计算机视觉技术,替代人工实现对苹果叶部病害的自动化、全天候、精准识别,已成为推动智慧农业转型升级的核心驱动力。

尽管现有的先进视觉诊断模型在实验室环境下已取得了令人瞩目的识别精度^[5-10],但其日益庞大的参数量与高昂的计算开销,使其在向资源极度受限的农业边缘设备迁移时面临着严峻的落地难问题。特别是在中国新疆阿克苏等集约化苹果种植基地,田间作业环境对算法的轻量化提出了极致要求:为满足果农手持诊断终端的硬件约束,模型内存占用需严格控制在 2 MB 以内;为保障无人机大田巡检实时飞行作业,单幅图像推理时间须低于 15 ms。同时,模型须具备在复杂背景下精准识别小于 20 像素的早期微小病斑的能力,这使得高性能模型的直接部署几乎不可能实现。

为此,学者们已尝试从多个维度探索轻量化路径,包括设计更高效的网络架构^[11-13]、优化卷积与注意力机制^[14-15]以及应用知识蒸馏技术^[16]。然而,现有方案普遍暴露出核心缺陷,本质上属于一种病灶不可知 (Lesion-agnostic) 的宏观压缩范式。该范式往往优先优化参数量、浮点运算量等全局统计指标,却忽视了对图像中关键病灶区域的特异性保护。例如,当常规剪枝算法应用于如 EAFP - Med ST^[17]这类精细化模型时,由于算法无法区分病灶前景与叶片背景,常因误裁关键的病斑响应通道,导致对早期小病斑漏检率激增至 15% 以上;即便是专为相似农业任务设计的 YOLO v8 - Corn 模型^[18],在迁移至苹果叶部病害诊断任务时,也因未能有效保持细微病灶特征,导致黑星病漏检率高达 8.7%。这种重宏观指标、轻病灶特征的压缩方式,无法在诊断精度、计算开销与部署速度之间找到平衡点,因此亟需研发一种能够主动感知并保留核心病灶信息的轻量化方案。

为解决上述问题,本文提出 L - EAFP (Lesion-aware progressive compression) 框架,旨在打破高性能医学诊断模型与边缘设备算力限制之间的壁垒。设计一种适配苹果叶部病害的病灶特征锚定蒸馏策略,通过构建分类损失与病斑区域特征损失的联合监督机制,试图突破传统全局特征对齐的局限,引导

轻量级模型在深层语义空间精准复现微小病灶的精细表征。在此基础上构建集自适应特征处理与渐进式模型压缩 (剪枝、蒸馏、量化) 于一体的端到端体系,以实现复杂模型向低算力设备的有效迁移,以期 为农业 AI 的实地部署提供一套兼具高精度与高效率的解决方案。

1 自适应教师模型 EAFP - Med ST

考虑到苹果病害分类任务需精准捕捉目标区域特征,首先对原始 EAFP - Med ST 进行针对性修改,使其适配苹果叶片病害分类场景,再将其作为知识迁移^[19]的教师模型,为轻量化压缩提供高性能特征表征基础。

1.1 EAFP - Med ST 适配性修改

原始 EAFP - Med ST 直接应用于 RGB 三通道、小尺度病斑、高纹理背景的苹果病害图像时,会出现特征捕捉偏差与计算冗余问题。因此,对该模型进行了 3 个方面的针对性修改,其适配后的结构如图 1 所示。EAFP - Med 模块接收 1 幅 RGB 苹果叶片图像作为输入,并通过 1 个三分支并行的架构来解构图像特征,同时由 1 个顶层控制器进行动态调节。

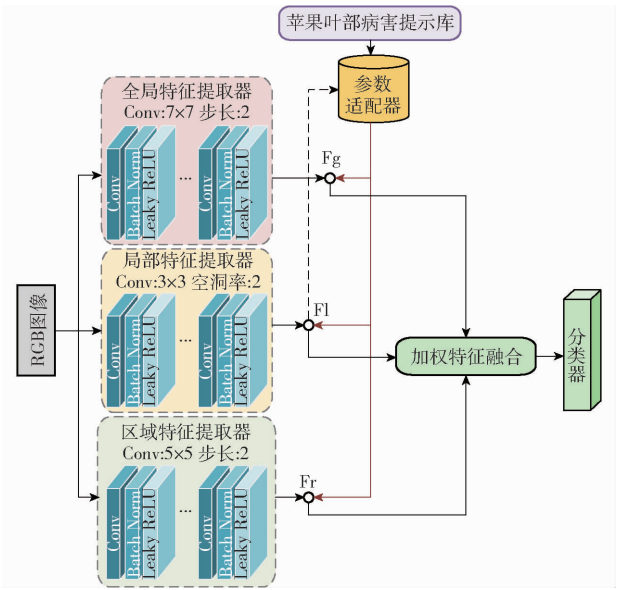


图 1 适配苹果叶部病害任务的 EAFP - Med 模块结构图
Fig. 1 Architecture of EAFP - Med module adapted for apple leaf disease tasks

1.1.1 多尺度特征提取器参数适配

原始 EAFP - Med 模块的三分支架构是为大尺寸医学病灶设计的,参数不适用于 10 ~ 50 像素且易与叶脉混淆的苹果病斑,存在误检风险。为精准匹配苹果病害从整体叶片到微小病斑的多样化尺度,对 3 个并行的特征提取器进行了参数优化 (图 1)。即:将全局提取器 (粉色模块) 的卷积核从 11 × 11

缩小至 7×7 (步长 s 为 2),以兼顾宏观状态与早期小病斑;为局部提取器(黄色模块)的 3×3 卷积核新增 $\text{dilation} = 2$ 的空洞卷积,以在不增加参数的前提下扩大感受野,增强对微小病斑与叶脉的区分能力;将区域提取器(绿色模块)的卷积核从 7×7 调整为 5×5 (s 为 2),使其更聚焦于病斑密集区域。

经过上述修改,3 个提取器能够协同工作,其感受野分别覆盖从宏观到微观的多个尺度,完美适配苹果病害从整体形态到病斑细节的诊断需求。

1.1.2 参数适配器任务重定义与动态权重分配

苹果病害分类需关注病斑颜色、纹理类型以及叶片背景 3 类关键信息,为使 EAFP-Med 模型能理解苹果病害专属特征,重新定义自适应机制。首先构建了 1 个苹果叶部病害害 4 类目标专属提示库,如图 1 顶部紫色模块所示。该库包含 12 条结构化提示,用于描述各类病害的核心视觉特征(如颜色、纹理、分布)。然后,这些提示被送入参数适配器。适配器解析提示后,动态生成一组权重。这些权重通过红色连接线,传递给全局、局部和区域 3 个特征提取器的输出端,实现动态加权融合。例如,黑星病提示会引导适配器提高局部特征权重。通过这种图文结合的自适应机制,模型能够根据不同病害类型,智能地调控其关注点,最终将融合后的特征送入分类器进行精准判别,从而确保了即使在压缩模型时,这些关键病灶感知能力也能被有效保留。

