

基于生成对抗网络和 Mask R-CNN 的 苹果早期变质检测

Early spoilage detection of apple based on generative adversarial network and Mask R-CNN

于琦龙¹ 赵晓东¹ 籍宇¹ 王春荣² 孙尧³

YU Qilong¹ ZHAO Xiaodong¹ JI Yu¹ WANG Chunrong² SUN Yao³

(1. 河北机电职业技术学院, 河北 邢台 054000; 2. 河北科技大学, 河北 石家庄 050018;

3. 河北农业大学, 河北 保定 071001)

(1. Hebei Institute of Mechanical and Electrical Technology, Xingtai, Hebei 054000, China;

2. Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China;

3. Hebei Agricultural University, Baoding, Hebei 071001, China)

摘要:[目的]提高苹果早期变质区的检测准确率。[方法]基于生成对抗网络和卷积神经网络技术的苹果变质区检测方法。利用 Pix2PixHD 模型生成包含采后早期变质区的贮藏苹果的近红外成像数据;使用 Mask R-CNN 模型对生成的近红外图像进行分割,以检测苹果中的变质区;在具有人工智能功能的低成本嵌入式系统上,利用生成的近红外成像数据,实施基于生成对抗网络和卷积神经网络技术的采后苹果的早期变质区域分割和预测。[结果]该方法对收获后苹果的早期变质检测平均准确率比其他 9 种方法高 1.825%~10.435%;Pix2PixHD 能以 17 帧/s 的速度从 RGB 图像生成了可视近红外图像,Mask R-CNN 能够以 4.2 帧/s 的速度对苹果图像中的变质区域进行分割。[结论]研究提出的方法有望促进低成本食品质量控制器的开发。

关键词:苹果;早期变质检测;生成对抗网络;卷积神经网络;图像转换

Abstract: [Objective] To improve the detection accuracy of early apple spoilage zone. [Methods] An apple spoilage detection method was proposed based on generative adversarial network and convolutional neural network. The Pix2PixHD model was used to generate near-infrared imaging data of stored apples in the early postharvest metamorphic area. The Mask R-CNN model

was used to segment the generated near Infrared image to detect the deterioration zone in the apple. Based on generative adversarial network and convolutional neural network technology, the early deterioration region segmentation and prediction of postharvest apples were implemented by using the generated near-infrared imaging data on a low-cost embedded system with artificial intelligence function. [Results] The average accuracy of this method was 1.825%~10.435% higher than that of the other nine methods. The Pix2PixHD generated a visible NIR image from an RGB image at 17 frames per second, and the Mask R-CNN was able to segment spoilage areas in an apple image at 4.2 frames per second. [Conclusion] The proposed method is expected to facilitate the development of low-cost food quality controllers.

Keywords: apple; early spoilage detection; generative adversarial network; convolutional neural network; image conversion

苹果采后阶段的任何损害都可能导致贮藏果实腐烂、霉变(变质)及变质区域的形成^[1-2]。变质区域在早期并不能很好地从视觉上观察到,但变质的生长过程是动态的^[3]。果实收获后应立即检测,因为贮藏作物有可见的腐烂区或霉变区(通过肉眼或 RGB 相机和传感器获取)可能预示着严重变质^[4]。

卷积神经网络(convolutional neural network, CNN)作为常用的水果收获后变质检测方法,其特有结构确保了 CNN 可使用 RGB(red-green-blue)成像数据高效完成水果分类、检测和分割任务。例如:王灵敏等^[5]搭建了基于 AlexNet 的自动香蕉分级系统,该系统引入经过微调

基金项目:河北省高等学校科学技术研究项目(编号:ZD2019123);邢台市科技计划自筹经费项目(编号:2023ZC013)

作者简介:于琦龙(1988—),男,河北机电职业技术学院讲师,硕士。E-mail:larrylong321@hotmail.com

收稿日期:2024-02-15 **改回日期:**2024-05-30

的 VGG-16 模型和 RGB 图像数据,利用皮质、大小和成熟度等特征对香蕉进行分类。Buyukarikan 等^[6]比较了 5 种 CNN 模型对贮藏苹果的生理失调分类效果;测试结果表明,VGG-19 模型结合 SVM 方法对苹果生理失调的分类平均准确率最高,分别为 96.11%和 96.09%。

然而,受限于 RGB 成像只能反映水平表面物理特征,以上方法只检测有明显缺陷的作物,而未考虑缺陷作物的变质阶段;此外,以上方法也未提及作物受到感染但未见明显变质特征的特殊情况。

与 RGB 成像相比,采用近红外高光谱和多光谱成像获得的图像数据可确保水果早期腐烂或霉变检测更准确。近红外高光谱和多光谱成像数据的获取需要高性能的光谱采集装置和设备,如采集近红外(near infrared, NIR)光谱的多光谱或高光谱照相机^[7]。近红外光谱仪覆盖了从 780~2 500 nm 人眼无法看到的光谱区域。最先进的高光谱数据采集相机和设备不仅能提供扫描样品的光谱信息,还能让用户在设备波段范围内获得扫描区域的图像。Ekramirad 等^[8]比较了线性判别分析、随机森林、支持向量机、kNN、梯度树增强和偏最小二乘判别等 CNN 方法对嘎啦、史密斯奶奶和富士等苹果的早期蠹蛾区进行检测;研究是在像素级使用 900~1 700 nm 范围内的近红外高光谱反射成像数据进行的,并对波长进行了优化选择;结果表明,梯度树增强在苹果早期蠹蛾区的像素级检测方面取得了更好的结果,梯度树增强对苹果早期蠹蛾区的检测准确率达到 97.4%。

但是,近红外高光谱和多光谱成像数据的获取一般需要复杂的设备,这些设备通常包括带波段的相机、成像摄谱仪、样品台、照明灯、闪电系统以及用于处理和捕获近红外数据和图像的辅助软件和设备^[9]。近红外高光谱和多光谱成像设备高昂的价格降低了农民和食品销售公司在收获后阶段进行食品质量控制时使用高光谱相机的可能性和使用率。

为提高苹果早期变质区的检测准确率并解决高光谱成像设备昂贵问题,提出一种基于生成对抗网络(generative adversarial network, GAN)和 Mask R-CNN 的苹果变质区检测方法。GAN 可用于图像到图像的转换任务,即将高质量图像从一个领域合成到另一个领域^[10]。所提方法通过引入 GAN,在不使用低成本高光谱系统的情况下开发近红外成像数据生成新方法的需求;利用可见光近红外成像数据,对采后阶段贮藏苹果中的腐烂区域进行分割和早期检测。

1 数据收集和采集方法

所提方法的架构如图 1 所示。

(1) 选择 GAN 模型,用于从输入的 RGB 数据生成近红外图像,所提方法将农业领域广泛应用的 Pix2Pix、CycleGAN 或 Pix2PixHD 作为候选 GAN。为确定 GAN

