

基于神经网络算法的水果分类技术及其实现

刘齐军¹, 罗建书^{2*}, 冯 秀², 黄林香²

¹山东协和学院计算机学院, 山东 济南

²湖南交通工程学院电气与信息工程学院, 湖南 衡阳

收稿日期: 2026年9月23日; 录用日期: 2026年10月20日; 发布日期: 2026年10月27日

摘 要

本文提出一种基于CBAM注意力机制改进的轻量级MobileNetV2水果分类系统。首先, 以MobileNetV2为骨干网络, 在其特征提取末端嵌入CBAM模块, 通过通道注意力和空间注意力增强模型对水果关键判别区域的特征表达能力; 其次, 在训练过程中动态引入MixUp和CutMix混合增强策略, 有效提升模型的泛化能力与鲁棒性; 最后, 基于Tkinter框架开发了可视化GUI演示程序。实验在10类常见水果数据集上进行, 结果表明, 改进后的模型在保持轻量化的同时取得了良好的分类准确率, 验证了所提方法的有效性。

关键词

水果分类, 神经网络, 轻量化模型, CBAM注意力机制, MobileNetV2

Fruit Classification Technology and Its Implementation Based on Neural Network Algorithms

Qijun Liu¹, Jianshu Luo^{2*}, Xiu Feng², Linxiang Huang²

¹School of Computer Science, Shandong Xiehe Institute, Jinan Shandong

²School of Electrical and Information Engineering, Hunan Institute of Traffic Engineering, Hengyang Hunan

Received: September 23, 2026; accepted: October 20, 2026; published: October 27, 2026

Abstract

This paper proposes a lightweight fruit classification system based on an improved MobileNetV2 with the CBAM attention mechanism. First, MobileNetV2 is adopted as the backbone network, and a

*通讯作者。

文章引用: 刘齐军, 罗建书, 冯秀, 黄林香. 基于神经网络算法的水果分类技术及其实现[J]. 计算机科学与应用, 2026, 16(10): 141-155. DOI: 10.12677/csa.2026.1610318

CBAM module is embedded at the end of its feature extraction pathway to enhance the model's ability to represent key discriminative regions of fruits through channel attention and spatial attention. Second, MixUp and CutMix data augmentation strategies are dynamically introduced during training to effectively improve the model's generalization ability and robustness. Finally, a visual GUI demonstration program is developed using the Tkinter framework. Experiments are conducted on a dataset containing ten common fruit categories. The results show that the improved model achieves good classification accuracy while maintaining a lightweight architecture, demonstrating the effectiveness of the proposed method.

Keywords

Fruit Classification, Neural Networks, Lightweight Model, CBAM Attention Mechanism, MobileNetV2

Copyright © 2026 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

关于 CNN 基础应用, Lu 等人基于 DenseNet-201 迁移学习方法展开了番茄智能分类系统研究, 实验证实了深度卷积神经网络在水果分类任务中的有效性, 验证了网络深度对分类性能的提升作用[1]。Shahi 等人面对水果分类任务中深度模型参数量大、存储和计算资源要求高的情况, 提出了基于注意力机制改进的 MobileNetV2 轻量化模型, 采用迁移学习方法验证了预训练模型在小样本条件下的有效性, 为解决资源受限场景下的水果识别问题提供了可行方法[2]。

Udefi 等人则将研究视角聚焦在西非本土水果的识别问题。利用卷积神经网络模型对相对稀缺的西非水果图像进行分类, 展示了深度学习于低资源区域农业物种识别中的潜力。这项工作扩大了水果分类研究的物种覆盖范围, 具有重要的区域应用价值[3]。

在轻量化和注意力机制方面, 虽然国外研究在 MobileNet、EfficientNet 等高效网络架构上进行了广泛探索, 但专门将 CBAM 模块嵌入轻量级网络并应用于水果分类的研究仍较为有限。现有工作主要集中在工程领域的通用图像分类任务, 针对农业场景的系统性优化尚不充分。此外, MixUp、CutMix 等混合增强策略在国外水果分类相关研究中已被证明能够有效提升泛化能力, 但将其与注意力机制联合优化的工作仍较少。

国内学者对于水果分类及检测方面的研究由早期传统方法到如今运用深度学习的技术转变过程, 最初的研究主要是利用人工提取的特征以及传统的机器学习算法, 何伟基于卷积神经网络设计并实现了果蔬识别与定位软件, 验证了深度学习模型在果蔬识别任务中的有效性, 为农业自动化提供技术支撑[4]。王士龙等人基于机器视觉和傅里叶变换方法实现了沙糖橘果皮缺陷检测, 推动了传统图像处理方法在水果品质分级中的应用[5]。

随着卷积神经网络的兴起, 国内研究者开始广泛应用深度学习模型进行水果图像分类。刘素娇等人基于改进鲸鱼优化卷积神经网络展开了红富士苹果外观分级研究, 有效地提升了水果外观品质分级的准确率, 验证了深度学习模型在该类任务中的有效性[6]。简钦等人提出了 FC-CNN 模型, 专门针对水果图像分类问题设计了卷积神经网络结构[7]。黄玉富等人从多尺度特征融合的角度出发, 研究了水果图像识别算法, 通过融合不同尺度的特征信息提升了识别精度[8]。

在轻量化网络方面,部分研究者开始关注模型的部署效率问题。刘敏聪等人基于改进 YOLOv5 算法开展了水果识别分类研究,在检测速度和准确率之间取得了较好的平衡,验证了 YOLO 系列算法在水果分类任务中的有效性[9]。毕好昌和黄宇提出 VisionIntelliNet 模型用于水果分类检测与实验,进一步推动了轻量化模型在该领域的应用探索[10]。李银银等人对卷积神经网络进行了改进,设计了专门的水果分类算法[11]。