1.1.3 SwinV2-T 骨干网络

为确立性能最优的基线教师模型,对多种主流的深度学习架构进行系统性对比。评估对象涵盖经典的卷积神经网络^[20]模型(如 Swin Transformer^[21]、ResNet50^[22]、VGG^[23]系列),以及现代高效模型(如 EfficientNetV2^[24]、MobileNetV3^[25] 和 ConvNeXtV2^[26])。通过对候选模型训练与评估,因 Swin Transformer V2 Tiny 在准确率、F1 值等关键指标上表现优异,被最终确立为基线教师模型。

如图 2 所示,针对苹果叶片 RGB 彩色图像的特性,对骨干网络进行了针对性优化。首先,将输入层通道数由原始的单通道调整为 3 通道,并去除首层冗余的下采样操作,以更充分保留原始图像中高分辨率空间细节,减少黑星病等早期微小病斑特征因过早压缩而削弱或丢失风险,为后续特征提取提供更加可靠的数据基础。其次,将原有固定的 7×7 窗口注意力机制改进为一种动态自适应策略,以增强模型对不同尺度病斑区域的表征能力。具体而言,模型引入 EigenCAM 算法实时评估输入图像中病灶激活区域的面积。设定 50 像素为尺度阈值:当激活区域面积小于 50 像素(判定为微小病)时,注意力窗口自动切换为更精细的 5×5 ,强制模型聚焦于局部微观纹理,提升对细粒度特征的捕捉能力;而对于面积大于 50 像素的大病或健康叶片,窗口则恢复为 7×7 ,以确保模型能捕捉到足够广阔的上下文语义信息。

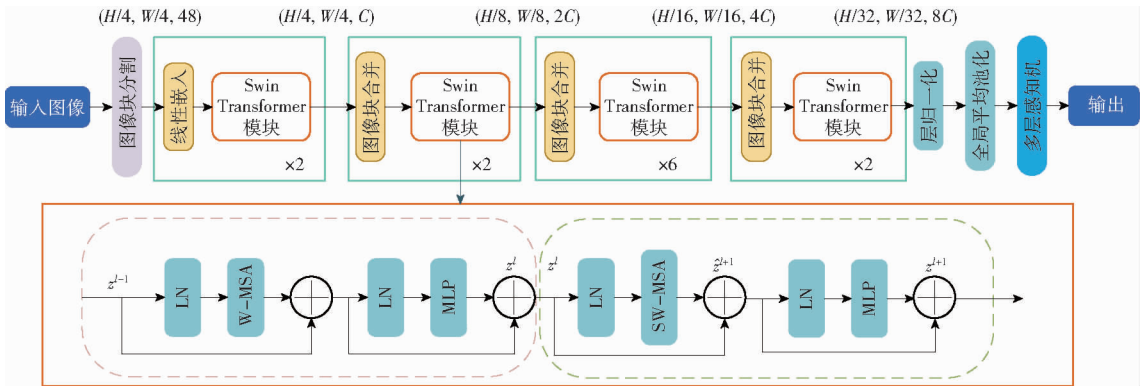


图 2 Swin Transformer V2 网络结构图

Fig. 2 Architecture of Swin Transformer V2

1.2 EAFP-Med ST 实际应用

1.2.1 数据集适配与预处理

采用 1 个双源构建的数据集(共 9 263 幅,涵盖 4 类目标:健康叶、黑腐病、锈病、黑星病),如图 3 所示,数据集结合 PlantVillage 公开数据(6 321 幅)与新疆阿克苏实地采集数据(2 942 幅)。这种构建方式确保数据多样性,既包含标准光照下的典型病害特征,也覆盖强光/阴天、早期微小病斑(小于 20 像

素)、叶片重叠及污物污染等复杂田间场景,保证了模型鲁棒性与泛化能力。数据集类别分布相对均衡(健康 2 598 幅,黑腐病 1 960 幅,锈病 2 753 幅,黑星病 1 952 幅)。

为使模型能够从纷繁复杂的背景中精准聚焦于细微的病灶差异,设计了一套多维度预处理方案:①应用 GrabCut 算法对原始图像进行前景提取,将叶片主体从复杂的土壤、枝干背景中分离。对于含

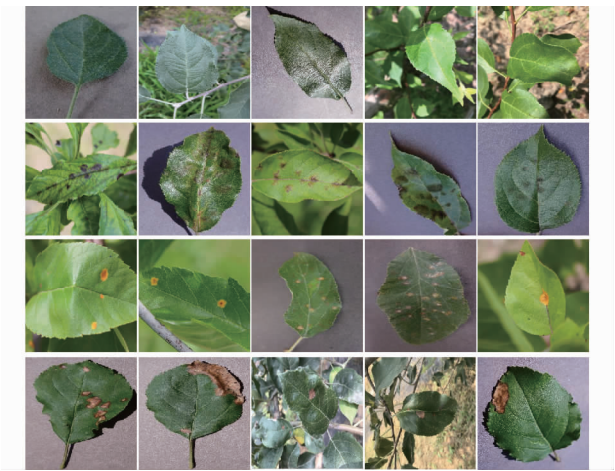


图 3 数据集样本示例

Fig. 3 Representative samples from dataset

有小病灶(<20 像素)的样本,采用局部放大与随机角度扰动($\pm 15^\circ$),在视觉上强化微小特征的显著性。②对于健康叶片,则采用纹理锐化处理,增强叶脉细节,防止模型将模糊的叶脉纹理误判为早期病斑。③将所有图像尺寸统一调整为 224 像素 $\times$ 224 像素,并利用 Z-Score 方法对 RGB 通道进行标准化,消除光照强度差异带来的数据分布偏移。

1.2.2 教师模型训练

以 4 类目标分类准确率与病灶特征表征一致性为双优化目标,对改进的 EAFP-Med ST 进行专项训练:采用交叉熵损失(权重 0.7)、特征熵损失(权重 0.3)加权组合,其中交叉熵损失通过类别权重调整降低黑星病与锈病的误判率,特征熵损失通过最小化 EAFP-Med 模块输出特征图信息熵,强化 3 类病害专属特征的稳定表征。训练 AdamW 优化器(初始学习率 1×10^{-4} ,每 15 轮衰减至前一轮的 0.5)进行 50 轮,最终模型在验证集整体 F1 值达 98.9%,为蒸馏提供精准特征知识。

然而,EAFP-Med ST 的强大性能以巨大的计算资源消耗为代价,其 SwinV2-T 骨干网络参数量约 2.758×10^7 ,叠加 EAFP-Med 模块后,总参数量超过 3.1×10^7 ,且伴随着极高的浮点运算量。这种高昂的计算成本完全无法满足农业边缘设备(如手持终端、无人机)对实时性与低功耗的严苛要求。因此,开展高效压缩研究,力求在保留核心自适应能力的同时实现极致轻量化。

2 病灶感知的渐进式 L-EAFP 框架

为解决高性能自适应模型 EAFP-Med ST 存在的参数量大、浮点运算量大而难以在边缘设备部署的瓶颈问题,提出了一种融合自适应特征与病灶感知的渐进式压缩框架,命名为 L-EAFP。该框架核心是 L-PMC (Lesion-aware progressive model

compression),其目标是将大型教师模型 EAFP-Med ST 所具备的强大自适应诊断能力,高效且高保真地迁移至一个轻量级的 ShuffleNetV2^[27] 学生模型中。L-EAFP 框架如图 4 所示,其将整个压缩过程分解为伪病灶掩码生成、病灶加权剪枝、病灶特征锚定蒸馏和自适应混合精度量化 4 个阶段。