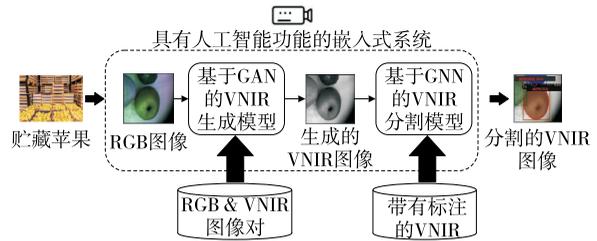


图 1 基于生成对抗网络和 Mask R-CNN 的苹果早期变质检测方法框图

Figure 1 Schematic diagram of early spoilage detection method for apples based on generative adversarial network and Mask R-CNN

模型,采用包含成对 RGB 和近红外图像的数据集对 Pix2Pix、CycleGAN 和 Pix2PixHD 模型的图像转换效果进行比较。

(2) 选择 CNN 模型,分割转换合成的可视近红外(visible NIR, VNIR)图像中的变质区域。所提方法选用 Mask R-CNN 模型,该模型利用特征金字塔网络(feature pyramid network, FPN)和 ResNet101 主干网实现,以便生成边界框(检测对象)和分割掩码(实例分割)。Stasenko 等^[11]将 Mask R-CNN 与 U-Net、Deeplab 等 CNN 应用模型进行了比较;结果表明,Mask R-CNN 在水果的收获后早期腐烂检测方面取得了最高的平均精确度。鉴于此,选用 Mask R-CNN 作为苹果早期变质区域的分割模型。

(3) 在具有人工智能功能的单板计算机上实施所提方法。该单板计算机是一款小巧且功能强大的计算机,配备了计算机视觉和深度学习应用的加速库,可被应用于农业领域的各种实时图像数据处理问题,例如,杂草控制、温室土壤测绘和收获产品检测等。

1.1 深度学习技术

1.1.1 Pix2Pix GAN 包括 2 个子神经网络模型,第 1 个网络称为生成器,其目标是创建可信样本;第 2 个网络称为判别器,验证创建的可信样本是真或伪。Pix2Pix 模型是一种 cGAN,其图像转换有效性已在一系列图像到图像的转换任务中得到验证,例如,将卫星图像转换为地图或将黑白照片转换为彩色图像。Pix2Pix 模型以源图像为条件。判别器涵盖观察到的源图像(区域 A)和目标图像(区域 B),判别器用于确定目标是否是源图像的合理变换。生成器通过对抗损失进行训练,以在目标域生成可信的图像。生成器还通过生成图像与预期输出图像之间测量的 L1 损失进行更新,这种额外的损失会促使生成器模型创建源图像的可信翻译。在数学上, Pix2Pix 的目标函数可定义为^[12]

$$L_{cGAN}(G, D) = \max_{x, y \sim p_{data}(x, y)} [\log D(x, y)] + E_{x, z \sim p_{data}(x, z)} [\log(1 - D(x, G(x, z)))] \quad (1)$$

式中:

x ——观测图像的像素点;

y ——目标图像的像素点;

z ——随机噪声向量的像素点;

G ——生成器的输出结果,即输出图像的像素值;

D ——判别器的输出结果,是一个 $0 \sim 1$ 范围内的实数,判别输入图像是真实图像的概率;

$p_{data(x,y)}$ ——真实图像的分布;

$p_{data(x,z)}$ ——生成图像的分布;

$E_{x,y \sim p_{data(x,y)}}$ ——真实图像数据输入到判别模型输出的计算值;

$E_{x,z \sim p_{data(x,z)}}$ ——生成图像数据输入判别器输出的计算量;

$L_{cGAN}(G,D)$ ——真实图像数据输入到判别模型输出的计算值和生成图像数据输入判别器输出的计算量之和最大化。

Pix2Pix 模型在训练过程中需要完全对齐配对图像。U-Net 模型被用作 Pix2Pix 的生成器;判别器选用 PatchGAN,它通过训练以区分区域 B 中的生成图像和区域 A 中的真实图像。Pix2Pix 的工作流程框图如图 2 所示。

1.1.2 CycleGAN CycleGAN 模型的目标是学习映射 $G: X \rightarrow Y$,使得来自 $G(X)$ 的图像分布与使用未配对的图像对集的分布 Y 无法区分。这个映射与一个逆映射 $F: Y \rightarrow X$ 耦合,并且引入一个循环一致性损失来强制 $F(G(X)) \approx X$ 。对于映射 $G: X \rightarrow Y$ 及其判别器 D_Y ,目标函数为^[13]

$$L_{GAN}(G, D_Y, X, Y) = \min_{G, F} \max_{D_X, D_Y} E_y [\log D_Y(y)] + E_x [\log(1 - D_Y(G(x)))] \quad (2)$$

式中:

X ——观测图像的像素点;

Y ——目标图像的像素点;

z ——随机噪声向量的像素点;

G ——生成器的输出结果,即输出图像的像素值;

D_Y ——判别器的输出结果,是一个 $0 \sim 1$ 范围内的实数,判别输入图像是真实图像的概率;

$E_y [\log D_Y(y)]$ ——真实图像数据输入到判别模型输出的计算值;

$E_x [\log(1 - D_Y(G(x)))]$ ——生成图像数据输入判别器输出的计算量;

$L_{GAN}(G, D_Y, X, Y)$ ——真实图像数据输入到判别模型输出的计算值和生成图像数据输入判别器输出的计算量之和最大化。

CycleGAN 模型在没有对齐配对图像的情况下学习转换映射,在对抗损失之外优化循环一致性损失。

CycleGAN 的工作流程框图如图 3 所示。

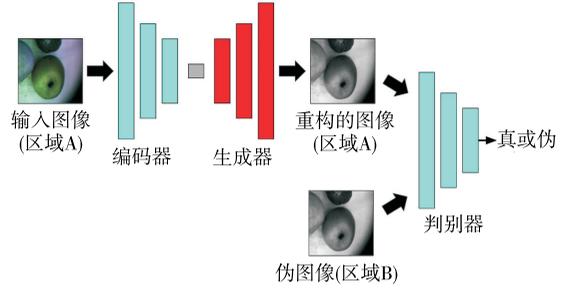


图 2 Pix2Pix 工作流程框图

Figure 2 Pix2Pix flow diagram

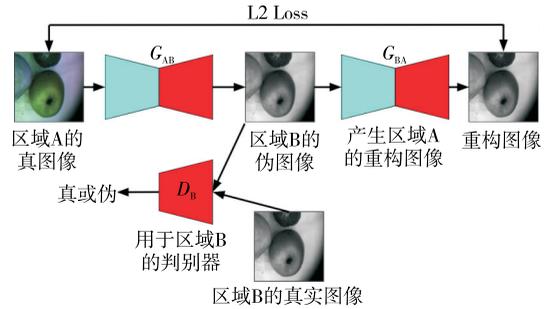


图 3 CycleGAN 的工作流程框图

Figure 3 CycleGAN flow diagram

1.1.3 Pix2PixHD Pix2PixHD 模型对 Pix2Pix 模型中的解决方案进行了改进,具体包括从粗到细生成器、多尺度判别器和对抗损失等多项改进。Pix2PixHD 通常由全局生成器 G_1 和局部增强器 G_2 组成(见图 4)。为了在多个尺度上有效捕捉细节,Pix2PixHD 使用了 3 个判别器。