注意力机制在水果分类中的应用也逐渐成为研究热点。吴迪等人在改进的 Res2Net 网络中引入了动态多尺度融合注意力模块,通过自适应加权机制增强了模型对水果图像判别性特征的关注能力[12]。李莹等人基于改进 YOLOv8s 模型开展了柑橘果实成熟度检测研究,在主干网络中引入混合注意力变换器模块,显著提升了模型对果实特征的识别精度[13]。

吴晓瑶等人提出了基于改进 YOLOv8n-seg 的樱桃叠果与瑕疵实时分割方法,通过融合颜色、形状和纹理等多维特征实现了水果表面缺陷的精准检测[14]。颜玉洁等人在《食品科学》上发表综述,系统总结了深度学习驱动的机器视觉技术在果蔬品质智能感知中的研究进展,涵盖了轻量化模型开发、多模态融合等前沿方向,为该领域的技术发展提供了系统性参考[15]。总体上看,国内研究已从早期的基础性工作发展到对模型的精细化改良及轻量化处理,但是在将 CBAM 注意力机制与 MobileNetV2 轻量级网络结合起来,并且用于水果分类方面还存在着研究的余地。

总之,国内外对于水果分类的研究都取得了一定的成绩,国内的学者对于网络改进、轻量化研究以及注意力机制都做了大量的研究工作,而国外的研究在迁移学习、地区性特殊物种分类等方面有一定的优势。但是,目前还没有很多工作能够把 CBAM 注意力机制及 MobileNetV2 轻量级网络很好地结合在一起,在加上 MixUp/CutMix 混合增强方法,还没有进行系统的轻量化评估的工作。本文就是在这样的背景下进行的研究,希望可以开发出一个准确率与速度并重的水果分类模型。

2. 理论和技术

2.1. 卷积神经网络基础

卷积神经网络(Convolutional Neural Network, CNN)是深度学习领域的基础模型之一,在生物视觉皮层的信息传播基础上建立起来的。CNN 通过局部连接、权值共享、池化操作,实现对图像层次化特征的提取。CNN 的典型基础结构见图 1。

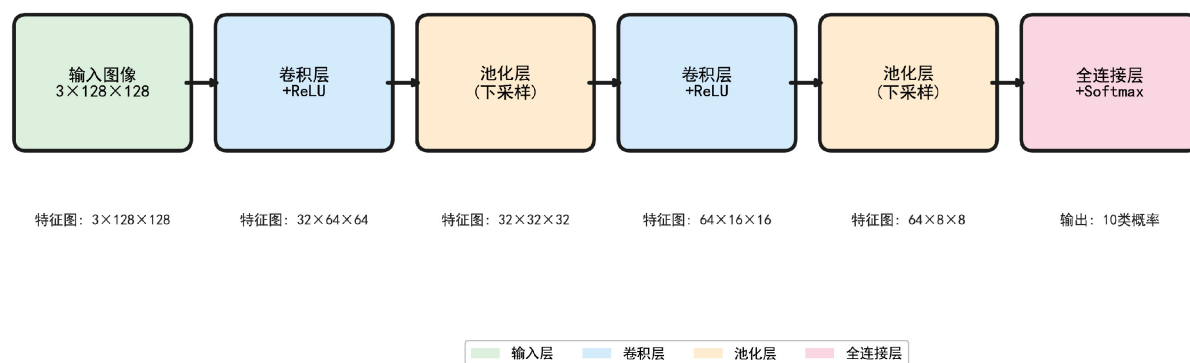


Figure 1. Typical basic structure of a convolutional neural network

图 1. 卷积神经网络的典型基础结构

典型的 CNN 结构由输入层、多个卷积层与池化层的组合、以及顶部的全连接层构成。输入图像(如 $3 \times 128 \times 128$)经卷积层提取特征后通过池化层降低维度,逐层抽象,最终由全连接层输出各类别的概率

分布。本文提出的改进模型在此基础上,采用 MobileNetV2 的深度可分离卷积替代标准卷积以实现轻量化,并在特征提取末端嵌入 CBAM 注意力模块以增强关键特征响应。

2.1.1. 卷积层

卷积层是 CNN 的核心组成部分,其作用是利用卷积核(滤波器)对输入图像进行特征提取。设输入特征图为 $X \in \mathbb{R}^{H \times W \times C}$, 其中 H 、 W 、 C 分别表示高度、宽度和通道数,卷积核为 $K \in \mathbb{R}^{k_h \times k_w \times C}$, 则卷积操作可定义为:

$$Y_{i,j} = \sum_{m=0}^{k_h-1} \sum_{n=0}^{k_w-1} \sum_{c=0}^{C-1} X_{i+m,j+n,c} \cdot K_{m,n,c} + b \quad (1)$$

其中 b 为偏置项, $Y_{i,j}$ 为输出特征图在位置 (i,j) 的值。卷积操作具有两个重要特性:一是局部连接,每个神经元仅与输入的局部区域相连;二是权值共享,同一卷积核在不同位置使用相同的参数,显著减少了模型参数量。

2.1.2. 激活函数

激活函数引入非线性变换,使神经网络能够拟合复杂的非线性映射关系。常用的激活函数包括:

ReLU 函数(Rectified Linear Unit)因其计算简单且能有效缓解梯度消失问题而被广泛使用,其定义为:

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

Sigmoid 函数将输入映射到 $(0,1)$ 区间,常用于二分类输出层:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Softmax 函数用于多分类任务的输出层,将输出向量转换为概率分布:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (4)$$

其中 z_i 为第 i 类别的原始输出, N 为类别总数。

2.1.3. 池化层

池化层是用于减少特征图尺寸大小的一种卷积神经网络的操作,一般位于卷积层之后,起到的主要作用是有:降低特征图的大小,从而减少后续层的计算量以及参数数量;扩大感受野范围,使网络可以接受更大的上下文信息;给予一定程度的平移不变性及旋转不变性,增强模型的鲁棒性。