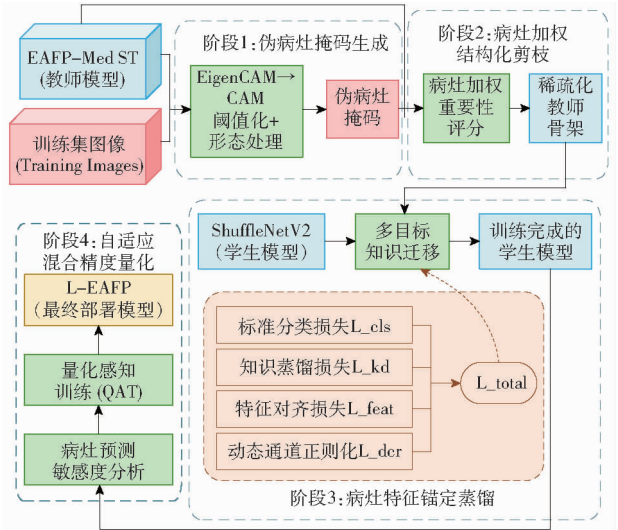


图 4 L-EAFP 框架总体流程图

Fig. 4 Overall flowchart of L-EAFP framework

2.1 基于 EigenCAM 的伪病灶掩码生成

首先以无监督方式生成病灶位置的先验信息,以指导病灶感知压缩。具体而言,基于已训练的高性能 EAFP-Med ST 教师模型,对训练集中的每幅图像应用 EigenCAM 算法^[28],生成高质量的类激活图。该方法生成的掩码与人工精细标注的平均交并比(IoU)达到 0.83,证明其能精准定位不同尺度的病灶。

这一自动化流程不仅有效规避了昂贵的像素级人工标注成本,更直接提取了教师模型对于致病特征的内在关注逻辑。随后,通过 0.6 的激活值阈值处理和 3×3 矩形核膨胀操作,将类激活图转换为二值化伪病灶掩码(Pseudo-lesion mask),用于引导后续所有病灶感知的压缩操作。这种显式空间约束机制能够屏蔽背景噪声干扰,强制模型在剪枝与量化过程中,优先保留对诊断决策起决定性作用的病理特征。

2.2 病灶加权结构化剪枝

在启动知识蒸馏过程前,对冗余的教师模型实施剪枝操作,目的为剔除网络中的非显著连接,识别并保留对病灶响应最为关键的神经元通道,从而为学生模型构建一个更精炼、信息密度更高的知识源。采用结构化剪枝^[29]策略,能够直接移除完整的卷积核,无需专用硬件支持即可实现加速。剪枝范围全面覆盖 SwinV2-T 骨干网络及 EAFP-Med 模块中

的所有卷积层。

经过对比实验,最终将剪枝比例设定为 30%。实验结果显示:若仅剪枝 20%,模型参数量仍高达 2.21×10^7 ,无法满足极致轻量化的需求;而当剪枝比例激进地提升至 40%时,黑星病漏检率从 5.8% 回升至 8.2%,表明关键特征遭到破坏。因此,30% 的剪枝比例在模型压缩与诊断性能保留之间取得了最佳平衡,成功使教师模型参数量从 3.157×10^7 降至 2.21×10^7 ,浮点运算量从 6.82×10^9 降至 4.17×10^9 。

为确保剪枝过程具备病灶感知能力,剪枝决策严格依据病灶加权重要性评分。对于网络中的第 j 个通道,其重要性评分计算式为

$$S_j = \|W_j\|_2 E_{(x,m) \in D} \left(\frac{1}{N_m} \sum_{p \in m} A_j(x)_p \right) \quad (1)$$

式中 W_j ——第 j 个输出通道对应的权重张量

D ——训练集

N_m ——掩码 m 所覆盖的像素总数

$A_j(x)_p$ ——第 j 个通道的激活图在像素位置 p 处的值

为引导模型聚焦于病灶区域,式(1)中通道通用重要性(左侧项)与病灶激活强度(右侧项)的权重比最终设定为 1:1.2。

2.3 多目标病灶特征锚定蒸馏

为将上一阶段通过稀疏化教师模型获取的深层诊断知识,高保真地传授给轻量级的 ShuffleNetV2 学生模型,设计了一种多目标病灶特征锚定蒸馏策略。学生模型中复合损失函数定义为

$$L_{total} = w_{ce} L_{CE} + w_k L_{KD} + w_{feat} L_{Feat} + w_{penalty} L_{Penalty} \quad (2)$$

式中 w_{ce} 、 w_k 、 w_{feat} 、 $w_{penalty}$ 为各项损失权重超参数,用以平衡多个学习目标。通过验证集网格搜索对比,最终选择最优权重组合为 0.3、0.7、0.5、1.0。式中各分项损失定义为:

(1) 标准分类损失 (L_{CE}): 衡量学生模型预测结果与病害样本真实标签 (Hard label) 之间的差异,确保模型能够从标注数据中直接习得苹果叶部病害的基础分类知识。

(2) 知识蒸馏损失 (L_{KD}): 通过引入软标签 (Soft label), 将教师模型对非目标类别的概率分布 (即暗知识, 如对模糊病斑的判断倾向) 传递给学生, 弥补了硬标签信息量单一的不足, 从而提升学生模型的泛化能力。

(3) 特征对齐损失 (L_{Feat}): 对齐师生模型在关键层的病斑特征表征, 其定义为

$$L_{Feat} = \frac{1}{HWC} \sum_{l \in S} \| \text{Up}(\text{Conv}(f_{S,l})) - f_{T,l} \|_2^2 \quad (3)$$

式中 S ——学生模型 (Student)

T ——教师模型 (Teacher)

$f_{S,l}$ ——学生模型在第 l 个层级提取的特征图

$f_{T,l}$ ——教师模型在第 l 个层级提取的特征图

特选取代表浅层纹理、中层局部和深层全局语义的教师模型关键层 (第 3、5、7 层), 使其与学生模型的对应层 (第 2、4、6 层) 进行特征对齐。这种跨尺度的特征锚定机制, 有助于学生模型完整学习小病斑的层级特征至关重要。

(4) 动态通道正则化损失 ($L_{Penalty}$): 通过惩罚冗余激活进一步提升学生模型的稀疏性与推理效率, 其定义为

$$L_{penalty} = \frac{1}{LB} \sum_{l \in U} \sum_{i=1}^B \sum_{j \in J_l} \|f_{S,l,i,j}\|_1 \quad (4)$$

式中 U ——非敏感层集合

L ——非敏感层总数

J_l ——第 l 层中被识别为非敏感的通道索引集合

$f_{S,l,i,j}$ ——学生模型第 l 层、第 i 个样本、第 j 个通道的特征激活值

2.4 自适应混合精度量化

为在资源受限的边缘硬件上获得极致的推理加速, 在蒸馏完成后, 对学生模型引入了一种自适应混合精度量化策略。

该策略通过逐层敏感度分析, 精准识别并保护那些对模型最终精度影响巨大的关键层。具体实施时, 逐一将模型中的单个层量化为 INT8 格式, 保持其他层不变, 测量由此导致的模型准确率下降幅度, 以此作为衡量该层量化敏感度的指标。

