

DOI: 10.13652/j.spjx.1003.5788.2025.80635

基于改进 ResNet18 的水果融合图像 表面缺陷检测模型

司海平^{1,2} 赵文沛^{1,2} 赵雨洋^{1,2} 李婷婷¹ Fernando Bacao^{2,3} 李艳玲^{1,2}

(1. 河南农业大学信息与管理科学学院, 河南 郑州 450046; 2. 河南省农业大数据与人工智能国际联合实验室, 河南 郑州 450046; 3. 新里斯本大学信息管理学院, 葡萄牙, 里斯本 1070-312)

摘要: [目的] 提升水果在运输过程中因轻微机械摩擦产生的表面缺陷检测准确性。[方法] 提出一种基于改进 ResNet18 的水果融合图像表面缺陷检测模型。为解决多尺度特征提取难题, 设计了 MELA 模块, 以更全面地捕捉图像细节。同时, 引入基于能量函数的正则化损失函数, 与交叉熵损失函数协同优化训练流程。此外, 采用迁移学习策略以减轻过拟合现象。[结果] 改进后的模型在水果表面缺陷检测任务中表现卓越, 准确率、精确度、召回率及 F_1 分数分别高达 99.53%, 99.60%, 99.58%, 99.58%。[结论] 基于改进 ResNet18 的水果融合图像表面缺陷检测模型显著提升了检测准确性, 适用于水果融合图像的表面缺陷检测。

关键词: 深度学习; 缺陷检测; 注意力机制; ResNet18; 融合图像

A surface defect detection model for fused images of fruits based on improved ResNet18

Si Haiping^{1,2} Zhao Wenrui^{1,2} Zhao Yuyang^{1,2} Li Tingting¹ Fernando Bacao^{2,3} Li Yanling^{1,2}

(1. College of Information and Management Science, Henan Agricultural University, Zhengzhou, Henan 450046, China; 2. Henan International Joint Laboratory of Agricultural Big Data and Artificial Intelligence, Zhengzhou, Henan 450046, China; 3. NOVA Information Management School, Universidade Nova de Lisboa, Lisboa 1070-312, Portugal)

Abstract: [Objective] To enhance the accuracy of detecting fruit surface defects caused by minor mechanical friction during transportation. [Methods] A surface defect detection model for fused images of fruits based on improved ResNet18 is proposed. To address the challenge of multi-scale feature extraction, the model incorporates a MELA module designed to capture image details more comprehensively. Additionally, an energy function-based regularization loss function is introduced to collaboratively optimize the training process alongside the cross-entropy loss function. Furthermore, a transfer learning strategy is employed to mitigate overfitting. [Results] Experimental results demonstrated that the improved model exhibits exceptional performance in fruit surface defect detection tasks, achieving the accuracy, precision, recall, and F_1 score as high as 99.53%, 99.60%, 99.58%, and 99.58%, respectively. [Conclusion] The surface defect detection model for fused images of fruits based on improved ResNet18 significantly enhances the detection accuracy and is suitable for surface defect detection in fused images of fruits.

Keywords: deep learning; defect detection; attention mechanism; ResNet18; fused image

基金项目: 河南省重点研发专项 (编号: 251111211300); 河南省科技攻关项目 (编号: 252102520037); 河南省杰出外籍科学家工作室 (编号: GZS2024006)

通信作者: 李艳玲 (1976—), 女, 河南农业大学副教授, 博士。E-mail: ly_lingling@163.com

收稿日期: 2025-08-06 改回日期: 2026-02-11

引用格式: 司海平, 赵文沛, 赵雨洋, 等. 基于改进 ResNet18 的水果融合图像表面缺陷检测模型[J]. 食品与机械, 2026, 42(7): 90-97.

Citation: Si Haiping, Zhao Wenrui, Zhao Yuyang, et al. A surface defect detection model for fused images of fruits based on improved ResNet18[J]. Food & Machinery, 2026, 42(7): 90-97.

水果中含有丰富的维生素和矿物质,具有丰富的营养价值,可以有效补充人体所需营养物质,促进新陈代谢和能量代谢,有益于身体健康。随着人们生活水平的不断提高,水果已成为中国继粮食、蔬菜之后的第三大农业种植产业,果园总面积和水果总产量常年稳居世界首位^[1]。

目前,红外与可见光图像融合技术已被应用于多种领域,如视觉跟踪^[2]、夜间导航^[3]、事故鉴定取证^[4]。但基于水果融合图像的表面缺陷检测领域的应用仍处于探索阶段,通过开展基于红外与可见光图像融合的水果表面缺陷检测研究,不仅有助于推动农业生产的智能化和自动化,还能为果农和供应链管理者提供科学决策支持,减少缺陷水果流入市场带来的经济损失。

