

基于深度学习的移动端水果识别

郭 健, 吴 薇*

杭州电子科技大学电子信息学院, 浙江 杭州

收稿日期: 2024年7月4日; 录用日期: 2024年11月21日; 发布日期: 2024年11月29日

摘 要

超市水果识别主要依赖人工, 计算机视觉成为一种解决方案。然而目前仍面临部分水果识别精度低、终端设备部署困难、误识别图片难处理等挑战。因此, 文章基于深度学习对移动端水果识别进行研究, 旨在替代人工识别。首先文章构建了包含49种水果的超市水果图像数据集DailyFruit-49。并针对细分类特征相似度高、包装遮挡、形状小量少的水果识别困难, 以及低算力设备模型部署问题, 筛选了满足部署要求的骨干模型。设计了新的注意力模块RMA, 改进了ViT Block以增强模型的细节识别能力和深层语义特征整合能力, 最终得到DenseRMA_ViT模型, 并基于Focal Loss改进损失函数。并在公开数据集Fruits-262上进行消融实验验证模型改进的有效性。最后结合实际设备, 实现水果识别系统, 满足实际使用。基于与用户的交互行为对误识别水果图像进行收集, 并基于误识别图像实现模型权重自动微调, 随使用时间延长, 系统收集更多图片, 提升模型识别精度与泛化能力, 以处理实际应用中误识别水果。

关键词

水果识别, 数据集构建, 改进注意力机制, ViT, 系统设计, 模型权重自更新

Mobile Fruit Recognition Based on Deep Learning

Jian Guo, Wei Wu*

School of Electronic Information, Hangzhou Dianzi University, Hangzhou Zhejiang

Received: Jul. 4th, 2024; accepted: Nov. 21st, 2024; published: Nov. 29th, 2024

Abstract

Supermarket fruit recognition mainly relies on manual processes, and computer vision has emerged as a solution. However, challenges remain, including low accuracy for some fruits, difficulties in deploying them on terminal devices, and handling misidentified images. Therefore, this paper

*通讯作者。

researches mobile fruit recognition based on deep learning, aiming to replace manual identification. First, the paper constructs the DailyFruit-49 dataset, which includes images of 49 types of fruits. Addressing the challenges of recognizing fruits with high feature similarity, packaging obstructions, and small shapes, as well as the deployment issues on low-compute devices, the backbone model meeting deployment requirements was selected. A new attention module, RMA, was designed, and the ViT Block was improved to enhance the model's detail recognition and deep semantic feature integration capabilities, resulting in the Dense RMA_ViT model. The loss function was also improved based on Focal Loss. Ablation experiments on the public dataset Fruits-262 verified the effectiveness of these improvements. Finally, a fruit recognition system was implemented on actual devices to meet practical usage needs. The system collects misidentified fruit images based on user interactions and automatically fine-tunes the model's weights based on these images. Over time, as the system collects more images, the model's recognition accuracy and generalization ability improve, effectively handling misidentified fruits in real-world applications.

Keywords

Fruit Recognition, Dataset Construction, Improved Attention Mechanism, ViT, System Design, Model Weight Self-Updating

Copyright © 2024 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

中国自古以来重视农业发展, 尤其在水果种植方面取得了显著成就, 品种繁多, 面积广泛覆盖, 居全球前列。随着经济增长和生活品质提升, 消费者更加注重饮食营养, 水果已成为生活中不可或缺的一部分, 使中国成为全球最大的水果消费地[1]。目前, 水果商户主要依赖人工识别水果种类并进行称重和结算, 这种方式不便且不环保。水果种类繁多, 外观相似, 增加了识别难度, 人工处理速度慢且易出错。中国劳动力成本不断上升, 传统服务模式浪费人力资源, 影响顾客满意度和超市盈利。因此, 探索自动水果识别结算方法十分重要[2]。

水果识别技术通过深度分析水果图像或视频, 实现智能判别, 广泛适用于水果商店和无人超市, 提升购物体验, 减少人力成本, 优化运营效果。加快智慧农业发展, 推动农业数字化、智能化发展, 实现高效可持续的智慧农业。推动智能城市发展, 在无人自助超市中, 智能水果识别系统能自动完成消费流程, 提供便捷高效的购物体验。

2. 相关工作

基于机器视觉的水果识别方法通常包括三个步骤。首先是图像预处理, 包括调整图像的亮度和对比度、颜色空间转换、直方图均衡化、以及像素值归一化[3]。其次是特征提取, 方法包括颜色直方图[4]、Gabor 滤波[5]、局部二值模式[6]、方向梯度直方图[7]、边缘特征如 Sobel [8]和 Canny [9]、尺度不变特征变换[10]以及主成分分析[11]。最后是分类, 常用方法有支持向量机[12]、K 最近邻[13]、决策树(Decision Tree) [14]和随机森林[15]。

近年来, 基于深度学习的水果识别方法取得了显著进展。Enciso 等人基于 AlexNet 模型在 320 张数据集上, 对柠檬的新鲜和腐烂分类分别达到了 98.25%和 93.73%的准确率[16]。孟欣欣等人基于计算机视觉的香梨果实目标检测模型, 使用 Resnet 模型在 9600 张水果数据上进行预训练, 并构建了 Mask R-CNN

模型, 在成熟香梨数据集上的平均分割准确率达到 98.02% [17]。Xue 等人提出了一种基于深度学习的混合水果分类方法, 构建了 CAE-ADN 框架, 结合了注意力模型和卷积自动编码器, 实现了高效的水果图像分类[18]。Lu 等人基于 DenseNet201 模型和迁移学习技术, 实现了番茄分类的高准确率, 即使在图像质量受干扰的情况下, 分类准确率仍达到 96.16% [19]。