基于上述分析结果, 制定了混合精度配置方案: 将敏感度较低的层 (准确率下降值小于阈值 0.1) 配置为 INT8 量化, 以最大化压缩比和推理速度; 而将严重影响精度的高敏感度层保留为 FP16 或 FP32 格式, 以维持特征表达的细腻度。

在确定此混合精度配置后, 采用量化感知训练 (Quantization-aware training, QAT) 对模型进行最终的微调, 并借助直通估计器 (Straight-through estimator, STE)^[30] 技术解决量化操作不可导的问题, 使梯度能够顺利回传。

3 实验结果与分析

3.1 实验设置

3.1.1 实验平台

实验硬件为 Intel i7-12700K CPU 与 NVIDIA RTX 4060 Ti (8 GB) GPU; 软件环境为 Windows 11, Python 3.9, PyTorch 2.1.0 及 OpenCV 4.8.0。

3.1.2 数据集与划分

所用数据集共 9 263 幅,按比例 7:1:2 划分为训练集(6 484 幅)、验证集(926 幅)和主测试集(1 853 幅)。为确保评估客观性,主测试集与训练集无场景重叠。为评估模型在真实田间环境下鲁棒性,另从 AppleLeaf9 数据集^[31]中构建 1 个独立的复杂干扰测试子集(CITS),包含 320 幅图像(图 5),涵盖重叠遮挡、极端光照与污物附着等典型干扰。

3.1.3 评价指标

评价指标主要有模型压缩效果、模型诊断性能及数据增强效果 3 个维度。模型压缩效果采用参数量(Params)、模型内存占用量、浮点运算量和压缩率进行衡量;模型诊断性能采用准确率(Acc)、F1 值(F1 score)、病灶区域准确率和病灶漏检率进行判定;数据增强效果采用峰值信噪比(PSNR)、结构相似性(SSIM)及生成图像质量(FID)指标进行综合评估。



图 5 复杂干扰测试子集(CITS)样本示例
Fig.5 Sample examples of complex interference test subset (CITS)

3.2 不同分类模型性能对比

为验证 EAFP – Med ST 教师模型的基线性能,选取经典 CNN 模型、农业领域轻量化模型作为对比,在苹果叶部病害测试集上评估核心指标,结果如表 1 所示。

表 1 不同分类模型在苹果叶部病害原始测试集上性能对比结果

Tab.1 Performance comparison of different classification models on original apple leaf disease test set							
网络模型	骨干网络	准确率/%	F1 值/%	病灶区域 准确率/%	黑星病漏 检率/%	参数量	浮点运算量
ResNet50	ResNet50	94.78	95.18	89.2	8.7	2.556×10^7	4.120×10^9
VGG16	VGG16	94.35	93.84	88.5	9.2	1.384×10^8	1.568×10^{10}
EfficientNetV2	EfficientNetV2	93.38	92.81	87.1	10.5	1.123×10^7	2.050×10^9
MobileNetV3	MobileNetV3	93.86	93.84	88.7	9.8	2.540×10^6	1.830×10^9
MobileNet + Faster R – CNN	MobileNet	93.28	93.15	86.3	12.0	1.500×10^6	6.200×10^8
YOLO v8 – Corn	YOLOv8n	92.50	91.80	88.2	8.7	4.400×10^6	1.200×10^9
ConvNeXtV2	ConvNeXtV2	97.56	97.85	95.2	3.5	1.527×10^7	3.260×10^9
EAFP – Med ST	SwinV2 – T	98.69	98.58	97.8	1.2	3.157×10^7	6.820×10^9

由表 1 可知,EAFP – Med ST 教师模型在各项核心诊断指标上均显著优于所有对比模型,其病灶区域准确率高达 97.8%,且在最具挑战性的黑星病识别任务中,其漏检率仅为 1.2%。表明本文模型有效克服了传统模型对早期微小病灶捕捉能力不足的固有缺陷。

3.3 L – EAFP 框架消融实验

为系统性地解构并验证 L – EAFP 框架中各病灶感知组件的独立贡献及其协同效应,设计了一组阶梯式消融实验,结果如表 2 所示。以不包含任何剪枝与特定量化操作的标准知识蒸馏模型作为基准对照组(P0),其蒸馏参数固定为温度 $T = 2.0$ 且损失权重均分。在此基础上,依次叠加病灶加权结构化剪枝策略(P1)与多目标病灶特征锚定蒸馏策略(P2),最终构建出完整的 L – EAFP 方法(P3)。

表 2 结果揭示了各组件显著的逐级增益特性,

引入病灶加权剪枝后,模型在保持参数量不变的情况下,黑星病漏检率从基线的 4.8% 下降至 3.2%,病灶区域准确率提升 1.3 个百分点。这一性能跃升主要归因于该剪枝策略改变了参数保留的逻辑:通过引入通道重要性评分机制,模型优先保留了对黑星病等微小病灶响应最强烈的神经元通道(保留率达 92%),从而有效避免了传统剪枝因无差别裁剪导致的关键细粒度特征丢失问题。在 P1 的基础上进一步引入病灶特征锚定蒸馏后,模型性能获得二次提升,黑星病漏检率进一步下降至 1.8%,整体准确率提升 0.79 个百分点。这一改进验证了所提出的特征对齐机制的有效性,成功修正了传统蒸馏仅侧重于全局 Logits 分布拟合,而忽视局部关键特征空间位置对齐的缺陷。最终的 L – EAFP 完整方法(P3)在仅有 3.5×10^5 参数量的超轻量级学生模型上,实现了高 F1 值(98.62%)与高的病灶区域准确率(97.2%)。相较于基线 P0,病灶区域准确率提升

表 2 L-EAFP 框架消融实验结果

Tab.2 Ablation study on components of L-EAFP framework

方案	策略	参数量	准确率/%	F1 值/%	黑星病	黑腐病	黑星病	病灶区域
					准确率/%	准确率/%	漏检率/%	准确率/%
P0	标准知识蒸馏模型	3.5×10^5	97.62	97.59	96.85	98.21	4.8	94.5
P1	+ 病灶加权剪枝	3.5×10^5	97.91	97.88	97.52	98.35	3.2	95.8
P2	+ 病灶特征锚定蒸馏	3.5×10^5	98.45	98.43	98.31	98.58	1.8	96.9
P3	L-EAFP	3.5×10^5	98.58	98.62	98.55	98.62	1.2	97.2

2.7 个百分点。充分证明了剪枝精炼与特征锚定 2 个组件之间存在显著的协同增效作用,即前者为后者提供了更纯净的特征通道,后者则最大化了保留通道的表达能力,从而精准解决了苹果叶部病害识别中小病灶难发现的核心痛点。