传统的水果表面缺陷检测主要依赖于特征提取与分类识别两个关键步骤。在特征提取阶段,多采用人工设计的特征描述方法,从水果图像中提取具有判别能力的视觉特征,包括形态特征、颜色特征以及纹理特征等^[5]。随后,将提取的单一特征或多维组合特征输入到机器学习模型中,从而实现水果表面缺陷检测。但该方法很大程度上依赖于先验知识和人工经验,所提特征的表达能力有限^[6]。深度学习方法通过自动学习图像中的多层次特征不仅有效简化了特征提取过程,还增强了检测的泛化能力和鲁棒性^[7-8]。Cai 等^[9]提出了一种 FastSegFormer 网络,实现了水果表面缺陷检测。该网络采用多尺度金字塔(MSP)模块架构和多分辨率知识提炼策略,为大规模脐橙缺陷检测提供了高效、准确的解决方案。侯文慧等^[10]提出了一种基于弱监督学习的 MCF-UNet 模型,结合全局与局部统计生成伪标签,通过引入 CBAM 注意力机制与特征金字塔模块提升网络边缘感知与特征表达能力,最终实现砀山梨的表面缺陷检测,并展现出了良好的鲁棒性。王云鹏等^[11]基于红外与可见光图像融合算法增强缺陷特征表达,结合卷积神经网络 AlexNet 进行识别,取得了 99.0% 的平均识别准确率。然而,现有方法仍存在一些局限性,比如深度学习的方法对多尺度缺陷特征捕捉不足、鲁棒性较差等。

试验拟设计一种轻量化的多尺度特征增强模块(MELA),有效融合局部细节与全局上下文信息;引入基于能量函数的正则化损失函数,提升模型对噪声和模糊等干扰的鲁棒性;结合迁移学习与图像融合策略,能够在有限数据下实现高精度、高效率的水果表面缺陷检测。

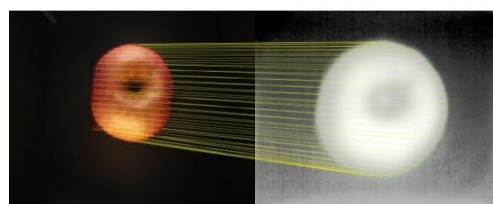
1 材料与方法

1.1 构建水果融合图像数据集

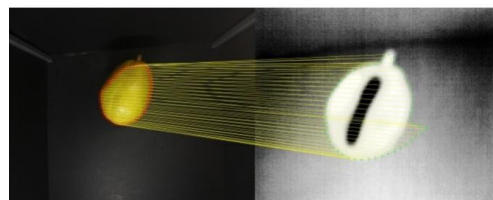
1.1.1 水果图像数据采集 以苹果和香梨作为试验对象,鉴于当前尚未有公开可用的苹果与香梨的红外和可见光图像数据集,自行构建与之匹配的水果图像数据集。

为生成带有表面缺陷的样本,模拟实际生产环境中尖锐物品对水果造成磨损的情况。借助大疆御 2 行业进阶版无人机搭载的双光相机进行图像采集工作,该相机能够同步获取可见光图像与红外图像,其中可见光图像分辨率为 8 000 像素×6 000 像素,红外图像分辨率为 640 像素×512 像素。采集过程中,水果在传送带上保持固定位置,当其进入图像采集区域时,利用上方设置好的双光相机进行图像采集。

1.1.2 水果图像数据配准 由于双光相机采集的红外图像与可见光图像在分辨率上存在差异,空间位置未对齐,无法直接进行图像融合。因此需要进行图像配准为后续融合提供数据支持。采取基于主动轮廓模型的图像配准算法^[12]来实现图像配准,如图 1 所示。



(a) 苹果图像配准



(b) 香梨图像配准

图 1 图像配准过程示例

Figure 1 Example of image alignment process

尽管大部分图像在配准后能够较好对齐,但仍有少量图像存在配准误差。针对存在偏差的图像,进一步采用手工调整的方式,以实现图像配准的半自动化处理。采用人工微调策略,通过手动进行偏移调整,使红外与可见光图像中的水果轮廓能够最大限度重合。经过上述步骤,共获得 847 对配准后的水果图像,作为后续图像融合试验的基础数据。图 2 为图像配准前后的对比效果。

1.1.3 水果融合图像数据集 在红外图像中,水果中的果梗与花萼容易混淆;在可见光图像中,水果表面缺陷与水果纹理不易区分。采用红外与可见光图像融合算法^[13]生成水果融合图像数据集。水果融合图像结合了红外图像与可见光图像的优势,能更全面地反映水果的表面特征。水果融合图像示例如图 3 所示。