此外, Chandel 等人基于 GoogLeNet 模型, 提出了一种用于识别农作物水分胁迫条件的方法, 在 1200 张包含玉米、秋葵和大豆的数据集上, 表现优异[20]。Kang 等人基于 ResNet 模型, 建立了一个能够区分水果新鲜度和种类的分类模型, 收集了 18,000 张包含 7 类水果的图片, 通过迁移学习技术进行模型训练, 新鲜度分类的识别准确率达到 98.50%, 种类分类的识别准确率为 97.43% [21]。Ismail 等人提出了一种基于 CNN 的实时水果等级分类系统, 使用 EfficientNet 模型对苹果和香蕉的数据集进行验证, 分类准确率达 99.2% [22]。Huang 等人提出了一种基于 ViT 的神经网络模型, 建立了一个包含 9375 次对 15 种水果的触觉数据集[23]。这些研究展示了深度学习在水果识别中的广泛应用和高效性能, 为进一步研究和应用提供了坚实的基础。

3. 方法

3.1. 基础模型选择

针对移动端的水果识别应用部署, 本文选用内存为 2 GB 的 Rockchip RK3399-mid 芯片, 因此需要考虑模型权重大小和模型计算量。因此本文对模型的筛选规则为模型计算量(FLOPS/10⁶)小于 10,000, 模型参数量小于 100 M。通过对目前主流的网络模型进行筛选, 本文选用 DenseNet-169 [30]为本文的骨干网络, 用于后续的研究, 其模型结构如图 1 所示。

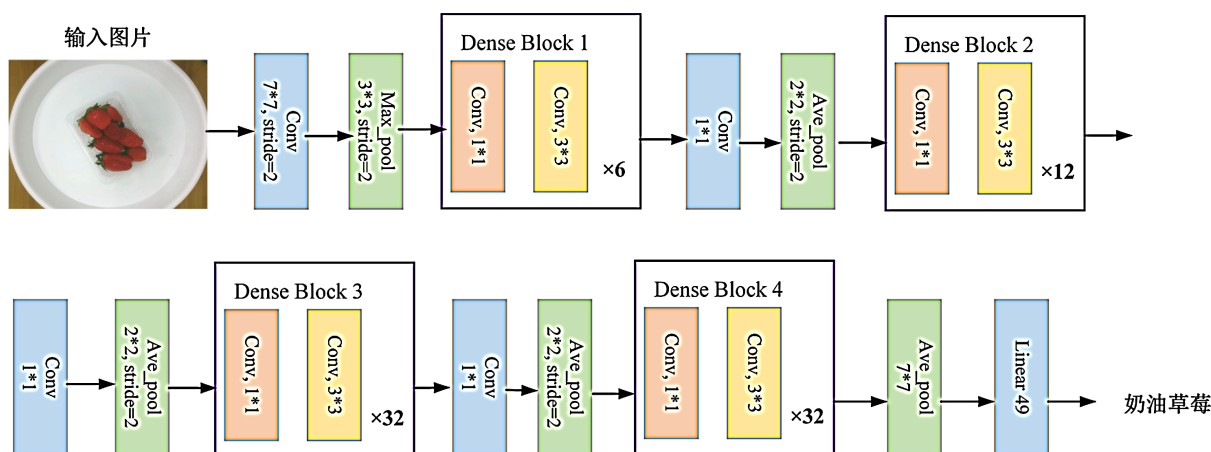


Figure 1. DenseNet-169 model architecture

图 1. DenseNet-169 模型结构

3.2. 注意力机制

为进一步提升模型性能。让模型更好地关注到水果的细节纹理, 小目标水果特征, 并缓解包装、摆放数量等干扰信息的影响, 本文尝试在骨干网络中引入注意力机制。结合 DenseNet-169 模型自身特点, 其优势在于稠密块连接, 如在稠密块中添加注意力将会破坏稠密结构, 因此, 本文主要在其稠密块连接处进行添加注意力。

本文设计新的注意力模块, 残差混合注意力(Residual Mixed attention, RMA)。同时考虑空间和通道信息, 使用多尺度卷积特征融合形式提取通道信息, 并添加残差连接, 相比于单一尺度特征提取, 多尺度

特征可以动态调整对目标的关注程度, 使得模型对于尺度变化具有更好的适应性, 综合利用水果的全局结构和局部细节信息提高模型对图像的识别准确性, 从而提高模型的鲁棒性。

其模型结构如图 2 所示, 首先分别通过 3×3 卷积核和 5×5 卷积核对输入特征进行卷积提取, 得到两个不同尺度的特征图, 然后通过相加操作同时结合两个不同尺度的特征图, 输入至全局平均池化层得到权重向量, 并通过一个线性层来映射全局特征向量, 再通过 Sigmoid 激活函数将其归一化到 0 到 1 之间, 得到全局特征向量, 然后分别对所提取的多尺度特征进行加权输出, 并通过 Concat 操作按照通道拼接, 然后对特征使用 1×1 卷积核进行通道融合并输入空间注意力模块。

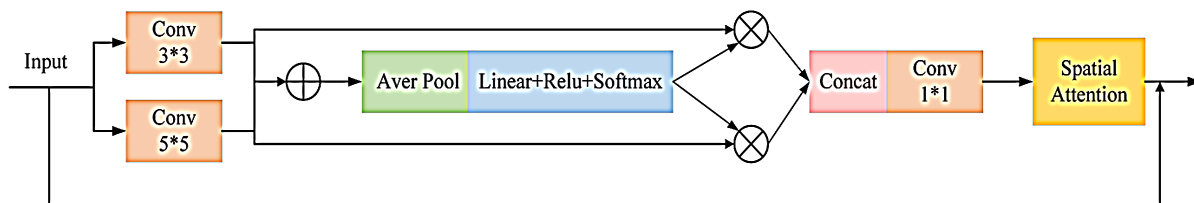


Figure 2. Residual Mixed Attention (RMA) module

图 2. 残差混合注意力 RMA 模块

3.3. 添加 ViT Block

深层特征是对输入数据进行多次非线性变换后得到的结果, 具有更高级的语义信息, 因此对深层特征的信息整合可以使得模型具有更好的泛化性, 提升识别精度。ViT [24] 架构使用自注意力机制来捕捉图像中的全局信息, 能够更好地理解图像整体结构和上下文关系, 具有更好的语义理解能力, 可以增强模型的泛化能力。