对损失项进行修改后,目标函数 $L_{cGAN}(G,D)$ 增加了额外项,特征匹配损失 L_{FM} 和感知损失 L_{VGG} 。特征匹配损失可以稳定训练结果,这是因为生成器必须在多个尺度上生成自然统计数据^[14]。

$$L_{FM}(G, D_k) = \max_{\lambda_{FM}} E_{y,x} \sum_{i=1}^{N_i} \frac{1}{N_i} [\| D_k^{(i)}(y, x) - D_k^{(i)}(y, G(y)) \|_1] \quad (3)$$

式中:

N_i ——多尺度判别器 D_k 的层数;

$D_k^{(i)}$ —— D_k 识别器第 i 层的输出;

λ_{FM} ——正则化参数;

$\| \cdot \|_1$ ——一阶范数。

$$L_{VGG} = \max_{\lambda_{VGG}} E_{y,x} \sum_{i=1}^{M_i} \frac{1}{M_i} [\| F^{(i)}(x) - F^{(i)}(G(y)) \|_1] \quad (4)$$

式中:

M_i ——VGG 网络的层数;

$F^{(i)}$ ——VGG 网络第 i 层的输出;

$D_k^{(i)}$ —— D_k 识别器第 i 层的输出;

λ_{vgg} ——正则化参数。

1.1.4 Mask R-CNN Mask R-CNN 可对图像中的各种物体进行实例分割,图像中的这些物体通常被称为感兴趣区域(region of interest, ROI)。Mask R-CNN 是一种用于目标检测和实例分割的深度学习模型。它扩展了 Faster R-CNN 的功能,能够在检测物体位置的同时生成物体的像素级分割掩码。Mask R-CNN 的基本原理和组成部分:① 基础网络:使用卷积神经网络(如 ResNet)作为基础网络,从输入图像中提取特征图;② 区域建议网络:基于特征图生成一系列的候选区域,即可能包含目标的矩形区域;RPN 通过滑动窗口的方式在特征图上生成锚框,然后通过分类和回归来调整锚框的位置和大小;③ RoI Align:对候选区域进行裁剪和重新调整大小,以便与特征图对齐;Mask R-CNN 采用 RoI Align 而非 RoI Pooling,以避免像素级的量化误差,从而提高分割的精度;④ 分类和边界框回归:对每个候选区域进行分类,以确定其类别;对候选区域进行边界框回归,以进一步精确定位目标;⑤ 掩码分支:在每个候选区域的基础上,生成目标物体的二进制掩码;掩码分支是一个全卷积网络,它独立于分类和边界框回归任务,并为每个候选区域预测一个像素级的分割掩码。

图 5 是 Mask R-CNN 模型的工作流程框图,在 Mask R-CNN 的模型中,ROI 通过 CNN 特征的选择性搜索来检测。

Mask R-CNN 的工作流程:① 输入图像:输入图像首

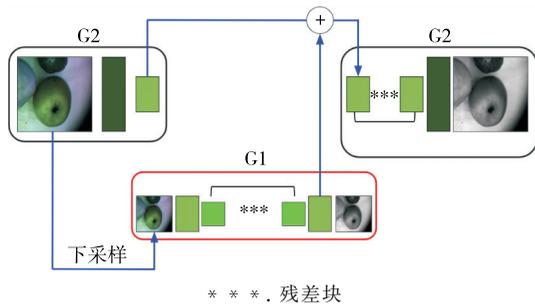


图 4 Pix2PixHD 的工作流程框图
Figure 4 Pix2PixHD flow diagram

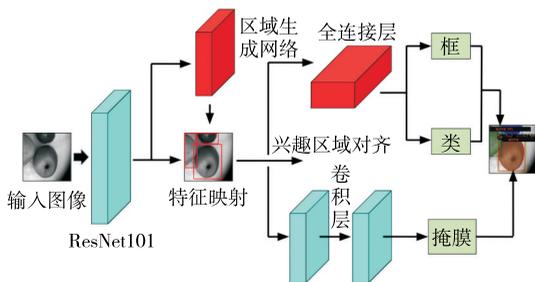


图 5 Mask R-CNN 的工作流程框图
Figure 5 Mask R-CNN flow diagram

先通过基础网络 ResNet101,提取特征图;② 生成候选区域:在特征图上生成锚框,并通过分类器筛选出高概率包含目标的候选区域;③ 区域对齐:使用 RoI Align 将这些候选区域在特征图上进行裁剪和对齐,确保像素级别的精度;④ 分类和回归:对齐后的特征区域分别通过分类器和回归网络,输出类别标签和精确的边界框位置;⑤ 生成掩码:掩码分支在对齐后的特征区域上生成每个目标的二进制分割掩码。

1.2 性能指标

为验证 Pix2PixHD 模型生成的 VNIR 图像的有效性,将 Pix2PixHD 模型生成的 VNIR 图像与原始 VNIR 图像进行对比,对比效果选用平均误差 M_{MAE} 、平均百分比误差 M_{MAPE} 、平均平方误差 M_{MSE} 、均方根误差 R_{RMSE} 、峰值信噪比 P_{PSNR} 、结构相似性指数 S_{SSIM} 和特征相似性指数 F_{FSIM} 等^[15] 作为评价指标。

$$M_{MAE} = \frac{1}{n} \sum_{i=1}^n | (y_i - x_i) |, \quad (5)$$

$$M_{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{(y_i - x_i)}{y_i} \right|, \quad (6)$$

$$M_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2, \quad (7)$$

$$R_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}, \quad (8)$$

$$P_{PSNR} = 10 \log_{10} \left(\frac{R^2}{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \right), \quad (9)$$

$$S_{SSIM} = [l(x_i, y_i)^a \cdot c(x_i, y_i)^\beta \cdot s(x_i, y_i)^\gamma], \quad (10)$$

$$F_{FSIM} = [S_{PC}(x_i, y_i)^a \cdot S_{GM}(x_i, y_i)^\beta], \quad (11)$$

- 式中:
- n —— 观测值的数量;
 - y_i —— 生成图像第 i 个像素值;
 - x_i —— 原始图像第 i 个像素值;
 - R —— 图像可能的最大像素值;
 - l —— 亮度;
 - c —— 对比度;
 - s —— 结构;
 - a, b, g —— 权重系数;
 - S_{PC} —— 图像中光线变化时的不变量;
 - S_{GM} —— 图像梯度的计算值。

在训练和验证阶段,使用精度 (precision)、召回率 (recall)、平均交叉率 (mean intersection over union, mIoU)、平均精度 (mean average precision, mAP) 和 F_1 分数^[16] 来验证 Mask R-CNN 模型对合成 VNIR 图像的处理效率。