基于各模态的特点,将水果融合数据集划分为果梗、果梗+缺陷、花萼、花萼+缺陷、缺陷、完好果 6 类(见图 4)。因采集数量有限,对图像进行随机旋转、翻转、亮

度调整、对比度调整等数据增强策略,将数据集扩增至原始数据的 5 倍。最终,水果融合图像数据集的类别与数量见表 1。

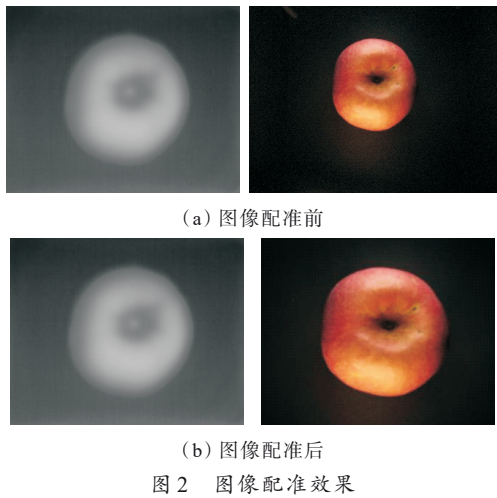


Figure 2 Image alignment effect

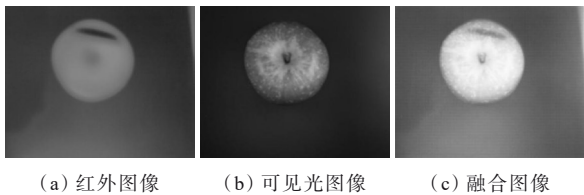


Figure 3 Examples of fused image categories

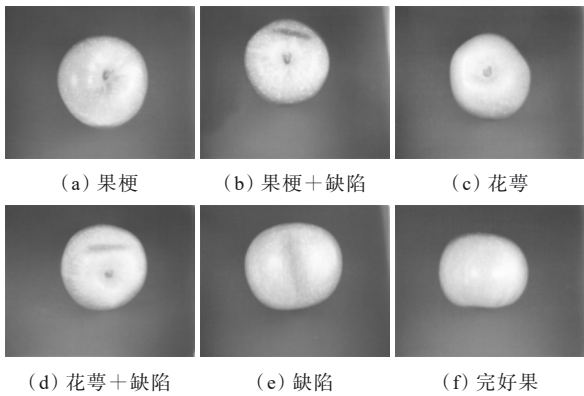


Figure 4 Examples of fused image categories

表 1 图像扩增后各类别数量			
Table 1 Number of categories after image amplification			
类别	数量	类别	数量
果梗	850	缺陷	735
果梗+缺陷	520	完好果	820
花萼	810	总计	4 235
花萼+缺陷	500		

1.2 水果表面缺陷检测模型

1.2.1 网络整体架构 基于 ResNet18^[14]进行改进来实现水果表面缺陷检测。在第一层残差块 (BasicBlock1) 之后引入了一个新模块 MELA, 该模块通过结合多尺度特征提取与注意力机制来更有效地捕捉局部细节和全局上下文信息, 优化特征交互, 使网络能够更充分地利用输入信息。改进后的网络架构如图 5 所示。

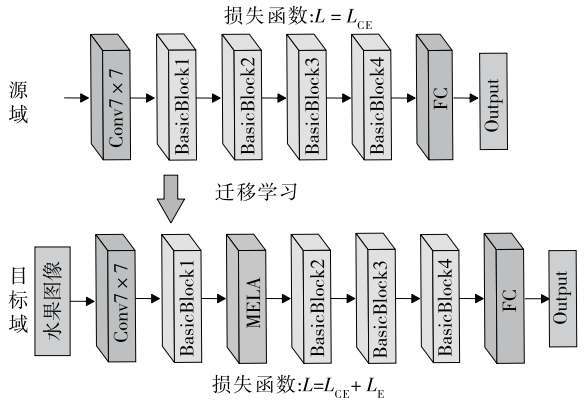


Figure 5 Overall network architecture

为了进一步提升模型性能,在训练过程中引入迁移学习的微调策略来快速适应水果表面缺陷检测任务。该策略在数据量有限的情况下,可以有效降低过拟合风险。此外,基于能量函数设计了能量正则化损失函数。该损失函数与传统的交叉熵函数相结合,能够更合理地引导输出特征的分布和目标优化过程,从而提升模型在目标任务上的鲁棒性与性能表现。

1.2.2 MELA 模块 由于获取的图像是融合图像,为了能够更好地提取多尺度特征,提出一种多尺度特征增强模块 MELA 来更好地保留图像的特征,具体架构如图 6 所示。

该模块采用残差结构,通过多尺度特征提取、注意力增强和自适应加权机制,实现了高效的特征提取。模块的核心设计包含两条并行路径:① 对输入特征进行自适应权重调整;② 通过多尺度处理机制提取多层次特征信息,并利用 ELA 注意力机制^[15]进一步增强特征的判别能力。最终,两条路径的特征通过加权融合与残差相加相结合的方式,实现信息的高效整合,提升模型的表达能力。

为了确保该模块的轻量化特性,在多尺度特征提取时采取深度可分离卷积^[16]实现卷积操作,避免引入过多的计算负担。输入特征首先分别通过 7×7 和 3×3 的深度可分离卷积进行处理。其中,7×7 卷积用于捕获全局范围内的结构信息,其输出特征进一步通过 3×3 卷积以增强特征的表达能力,并通过平均池化实现下采样,有效