3.4 模型压缩与部署性能

L-EAFP 框架在追求极致的模型轻量化与维持高性能诊断之间取得了较优的平衡。

在模型物理属性方面,L-EAFP 实现了对教师模型的大幅精简。模型参数量从原始 3.157×10^7 压缩至 3.5×10^5 ,压缩率高达 98.9%;浮点运算量从 6.82×10^9 降至 5×10^8 ,计算负载削减 92.7%;模

型内存占用量从 107.06 MB 减至 1.35 MB。这种量级上的缩减,使得压缩模型能够轻松适配存储空间与算力极其受限的农业边缘设备(如手持终端与无人机嵌入式芯片),完全满足了现场部署的硬件需求。在诊断性能方面压缩后模型准确率仅下降 0.11 个百分点。其病灶区域准确率保留度高达 99.2%,且黑星病漏检率维持在 1.2% 的极低水平。此外,与现有的主流压缩技术相比,L-EAFP 的优势十分显著。如表 3 所示,其准确率较传统结构化剪枝高 4.28 个百分点,较普通知识蒸馏高 0.96 个百分点。而在病灶区域准确率上,L-EAFP 分别提升 10.1、2.7 个百分点。

表 3 模型性能综合对比实验

Tab.3 Comprehensive comparison of model compression and performance

模型	骨干网络	准确率/%	F1 值/%	病灶区域准确率/%	参数量	浮点运算量	压缩率/%
EAFP-Med ST	SwinV2-T	98.69	98.58	97.8	3.157×10^7	6.820×10^9	
ShuffleNetV2	ShuffleNetV2	93.76	93.21	86.3	3.500×10^5	5.000×10^8	
传统结构化剪枝	ShuffleNetV2	94.30	93.71	87.1	3.500×10^5	4.200×10^8	98.9
普通知识蒸馏	ShuffleNetV2	97.62	97.58	94.5	3.500×10^5	5.000×10^8	98.9
固定 8bit 量化	ShuffleNetV2	97.55	97.47	94.2	3.500×10^5	5.000×10^8	98.9
L-EAFP	ShuffleNetV2	98.58	98.62	97.2	3.500×10^5	5.000×10^8	98.9

3.5 关键超参数影响实验

针对 L-EAFP 框架中 3 个核心超参数(剪枝权重 ω 、蒸馏温度 T 、量化敏感度阈值 Δ),通过控制变量法验证其对模型性能的影响,在分析某一参数时,其余参数固定为预实验确定的较优取值,确保超参数选择的合理性与稳定性。

3.5.1 剪枝权重 ω 的影响

ω 用于平衡通道通用重要性与病灶区域激活强度。实验固定蒸馏温度 T 为 3.0、量化敏感度阈值 Δ 为 0.1,测试不同剪枝权重对性能的影响。结果表明,当 ω 为 0.5 时最优。当 ω 过小(小于 0.5)时,对病灶激活的关注不足,导致部分小病灶关键通道被错误裁剪,黑星病漏检率升至 3.5%;当 ω 过大(大于 0.7)时,过度侧重病灶激活而忽略了通用特征,可能导致冗余通道裁剪不彻底,模型泛化性略有下降。

3.5.2 蒸馏温度 T 的影响

蒸馏温度 T 是知识蒸馏中的重要超参数,用

于调节教师模型输出概率分布的平滑程度。设教师模型输出的 logit 为 Z_i ,则第 i 类对应的软标签概率为

$$p_i = \frac{\exp(Z_i/T)}{\sum_{j=1}^J \exp(Z_j/T)}$$

当 $T > 1$ 时,概率分布更加平滑,有助于挖掘类别间的潜在关系;当 T 过大时,则可能削弱关键信息。本文在 $\omega = 0.5$ 、 $\Delta = 0.1$ 条件下,设置 $T \in \{1.0, 2.0, 3.0, 4.0\}$ 进行对比实验,结果表明 $T = 3.0$ 时模型综合最优。

3.5.3 量化敏感度阈值 Δ 的影响

量化敏感度阈值 Δ 决定混合精度量化中层的精度分配策略,即允许的准确率下降阈值。实验固定剪枝权重 ω 为 0.5、蒸馏温度 T 为 3.0,测试不同 Δ 值对性能的影响。结果表明,当 Δ 为 0.1 时最优。当 Δ 为 0.05 时,阈值过于严苛,仅 42% 的层被量

化,压缩收益有限;当 Δ 为 0.15 时,阈值过于宽松,导致 88% 的层被量化,部分关键敏感层精度下降,造成黑星病准确率下降 1.2 个百分点。 Δ 为 0.1 则能在量化 65% 的层与保持精度之间取得最佳平衡。

3.5.4 特征对齐层选择实验

为验证 2.3 节中病灶特征锚定蒸馏所选对齐层(教师模型第 3、5、7 层)的合理性,设计对比实验,探讨不同语义深度的特征层组合对模型性能的影响。固定其他超参数(ω 为 0.5, T 为 3.0, Δ 为 0.1)不变,结果如表 4 所示。

表 4 特征对齐层选择实验语义覆盖与性能对比结果
Tab.4 Comparison of semantic coverage and performance in feature alignment layer selection experiment %

对齐层组合 (教师层)	语义覆盖	准确率	F1 值	病灶区域 准确率
浅层(1, 2, 3)	侧重纹理	98.15	98.12	96.1
中层(3, 4, 5)	纹理与局部	98.32	98.29	96.6
深层(5, 6, 7)	局部与全局	98.40	98.37	96.9
本文(3, 5, 7)	纹理 + 局部 + 全局	98.58	98.62	97.2

由表 4 可知,仅使用浅层对齐时,虽模型能够捕捉到丰富的纹理细节,但由于缺乏深层语义信息的引导,模型在复杂背景下的类别判定能力受限,导致 F1 值较低(98.12%)。仅使用深层对齐时,虽整体分类准确率有所提升,但由于深层网络经过多次下采样,丢失了对微小病灶至关重要的高频空间信息(如边缘锐利度),导致病灶区域准确率下降至 96.9%。

表 5 L-EAFP 框架各阶段模型性能演进与资源开销对比

Tab.5 Performance evolution and resource cost comparison of models at each stage of L-EAFP framework							
模型	准确率/%	F1 值/%	病灶区域准确率/%	黑星病漏检率/%	参数量	浮点运算量	内存占用量/MB
教师模型	98.69	98.58	97.8	1.2	3.157×10^7	6.820×10^9	107.06
剪枝后教师模型	98.09	98.06	97.1	1.5	2.210×10^7	4.170×10^9	88.40
蒸馏后学生模型	98.52	98.57	97.3	1.3	3.500×10^5	5.000×10^8	1.35
量化后学生模型	98.58	98.62	97.2	1.2	3.500×10^5	5.000×10^8	1.35

3.7 病灶注意力可视化与量化验证

为定性并定量验证 L-EAFP 框架对病灶特征的保留效果,采用 EigenCAM 生成各阶段模型的病灶注意力热图,并计算注意力区域与人工标注病灶的 IoU 值,结果如图 6 所示。

3.7.1 注意力热图可视化

如图 6 所示,教师模型(第 1 行)能精准锁定病灶,证明了其是可靠的知识源。剪枝后教师(第 2 行)热图与教师高度一致,表明剪枝保留了关键神经元。蒸馏后学生(第 3 行)为验证核心,尽管学生