1.3 数据采集

用于数据采集的试验平台:① 用于获取数据集 1 的试验平台,生成包含贮藏苹果的成对 RGB 和 VNIR 图像;② 用于收集数据集 2 的试验平台,由多光谱相机获取 VNIR 图像。第 1 个试验平台用于收集成对的 RGB 和 VNIR 图像,以训练和验证 GAN 模型,从 RGB 图像转换 VNIR 图像。第 2 个试验平台用于收集贮藏的苹果 VNIR 图像,并将其用于训练和验证基于 Mask R-CNN 的苹果采后早期变质区分割和检测模型。

1.3.1 数据集 1 收集 选取了 4 种苹果(“美味”“富士”“嘎啦”“Reinette Simirenko”)共 16 个(如图 6 所示),按照苹果的种类分成 4 行(每列对应每种苹果)。每行包含 4 个不同种类的苹果,每个苹果都经过不同的处理,从左到右依次为:未处理的苹果(作为参考)、彻底清洗和擦拭的苹果(苹果的天然保护蜡层已被去除)、机械损伤的苹果和低于 $-20\text{ }^{\circ}\text{C}$ 的温度环境中过冷冲击冷冻的苹果。

数据集 1 在室内存储条件下收集。室内温度范围为 $25\sim 32\text{ }^{\circ}\text{C}$,相对湿度为 34% 。数据收集台包含一个铝制框架,长 1 m 、宽 1 m 、高 1.7 m 。苹果放在离地面 1.3 m 高的桌子上,桌子上有一个白色托盘。将单反相机 Canon M50 和多光谱相机 CMS-V1CMS18100073(CMS-V)安装在框架的中间顶部,并通过 USB 集线器连接到 PC 笔记本电脑。放有苹果的桌子与相机之间的距离为 500 mm 。在全光照和部分光照条件下采集图像。



图 6 选定用于收集 RGB 和 VNIR 成对成像数据的部分苹果

Figure 6 Selected apple used to collect RGB and VNIR paired imaging data

多光谱相机 CMS-V 可以获取 $561\sim 838\text{ nm}$ 波段(包括可见光和近红外波段)的图像。该相机成像器的特点是采用改良贝叶矩阵,由一组 3×3 像素组成,过滤 3×3 个光谱带。相机提供的原始图像由 9 个交错光谱子图像(8 种颜色+1 个全色)组成,分辨率为 $1\ 280\text{ 像素}\times 1\ 024\text{ 像素}$ 。每幅 RGB 图像涉及以下光谱带的 9 幅图像:通道 0 获取 561 nm 图像,通道 1 获取 597 nm 图像,通道 2 获取 635 nm 图像,通道 3 获取 673 nm 图像,通道 4 获取 724 nm 图像,通道 5 获取 762 nm 图像,通道 6 获

取 802 nm 图像,通道 7 获取 838 nm 图像,通道 8 为全色通道。9 个子图像的分辨率均为 $426\text{ 像素}\times 339\text{ 像素}$ 。通过试验平台获取了 $1\ 305$ 幅连续的 RGB 图像和 $1\ 305$ 幅 838 nm 范围内的 VNIR 图像,以观察所呈现的苹果的衰变动态。图像示例如图 7 所示。

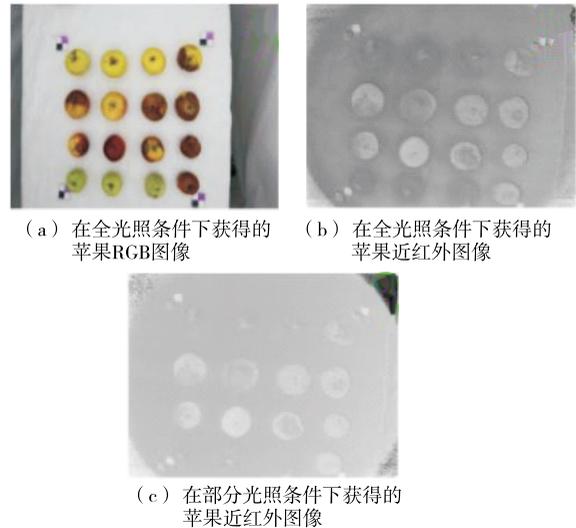


图 7 获得的苹果 RGB 和 VNIR 图像

Figure 7 Apple RGB and VNIR images

1.3.2 数据集 2 收集 选择了 22 个美味、富士、嘎啦、Reinette Simirenko 等季节性类型的苹果进行近红外成像数据(数据集 2)采集。这些苹果的直径在 $8\sim 10\text{ cm}$,大部分都是红黄相间的多色苹果。还有一些苹果含有灰褐色发霉区。使用这些苹果是为了增加数据代表性,以便利用近红外成像数据对贮藏苹果进行采后早期变质检测。用于近红外图像数据采集的苹果如图 8 所示。近红外成像数据收集的平台是一个温室,包括铝制框架和 5 个架子、塑料膜、多光谱照相机、10 个带红/蓝二极管的 LED 条形灯、一个用于控制 LED 的电源(总功率为 150 W)、一个记录仪和一个装有苹果的托盘。该平台可用于模拟各种环境条件下与苹果贮藏有关的不同过程,包括极端干燥或潮湿模式。试验台的温度和湿度调节是通过 LED 条形灯、塑料膜和位于底部 3 个独立架子上的几个托盘来实现的。



图 8 用于近红外图像数据采集的苹果

Figure 8 Apples for near infrared image data acquisition

硅胶框架高 170 cm、长 48 cm、宽 67 cm。每个架子上固定 2 个条形灯,而多光谱相机和装苹果的托盘则分别固定在不同的架子上。每个选定的灯条上有 60 个 LED,波长为 650~660 nm(红光 LED)和 455~465 nm(蓝光 LED)。依靠电源(12 V DC、150 W)作为 SMD 5050 LED 灯条的能量,Graphtech 公司的 GL100-N/GL100-WL 记录器(配备 GS-TH 传感器模块)在数据采集过程中记录温度和湿度值。为了获取近红外图像,选用多光谱相机 CMS-V。该相机通过 USB-A 线连接到 HP Elite Book 820 G3 笔记本电脑(CPU 为英特尔酷睿 i3-6100,2.30GHz)上,所有图像均在此采集并保存为 426 像素×339 像素的 JPG 文件。

从 CMS-V 相机的通道 7 中获得了 1 029 幅 838 nm 范围内的连续 VNIR 图像。这些图像是在温度 35~40 °C、相对湿度 70% 的范围内采集的,目的是模拟选定苹果贮藏过程中可能出现的非规范贮藏现象。这种贮藏条件会加速苹果的衰变过程。此外,还收集了 100 幅 RGB 连续图像,用于 CNN 模型训练和验证。RGB 图像的尺寸为 339 像素×426 像素×3 个通道。

1.4 数据注释

为了将所提方法用于图像实例分割,使用 Supervisely Ecosystem 对数据集 2 中的 VNIR 图像进行注释和标记,这些图像用于对苹果采后早期变质检测和分割的 Mask R-CNN 模型进行训练和验证。图像中的对象分为 3 类:健康苹果、早期变质苹果和严重变质苹果。健康苹果指的是图像中没有任何明显损坏或变质区域的苹果。苹果采后腐烂的深灰色区域表示腐烂,霉变是指苹果中的白色霉变区。在此,将采后腐烂区域和霉变区域标记为变质区域。