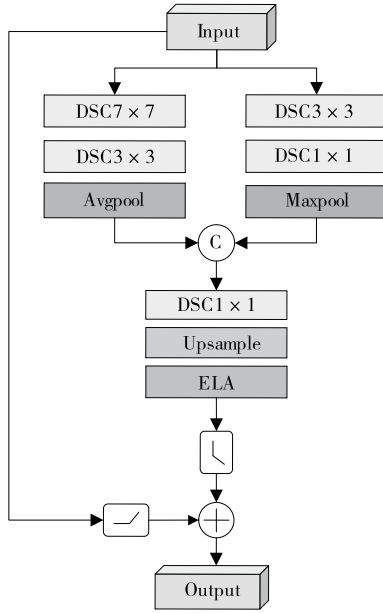


图6 MELA 模块

Figure 6 MELA module

保留全局信息。而 3×3 卷积侧重于提取局部细节特征,其输出特征首先通过 1×1 卷积调整通道维度,再通过最大池化完成下采样,从而在保留空间显著信息的同时降低计算复杂度。随后,两条路径生成的特征图在通道维度上进行拼接,并通过 1×1 卷积实现多尺度特征融合,以确保通道数的一致性。由于池化操作会导致特征图尺寸减半,为了保持空间维度的对齐,该模块利用转置卷积进行上采样,从而恢复至原始分辨率。这在确保多尺度特征在空间维度上对齐的同时,还有效保留了全局和局部特征的完整性。

为进一步增强特征的判别能力,多尺度分支的输出被送入 ELA 注意力机制。通过 ELA 注意力机制,能够聚焦特征图中信息量更高的局部区域,从而提升特征的语义表达能力。此外,在双分支特征融合中,模块引入了一种门控机制。具体来说,多尺度分支的输出 m 和原始输入 x 分别通过一个 1×1 卷积层生成自适应权重矩阵来动态实现特征融合,具体计算式为:

$$\text{out} = (\text{gate}(m)) \cdot m + (\text{gate}(x)) \cdot x, \quad (1)$$

式中:

$\sigma(\cdot)$ ——Sigmoid 函数;

$\text{gate}(\cdot)$ ——由 1×1 卷积层实现。

这种特征融合方式能够对原始输入和多尺度特征的贡献进行自适应调控,实现信息流的动态重塑。在保留关键特征的同时有效抑制冗余信息,从而优化特征分布。

1.2.3 损失函数 在深度学习模型训练过程中,损失函数的设计对模型的性能具有重要的影响。在传统的交叉

熵损失函数的基础上,提出一种基于能量函数的正则化损失函数,用于约束模型输出特征的分布,并将其应用于 ResNet18 模型来提升模型的鲁棒性。

损失函数具体由两部分组成:交叉熵损失函数与能量正则化损失函数。

$$L = L_{\text{CE}} + \lambda \cdot L_{\text{E}}, \quad (2)$$

式中:

L_{CE} ——交叉熵损失函数(用于衡量模型预测与真实标签之间的差异);

L_{E} ——能量正则化项(用于约束模型输出特征的分布);

λ ——权重系数(用来调节 L_{E})。

L_{CE} 是最常用的损失函数之一,定义为:

$$L_{\text{CE}} = -\frac{1}{N} \sum_{x=1}^N \sum_{c=1}^C y_{x,c} \cdot \log(p_{x,c}), \quad (3)$$

式中:

N ——样本数量;

C ——类别数量;

$y_{x,c}$ ——样本 x 的真实标签;

$\log(p_{x,c})$ ——样本 x 属于类别 C 的预测概率。

能量函数将每个样本映射为能量,该能量越高,表明模型对样本的预测越不确定;能量越低,表明模型对样本的预测越自信。试验基于分类器构建基于能量的模型^[17],将能量函数定义为对数之和的负求和。基于该能量函数,提出一种能量正则化损失函数,其目标是通过约束模型输出特征的分布,避免模型对某些样本的预测过于自信或者不确定。通过能量正则化损失函数可以有效降低模型对某些样本的极端预测风险,缓解过拟合问题,提升模型的鲁棒性。具体计算过程为:

$$L_{\text{E}} = \frac{1}{N} \sum_{x=1}^N \left| E_x - \frac{1}{N} \sum_{x=1}^N E_x \right|, \quad (4)$$

$$E_x = -\log \left(\sum_{c=1}^C \exp(z_{x,c}) \right), \quad (5)$$

式中:

E_x ——样本 x 的能量得分;

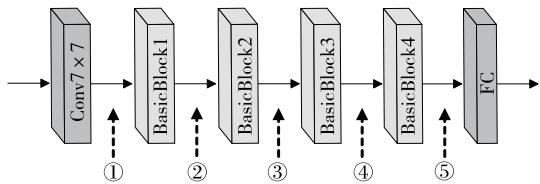
$z_{x,c}$ ——样本 x 在各类别上的置信度得分。

2 结果与分析

网络模型的训练与测试工作均在配备 NVIDIA GeForce RTX 2080Ti 的设备上进行,数据集按照 8:1:1 划分为训练集、验证集和测试集。

2.1 MELA 模块嵌入分析

以 ResNet18 网络作为基线模型,提出一种 MELA 模块。为了检验 MELA 模块对于提升 ResNet18 网络性能的有效性,分析 MELA 模块在不同嵌入位置对模型性能的影响,结果如图 7 和表 2 所示。