最终选用组合为基于对教师模型(Swin Transformer)层级架构的深度考量:对应 Stage 1 输出,保留了高分辨率特征,负责传递病斑边缘的锐利度与高频纹理,这对区分早期黑星病微小病斑与叶片灰尘至关重要。对应 Stage 2 输出,包含了病斑的局部几何形状信息(如区分圆形的黑星病与不规则的黑腐病)。对应 Stage 3 输出,聚合了长距离的全局上下文,有助于判断病斑在叶片上的分布规律(如锈病倾向于沿叶脉分布的特征)。

结果表明,采用跨尺度特征锚定策略,使模型在参数量极低情况下,仍能兼顾对微小病灶的敏锐感知与对整体病害类别的准确判断,从而取得了最佳的综合性能(F1 值为 98.62%,病灶区域准确率为 97.2%)。

3.6 各阶段模型性能演进分析

为直观呈现 L-EAFP 框架剪枝、蒸馏、量化渐进式压缩对性能的影响,跟踪各阶段模型的关键指标变化,结果如表 5 所示。L-EAFP 框架各阶段的演进逻辑清晰地揭示了其有效性:剪枝阶段参数量显著减少近 30%,而准确率仅降低 0.6 个百分点(病灶区域准确率保留 99.1%),成功构建了一个更精炼、高效的知识源。蒸馏阶段模型参数量降至 3.5×10^5 ,但准确率高 0.43 个百分点。证明了病灶特征锚定蒸馏策略使轻量级学生模型有效复现了教师的病灶识别能力。量化阶段准确率和 F1 值分别获得 0.06、0.05 个百分点的微小提升。这得益于 QAT 的正则化作用与自适应量化策略的误差补偿,最终实现了精度无损甚至略有增益的优化效果。

模型结构远比教师简单,但其注意力模式完美复现了教师的行为,直观地印证了病灶特征锚定蒸馏的卓越成效。量化后学生(第 4 行)的热图与蒸馏后几乎无差别,表明自适应量化未损害模型已习得的病灶特征表达,确保了部署可靠性。

3.7.2 注意力区域 IoU 量化

基于 100 幅标注样本的 IoU 计算结果(表 6)显示,各阶段模型平均 IoU 稳定在 0.81,黑星病小病灶 IoU 仅下降 0.03,表明 L-EAFP 在极限压缩后完整保留了教师模型的病灶注意力能力,确保了压缩

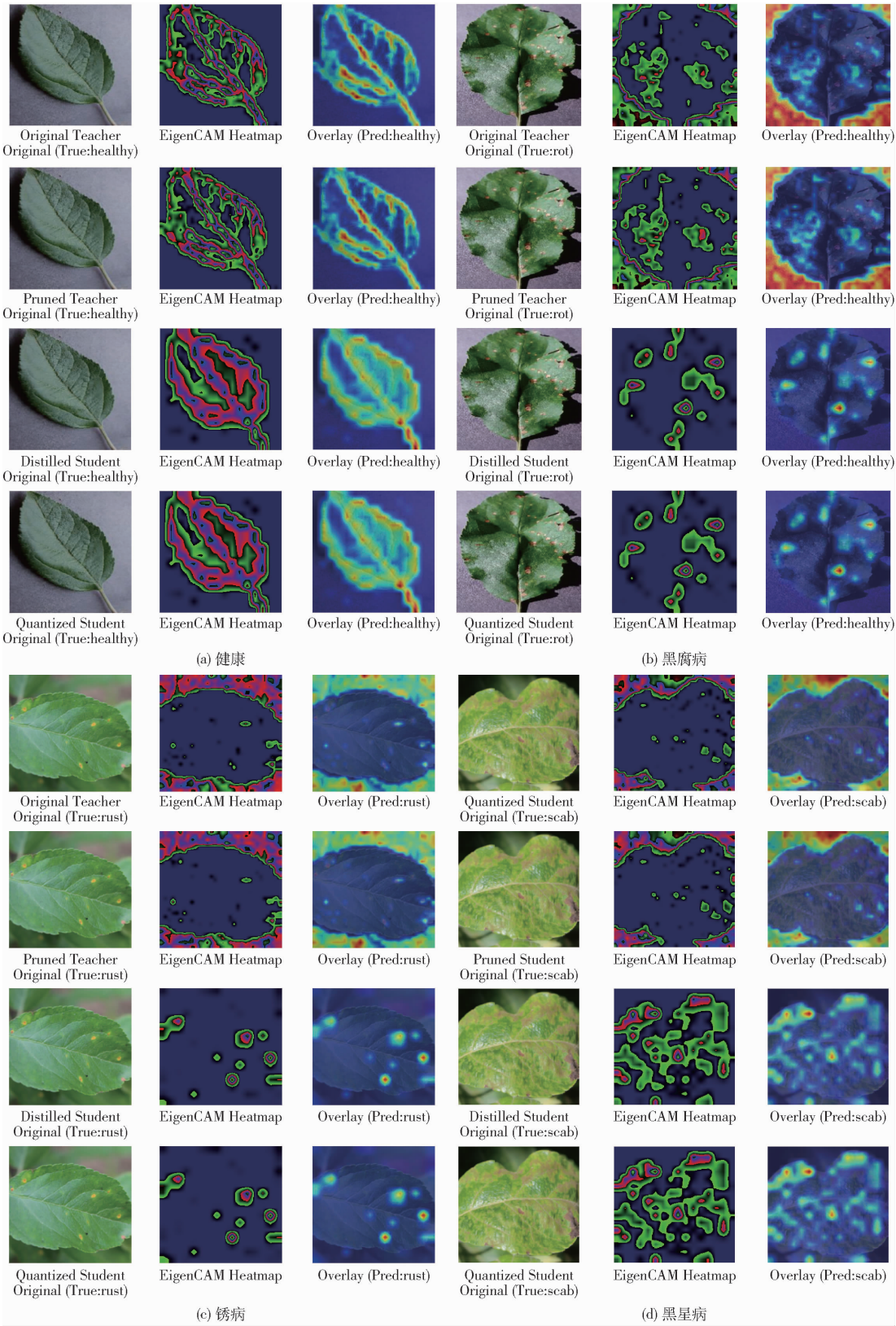


图 6 不同压缩阶段模型的病灶注意力可视化结果

Fig. 6 Visualization results of lesion attention for models at different compression stages

模型在农业病害细粒度识别中的高精度与可靠性。

3.8 模型抗干扰测试

使用 CITS 子集对本文 L-EAFP 模型及表 3 中准确率较高的模型进行抗干扰能力测试,结果如表 7 所示。

在面临重叠、附着物和光照干扰时,所有模型性能均下降,但 L-EAFP 表现最佳。相较于普通知识蒸馏,L-EAFP 准确率高 2.24 个百分点,且最接近教师模型。这归功于其病灶感知压缩机制,特别是病灶特征锚定蒸馏,其强制学生模型学习教师对病

表 6 各阶段模型注意力区域与真实病灶 IoU 对比
Tab.6 IoU comparison between attention regions and
real lesions