如果有霉变类的苹果对象,这意味着该苹果是在异常的贮藏条件下贮藏的,例如极端的温度或湿度导致苹果全部变质。定义了“严重变质苹果”类别:贮藏的苹果中变质区域(腐烂对象)或发霉区域(真菌对象)覆盖率超过 50%。

2 仿真测试与结果分析

2.1 将 RGB 图像转化为 VNIR 图像的效果分析

首先分析基于生成式对抗网络的深度学习模型将 RGB 图像转换为 VNIR 图像的效果。对数据集 1 中的 RGB 图像转换成的 VNIR 图像与 838 nm 范围内采集的与 RGB 图像相对应的 VNIR 图像进行了比较。为了估算 RGB 图像转换为 VNIR 图像的性能,将数据分为训练集(80%)和验证集(20%)。采用了随机旋转、移位、缩放和翻转等增强技术提高数据的代表性。不使用对比度/亮度调整等变换,因为它们可能会导致获取的近红外成像数据信息丢失。考虑到图像的平移也称为从区域 A 到区域 B 的平移(A-to-B),有必要在获取的成对数据集中标注区域 B 和区域 A 图像。将 RGB 图像确定为区域 A,

将 VNIR 图像确定为区域 B。所有模型均通过 200 次迭代进行评估,其中,前 100 次迭代采用恒定学习率,其余 100 次迭代采用线性递减至零的学习率。模型的训练和验证是通过 GoogleColab 中的 Python 脚本实现。

在 CycleGAN 模型中,使用由 2 个下采样层、6 个 ResNe 块和 2 个上采样层组成的 ResNet 编码器—解码器架构。还采用了学习率为 0.000 2、动量参数 $\beta_1 = 0.5$ 和 $\beta_2 = 0.999$ 的 Adam 优化器。对于 Pix2Pix 模型的训练,参数为批量大小 1、 $\beta_1 = 0.5$ 、 $\beta_2 = 0.999$ 、学习率 0.000 2。U-Net 生成器有 4 个下采样块。优化步骤包括生成器损失优化步骤和判别器损失优化步骤。正则化参数为 $\lambda_{VGG} = \lambda_{FM} = 10$ 。对于 Pix2PixHD 模型,Adam 优化器采用了以下参数:批量大小 1、 $\beta_1 = 0.5$ 、 $\beta_2 = 0.999$ 、学习率 0.000 2。图 9 给出了 CycleGAN、Pix2Pix 和 Pix2PixHD 模型在训练阶段的损失值。

对于选定的 GAN 模型,训练起始阶段并不稳定,但随着时间推移,判别器损失不断减小。与 CycleGAN 和 Pix2Pix 相比,Pix2PixHD 的损失值最低。为了验证图像转换的有效性,使用训练期间获得的模型权重重建了

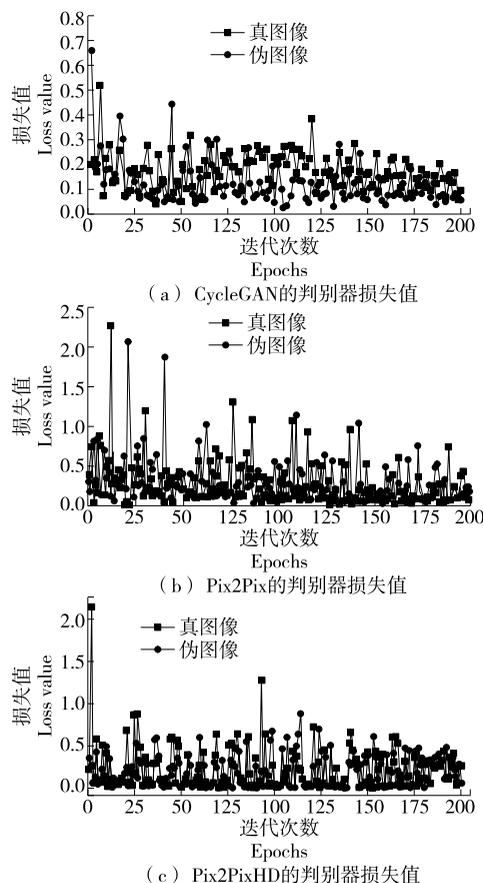
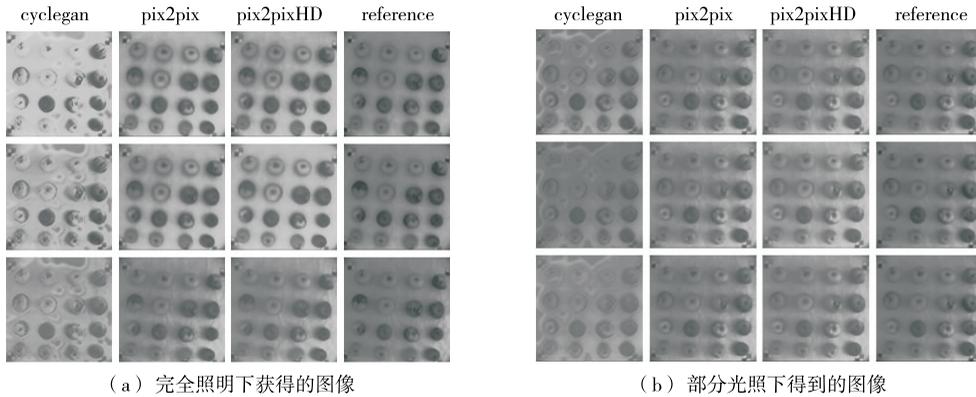


图 9 GAN 模型训练期间的判别器损失值
Figure 9 Loss values of discriminators during GAN model training

VNIR 图像。使用 MAE、MAPE、MSE、PSNR 和 SSIM 指标来评估 VNIR 重建图像与原始 VNIR 图像的差异。通过 Python 可视化工具将 VNIR 重建图像与原始 VNIR

图像(“参考”标签)的对比情况如图 10 所示,图中的“Cyclegan”“pix2pix”和“pix2pixHD”标签分别代表所采用的图像转换方法。“reference”表示原始 VNIR 图像。



(a) 完全照明下获得的图像 (b) 部分光照下得到的图像

图 10 VNIR 重建图像与原始 VNIR 图像

Figure 10 VNIR reconstruction images and original VNIR images

表 1 总结了不同 GAN 模型的图像转换效果,其中 Pix2PixHD 模型的结果用加黑突出显示。CycleGAN 模型合成的苹果图像遗漏了很多腐烂区, Pix2Pix 和 Pix2PixHD 模型的表现要好得多,对腐烂区域保存得相对较好。Pix2PixHD 模型产生了良好的感知图像,保留了任务特征的重要性,平均误差水平为 0.6%。对于 PSNR 和 SSIM, Pix2PixHD 模型比 Pix2Pix 显示出更高的值。鉴于表 1 所示比较结果,决定在后续阶段使用 Pix2PixHD 模型实现从 RGB 图像到 VNIR 图像转换。