虚线箭头表示模块在 ResNet18 网络中的具体空间嵌入方式

图 7 MELA 模块不同嵌入位置

Figure 7 Different embedding positions of MELA module

由表 2 可知, MELA 模块的不同嵌入方式对模型效果产生了不同影响。将 MELA 模块嵌入 BasicBlock1 与 BasicBlock2 之间时, 模型效果最佳。相较于原模型, 改进后的模型准确率提升了 0.94%, 精确率提升了 0.70%, 召回率提升了 0.83%, F_1 分数提升了 0.78%, 后续改进模型默认采用此种嵌入方式。此外, 从参数量对比情况来看, 该模块具备轻量化特性, 相较于原模型, 参数量仅增加了 0.04 M, 不会给模型带来较大的计算负担。

表 2 MELA 不同嵌入位置的评价指标结果

Table 2 Evaluation results for different embedding locations of MELA

嵌入方式	准确率/ %	精确度/ %	召回率/ %	F_1 分 数/%	参数量/ M
无	95.52	95.54	95.08	95.28	11.69
BasicBlock1 之前	95.05	94.84	94.67	94.74	11.73
BasicBlock2 之前	96.46	96.24	95.91	96.06	11.73
BasicBlock3 之前	95.05	95.26	94.91	95.05	11.73
BasicBlock4 之前	95.52	95.45	94.97	95.18	11.73
BasicBlock4 之后	95.75	95.66	94.97	95.27	11.73

2.2 模型训练策略

为进一步提升模型的泛化能力, 在模型训练阶段引入迁移学习策略。迁移学习不仅能够提高模型性能、加速模型收敛, 还能显著改善模型在小数据集上的表现。此外, 通过引入能量正则化损失函数对网络性能进行优化, 进而提升模型的准确率。

为验证上述改进策略的有效性, 一方面, 通过对比是否采用迁移学习的模型性能, 证明迁移学习的有效性; 另一方面, 依次解冻 MELA 模块前的模型层, 进一步筛选出最优模型, 结果见表 3。由表 3 可知, 采用迁移学习后模型的 4 个评估指标均有显著提升, 证明迁移学习能够有效提高模型对水果表面缺陷的检测效果。其中, 迁移一无冻结的模型表现最为突出, 准确率、精确度、召回率和 F_1 分数分别为 98.82%, 98.89%, 98.85%, 98.91%。因此, 模型默认采用无冻结的微调策略, 以确保获得更佳的性能表现。

表 3 迁移学习的消融试验

Table 3 Ablation experiments for transfer learning %

是否采用迁移	准确率	精确度	召回率	F_1 分数
无迁移	96.46	96.24	95.91	96.06
迁移一无冻结	98.82	98.89	98.95	98.91
迁移一冻结 BasicBlock1 前	98.35	98.45	98.15	98.29
迁移一冻结 MELA 前	98.82	98.85	98.94	98.89

将交叉熵损失函数与能量正则化损失函数相结合, 用于模型训练。由图 8 可知, 改进后的模型在各种权重设置下均展现出良好的收敛性能。

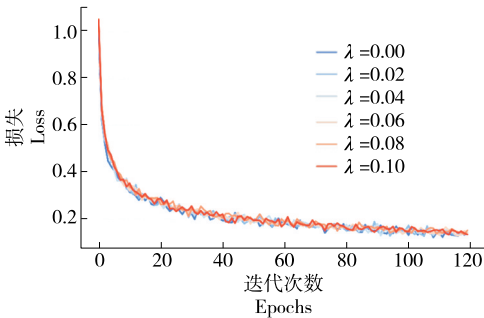


图 8 不同权重取值的损失函数图

Figure 8 Loss function plot for different values of weights

由表 4 可知, 相较于仅采用交叉熵损失函数的情况, 试验采用的总损失函数能够进一步优化模型性能, 当权重参数为 0.08 时, 模型性能达到最优状态。

表 4 不同权重的评价指标结果

Table 4 Results of evaluation indicators with different weights

λ	准确率/%	精确度/%	召回率/%	F_1 分数/%
0.00	98.82	98.89	98.95	98.91
0.02	98.82	98.89	98.79	98.83
0.04	99.06	99.08	99.02	99.04
0.06	99.29	99.38	99.24	99.30
0.08	99.53	99.60	99.58	99.58
0.10	99.06	99.07	98.92	98.99