因此为提升模型深层语义提取能力, 增强泛化性。本文实现适合在 CNN 模型中添加的 ViT Block, 其模型结构如图 3 所示, 通过对三维输入特征按照形状依次展平变成二维特征, 并进行线性映射实现通道特征融合得到新的长度为 D 的二维特征, 然后对特征进行位置编码, 输入 Encoder 块, 进行深层特征抽取, 并输出特征。

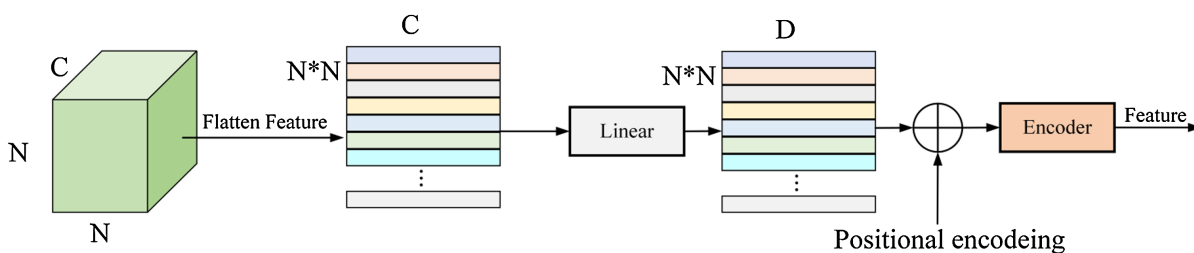


Figure 3. ViT Block module

图 3. ViT Block 结构图

本文尝试在 DenseNet-169 模型中添加基于 ViT 架构的语义提取模块。通过分析认为 CNN 在处理图像浅层特征时具有先验知识即归纳偏置(Inductive Bias)例如平移等变性、局部连接性, 因此能够更好地利用图像本身的空间信息, 而 ViT 结构缺少了这种归纳偏置, 并且在前期 ViT 架构计算复杂度也较高。因此对于低级特征处理阶段 CNN 更具优势, 而随着模型层数的加深, 多重卷积操作使得模型特征逐渐深层化, 使得特征更加具有语义性, 且特征尺寸有所下降, 此时 ViT Block 更具有优势。最终的模型结构如图 4 所示。

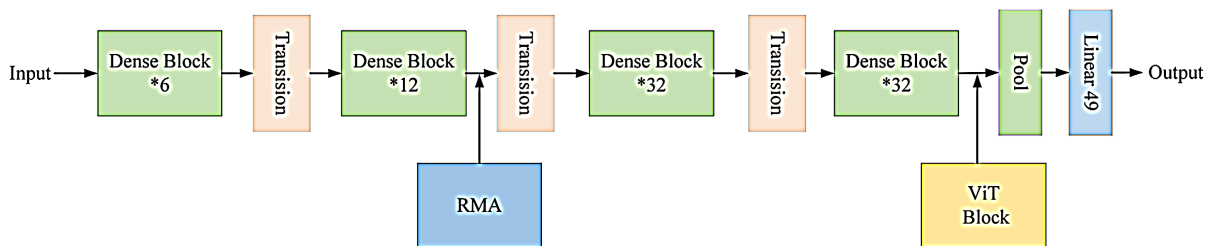


Figure 4. DenseRMA_ViT model architecture

图 4. DenseRMA_ViT 模型结构图

3.4. 改进损失函数

损失函数在分类模型中扮演着非常重要的角色, 用于衡量模型预测与真实标签的差距, 模型根据损失来调整参数, 模型的训练过程旨在最小化损失函数, 使得模型能够从训练数据中学到合适的特征和规律, 以便正确分类样本。因此选择合适的损失函数可以帮助提高模型的性能、泛化能力以及对特定问题的适应性。常规分类损失函数为交叉熵损失(Cross-Entropy Loss), 计算公式如下:

$$L_{CE}(p, y) = -\sum_{i=1}^K y_i \log(p_i) \quad (1)$$

其中 K 为水果类别数, P 为预测概率类别向量, y 为真实标签独热编码向量。

通过分析模型训练损失曲线图和数据集特点发现, 数据集样本中简单样本居多, 困难样本相对较少。而交叉熵损失函数没有对简单样本和困难样本的损失进行区分。虽然困难样本的损失较大, 但是简单样本数量多, 这将会导致模型过多地关注简单样本的损失, 模型难以学习困难样本的特征。同时本文水果识别系统在后续使用中会对识别错误的水果图片进行收集, 并进行模型微调, 然而在实际使用中, 对于热销水果收集到的图片必定更多, 这对于后续模型微调将会造成类别样本不均衡问题。基于上述两个问题, 本文基于 Focal Loss [25] 对损失函数进行改进, 计算公式如下:

$$L_F(p_t) = -w_t (1 - p_t)^2 \log(p_t) \quad (2)$$

其中 t 为对应类别序号, w_t 为对应类别损失权重, 在训练之前设定, 统计训练样本类别数量, 设定样本数最高类别的损失权重为 1, 其他类别的损失权重为与样本数最高类别的比值的倒数, 即数量越少权重越大, 因此相对地提升小样本数量类别的损失权重。 P_t 为对应预测类别概率。当 P_t 越接近 1, 即样本为简单样本, $(1 - P_t)^2$ 则相对越小, 模型对简单样本的损失相对权重越小。当 P_t 越接近 0, 即样本为困难样本, $(1 - P_t)^2$ 则相对越大, 模型对困难样本的损失相对权重越大。