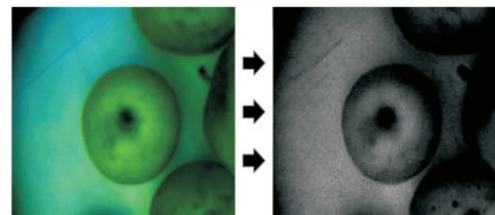
表 1 不同 GAN 模型的图像转换效果

Table 1 Image conversion effects of different GAN models

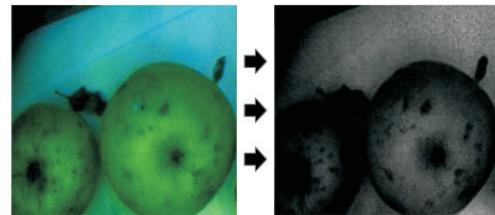
模型	MAE	MAPE	MSE	PSNR	SSIM
CycleGAN	0.067	0.105	0.011 27	27.375	0.856
Pix2Pix	0.004	0.006	0.000 03	46.433	0.955
Pix2PixHD	0.004	0.006	0.000 03	46.859	0.972

2.2 用于苹果采后早期腐烂检测的 VNIR 图像分割

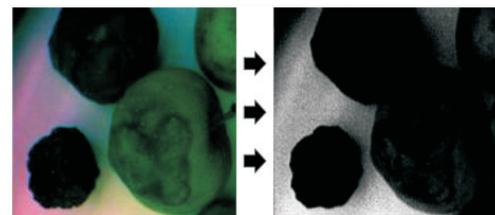
使用 Pix2PixHD 模型生成 VNIR 图像。456 张苹果图像被用于 Pix2PixHD 模型训练。由相应的输入 RGB 图像合成 VNIR 图像的示例如图 11 所示。将新图像的质量与 Pix2PixHD 训练阶段合成的图像进行比较, PSNR 和 SSIM 值分别从 46.859 和 0.972 提高到 52.876 和 0.994。Mask R-CNN 作为图像实例分割模型。在将 Mask R-CNN 应用于 Pix2PixHD 合成的图像之前,需要在真实的 VNIR 图像上训练 Mask R-CNN, 以使其具有检测和分割贮藏苹果中腐烂区域的能力。使用包含 1 029 张 VNIR 图像的标记数据集进行 Mask R-CNN 模型的训练和验证。



输入RGB图像 合成的近红外图像
(a) 正常苹果的RGB图像和合成的近红外图像



输入RGB图像 合成的近红外图像
(b) 表面早期变质苹果的RGB图像和合成的近红外图像



输入RGB图像 合成的近红外图像
(c) 表面严重变质苹果的RGB图像和合成的近红外图像

图 11 使用 Pix2PixHD 模型加权合成的近红外图像示例

Figure 11 Examples of a near-infrared image weighted by the Pix2PixHD model

在该研究中, Mask R-CNN 选用 L1 作为损失函数, ResNet50 作为主干网络, 随机梯度下降作为优化器, 并使用 COCO 权重来使用 Detectron2 库。采用 GaussianNoise、RandomGamma、RandomBrightness 和

HorizontalFlip 作为数据增强函数,以保证模型在训练和验证阶段的有效性。Mask R-CNN 模型采用 Python 开发,所有计算都是在 Google Colab 中实现。

在包含 VNIR 图像的数据集上对 Mask R-CNN 模型进行训练和交叉验证。在测试例子中,处理对象是序列图像,即一个苹果可以在许多图像中定位而不改变位置,这有助于在训练过程中降低损失值。在交叉验证过程中,数据通常被分成几组,称为折叠,每组用于逐个进行训练和验证。例如,数据集折叠可从 3 种折叠方式中选择:① 第 1 次折叠是 1 个验证集,第 2 次和第 3 次折叠形成训练集;② 第 1 次和第 3 次折叠形成训练集,第 2 次折叠形成验证集;③ 第 1 次和第 2 次折叠形成训练集,第 3 次折叠形成验证集。该设置也适用于四折及更高折数分布的交叉验证。默认情况下,折叠数(也称为 k -folds)通常设置为 5 或 10。在该研究中,候选折叠数设为 2,3,6,9。

经过训练的 Mask R-CNN 模型允许对早期变质苹果进行准确分割和检测,经过训练的 Mask R-CNN 模型对早期变质苹果检测的 mAP 和 F_1 分数分别为 97.587 和 97.829(见表 2 和表 3)。

2.3 利用嵌入式系统生成的近红外成像数据检测贮藏苹果的采后早期腐烂情况

为评估所提方法在现实场景中的适用性,使用 NVIDIA Jetson Nano 嵌入式系统进行了试验测试。

表 2 k 取不同值时 Mask R-CNN 模型分割苹果图像的 mAP 值

Table 2 The mAP values of Mask R-CNN model for segmenting apple images when k takes different values

k 值	健康苹果	严重变质苹果	早期变质苹果
2	94.785	94.839	92.090
3	95.154	95.567	93.117
6	97.951	98.678	96.905
9	98.350	98.997	97.587

表 3 k 取不同值时 Mask R-CNN 模型分割苹果图像的 F_1 分数

Table 3 The F_1 scores of Mask R-CNN model for segmenting apple images when k takes different values

k 值	健康苹果	严重变质苹果	早期变质苹果
2	95.640	92.122	91.015
3	95.589	93.134	92.460
6	96.299	95.689	94.976
9	98.375	97.800	97.829

处理对象为 100 张 RGB 苹果图像,图像大小为 256 像素。使用 GAN 模型从输入图像生成 VNIR 图像,并以平均 17.00 帧/s 的速度处理苹果图像。然后用 Mask R-CNN 对生成的图像进行测试,平均处理速度为 4.20 帧/s。与 Pix2PixHD 相比,Mask R-CNN 的低每秒帧数源于其运算复杂性。由于两阶段检测模型通过检测对象和为每个对象生成像素级掩码来执行实例分割,因此需要更多的计算资源,可以通过引入更高处理效率的计算资源解决该问题。

2.4 与已有方法对比

所提方法通过引入 cGAN,在不使用高成本高光谱系统的情况下开发近红外成像数据生成新方法的需求。该方法利用可见光近红外成像数据,对采后阶段贮藏苹果中的腐烂区域进行分割和早期检测。为验证所提方法的有效性,将所提方法的苹果早期变质检测结果与应用近红外成像数据和深度学习技术预测贮藏苹果采后早期变质区的相关研究成果进行比较,对比结果如表 4 所示。

表 4 不同方法对苹果早期变质检测效果对比

Table 4 Comparison of the effects of different methods on early apple spoilage detection

方法	mAP	方法	mAP
YOLOv4 ^[17]	92.043	YOLO5s ^[19]	94.495
U-Net ^[18]	87.152	CARS-PLS-DA ^[20]	95.088
Improved U-Net ^[18]	91.537	梯度树 ^[21]	95.762
Faster R-CNN ^[19]	94.925	BP-ANN ^[22]	95.167
YOLO v3 ^[18]	95.003	所提方法	97.587