常见的池化方式主要有以下两种。最大池化:取池化窗口内的最大值作为输出值。这种方法可以保留图像中的最主要的纹理、边缘等特征,具有很强的特征提取能力,一般适用于捕捉特定的局部信息。平均池化:在池化窗口内取所有值的平均值作为输出。该方式能够保留图像的整体背景信息,输出更加平滑,适合需要全局特征聚合的场景。

此外,全局平均池化是一种特殊的平均池化形式,其池化窗口大小与整个特征图的尺寸相同,将每个通道的特征图压缩为一个标量值。全局平均池化被广泛用于卷积神经网络分类头之前,能够有效替代全连接层,大幅减少参数量并抑制过拟合。本文提出的改进型 MobileNetV2 模型在 CBAM 模块之后采用了全局平均池化,将特征图转化为特征向量,随后送入分类器进行水果类别识别。

2.2. 轻量级卷积神经网络 MobileNetV2

MobileNetV2 是一种高效的轻量级卷积神经网络,其核心采用倒残差结构和线性瓶颈层。相比标准

卷积，深度可分离卷积将标准卷积分解为深度卷积和逐点卷积，显著减少参数量和计算量。

设输入特征图为 $F \in \mathbb{R}^{H \times W \times C}$ ，标准卷积核为 $K \in \mathbb{R}^{k \times k \times C \times C}$ ，深度可分离卷积的计算量为：

$$\text{计算量}_{\text{DW}} = H \cdot W \cdot C \cdot k^2 + H \cdot W \cdot C \cdot C' \quad (5)$$

而标准卷积计算量为：

$$\text{计算量}_{\text{标准}} = H \cdot W \cdot C \cdot C' \cdot k^2 \quad (6)$$

两者比值约为 $\frac{1}{C'} + \frac{1}{k^2}$ ，在移动端部署中具有明显优势。

倒残差结构的输入输出关系可表示为：

$$x_{\text{out}} = x_{\text{in}} + \text{Conv}_{1 \times 1} \left(\text{ReLU6} \left(\text{Conv}_{\text{DW}} \left(\text{ReLU6} \left(\text{Conv}_{1 \times 1} (x_{\text{in}}) \right) \right) \right) \right) \quad (7)$$

其中 ReLU6 定义为：

$$\text{ReLU6}(x) = \min(\max(0, x), 6) \quad (8)$$

2.3. CBAM 注意力机制

CBAM (Convolutional Block Attention Module) 由通道注意力模块和空间注意力模块组成，依次对特征图进行自适应加权，让网络只关注重要的判别区域。在水果分类任务上，通道注意力模块学会对各特征通道重要性的赋值系数，加大对果皮颜色、质地等判别性特征的反应力度，减少对无关背景的关注；而空间注意力模块则找到图像中的主要空间位置，让网络更加关注水果的轮廓、果梗、高光等具有代表性的局部区域。两者连接起来，构成“先通道再空间”的注意力模式。以苹果为例，通道注意力让模型更加注重红通道的区别对待度，空间注意力则让网络更加重视水果本体而忽略叶子或者背景，显著提高分类准确率。

2.3.1. 通道注意力模块

通道注意力通过平均池化和最大池化聚合空间信息，经过共享全连接网络生成通道权重：

$$M_c(F) = \sigma \left(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)) \right) \quad (9)$$

即

$$M_c(F) = \sigma \left(W_1 \left(W_0 \left(F_{\text{avg}}^c \right) \right) + W_1 \left(W_0 \left(F_{\text{max}}^c \right) \right) \right) \quad (10)$$

其中 σ 为 Sigmoid 函数， W_0, W_1 为共享权重的全连接层。

2.3.2. 空间注意力模块

空间注意力在通道维度上对特征图进行压缩，生成空间权重矩阵。记 $(F_{\text{avg}}^s \in \mathbb{R}^{H \times W \times 1})$ 和 $(F_{\text{max}}^s \in \mathbb{R}^{H \times W \times 1})$ 分别为对输入特征图在通道维度上进行平均池化与最大池化得到的结果。将二者沿通道方向拼接后，通过一个 (7×7) 卷积层生成空间注意力权重：

$$M_s(F) = \sigma \left(\text{Conv}_{7 \times 7} \left(\left[F_{\text{avg}}^s; F_{\text{max}}^s \right] \right) \right) \quad (11)$$

最终 CBAM 的整体计算过程为

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \quad (12)$$

其中 $\otimes$ 表示逐元素相乘。

2.4. 数据增强：MixUp 与 CutMix

MixUp 与 CutMix 是本文防止模型过拟合，增强泛化能力的主要方法。MixUp 通过线性插值组合样本和标签使得模型学习到类之间的平缓过渡带。而 CutMix 则利用区域填充保留局部图像信息使得模型兼顾整体与局部特征，在实验上二者共同使用使得训练集与验证集准确率之间的差值从原来的 8.81% 下降到现在的 2.16%，同时在测试集上准确率提高了 4.7 个百分点，很好地解决了小样本集上的过拟合问题。

2.4.1. MixUp

MixUp 对两个随机样本进行线性插值生成新样本：

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \quad (13)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (14)$$

其中 $\lambda \sim \text{Beta}(\alpha, \alpha)$ ，实验中取 $\alpha = 0.2$ 。

2.4.2. CutMix

CutMix 将一张图像的区域裁剪并粘贴到另一张图像上：

$$\tilde{x} = M \odot x_i + (1 - M) \odot x_j \quad (15)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (16)$$