2.3 模型鲁棒性分析

在改进模型中, 引入基于能量函数的正则化损失函数, 能够显著提升模型应对数据波动的能力。为进一步验证改进模型的鲁棒性, 在数据处理环节引入高斯噪声、散粒噪声、脉冲噪声以及均值模糊 4 种数据干扰策略, 并通过将 ResNet18 模型与改进模型进行对比来评估模型的鲁棒性。具体而言, 高斯噪声是一种符合正态分布的随机噪声; 散粒噪声遵循泊松分布; 脉冲噪声表现为图像

中随机出现的极端值干扰;均值模糊则是通过计算局部像素的平均值来模拟成像时的离焦模糊现象。由表 5 可知,在面对数据波动时,改进模型展现出更为优异的性能。特别是在均值模糊干扰下,改进模型的准确率相较于 ResNet18 模型提升了 3.07%,表明改进模型有效增强了模型对数据波动的应对能力,具备更强的鲁棒性。

表 5 模型鲁棒性测试结果								
Table 5 Model robustness test results								
干扰策略	ResNet18 模型				改进模型			
	准确率	精确率	召回率	F ₁ 分数	准确率	精确率	召回率	F ₁ 分数
高斯噪声	88.21	88.05	87.30	87.20	89.15	88.64	89.02	88.70
散粒噪声	85.38	84.31	83.67	83.87	86.56	85.81	85.54	85.55
脉冲噪声	83.96	83.21	83.02	82.88	85.85	85.33	84.24	84.47
均值模糊	77.83	77.92	76.99	77.07	80.90	81.22	79.74	80.19

2.4 分类模型对比

为了验证试验提出的模型在水果融合图像上的性能,将其与 Alex^[18]、VGG16^[19]、ResNet18、ResNet50、EfficientNetV2^[20]、MobileNetV4^[21]、ViT^[22]、RepViT^[23]、Vision Mamba^[24]进行了对比试验,各模型均采用相同的迁移学习策略,在同一条件下进行试验,结果见表 6。

表 6 不同模型在测试集上的结果					
Table 6 Comparison of experimental results of different models on the test set					
模型	准确率/%	精确度/%	召回率/%	F ₁ 分数/%	参数量/M
Alex	91.75	91.85	91.75	91.66	62.38
VGG16	98.35	98.36	98.27	98.30	138.36
ResNet18	98.35	98.12	98.50	98.29	11.69
ResNet50	98.58	98.69	98.58	98.62	25.60
EfficientNetV2	98.35	98.22	98.52	98.36	54.14
MobileNetV4	94.58	94.75	93.98	94.30	9.72
ViT	96.46	96.08	96.28	96.17	86.57
RepViT	98.82	99.01	98.46	98.72	14.13
Vision Mamba	99.29	99.29	99.16	99.22	25.80
试验模型	99.53	99.60	99.58	99.58	11.73

由表 6 可知,试验模型在多个关键性能指标上均达到了最优水平。具体而言,准确率、精确度、召回率以及 F₁ 分数分别高达 99.53%,99.60%,99.58%,99.58%,充分彰显了其卓越的性能表现。与基准模型 ResNet18 相比,试验模型仅增加了 0.04 M 的参数量,在保持轻量化优势的同时,显著提升了模型的准确率。

在与其他主流模型的对比中,ResNet50 在准确率方面虽略优于 ResNet18,但其参数量约为试验模型的 2 倍。AlexNet 和 VGG16 在特征提取方面具备一定的优势,但其参数量较大,导致计算成本较高。EfficientNetV2 在优

化计算效率方面取得了一定进展,但仍存在较高的参数需求。MobileNetV4 通过减少参数量来降低模型复杂度,但也降低了准确率。ViT 作为一种基于 Transformer 架构的深度学习模型,在该任务中的表现欠佳。RepViT 将 ViT 整合到标准轻量级 CNN 中,实现了一定程度的优化,但其准确率仍低于试验模型。Vision Mamba 作为 Mamba-Transformer 的混合网络,在准确率上接近试验模型,但参数量相对更大。

综上,试验模型在水果表面缺陷检测任务中表现出色,不仅在各项指标上取得了最优值,还保持了较低的参数量,为水果表面缺陷检测提供了一种高效且可行的解决方案。

2.5 基于改进模型的多模态性能评估

为了进一步验证基于改进模型的水果表面缺陷检测方法的有效性,在红外图像、可见光图像以及融合图像 3 种模态上进行对比试验,结果见表 7。由表 7 可知,基于改进模型的水果表面缺陷检测方法在各评价指标上均取得了最优值,且优于单一模态的结果。

此外,以混淆矩阵的形式对改进模型的多模态结果进行了可视化展示,如图 9 所示。在红外图像上,果梗类别与花萼类别之间存在较高的混淆率;在可见光图像上,缺陷类别与完好果类别容易混淆。相比之下,基于改进

表 7 多模态在测试集上的结果对比					
Table 7 Comparison of experimental results for multi-modal on the test set					
模型	模态	准确率	精确度	召回率	F ₁ 分数
ResNet18	红外图像	94.81	94.82	94.85	94.81
	可见光图像	93.16	93.63	92.98	93.16
	融合图像	95.52	95.54	95.08	95.28
改进 ResNet18	红外图像	96.70	96.48	96.16	96.27
	可见光图像	96.46	97.18	96.14	96.52
	融合图像	99.53	99.60	99.58	99.58