4. 实验

4.1. 数据集

目前常见的水果数据集包括 FruitVeg-81 [26]、Fruits-262 [27]、Fruits-360 [28]等, 其中数字代表数据集中水果种类数量, 图 5 展示了部分上述水果数据集的图片。可以看出尽管 Fruits-360 的水果种类众多, 但相对于其他数据集过于简单, 不适合实际超市场景的应用。FruitVeg-81 数据集除了水果图片, 还包含了蔬菜图片, 并且果蔬类别数量相较于 Fruits-262 较少, 因此在公开数据集中, Fruits-262 数据集更加适合作为水果识别系统模型训练数据集。

尽管 Fruits-262 包含大量水果种类和图片, 但与本地超市的水果种类仍有差距。例如, 本地超市的苹果种类包括树顶红富士、优选红富士、脆心甜红富士、甜心小苹果, 而 Fruits-262 中有 apple (苹果)、elephant

apple (象苹果)、malay apple (马来苹果)、otaheite apple (塔希提苹果), 相关图片如图 6 所示。虽然其他三类水果名称与 apple 相关, 但与本地常见苹果差异较大, 难以视为同一类。类似情况在其他水果中也存在。因此, Fruits-262 尽管种类多, 但未细分类, 也未覆盖本地超市的水果种类。因此, 构建适用于本地超市的水果数据集对于水果识别系统至关重要。



Figure 5. Sample Images from the public dataset

图 5. 公开数据集部分图片



Figure 6. Images of some apple varieties from the fruits-262 dataset

图 6. Fruits-262 部分相关苹果种类图片

本文通过自制数据集用于训练模型得到权重, 用于实际超市水果识别系统, 并且使用公开数据集 Fruits-262 验证模型在算法上改进的有效性。在实际超市应用场景中, 水果种类繁多同时存在较多细分类, 因此为贴合超市的实际使用, 在拍摄时覆盖超市所有种类水果, 确保数据集的多样性。本文拍摄 49 种超市水果, 在实际分类模型中, 模型依照序号对类别预测概率进行输出。由于水果包装状态存在差异, 包括袋装盒装和散装等, 因此本文考虑到实际使用针对不同水果的不同包装进行拍摄, 提高模型的泛化性。针对不同水果的状态拍摄, 如图 7 为小番茄不同状态下的拍摄图片。

数据集扩增有助于提升模型泛化性和识别精度, 防止过拟合。Meta 开源的 SAM (Segment Anything Model) [29] 实现了无需标注即可对任意图像中的任何物体进行分割。本文基于 SAM 设计了水果数据集扩增方法。将图片传入 SAM, 随机生成多个提示输入, 对图像进行分割, 得到分割对象。然后对分割对象

进行 1 到 4 次随机复制, 并随机翻转、旋转、缩放, 在背景图像上随机粘贴, 完成数据扩增。如图 8 所示, 对同一张草莓图像随机生成提示点进行扩充, 左列为随机提示点, 中间为分割实例, 右列为最终扩充图。最终, 每种水果图片扩增至 6000 张, 总计 29.4 万张, 命名为 DailyFruit-49。