模型阶段	健康叶	黑腐病	锈病	黑星病	平均
	IoU	IoU	IoU	IoU	IoU
教师模型	0.82	0.89	0.86	0.78	0.84
剪枝后教师模型	0.81	0.88	0.85	0.77	0.83
蒸馏后学生模型	0.80	0.87	0.84	0.76	0.82
量化后学生模型	0.80	0.86	0.83	0.75	0.81

表 7 不同模型在复杂干扰子集 (CITS) 上性能对比
Tab.7 Performance comparison of different models on
complex interference test set (CITS) %

模型	骨干网络	准确率	F1 值	病灶区域 准确率
EAFP - Med ST	SwinV2 - T	97.05	96.98	95.8
普通知识蒸馏	ShuffleNetV2	93.88	93.52	90.1
L - EAFP	ShuffleNetV2	96.12	96.05	93.5

灶的稳健特征表达。因此,即使在病灶信息被噪声干扰时,L - EAFP 仍能做出准确判断,展现了更强的田间鲁棒性。

3.9 边缘设备部署方案

将轻量化后的模型真正应用农业生产一线,实现在资源极其受限的边缘设备(如手持式叶部病害诊断仪、农用无人机机载计算平台)上的高效部署。设计了一套通用的端到端部署流程(图 7),主要包含模型导出、中间表示转换与特定硬件加速 3 个核心环节。

在 PyTorch 深度学习框架下完成 L - EAFP 学生模型的量化感知训练(QAT),确保模型权重已适应低精度计算。利用 torch.onnx 工具将训练好的模型权重转换为开放式神经网络交换格式(Open neural network exchange, ONNX)。ONNX 作为一种通用的中间表示(Intermediate representation, IR),不仅保留了模型的网络结构与权重信息,还有效解耦了上层训练框架与底层推理引擎,为后续的跨平台移植奠定了基础。针对农业场景中异构硬件并存的现状,采用差异化的推理加速策略,以最大化挖掘硬件潜能。针对 NVIDIA 嵌入式平台(如 Jetson Nano/Orin 系列),采用 NVIDIA 官方的高性能推理

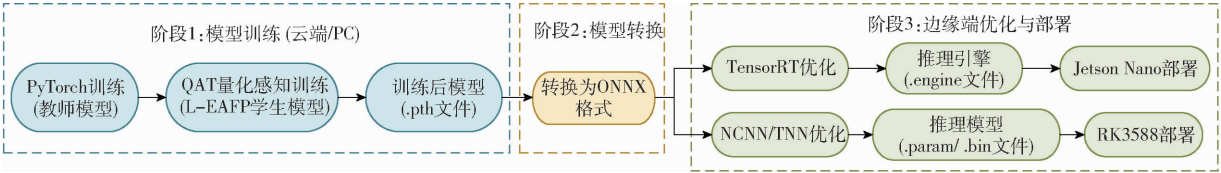


图 7 模型边缘端实时性部署流程示意图
Fig.7 Schematic of real-time deployment process for edge-side models

引擎 TensorRT 进行部署。TensorRT 能够对 ONNX 模型进行深度的图层级优化,包括算子融合(Layer fusion)、内核自动调优(Kernel auto-tuning)以及显存优化。通过将模型编译为针对特定 GPU 架构优化的 .engine 文件,能够显著降低推理延迟并提升吞吐量,适合无人机等对实时性要求极高(小于 15 ms)的场景。针对 ARM 架构移动端平台(如瑞芯微 RK3588、树莓派等),采用 NCNN 或 TNN 等轻量级推理框架。通过上述流程,L - EAFP 模型不仅在理论上实现了参数量的压缩,更在工程实践中打通了从云端训练到边缘端部署的完整路径,确保模型在实际田间作业中具备高实时性、低功耗与高稳定性的优异性能。

4 结论

(1)针对高性能病害诊断模型难以在农业边缘设备上高效部署的问题,提出了基于 EAFP - Med

ST 教师模型与 L - EAFP 渐进式压缩框架的苹果叶部病害高效识别方法。

(2)提出适配苹果病害的高性能教师模型 EAFP - Med ST,通过深度适配多尺度特征提取与动态机制,精准捕捉病害特征,为后续压缩提供了卓越的知识源。实验结果表明,其诊断准确率达 98.69%,对早期微小病灶漏检率仅 1.2%,性能显著优于基线模型。

(3)提出病灶感知的渐进式压缩框架 L - EAFP,融合病灶加权剪枝、特征锚定蒸馏与自适应量化,将教师模型的诊断能力高效迁移至轻量级学生模型。实验结果表明,最终模型参数量压缩至 3.5×10^5 (压缩率 98.9%),而 F1 值仍高达 98.62%,实现了轻量化与高性能的平衡。构建了从模型深度适配到病灶感知压缩的完整技术路径,为农业边缘 AI 部署提供了兼具精度与效率的解决方案。

参 考 文 献

[1] 王惠,黄明雪,吴惠芳.中国数字农业的研究进程与展望[J].中国农业大学学报,2025,30(10):364 - 380.

WANG Hui,HUANG Mingxue,WU Huifang. Research progress and prospects of digital agriculture in China[J]. Journal of China Agricultural University,2025,30(10):364–380. (in Chinese)

[2] 赵晓,杨梦婷,张懿丹. 基于改进 YOLO v10n 的苹果叶片病害检测方法[J]. 农业机械学报,2025,56(9):657–666, 728. ZHAO Xiao, YANG Mengting, ZHANG Yidan. Apple leaf disease detection method based on improved YOLO v10n[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(9):657–666, 728. (in Chinese)

[3] 李大湘,张雯凯,刘颖. 融合 Transformer 与原型自监督的苹果叶部病害识别[J]. 农业工程学报,2024,40(23):208–216. LI Daxiang, ZHANG Wenkai, LIU Ying. Identifying apple leaf diseases by integrating Transformer and prototype self-supervised learning[J]. Transactions of the CSAE, 2024, 40(23): 208–216. (in Chinese)

[4] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527–1554.

[5] 祝诗平,张越,唐茂杰,等. 基于改进 YOLO v10n 的百香果成熟果实检测方法[J]. 农业机械学报,2025,56(10):549–557, 595. ZHU Shiping, ZHANG Yue, TANG Maojie, et al. Ripe fruit detection model of passion fruit based on improved YOLO v10n[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(10):549–557, 595. (in Chinese)

[6] JIANG P, CHEN Y H, LIU B, et al. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks[J]. IEEE Access, 2019, 7: 59069–59080.

[7] GULSOY T, KABLAN E B. FocalNeXt: a ConvNeXt augmented FocalNet architecture for lung cancer classification from CT-scan images[J]. Expert Systems with Applications, 2025, 261: 125553.

[8] 朱锐,夏宇,季凡,等. 基于图像处理和卷积神经网络的大米加工精度识别方法研究[J]. 农业机械学报,2025,56(9):751–758. ZHU Rui, XIA Yu, JI Fan, et al. Rice processing precision identification method based on image processing and convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(9):751–758. (in Chinese)

[9] 王泰华,郭亚州,张家乐,等. 基于改进 YOLO v5s 的水稻害虫识别研究[J]. 农业机械学报,2024,55(11):39–48. WANG Taihua, GUO Yazhou, ZHANG Jiale, et al. Rice pest identification based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(11):39–48. (in Chinese)

[10] 王林惠,兰玉彬,刘志壮,等. 便携式柑橘虫害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 CSAE, 2021, 37(9): 282–288. (in Chinese)

[11] TOO E C, LI Y, NJUKI S, et al. A comparative study of fine-tuning deep learning models for plant disease identification[J]. Computers and Electronics in Agriculture, 2019, 161: 272–279.