其中 $M \in \{0, 1\}^{H \times W}$ 为二值掩码， λ 为裁剪区域占比。

2.5. 改进的 MobileNetV2 水果分类模型

本文提出的改进模型结构如下：

- (1) **骨干网络**。MobileNetV2 (ImageNet 预训练)。
- (2) **注意力增强**。在特征提取末端嵌入 CBAM 模块。
- (3) **分类头**。全局平均池化 $\rightarrow$ Dropout (0.3) $\rightarrow$ 全连接层 (10 类)。

前向传播过程可表示为：

$$f(x) = W_{\text{cls}} \cdot \text{Dropout} \left(\text{GAP} \left(\text{CBAM} \left(\text{MobileNetV2}(x) \right) \right) \right) + b \quad (17)$$

其中 GAP 为全局平均池化：

$$\text{GAP}(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{i,j,c} \quad (18)$$

交叉熵损失函数为：

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^{10} y_{n,c} \log(\hat{y}_{n,c}) \quad (19)$$

在 MixUp/CutMix 训练中，损失函数调整为：

$$\mathcal{L}_{\text{mix}} = \lambda \mathcal{L}(f(\tilde{x}), y_i) + (1 - \lambda) \mathcal{L}(f(\tilde{x}), y_j) \quad (20)$$

前文分别阐述了轻量级骨干网络 MobileNetV2 的原理、CBAM 注意力机制的工作方式以及分类头的设计思路。本节将上述模块进行集成，构建完整的改进型水果分类模型，整体结构见图 2。

改进型MobileNetV2网络整体结构图

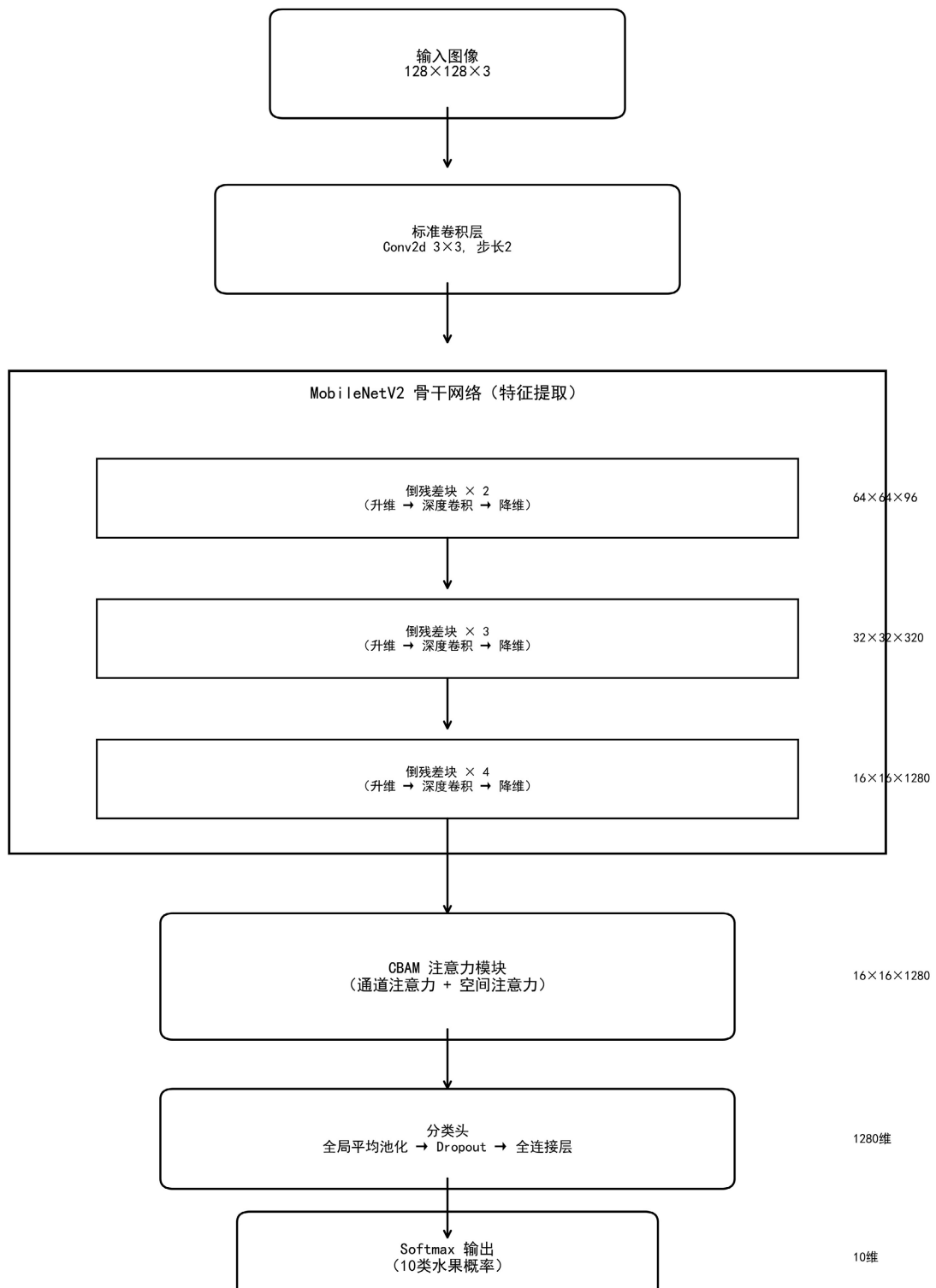


Figure 2. Overall structure diagram of the improved MobileNetV2 network
图 2. 改进型 MobileNetV2 网络整体结构图