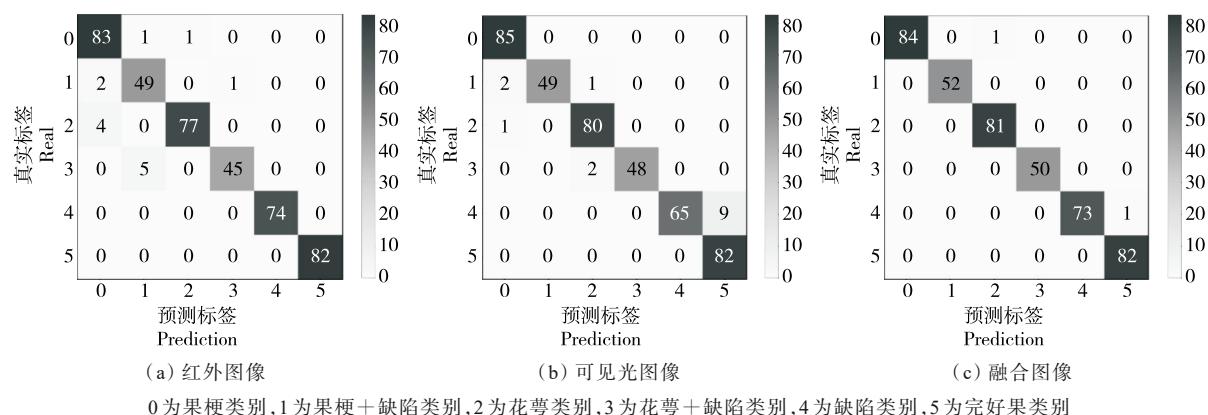


图9 多模态测试结果可视化效果

Figure 9 Visualization effect of multi-modal test results

ResNet18的水果融合图像表面缺陷检测模型能够有效降低类别间的混淆率,提升缺陷检测准确性。

3 结论

针对水果在运输过程中因轻微机械摩擦产生的表面缺陷检测问题,提出了一种基于改进 ResNet18 的水果融合图像表面缺陷检测模型。结果表明,通过引入多尺度特征增强模块(MELA)与基于能量函数的正则化损失函数,结合迁移学习策略,显著提升了水果表面缺陷检测的准确性与鲁棒性。改进后的模型在水果表面缺陷检测任务中表现卓越,准确率、精确度、召回率及 F_1 分数分别高达 99.53%, 99.60%, 99.58%, 99.58%, 不仅验证了 MELA 模块在多尺度特征提取方面的有效性,也证明了能量正则化损失函数在优化模型训练过程中的重要作用。通过 MELA 模块的引入,模型能够更全面地捕捉图像细节,提升特征表达能力;而能量正则化损失函数则有效约束了模型输出特征的分布,增强了模型的鲁棒性。在模型训练策略上,试验采用迁移学习策略,充分利用了预训练模型的知识,加速了模型收敛,降低了过拟合风险。特别是在数据集较小的情况下,迁移学习策略显著提升了模型的泛化能力,使模型在未见过的数据上也能保持优异的性能。

此外,通过对比试验验证了改进模型在不同数据干扰策略下的鲁棒性。在引入高斯噪声、散粒噪声、脉冲噪声及均值模糊等干扰后,改进模型仍能保持较高的准确率,特别是在均值模糊干扰下,改进模型的准确率相较于基准模型 ResNet18 提升了 3.07%,充分证明了其强大的抗干扰能力。在与其他主流模型的对比中,试验提出的模型在保持轻量化的同时,显著提升了检测准确率。与 ResNet50、AlexNet、VGG16 等模型相比,试验模型在参数量更小的情况下取得了更高的准确率;与 EfficientNetV2、MobileNetV4 等轻量级模型相比,试验模型则在准确率和

参数量之间取得了更好的平衡;与 ViT、RepViT 等基于 Transformer 架构的模型相比,试验模型在水果表面缺陷检测任务中表现更为优异;与 Vision Mamba 等混合网络模型相比,试验模型则在准确率和参数量上均展现出优势。

综上,试验提出的基于改进 ResNet18 的水果融合图像表面缺陷检测模型在检测准确性、鲁棒性及轻量化方面均表现出色,为水果表面缺陷检测提供了一种高效且可行的解决方案。未来工作将进一步优化模型结构,探索更多模态的图像融合策略,以提升模型在复杂场景下的检测能力。