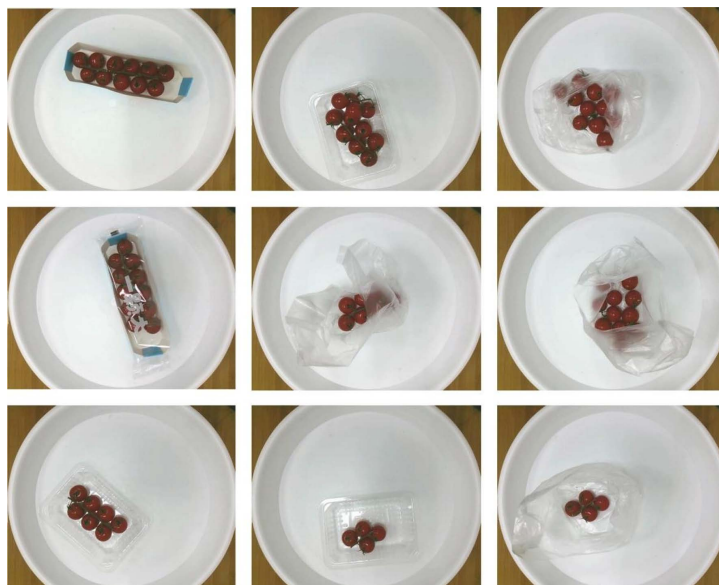


Figure 7. Photoshoot of cherry tomatoes in different states

图 7. 小番茄不同状态下的图片拍摄

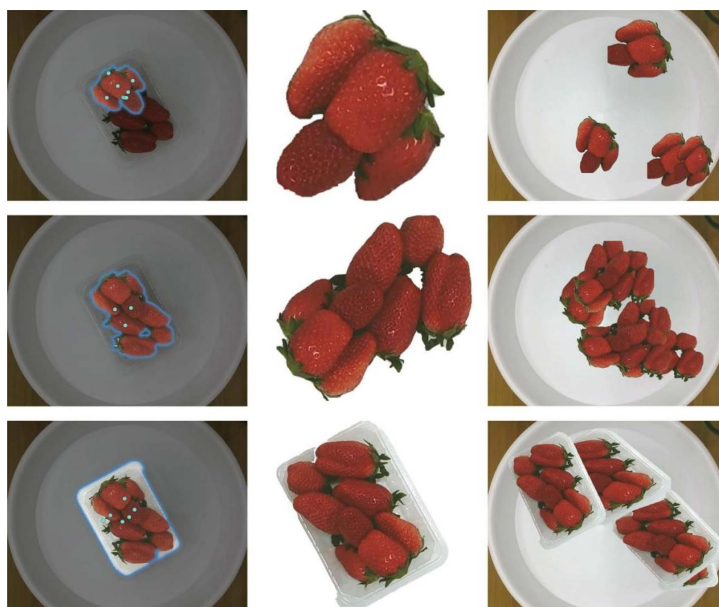


Figure 8. Fruit image augmentation concept based on SAM

图 8. 基于 SAM 的水果图片扩增示意图

4.2. 评价指标

为了全面客观地评价模型对水果识别的性能, 本文采用多种图像分类评价指标, 包括准确率 (Accuracy)、精确率(Precision)、召回率(Recall)、F1 分数(F1 Score, F1)。准确率、精确率、召回率的数学

表达式如式(3)、(4)、(5)所示。表 1 解释了式中元素具体情况。

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$pre = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \times pre \times recall}{pre + recall} \quad (6)$$

Table 1. Table of precision and recall explanation

表 1. 精确率和召回率解释表

区分	预测 Positive	预测 Negative	合计
实际 Positive	TP (True Positive)	FN (False Negative)	TP + FN
实际 Negative	FP (False Positive)	TN (True Negative)	FP + TN
合计	TP + FP	FN + TN	TP + FP + FN + TN

F1 分数是精确率和召回率的调和平均, 综合考虑了分类器的精确性和召回率, 是一个综合性的评价指标, 可以更加综合地反映了分类器的性能。其数学表达式如式(4)所示。

4.3. 消融实验

为验证改进模型的有效性, 本文使用公开数据集 Fruits-262 进行消融实验, 验证算法改进的有效性, 尽管 Fruits-262 数据与实际应用数据相差较大, 但是从算法角度, 复杂数据依然存在参考价值。为证明本文对模型改进的有效性, 对 RMA 模块、ViT Block、Focal Loss 进行消融实验, 并统计结果如表 2 所示, 由表可知相较于基础骨干网络, 3 种改进方案的有效性, 模型最终 Accuracy、Precision、Recall、F1 为 75.28%、75.84、75.37%、75.60%, 相较于仅添加 RMA 模块, 虽然精确率有所降低, 但是召回率和综合性能 F1 有所提升, 说明模型有效缓解了样本不均衡, 提升了模型的综合性能, 证明 Focal Loss 改进的有效性。同时对比原始模型 1 号 DenseNet-169, 添加改进模块的模型性能均有所提升, 证明 DenseRAM_ViT 改进的有效性。

Table 2. Ablation study results on the Fruits-262 dataset

表 2. 在 Fruits-262 数据集上的消融实验结果

编号	RMA	ViT Block	Focal	Accuracy/%	Precision/%	Recall/%	F1/%
1	-	-	-	73.63	74.94	74.21	74.57
2	√	-	-	74.64	75.90	74.75	75.32
3	-	√	-	74.63	75.77	74.76	75.26
4	-	-	√	74.02	75.12	74.48	74.80
5	√	√	√	75.28	75.84	75.37	75.60

同时为了直观展现添加 RMA 的效果对比, 如图 9 使用梯度类激活热力图对 Fruits-262 数据集中部分图像各区域受关注程度做可视化展示, 由图第一行对比可知, 相对于原始模型 DenseNet-169, 添加了 RMA

注意力模型减少了对背景的关注, 并且加大了对水果目标的关注, 由图第二行对比可知, 改进模型增大了对水果目标的关注程度与范围, 因此证明了本文改进注意力 RMA 的有效性。

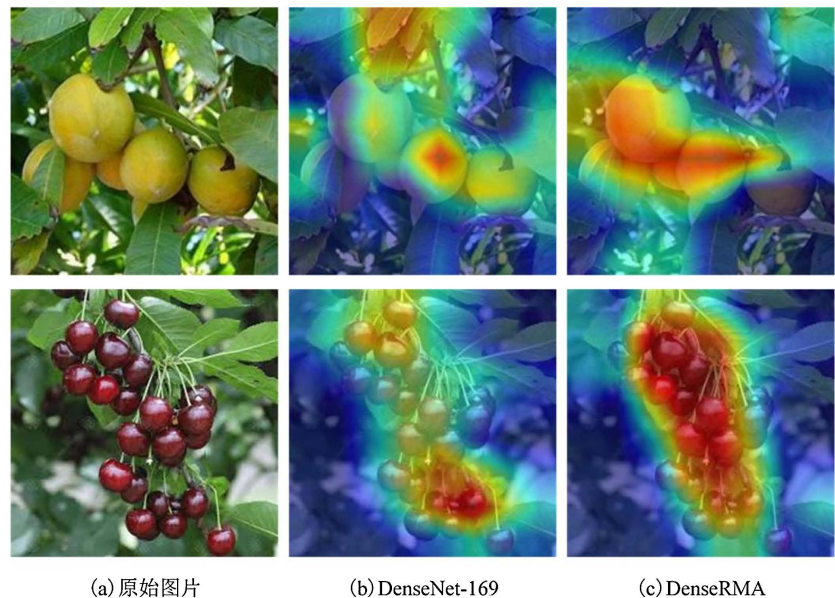


Figure 9. Comparison of heatmaps with and without the Introduction of the RMA module
图 9. 引入 RMA 模块热力图对比

4.4. 结果对比

为了说明本文模型设计的有效性, 列举具有代表性的同类水果识别研究进行对比, 结果如表 3 所示, 尽管有些研究识别性能很好, 但是他们的研究水果种类与数量较少, 不具有代表性。在公开数据集 Fruits-262 上, 本文最终实现模型 DenseRMA_ViT 性能相对于数据集发布者 Minut [28]有较大提升。在自制数