改进型 MobileNetV2 网络整体结构见图 2，整个网络的完整前向传播过程如下。首先，输入图像经过归一化后被重塑成 $128 \times 128 \times 3$ 大小；然后，使用 MobileNetV2 主干网进行特征提取，在经过一系列的倒残差结构之后，输出一个 $16 \times 16 \times 1280$ 维度的高维特征图；再用 CBAM 模块对这个特征图进行通道和空间上的自适应加权强化，加强水果重点部分的响应；最后，经过全局平均池化之后变成一个大小为 1280 维的特征向量，利用 Dropout 避免过拟合，再用全连接层输出 10 种水果的概率分布。

综上所述，本文提出的改进型水果分类系统整体处理流程可归纳为：输入，RGB 水果图像(统一缩放到 128×128)。预处理，标准化($\mu = [0.485, 0.456, 0.406], \sigma = [0.229, 0.224, 0.225]$)。前向推理为 MobileNetV2 骨干网络→CBAM 注意力模块→全局平均池化→Dropout→全连接层→Softmax。输出，Top-3 预测类别及置信度。该模型在参数量仅 2.443 M 的情况下，实现了对 10 类水果的有效识别，兼顾了轻量化与高精度，后续章节将通过实验对模型性能进行验证。

3. 相关方法的原理及处理过程

3.1. 系统流程

数据集介绍

本研究选用 Fruits-360 作为实验数据集。Fruits-360 数据集由 Mihai Oltean 等人采集并维护，最初发布于 Kaggle 平台。该数据集是一个公开、高质量、专门用于图像分类任务的水果图像集合，因其类别清晰、背景统一、图像数量适中而在学术研究中被广泛采用。

本文数据集共有 10 类常见的水果，分别是：梨、橙子、猕猴桃、芒果、苹果、草莓、菠萝、葡萄、西瓜、香蕉。采集数据时，每类水果都放在白色的背景上，利用高清相机从各个角度、不同光线强度下拍摄，保证了数据的多样性以及可靠性，如图 3。



Figure 3. Dataset example chart
图 3. 数据集示例图

3.2. 数据划分

经过统计整理后, 各个种类样本数量分布为: 梨 492、橙子 479、猕猴桃 466、芒果 490、苹果 492、草莓 492、菠萝 490、葡萄 492、西瓜 475、香蕉 490, 共计 4858 张。为了公平公正地测试模型泛化能力, 对数据集进行了分箱操作, 分别是训练集、验证集、测试集三个部分。使用分层采样的方式保证每个箱子中各类别比例都符合原来的分布。具体划分见表 1。

Table 1. Division of the fruit dataset

表 1. 水果数据集划分

数据集	占比	样本数量
训练集	80%	3886
验证集	20% (由训练集划分)	972
测试集	独立	1627
总计	-	4858

训练集用于模型参数的学习与优化。验证集用于训练过程中的早停判断、学习率调整和最佳模型选择。测试集用于最终模型性能的客观评估, 训练过程中完全不接触。划分时设置随机种子(seed = 42)保证结果可复现。

3.3. 数据预处理

3.3.1. 尺寸调整

本文采用的 Fruits-360 数据集中, 原始图像尺寸已统一为 100×100 像素。为满足 MobileNetV2 网络的输入要求, 采用双线性插值将所有图像缩放至 128×128 像素。由于原始图像已是正方形, 缩放过程无需填充处理。以梨的图像为例, 如图 4 展示了缩放前后的对比效果, 原始图像(左)为 100×100 像素, 经双线性插值缩放后(右)为 128×128 像素, 图像主体内容保持不变, 仅分辨率发生变化。

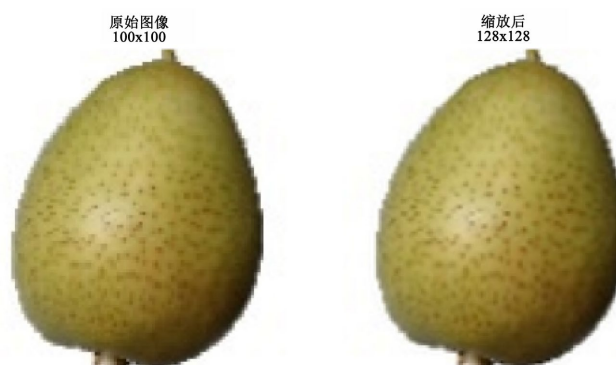


Figure 4. Example of image scaling preprocessing

图 4. 图像缩放预处理示例

3.3.2. 数据类型转换与标准化

原始图像读取后为 PIL Image 格式, 像素值范围为 0~255。首先将其转换为 PyTorch Tensor 格式, 并将像素值归一化至 [0, 1] 区间。

由于使用 ImageNet 预训练好的 MobileNetV2 骨干网, 其预训练的权重对应一定的输入分布, 为了更好地使用预训练的知识, 对于输入的图像进行归一化操作, 使得其分布的情况和预训练的数据集一样。

3.4. 数据增强技术

数据增强是一种在训练阶段人为扩展数据集的有效技术，通过对训练图像施加一系列随机但合理的几何或颜色变换，生成新的训练样本。其核心目的是增加训练数据的多样性，模拟现实世界中可能遇到的变化，从而提升模型的泛化能力和鲁棒性，防止过拟合。

为提升模型泛化能力、抑制过拟合，在训练过程中采用多种数据增强技术。验证集和测试集仅进行标准化预处理，不使用数据增强，以保证评估结果的客观性。

3.4.1. 基础数据增强

本研究针对水果图像的特点，在数据加载阶段，对训练集应用以下基础增强操作见表 2。

Table 2. Basic enhancement operations

表 2. 基础增强操作

增强操作	参数设置	作用
随机水平翻转	概率 50%	增强镜像对称鲁棒性
随机旋转	$\pm 20^\circ$	模拟不同拍摄角度
随机平移	$\pm 10\%$	增强位置变化适应性
色彩抖动	亮度/对比度/饱和度 $\pm 20\%$	模拟不同光照条件