由表 4 可知,在同样的仿真环境和参数下,所提方法比其他 9 种具有准确的苹果早期变质检测效果,该对比结果充分说明了所提方法的有效性。

3 结论

研究提出了基于 cGAN 和 Mask R-CNN 的苹果早期变质检测方法,该方法利用合成和分割的 VNIR 图像对收获后苹果的早期变质区进行检测。结果表明,该方法优于已有的机器学习和深度学习算法。通过 Pix2PixHD 生成的 VNIR 图像的 SSIM 指标达到了 0.972;经过训练的 Mask R-CNN 模型对早期变质苹果检测的平均精度和 F_1 分数分别为 97.587 和 97.829。此外,在具有人工智能功能的嵌入式系统上测试所提出方法,其中, Pix2PixHD 生成 VNIR 图像的速度为 17.00 帧/s,所用 Mask R-CNN 的平均处理速度为 4.20 帧/s。虽然 Mask R-CNN 需要复杂的运算,但 cGAN 的采用确保了该方法是一种很有前途的解决方案,可以替代昂贵的高光谱成像设备,用于采后食品质量控制的早期腐烂预测任务。

参考文献

- [1] 赵敏, 王成荣, 李苒. 联合图像最优特征提取及改进 RBF 神经网络的苹果质量估计[J]. 食品与机械, 2024, 40(2): 125-130, 183. ZHAO M, WANG C R, LI R. Apple weight estimation based on joint image optimal feature extraction and improved RBF neural network[J]. Food & Machinery, 2024, 40(2): 125-130, 183.
- [2] 符灵梅, 王玉健, 徐莉, 等. 超高效液相色谱—串联质谱法快速分析氟霜唑及其代谢物 CCIM 在水果中的残留[J]. 食品与生物技术学报, 2022, 41(9): 93-103. FU L M, WANG Y J, XU L, et al. Rapid analysis of cyazofamid and CCIM residues in fruits by ultra-high performance liquid chromatography tandem mass-spectrometry [J]. Journal of Food Science and Biotechnology, 2022, 41(9): 93-103.
- [3] 关洪浦, 耿明阳, 周逸博, 等. 基于 SERS 的苹果树腐烂病原菌早期感染检测[J]. 农业工程学报, 2024, 40(5): 224-230. GUAN H P, GENG M Y, ZHOU Y B, et al. Early infection detection of applevalsa canker pathogens based on SERS[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(5): 224-230.
- [4] 代云, 卢明, 何婷, 等. 基于改进型 YOLO 的密集环境下槟榔果实的快速识别方法[J]. 食品与机械, 2023, 39(4): 83-88. DAI Y, LU M, HE T, et al. Fast recognition method for betel nut in dense environments based on improved YOLO [J]. Food & Machinery, 2023, 39(4): 83-88.
- [5] 王灵敏, 蒋瑜. 基于深度学习的香蕉成熟度自动分级[J]. 食品与机械, 2022, 38(11): 149-154. WANG L M, JIANG Y. Automatic classification of banana ripeness based on deep learning [J]. Food & Machinery, 2022, 38(11): 149-154.
- [6] BUYUKARIKAN B, ULKER E. Classification of physiological disorders in apples fruit using a hybrid model based on convolutional neural network and machine learning methods [J]. Neural Computing and Applications, 2022, 34(19): 16 973-16 988.
- [7] 李鑫, 沈晓君, 王媛媛, 等. 近红外光谱法快速测定晒青毛茶三种儿茶素组分含量[J]. 现代食品科技, 2024, 40(3): 326-332. LI X, SHEN X J, WANG Y Y, et al. Rapid determination of three catechin components in crude Pu-erh tea by near infrared spectroscopy[J]. Modern Food Science and Technology, 2024, 40(3): 326-332.
- [8] EKRAMIRAD N, KHALED A Y, DOYLE L E, et al. Nondestructive detection of codling moth infestation in apples using pixel-based nir hyperspectral imaging with machine learning and feature selection [J]. Foods, 2021, 11(1): 8.
- [9] 陈龙跃, 段丹丹, 王凡, 等. 高光谱成像技术在水果品质无损检测中的研究进展[J]. 广东农业科学, 2023, 50(7): 83-94. CHEN L Y, DUAN D D, WANG F, et al. Research progress of non-destructive testing of fruit quality by hyperspectral imaging technology [J]. Guangdong Agricultural Sciences, 2023, 50(7): 83-94.
- [10] SA I, LIM J Y, AHN H S, et al. DeepNIR: datasets for generating synthetic NIR images and improved fruit detection system using deep learning techniques[J]. Sensors, 2022, 22(13): 4 721.
- [11] STASENKO N, SAVINOV M, BURLUSKIY V, et al. Deep learning for postharvest decay prediction in apples[C]// IECON 2021-47th Annual Conference of the IEEE Industrial Electronics Society. [S.l.]: IEEE, 2021: 1-6.
- [12] 朱赫鹏, 丁喆, 张严, 等. 基于生成对抗网络的跨分辨率拓扑优化方法的比较研究[J]. 振动与冲击, 2024, 43(6): 132-140. ZHU H P, DING Z, ZHANG Y, et al. Comparative study on cross-resolution topology optimization methods based on generative adversarial networks[J]. Journal of Vibration and Shock, 2024, 43(6): 132-140.
- [13] 陈嘉欣, 孙传猛, 葛耀栋, 等. 基于 CycleGAN 和 Pix2Pix 的轨面缺陷图像智能生成技术[J]. 铁道学报, 2024, 46(2): 122-130. CHEN J X, SUN C M, GE Y D, et al. Intelligent generation of rail surface defect images based on CycleGAN and Pix2Pix[J]. Journal of the China Railway Society, 2024, 46(2): 122-130.
- [14] 高志军, 冀远明, 史二美. 基于改进 pix2pixHD 模型的 COVID-19 CT 图像生成方法研究[J]. 智能计算机与应用, 2023, 13(5): 82-89, 97. GAO Z Y, YI Y M, SHI E M. Generation method of the COVID-19 CT image based on improved pix2pixHD model[J]. Intelligent Computer and Applications, 2023, 13(5): 82-89, 97.
- [15] 赵振宇, 杨天豪, 蒋汶乘, 等. 基于机器学习的多压多温多参标准单元延迟快速计算方法[J]. 计算机工程与科学, 2023, 45(8): 1 331-1 338. ZHAO Z Y, YANG T H, JIANG W C, et al. A machine learning-based fast calculation method of multi-voltage, multi-temperature and multi-parameter standard cell delay[J]. Computer Engineering & Science, 2023, 45(8): 1 331-1 338.
- [16] 夏军勇, 王康宇, 周宏娣. 基于改进 Faster R-CNN 的食品包装缺陷检测[J]. 食品与机械, 2023, 39(11): 131-136, 151. XIA J Y, WANG K Y, ZHOU H D, et al. Food packaging defect detection by improved network model of Faster R-CNN[J]. Food & Machinery, 2023, 39(11): 131-136, 151.
- [17] FAN S X, LIANG X T, HUANG W Q, et al. Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network[J]. Computers and Electronics in Agriculture, 2022, 193: 106715.
- [18] TANG Y, BAI H Y, SUN L J, et al. Multi-band-image based detection of apple surface defect using machine vision and deep learning[J]. Horticulturae, 2022, 8(7): 666.
- [19] TIAN M M, ZHANG J C, YANG Z R, et al. Detection of early bruises on apples using near-infrared camera imaging technology combined with adaptive threshold segmentation algorithm [J]. Journal of Food Process Engineering, 2024, 47(1): e14500.