Table 3. Comparison of related studies on fruit recognition
表 3. 水果识别相关研究对比

文献	年份	数据集	水果种类	样本数量	Top1 Acc/%	Top5 Acc/%
Enciso [16]	2018	自制	1	320	98.25	—
Lu [19]	2021	自制	9	6100	96.16	—
Kang [21]	2022	自制	7	18000	97.43	—
Ismail [22]	2022	自制	2	8791	96.7%	—
Minut [28]	2021	Fruits-262	262	225,640	58%	—
DenseNet [30]	2017	Fruits-262	262	225,640	73.63%	—
DenseRMA_ViT (our)	2024	Fruits-262	262	225,640	75.28%	—
Mobile v2 [31]	2018	DailyFruit-49	49	294,000	83.65	92.64
ResNet [32]	2015	DailyFruit-49	49	294,000	85.75	93.57
DenseNet [30]	2017	DailyFruit-49	49	294,000	87.53	98.83
ViT [24]	2020	DailyFruit-49	49	294,000	87.95	97.07
DenseRMA_ViT (our)	2024	DailyFruit-49	49	294,000	92.24	99.73

据集 DailyFruit-49 上, 本文对比了深度学习经典分类模型 Mobile v2、ResNet、DenseNet、ViT, 由此证明模型的改进有效性, 同时本实验覆盖的水果种类范围大, 前 5 类识别精度达到 99.73%, 识别率能够满足实际使用需求。综上可知本实验设计的水果识别算法在水果识别领域具有应用价值。

4.5. 系统实现

系统的终端为 Android 系统, 深度学习模型基于 Pytorch 框架和 Python 语言实现, 但无法直接在终端部署。因此, 需要将模型转换为适合终端部署的格式。本文采用 TorchScript 进行模型格式转化。TorchScript [33] 具有运行独立性、高性能、跨平台支持等优点, 可以将 PyTorch 模型转换为独立格式, 使其无需依赖 Python 解释器, 可以在移动端和嵌入式设备上运行, 并且无需重新训练模型。同时, TorchScript 的静态图表示可以优化模型性能和效率, 如进行模型剪枝等。经过测试, 水果识别终端设备不需要高性能 CPU 或 GPU, 在瑞芯微的 RK3399 芯片上可以流畅运行, 其他设备如摄像头也不需要高端配置。

本文系统设计框架如图 10 所示, 分为四个模块: 水果检测模块、前端交互模块、后台自更新模块、终端更新模块。四个模块相互连接, 形成闭环, 实现系统自更新。系统通过摄像头实时拍摄用户购买的水果图像, 使用模型进行识别, 并在终端界面展示供用户交互。用户可在终端界面更正识别错误的水果种类, 并将正确信息上传至后台服务器。后台系统收集误识别图片, 管理人员进行检测与修正。当误识别图片达到一定数量, 后台启动迁移学习进行模型微调, 并发布新权重。终端设备每日重启时会检查新权重并进行更新。整个系统实现了水果图像自动化识别与模型权重自优化, 随着使用时间的推移, 系统将获得更大的水果图像数据集和更优的模型权重。

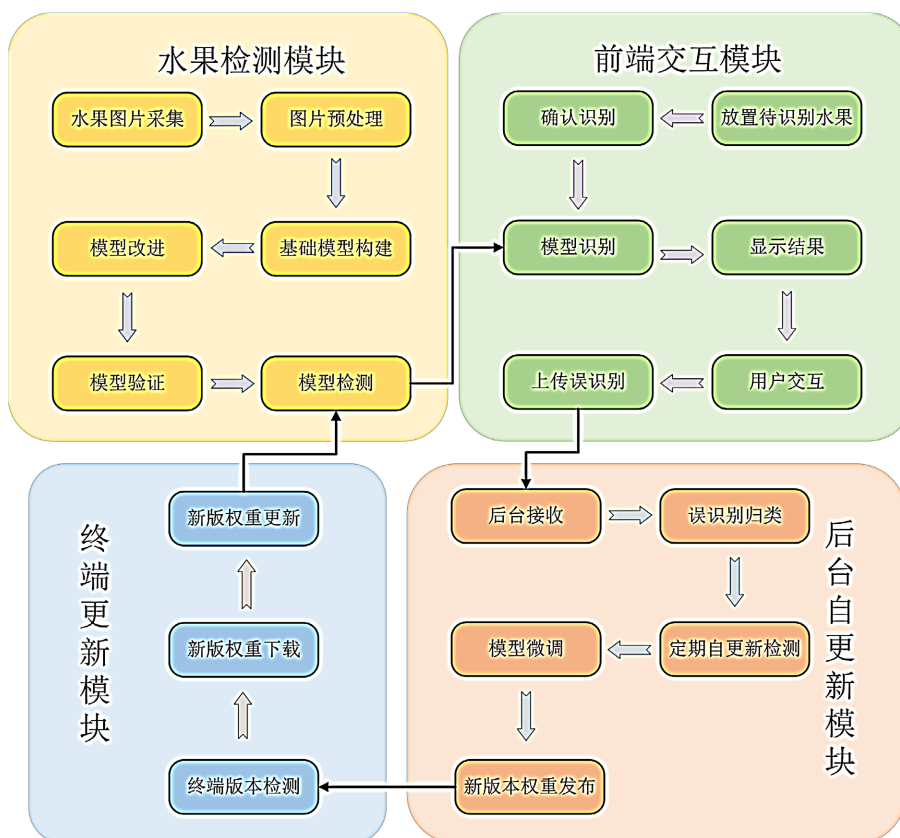


Figure 10. Framework diagram of deep learning-based fruit detection system

图 10. 基于深度学习的水果检测系统设计框架图

水果检测模块为主要功能, 涉及整个装置, 其整体结果如图 11 所示, 该模块包括摄像头拍摄模块、水果展示台、终端处理模块。由用户将待购买商品放入展示台, 并通过摄像头进行拍摄传入终端, 终端设备部署识别程序, 对水果进行识别。模型的部署包括对所得到的 Pytorch 模型权重 pth 文件进行转换为 TorchScript 格式, 再结合 Android 应用程序进行模型嵌入转换好的模型完成部署。