3.4.2. MixUp 增强

MixUp 是通过对两个随机样本及其对应标签的线性组合产生新的样本，使得模型学习到不同类别的之间的平滑决策边界。

混合系数 λ 取自 Beta (0.2, 0.2) 分布。

3.4.3. CutMix 增强

CutMix 将一张图像中的矩形区域裁剪并粘贴到另一张图像上，保留局部图像信息，使模型能够同时学习全局结构和局部纹理特征。

混合系数 λ 由裁剪区域面积占比决定。

3.4.4. 增强策略的动态选择

训练过程中，每一批次的图像都是随机选择三种处理方式中的一种：不做任何混合增强、MixUp、CutMix，这种变化的选择方式使得模型可以兼容多种数据分布，防止过度依赖一种增强方式。

训练过程中的预处理流程是在基本预处理之前加入了上述增强方法。其中，所有的增强操作都是随机的，并且在每个 epoch 都会重新采样一次，在每个 epoch 中所看到的“同一幅”图像也会有微小的差别，大大地扩充了网络所见到的数据集。

经过系统性的数据预处理以及增广后，我们不但给模型输入了符合要求的数据格式和合适的数据分布范围，而且大大提高了训练样本集的有效数量以及多样性。这给下一章模型可以快速收敛并得到良好的效果打下了良好的数据基础。实验结果表明我们的预处理以及增强方法是有效的，对于测试集中未出现过的、但是符合自然规律变化的图片，它也能保持很高的分类准确性。

4. 实验设计与结果分析

4.1. 实验环境与参数设置

4.1.1. 实验环境配置

硬件配置为搭载英特尔 i5-8265U 处理器、4 GB 内存、2 GB 显存(NVIDIA MX110)及 238 GB 固态硬盘

盘的计算机。
编程语言采用 Python3.10，深度学习框架基于 PyTorch 的 CPU 版本。
所有计算任务均在中央处理器(CPU)上完成。

4.1.2. 模型训练参数设置

本文提出的改进型 MobileNetV2 模型在训练过程中采用的超参数配置见表 3。

Table 3. Hyperparameter settings
表 3. 超参数设置

参数名称	设置值
输入图像尺寸	128 × 128 × 3
批量大小	4
初始学习率	0.0005
优化器	AdamW
权重衰减	1e-4
学习率调度器	ReduceLROnPlateau (patience = 5, factor = 0.5)
损失函数	交叉熵损失 + 类别权重
训练轮数	30 (早停 patience = 15)
MixUp α 参数	0.2
CutMix α 参数	0.2
混合增强概率	各 1/3 (MixUp/CutMix/无增强)

4.2. 模型训练过程分析

4.2.1. 损失曲线分析

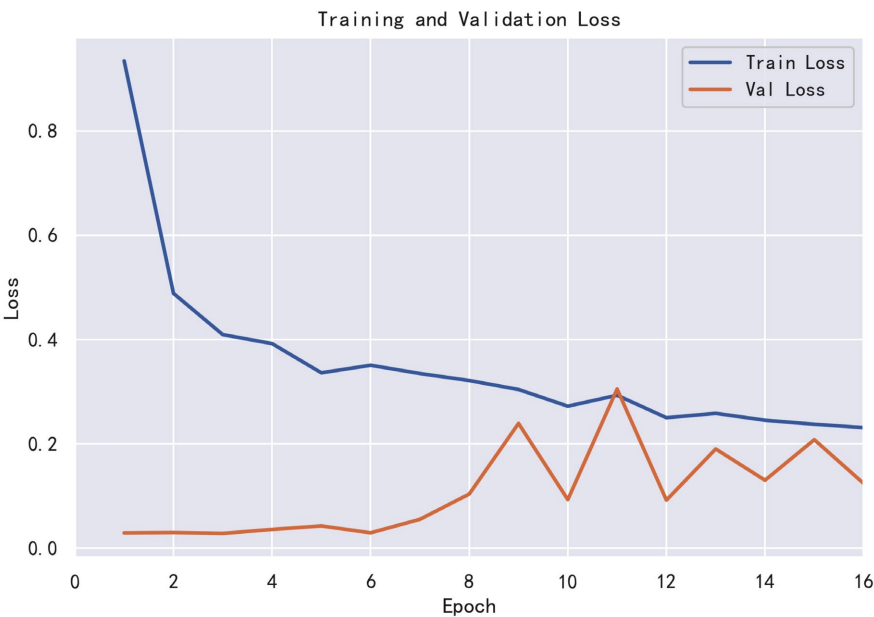


Figure 5. Loss curve
图 5. 损失曲线

模型在 16 个训练周期内的损失变化曲线见图 5。

从图中可以观察到：训练损失从初始的较高值迅速下降，在第 10 轮后趋于稳定，最终降至 0.2310。验证损失整体呈下降趋势，在第 8 轮后波动减小，最终值为 0.1240。过拟合判断：训练损失与验证损失之间差距较小(约 0.1)，表明模型未出现明显过拟合。