[12] 时雷,雷镜楷,王健,等. 基于改进 FasterNet 的轻量化小麦生育期识别模型[J]. 农业机械学报,2024,55(5):226–234. SHI Lei, LEI Jingkai, WANG Jian, et al. Lightweight wheat growth stage identification model based on improved FasterNet[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(5):226–234. (in Chinese)

[13] 徐相华,周德靖,余超然,等. 基于 DCA-YOLO 的受病虫害侵染树木农业无人机低空检测模型[J]. 农业机械学报,2025,56(10):479–491. XU Xianghua, ZHOU Dejing, YU Chaoran, et al. Improved model for low altitude detection of trees infected by pests and diseases using agricultural drones based on DCA-YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(10):479–491. (in Chinese)

[14] 任龙龙,杜永辉,李玉强,等. 基于 CEG-YOLO 的轻量化设施环境樱花检测花检测方法[J]. 农业机械学报,2025,56(9):648–656. REN Longlong, DU Yonghui, LI Yuqiang, et al. Lightweight cherry blossom detection method in facility environments based on CEG-YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(9):648–656. (in Chinese)

[15] 丁志康,兰玉彬,韩鑫,等. 基于改进 YOLO v7-tiny 的农药残留消解过程多品种果蔬精准识别方法[J]. 农业机械学报,2025,56(10):725–735. DING Zhikang, LAN Yubin, HAN Xin, et al. Precise multi-variety fruit and vegetable identification method in process of pesticide residues degradation based on improved YOLO v7-tiny[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(10):725–735. (in Chinese)

[16] 李龙海,王铁哲,吴禹竺,等. 基于剪枝与蒸馏的轻量化玉米病害检测方法[J]. 农业工程学报,2025,41(17):194–202. LI Longhai, WANG Yizhe, WU Yuzhu, et al. Method for lightweight corn leaf diseases based on pruning and distillation[J]. Transactions of the CSAE, 2025, 41(17): 194–202. (in Chinese)

[17] LI X, LAN L, LAHZA H, et al. EAFP-Med: an efficient adaptive feature processing module based on prompts for medical image detection[J]. Expert Systems with Applications, 2024, 247: 123334.

[18] 卞佳,王红霞,李泉,等. 人工智能辅助诊断系统在医学影像诊断学实践教学中的应用研究[J]. 中国高等医学教育, 2024(11):103–104.

[19] ZHUANG F Z, QI Z Y, DUAN K Y, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2021, 109(1): 43–76.

[20] REN S, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.

[19] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZALEZ J A, et al. A comprehensive review of YOLO architectures in computer vision; from YOLOv1 to YOLOv8 and yolo-nas[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680 – 1716.

[20] ZHANG J, LI X, LI J, et al. Rethinking mobile block for efficient attention-based models[C] //2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2023: 1389 – 1400.

[21] SOHAN M, SAI Ram T, REDDY R, et al. A review on yolov8 and its advancements[C] //International Conference on Data Intelligence and Cognitive Informatics, 2024: 529 – 545.

[22] HUSSAIN M. Yolov1 to v8; unveiling each variant—a comprehensive review of YOLO[J]. IEEE Access, 2024, 12: 42816 – 42833.

[23] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention; rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.

[24] YU Z, HUANG H, CHEN W, et al. Yolo – facev2: a scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.

[25] LI Y, ZHANG Y, TIMOFTE R, et al. NTIRE 2023 challenge on image denoising: methods and results[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 1905 – 1921.

[26] VASLUIANU F A, SEIZINGER T, ZHOU Z, et al. NTIRE 2024 image shadow removal challenge report[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 6547 – 6570.

[27] HUANG H, CHEN Z, ZOU Y, et al. Channel prior convolutional attention for medical image segmentation[J]. Computers in Biology and Medicine, 2024, 178: 108784.

[28] 陈冲, 陈杰, 张慧, 等. 深度学习可解释性综述[J]. 计算机科学, 2023, 50(5):52 – 63.
CHEN Chong, CHEN Jie, ZHANG Hui, et al. Review interpretability of deep learning[J]. Computer Science, 2023, 50(5): 52 – 63. (in Chinese)

[29] 路晓辰, 杨立明, 杨兴悦, 等. 深度学习方法在地震事件分类中的应用及可解释性研究[J]. 地震工程学报, 2023, 45(2):474 – 482.
LU Xiaochen, YANG Liming, YANG Xingyue, et al. Application and interpretability of deep learning methods in seismic event classification[J]. Journal of Earthquake Engineering, 2023, 45(2):474 – 482. (in Chinese)

(上接第 245 页)

[21] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84 – 90.

[22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas; IEEE, 2016: 770 – 778.

[23] PAYMODE A S, MALODE V B. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG[J]. Artificial Intelligence in Agriculture, 2022, 6: 23 – 33.

[24] GOMROKI M, HASANLOU M, REINARTZ P. STCD – EffV2T Unet; semi transfer learning EfficientNetV2 T – Unet network for urban/land cover change detection using Sentinel – 2 satellite images[J]. Remote Sensing, 2023, 15(5): 1232.

[25] ZHOU G Q, LIU W G, ZHU Q, et al. ECA – MobileNetV3 (large) + SegNet model for binary sugarcane classification of remotely sensed images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1 – 14.

[26] RETTENBERGER L, SZYMANSKI N J, GIUNTO A, et al. Leveraging unlabeled SEM datasets with self-supervised learning for enhanced particle segmentation[J]. npj Computational Materials, 2025, 11(1): 289.

[27] LIU D, DU J C, LI C H, et al. Multi-unit stacked architecture; an urban scene segmentation network based on UNet and ShuffleNetv2[J]. Applied Soft Computing, 2024, 165: 112065.

[28] STEFENON S F, SEMAN L O, KLAAR A C R, et al. Hypertuned – YOLO for interpretable distribution power grid fault location based on EigenCAM[J]. Ain Shams Engineering Journal, 2024, 15(6): 102722.

[29] 褚文博, 甘露, 李国法, 等. 面向自动驾驶的大模型高效压缩技术:综述[J]. 机械工程学报, 2024, 60(22):224 – 240.
CHU Wenbo, GAN Lu, LI Guofa, et al. Large models efficient compression technology for autonomous driving: a review[J]. Journal of Mechanical Engineering, 2024, 60(22): 224 – 240. (in Chinese)

[30] JIANG Y, ZHANG T, GUO J, et al. Clustering-based nonlinear training algorithm for precision constrained photonic micro-ring convolution chip[J]. Journal of Lightwave Technology, 2024, 42(22): 1 – 8.

[31] YANG Q, DUAN S K, WANG L D. Efficient identification of apple leaf diseases in the wild using convolutional neural networks[J]. Agronomy, 2022, 12(11): 2784.