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.
Wu Z Y, Li Y R, Dong Z D. Research on the current situation and countermeasures of fruit industry in China[J]. China Fruit & Vegetable, 2023, 43(11): 79-83, 87.
- [2] Meher B, Agrawal S, Panda R, et al. A survey on region based image fusion methods[J]. Information Fusion, 2019, 48: 119-132.
- [3] Zhang X C, Demiris Y. Visible and infrared image fusion using deep learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10 535-10 554.
- [4] Li S T, Kang X D, Fang L Y, et al. Pixel-level image fusion: a survey of the state of the art[J]. Information Fusion, 2017, 33: 100-112.
- [5] 褚璇, 洪嘉隆, 冯耿鑫, 等. 基于机器视觉的百香果品质多指标在线检测与分选[J]. 食品与机械, 2024, 40(6): 130-137, 142.
Chu X, Hong J L, Feng G X, et al. Online detection and sorting of passion fruit quality based on machine vision using multi-indicator[J]. Food & Machinery, 2024, 40(6): 130-137, 142.
- [6] Selvaraj A, Shebiah N, Nidhyanandhan S, et al. Fruit recognition using color and texture features[J]. Journal of

- Emerging Trends in Computing and Information Sciences, 2010, 1(2): 90-94.
- [7] 王伟, 杨世忠, 宫钰程, 等. EfficientNet V2 算法融合 GCN 和 CA-Transformer 的腐烂草莓分类方法[J]. 食品与机械, 2024, 40(12): 81-88.
- Wang W, Yang S Z, Gong Y C, et al. Rotten strawberry classification based on EfficientNet V2 algorithm fused with GCN and CA-Transformer[J]. Food & Machinery, 2024, 40(12): 81-88.
- [8] 王丹, 刘芳菲, 纪晓峰. 基于双通道改进卷积神经网络的新鲜葡萄品质检测分析[J]. 食品与机械, 2024, 40(12): 89-94.
- Wang D, Liu F F, Ji X F. Fresh grape quality inspection based on dual-channel improved convolutional neural network[J]. Food & Machinery, 2024, 40(12): 89-94.
- [9] Cai X J, Zhu Y, Liu S W, et al. FastSegFormer: a knowledge distillation-based method for real-time semantic segmentation of surface defects in navel oranges[J]. Computers and Electronics in Agriculture, 2024, 217: 108604.
- [10] 侯文慧, 郭丹丹, 周传起, 等. 基于弱监督语义分割的砀山梨表面缺陷识别方法[J]. 农业工程学报, 2025, 41(7): 173-181.
- Hou W H, Guo D D, Zhou C X, et al. Recognizing the surface defects of Dangshan pears using weakly supervised semantic segmentation network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(7): 173-181.
- [11] 王云鹏, 司海平, 宋佳珍, 等. 基于红外与可见光图像融合的水果表面缺陷检测方法[J]. 食品与机械, 2021, 37(12): 127-131.
- Wang Y P, Si H P, Song J Z, et al. Apple surface defect detection method based on fusion of infrared and visible images[J]. Food & Machinery, 2021, 37(12): 127-131.
- [12] Si H P, Wang Y P, Liu Q, et al. A registration algorithm for the infrared and visible images of apple based on active contour model[J]. The Visual Computer, 2024, 40(4): 2 833-2 855.
- [13] 司海平, 赵文沛, 李婷婷, 等. BDMFuse: 基于红外与可见光图像基础特征和细节特征的多尺度融合[J]. 红外与毫米波学报, 2025, 44(2): 289-298.
- Si H P, Zhao W R, Li T T, et al. BDMFuse: multi-scale network fusion for infrared and visible images based on base and detail features[J]. Journal of Infrared and Millimeter Waves, 2025, 44(2): 289-298.
- [14] 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). Las Vegas, NV: IEEE, 2016: 770-778.
- [15] Xu W, Wan Y. ELA: efficient local attention for deep convolutional neural networks[J/OL]. Computer Science. (2024-03-02) [2025-07-11]. <https://arxiv.org/abs/2403.01123>.
- [16] Howard A G, Zhu M L, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J/OL]. Computer Science. (2017-04-17) [2025-06-18]. <https://arxiv.org/abs/1704.04861>.
- [17] Yuan Y G, Xu B B, Hou L, et al. TEA: test-time energy adaptation[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA: IEEE, 2024: 23 901-23 911.
- [18] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. Computer Science. (2014-11-04) [2025-06-22]. <https://arxiv.org/abs/1409.1556>.
- [20] Tan M X, Le Q V. Efficientnetv2: smaller models and faster training[J/OL]. Computer Science. (2021-04-01) [2025-07-08]. <https://arxiv.org/abs/2104.00298>.
- [21] Qin D F, Lechner C, Delakis M, et al. MobileNetV4: universal models for the Mobile ecosystem[C]// Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2025: 78-96.
- [22] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J/OL]. Computer Science. (2020-10-22) [2025-07-02]. <https://arxiv.org/abs/2010.1129>.
- [23] Wang A, Chen H, Lin Z J, et al. Rep ViT: revisiting mobile CNN from ViT perspective[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA: IEEE, 2024: 15 909-15 920.
- [24] Zhu L H, Liao B C, Zhang Q, et al. Vision mamba: efficient visual representation learning with bidirectional state space model[J/OL]. Computer Science. (2024-01-17) [2025-06-29]. <https://arxiv.org/abs/2401.09417>.