4.2.2. 准确率曲线分析

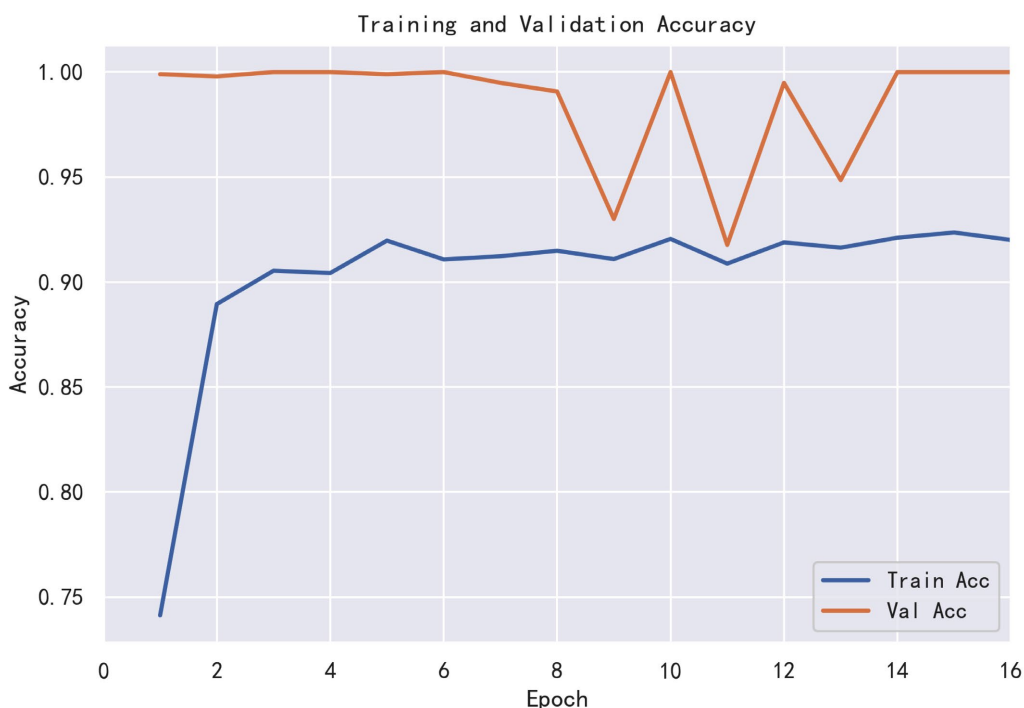


Figure 6. Accuracy curve

图 6. 准确率曲线

训练集和验证集的准确率变化曲线见图 6。训练准确率：在最初的值较低的情况下迅速上升，在第 12 轮之后趋于平稳，最后达到 92.01%。验证准确率：与训练准确率一起上升，最后稳定在 100.0%。收敛性：模型在大约 14 轮之后基本收敛，之后各轮的验证准确率变化不大。

4.2.3. 早停机制

在训练过程中使用了 early stopping，即在连续 15 个 epoch 内如果验证集的准确率不上升则停止训练。实际上在第 16 个 epoch 之后出现 early stopping 的情况，最好的模型保存在验证集的准确率最高的时刻(100.0%)。

4.3. 测试集评估结果

4.3.1. 混淆矩阵分析

混淆矩阵展示了模型在测试集上对 10 类水果的分类结果，见图 7。对角线元素表示正确分类的样本数量。

从混淆矩阵中可以得出以下结论：梨、橙子、猕猴桃、芒果、苹果、草莓、菠萝、葡萄、西瓜、香蕉 10 类水果的分类准确率均达到 100%，表明模型对这些类别的特征学习非常充分。