(下转第 169 页)

- Food Science, 2017, 38(4): 146-154.
- [14] 黄秋颜, 李斌, 林晓蓉, 等. 基于高分辨质谱和网络药理学的可可茶多酚降血糖活性研究[J]. 食品与机械, 2023, 39(9): 4-11.
HUANG Q Y, LI B, LIN X R, et al. Studying on the hypoglycemic activity of green tea polyphenols from *Camellia ptilophylla* Chang using high-resolution mass spectrometry and network pharmacology[J]. Food & Machinery, 2023, 39(9): 4-11.
- [15] 孙梦佳, 王焕军, 丁富娟, 等. 基于 UPLC-Q-Exactive Orbitrap MS 和质量亏损过滤技术的三叶青黄酮类化合物定性分析[J]. 质谱学报, 2020, 41(4): 359-367.
SUN M J, WANG H J, DING F J, et al. Qualitative analysis of flavonoids in *tertastigma hemsleyanum* based on UPLC-Quadrupole/exactive orbitrap mass spectrometry and mass defect filter method[J]. Journal of Chinese Mass Spectrometry Society, 2020, 41(4): 359-367.
- [16] 孙燕, 冯峰, 黄特辉, 等. 基于 UPLC-Q-Exactive 技术结合 OTCML 数据库快速分析沙棘的功效成分[J]. 天然产物研究与开发, 2019, 31(7): 1 192-1 202.
SUN Y, FENG F, HUANG T H, et al. Rapid analysis of functional components of *Hippophae rhamnoides* by UPLC-Q-Exactive and OTCML database[J]. Natural Product Research and Development, 2019, 31(7): 1 192-1 202.
- [17] 张巍, 吴亚猛, 陈思瑶, 等. 胡桃醌对宫颈癌 SiHa 细胞增殖及细胞周期的影响[J]. 吉林大学学报(医学版), 2015, 41(4): 697-700.
ZHANG W, WU Y M, CHEN S Y, et al. Effects of Juglone on proliferation and cell cycle of human cervical cancer SiHa cells[J]. Journal of Jilin University (Medicine Edition), 2015, 41(4): 697-700.
- [18] 邹翔, 曲中原, 季宇彬. 胡桃醌经由活性氧介导 JNK、p38 通路诱导 SGC-7901 细胞凋亡[J]. 中国药理学杂志, 2016, 51(7): 544-549.
ZOU X, QU Z Y, JI Y B. Juglone induces apoptosis of human gastric cancer SGC-7901 cells through reactive oxygen species mediated JNK and p38 pathway[J]. Chinese Pharmaceutical Journal, 2016, 51(7): 544-549.
- [19] 李晓颖, 武军凯, 王海静. 欧李果实发育期内挥发性成分变化特征[J]. 中国农业科学, 2021, 54(9): 1 964-1 980.
LI X Y, WU J K, WANG H J. Characterization of volatiles changes in Chinese dwarf cherry fruit during its development[J]. Scientia Agricultura Sinica, 2021, 54(9): 1 964-1 980.
- [20] 江珊珊, 邓青, 范佳. 麦长管蚜对 *E-β*-法尼烯的嗅觉行为反应[J]. 昆虫学报, 2015, 58(7): 776-782.
JIANG S S, DENG Q, FAN J. Behavioral responses of *Sitobion avenae* (Hemiptera: Aphididae) to *E-β*-farnesene [J]. Scientia Agricultura Sinica, 2015, 58(7): 776-782.
- [21] 张书锋, 董庆峰, 胡聪, 等. 石家庄野生荆条花挥发油的化学成分[J]. 中国实验方剂学杂志, 2013, 19(23): 151-153.
ZHANG S F, DONG Q F, HU C, et al. Chemical composition of flower volatile oil of wild vitex negundo var heterophylla from Shijiazhuang [J]. Chinese Journal of Experimental Traditional Medical Formulae, 2013, 19(23): 151-153.
- [22] 赵杨, 程力, 杜庭, 等. 基于 HS-SPME/GC/MS 的姜花不同部位挥发性成分分析[J]. 食品与机械, 2020, 36(3): 43-47.
ZHAO Y, CHENG L, DU T, et al. Analysis on volatile components in different parts of *Hedychium coroarium* by HS-SPME/GC/MS [J]. Food & Machinery, 2020, 36(3): 43-47.
- [23] 宋诗清, 童彦尊, 冯涛, 等. 金佛手香气物质的多维分析及其特征香气物质的确定[J]. 食品科学, 2017, 38(24): 94-100.
SONG S Q, TONG Y Z, FENG T, et al. Multidimensional analysis of odorous compounds in finger citron fruit (*Citrus medica* L. var. *sarcodactylis* Swingle) and Identification of key aroma compounds [J]. Food Science, 2017, 38(24): 94-100.
- [24] 高夏洁, 高海燕, 赵镭. SPME-GC-MS 结合 OAV 分析不同产区花椒炸花椒油的关键香气物质[J]. 食品科学, 2022, 43(4): 208-214.
GAO X J, GAO H Y, ZHAO L. Analysis of the key aroma components of fried *Zanthoxylum bungeanum* oil from different producing areas based on SPME-GC-MS and OAV [J]. Food Science, 2022, 43(4): 208-214.
- [25] 王艳, 周长远, 杜爱玲. 硅胶-硝酸银硅胶柱层析分离纯化 β -倍半水芹烯[J]. 食品与机械, 2016, 32(5): 165-167.
WANG Y, ZHOU C Y, DU A L. Separation of β -sesquiphellandrene by silica gel and argentation column chromatography[J]. Food & Machinery, 2016, 32(5): 165-167.
- [26] 潘亚琴, 李全清, 张森尧, 等. 山核桃叶与美国山核桃叶中化学成分的初步分析[J]. 中华中医药学刊, 2008, 26(11): 2 517-2 519.
PAN Y Q, LI Q Q, ZHANG S Y, et al. Study on chemical components in the leaves of *Carya cathayensis* and *C. illinensis*[J]. Chinese Archives of Traditional Chinese Medicine, 2008, 26(11): 2 517-2 519.

(上接第 151 页)

- [20] YUAN R R, LIU G S, HE J G, et al. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis[J]. Computers and Electronics in Agriculture, 2021, 182: 106043.
- [21] YANG Z R, YUAN Y H, ZHENG J H, et al. Early apple bruise recognition based on near-infrared imaging and grayscale gradient images[J]. Journal of Food Measurement and Characterization, 2023, 17(3): 2 841-2 849.
- [22] HU Q X, TIAN J, FANG Y. Detection of moldy cores in apples with near-infrared transmission spectroscopy based on wavelet and BP network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(12): 1950020.