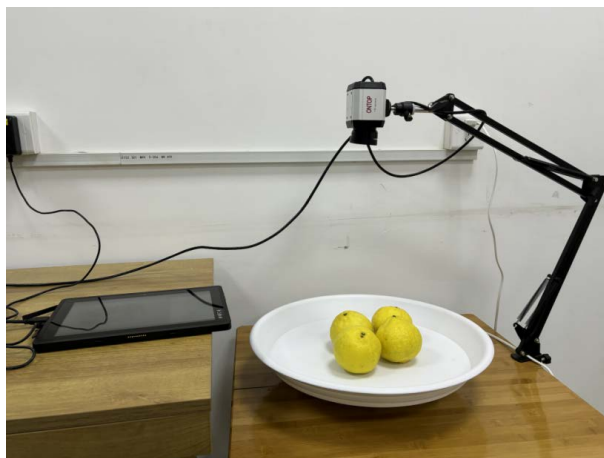


Figure 11. Fruit photography device
图 11. 水果拍摄装置

4.6. 实际超市水果验证

为验证 DenseRMA_ViT 水果识别算法的有效性, 在本地超市进行了实地拍摄和识别测试。由于季节原因, 部分水果未在售, 对 28 种在售水果进行拍摄, 每种 4 张, 分别以袋装和散装 1:3 拍摄, 共 112 张。拍摄过程中不固定光照、角度和距离, 以增加难度。识别结果显示, Top5 全部正确, 但有些不足, 如奇异果的 Top1 识别错误。整体上, 模型 Top1 识别准确率为 84.82%, Top5 为 96.43%。相比基础模型 DenseNet-169 的 Top1 准确率 79.46%、Top5 准确率 90.18%, 改进模型分别提升了 5.36% 和 6.25%。尽管存在误识别, 系统可以收集误识别图像, 更新数据集并微调模型权重, 从而提升精度。

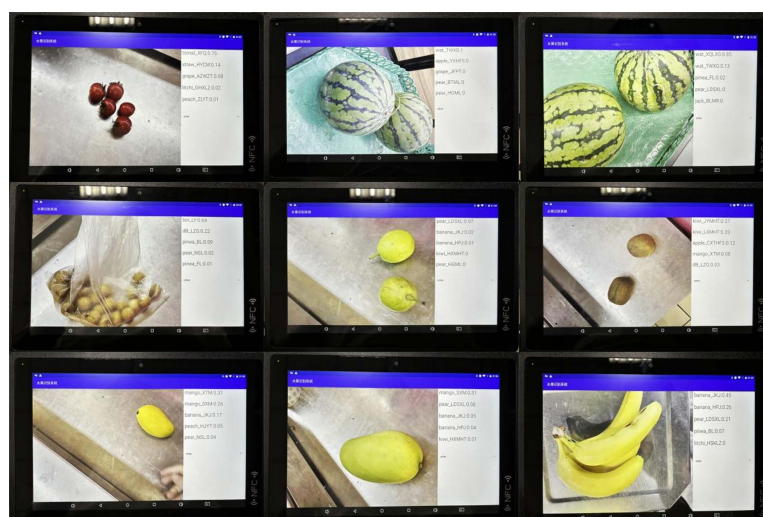


Figure 12. Supermarket real-world testing
图 12. 超市实拍测试

5. 结论

水果自动识别提升商店效益, 减少标签, 环保, 是智慧农业关键。本文为本地超市构建了新水果数据集, 设计了基于深度学习的识别系统, 实现自动识别, 加快智慧农业推广。拍摄超市各种水果, 考虑盒装、袋装、散装等不同包装。首次应用 SAM 进行数据扩增, 构建包含 49 种水果、29.4 万张图片的 DailyFruit-49 数据集。选择 DenseNet 为基础模型, 结合通道与空间注意力机制, 设计 RMA 注意力机制, 改进 ViT Block 用于深层语义特征提取, 得到 DenseRMA_ViT 模型。基于 Focal Loss 改进交叉熵损失函数, 增强对困难和少样本水果的关注, 提升模型泛化性。在低算力终端部署识别系统, 并设计自动微调模型权重的机制, 通过收集误识别图片不断优化模型, 提升识别率, 验证超市实际效果。