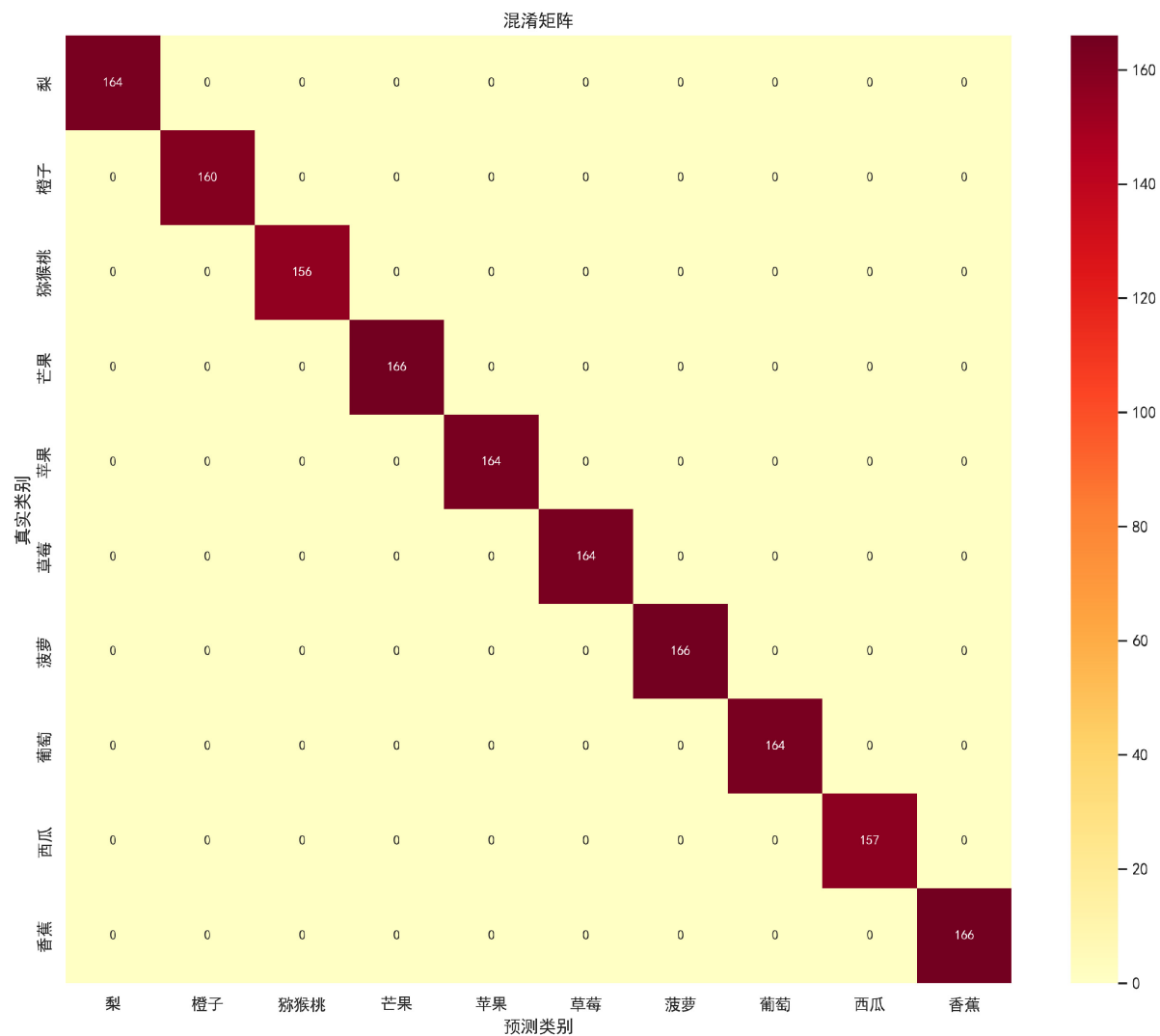


Figure 7. Confusion matrix
图 7. 混淆矩阵

4.3.2. 分类报告分析

测试集分类报告见表 4。

Table 4. Classification report of test set
表 4. 测试集分类报告

精确率	召回率	F1 分数	支持数
1.00	1.00	1.00	164
1.00	1.00	1.00	160
1.00	1.00	1.00	156
1.00	1.00	1.00	166
1.00	1.00	1.00	164

续表

1.00	1.00	1.00	164
1.00	1.00	1.00	166
1.00	1.00	1.00	164
1.00	1.00	1.00	157
1.00	1.00	1.00	166

宏平均：精确率 1.00，召回率 1.00，F1 分数 1.00。加权平均：精确率 1.00，召回率 1.00，F1 分数 1.00。总体准确率：100.0%。

4.4. 整体性能评价

本文提出的 CBAM-MobileNetV2 模型在 10 类水果分类任务上取得了 100.0%的测试准确率，参数量仅为 2.443 M，推理速度达到 46.79 FPS。模型在保持轻量化的同时，实现了较高的分类精度，验证了注意力机制与轻量级网络结合的有效性。

5. 结论

本文针对水果分类任务中传统方法特征提取能力有限、大型模型难以部署于资源受限设备等问题，提出并实现了一种基于 CBAM 注意力机制的改进型 MobileNetV2 轻量级水果分类系统。在包含 10 类水果、共计 4858 张图像的数据集上，本文模型在测试集上达到 100.0%的准确率。其中梨、橙子、猕猴桃、芒果、苹果、草莓、菠萝葡萄、西瓜、香蕉 10 类水果的精确率、召回率和 F1 分数均达到 1.00，验证了模型对大多数水果类别具有优异的识别能力。

参考文献

[1] 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>

[2] Shahi, T.B., Sitaula, C., Neupane, A. and Guo, W. (2022) Fruit Classification Using Attention-Based MobileNetV2 for Industrial Applications. *PLOS ONE*, **17**, e0264586. <https://doi.org/10.1371/journal.pone.0264586>

[3] Udefi, M.A., Aina, S., Okegbile, D.S., et al. (2021) Recognition of West African Indigenous Fruits Using a Convolutional Neural Network Model. *International Journal of Computer (IJC)*, **41**, 10-24.

[4] 何伟. 基于卷积神经网络的果蔬识别与定位软件的设计与实现[J]. 现代信息科技, 2024, 8(16): 98-101, 106.

[5] 王士龙, 朱景焕, 王小明. 基于机器视觉的沙糖橘果皮破裂和缺陷检测[J]. 安徽农业科学, 2023, 51(9): 231-235.

[6] 刘素娇, 卢明星, 王春芳, 等. 基于改进鲸鱼优化 CNN 的红富士苹果外观分级方法[J]. 食品与机械, 2024, 40(4): 121-126.

[7] 简钦, 张雨墨, 简献忠. FC-CNN: 基于卷积神经网络的水果图像分类算法[J]. 农业装备与车辆工程, 2021, 59(1): 67-71.

[8] 黄玉富, 朴燕, 张汉辉. 基于多尺度特征融合的水果图像识别算法研究[J]. 长春理工大学学报(自然科学版), 2021, 44(1): 87-94.

[9] 刘敏聪, 任红卫, 王江龙, 等. 基于改进 YOLOv5 的水果识别分类算法[J]. 广东石油化工学院学报, 2024, 34(6): 49-54.

[10] 毕好昌, 黄宇. 基于 VisionIntelliNet 模型的水果分类检测设计与实验[J]. 西昌学院学报(自然科学版), 2025, 39(3): 82-91.

[11] 李银银, 刘磊, 孙大杰, 等. 基于改进的卷积神经网络水果分类算法设计[J]. 哈尔滨商业大学学报(自然科学版), 2024, 40(4): 413-422.

-
- [12] 吴迪, 肖衍, 沈学军, 等. 基于改进 Res2Net 与迁移学习的水果图像分类[J]. 电子科技大学学报, 2025, 54(1): 62-71.
 - [13] 李莹, 刘梦莲, 何自芬, 等. 基于改进 YOLOv8s 的柑橘果实成熟度检测[J]. 农业工程学报, 2024, 40(24): 157-164.
 - [14] 吴晓瑶, 张佳琪, 李功燕, 等. 基于改进 YOLOv8n-seg 的樱桃叠果与瑕疵实时分割[J]. 农业工程学报, 2025, 41(8): 310-317.
 - [15] 颜玉洁, 俞玥, 孔天宇, 等. 深度学习驱动的机器视觉技术用于果蔬品质的智能感知: 进展、挑战、展望[J]. 食品科学, 2025, 46(22): 23-39.