参考文献

- [1] 吴中勇, 李延荣, 董中丹. 我国水果市场发展现状及对策研究[J]. 中国果菜, 2023, 43(11): 79-83+87.
- [2] 中研普华公司, 2022-2027 年中国果蔬行业市场全面分析及发展趋势调研报告[R]. 深圳: 中国行业研究网, 2022.
- [3] Jana, S., Basak, S. and Parekh, R. (2017) Automatic Fruit Recognition from Natural Images Using Color and Texture Features. 2017 *Devices for Integrated Circuit (DevIC)*, Kalyani, 23-24 March 2017, 620-624.
<https://doi.org/10.1109/devic.2017.8074025>
- [4] Novak, C.L. and Shafer, S.A. (1992) Anatomy of a Color Histogram. *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, 15-18 June 1992, 559-605.
<https://doi.org/10.1109/cvpr.1992.223129>
- [5] Movellan, J.R. (2002) Tutorial on Gabor filters. *Open Source Document*, **40**, 1-23.
- [6] Pietikäinen, M. (2010) Local Binary Patterns. *Scholarpedia*, **5**, 9775. <https://doi.org/10.4249/scholarpedia.9775>
- [7] Tomasi, C. (2012) Histograms of Oriented Gradients. *Computer Vision Sampler*, **1**, 1-6.
- [8] Gao, W.S., Zhang, X.G., Yang, L. and Liu, H.Z. (2010) An Improved Sobel Edge Detection. 2010 *3rd International Conference on Computer Science and Information Technology*, Chengdu, 9-11 July 2010, 67-71.
<https://doi.org/10.1109/iccsit.2010.5563693>
- [9] Ding, L. and Goshtasby, A. (2001) On the Canny Edge Detector. *Pattern Recognition*, **34**, 721-725.
[https://doi.org/10.1016/s0031-3203\(00\)00023-6](https://doi.org/10.1016/s0031-3203(00)00023-6)
- [10] Cruz-Mota, J., Bogdanova, I., Paquier, B., Bierlaire, M. and Thiran, J. (2011) Scale Invariant Feature Transform on the Sphere: Theory and Applications. *International Journal of Computer Vision*, **98**, 217-241.
<https://doi.org/10.1007/s11263-011-0505-4>
- [11] Wold, S., Esbensen, K. and Geladi, P. (1987) Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [12] Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. and Scholkopf, B. (1998) Support Vector Machines. *IEEE Intelligent Systems and their Applications*, **13**, 18-28. <https://doi.org/10.1109/5254.708428>
- [13] Kramer, O. (2013) K-Nearest Neighbors. In: Kramer, O., Ed., *Dimensionality Reduction with Unsupervised Nearest Neighbors*, Springer Berlin Heidelberg, 13-23. https://doi.org/10.1007/978-3-642-38652-7_2
- [14] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [15] Song, Y.Y. and Ying, L.U. (2015) Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Archives of Psychiatry*, **27**, 130.
- [16] Enciso-Aragón, C.J., Pachón-Suescún, C.G. and Jimenez-Moreno, R. (2018) Quality Control System by Means of CNN and Fuzzy Systems. *International Journal of Applied Engineering Research*, **13**, 12846-12853.
- [17] 孟欣欣, 阿里甫·库尔班, 吕情深, 等. 基于迁移学习的自然环境下香梨目标识别研究[J]. 新疆大学学报(自然科学版), 2019, 36(4): 461-467.
- [18] Xue, G., Liu, S. and Ma, Y. (2020) A Hybrid Deep Learning-Based Fruit Classification Using Attention Model and Convolution Autoencoder. *Complex & Intelligent Systems*, **9**, 2209-2219. <https://doi.org/10.1007/s40747-020-00192-x>
- [19] Lu, T., Han, B., Chen, L., Yu, F. and Xue, C. (2021) A Generic Intelligent Tomato Classification System for Practical Applications Using Densenet-201 with Transfer Learning. *Scientific Reports*, **11**, Article No. 15824.
<https://doi.org/10.1038/s41598-021-95218-w>
- [20] Chandel, N.S., Chakraborty, S.K., Rajwade, Y.A., Dubey, K., Tiwari, M.K. and Jat, D. (2020) Identifying Crop Water

- Stress Using Deep Learning Models. *Neural Computing and Applications*, **33**, 5353-5367. <https://doi.org/10.1007/s00521-020-05325-4>
- [21] Kang, J. and Gwak, J. (2021) Ensemble of Multi-Task Deep Convolutional Neural Networks Using Transfer Learning for Fruit Freshness Classification. *Multimedia Tools and Applications*, **81**, 22355-22377. <https://doi.org/10.1007/s11042-021-11282-4>
 - [22] Ismail, N. and Malik, O.A. (2022) Real-time Visual Inspection System for Grading Fruits Using Computer Vision and Deep Learning Techniques. *Information Processing in Agriculture*, **9**, 24-37. <https://doi.org/10.1016/j.inpa.2021.01.005>
 - [23] Huang, R., Zheng, W., Zhang, B., Zhou, J., Cui, Z. and Zhang, Z. (2023) Deep Learning with Tactile Sequences Enables Fruit Recognition and Force Prediction for Damage-Free Grasping. *Computers and Electronics in Agriculture*, **211**, Article ID: 107985. <https://doi.org/10.1016/j.compag.2023.107985>
 - [24] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
 - [25] Lin, T., Goyal, P., Girshick, R., He, K. and Dollar, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/iccv.2017.324>
 - [26] Waltner, G., Schwarz, M., Ladstätter, S., Weber, A., Luley, P., Lindschinger, M., *et al.* (2017) Personalized Dietary Self-Management Using Mobile Vision-Based Assistance. In: Battiato, S., Farinella, G., Leo, M. and Gallo, G., Eds., *New Trends in Image Analysis and Processing—ICIAP 2017*, Springer International Publishing, 385-393. https://doi.org/10.1007/978-3-319-70742-6_36
 - [27] Minut, M. and Iftene, A. (2021) Creating a Dataset and Models Based on Convolutional Neural Networks to Improve Fruit Classification. 2021 23rd *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Timisoara, 7-10 December 2021, 155-162. <https://doi.org/10.1109/synasc54541.2021.00035>
 - [28] Mureşan, H. and Oltean, M. (2018) Fruit Recognition from Images Using Deep Learning. *Acta Universitatis Sapientiae, Informatica*, **10**, 26-42. <https://doi.org/10.2478/ausi-2018-0002>
 - [29] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
 - [30] 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>
 - [31] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/cvpr.2018.00474>
 - [32] 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>
 - [33] Reed, J., DeVito, Z., He, H., *et al.* (2022) Torch. fx: Practical Program Capture and Transformation for Deep Learning in Python. *Proceedings of Machine Learning and Systems*, **4**, 638-651.