doi:10.6041/j.issn.1000-1298.2025.12.041

基于机器视觉与高光谱成像的西瓜考种系统研究

赵泽华[1,2]　王庆艳[2,3]　陈俊杰[2,3]　黄文倩[2,3]

(1.新疆农业大学机电工程学院,乌鲁木齐 830052;2.北京市农林科学院智能装备技术研究中心,北京 100097;
3.国家农业智能装备工程技术研究中心,北京 100097)

摘要:传统的西瓜考种方式存在效率低、工作量大、主观性强,易导致不同育种工作人员对相同品质性状的瓜评价不统一等问题。为此,提出了一种基于机器视觉与高光谱成像技术的西瓜考种系统。设计了西瓜图像自动采集装置,包含一台 GigE 视觉相机和一个称量平台来获取西瓜的表型数据,并对西瓜切面选择了 10 个不同的感兴趣区域(ROI)进行糖度建模,利用高光谱成像技术实现西瓜切面各个位置糖度分布可视化。通过最小外接矩形法对西瓜的横径和纵径做出测量,得到最小外接矩形的长和宽即为西瓜的横径和纵径。应用 Canny 边缘检测算法实现对瓜皮轮廓的提取,并取瓜皮轮廓的瓜蒂、瓜脐和两个中间位置,这 4 处瓜皮厚度的平均值作为皮厚真实值。采用卷积平滑(SGS)算法分别结合多元散射校正(MSC)、标准正态变换(SNV)、单位矢量归一化(UVN)3 种算法对光谱数据进行预处理,对经过最佳预处理后的光谱数据采用竞争性自适应重加权算法(CARS)、连续投影算法(SPA)和一次组合降维算法(CARS + SPA)进行特征波长筛选,最后用筛选后的光谱数据建立偏最小二乘回归(PLSR)模型并进行分析比较。研究结果表明,该系统对西瓜横径、纵径、瓜皮厚度的测量值与人工测量值相比准确率最高分别为 98.68%、98.82%、93.81%,均方根误差分别为 2.43、2.08、1.63 mm。对西瓜糖度检测效果最好的模型为(SGS + UVN) - CARS - PLSR,其检测相关系数 R_c 和 R_p 分别为 0.920 4 和 0.912 7,均方根误差 RMSEC 和 RMSEP 分别为 0.376 0°Brix 和 0.466 8°Brix,残差预测偏差 RPD 为 2.490 0。

关键词:西瓜考种;表型数据;糖度可视化;机器视觉;高光谱成像

中图分类号:S237;O433.4　　**文献标识码**:A　　**文章编号**:1000-1298(2025)12-0450-10　　**OSID**:

Watermelon Breeding System Based on Machine Vision and Hyperspectral Imaging

ZHAO Zehua[1,2]　WANG Qingyan[2,3]　CHEN Junjie[2,3]　HUANG Wenqian[2,3]

(1. *Mechanical and Electronic Engineering Institute, Xinjiang Agricultural University, Urumqi* 830052, *China*
2. *Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing* 100097, *China*
3. *National Research Center of Intelligent Equipment for Agriculture, Beijing* 100097, *China*)

Abstract: The watermelon breed testing system combining machine vision and hyperspectral imaging was developed to enhance efficiency, reduce labor, and improve consistency in quality evaluation. The system included an automatic image acquisition device, featuring a Gige camera and a weighing platform, to capture watermelon phenotypic data. It used hyperspectral imaging on selected regions of interest to model SSC distribution within the watermelon, with measurements of transverse and longitudinal diameters achieved through minimum external rectangle fitting and rind thickness estimated using Canny edge detection. Convolutional smoothing (SGS) algorithm was used to preprocess the spectral data by combining three algorithms, namely multivariate scattering correction (MSC), standard normal transform (SNV), and unit vector normalisation (UVN), respectively, and then the best preprocessed spectral data were filtered by competitive adaptive reweighting algorithm (CARS), successive projection algorithm (SPA), and one-time combined dimensionality reduction algorithm (CARS + SPA) for feature wavelength screening, and finally the screened spectral data were used to build PLSR models and

收稿日期:2024-09-01　修回日期:2024-10-28

基金项目:国家重点研发计划项目(2022YFD2002202)和财政部和农业农村部:国家西甜瓜产业技术体系项目(CARS-25-2024-G23)

作者简介:赵泽华(1999—),男,硕士生,主要从事果蔬品质无损检测研究,E-mail:2534316007@qq.com

通信作者:黄文倩(1980—),男,研究员,博士,主要从事农产品质量与安全快速无损检测研究,E-mail:huangwq@nercita.org.cn

analyzed for comparison. The results showed that the system had the highest accuracy of 98.68%, 98.82% and 93.81% for the measured values of transverse diameter, longitudinal diameter and rind thickness of watermelon, respectively, with the root mean square error of 2.43 mm, 2.08 mm and 1.63 mm, respectively, as compared with the manual measurements. The best model for predicting watermelon brix was (SGS + UVN) − CARS − PLSR, with prediction correlation coefficients and of 0.920 4 and 0.912 7, root mean square errors RMSEC and RMSEP of 0.376 0° Brix and 0.466 8° Brix, respectively, and a relative analytical error of prediction RPD of 2.49.

Key words: watermelon breeding; phenotypic data; brix visualization; machine vision; hyperspectral imaging

0　引言

西瓜是我国产量最高的水果之一[1-2]，通过西瓜育种，提高西瓜产量、改善风味和增强抗环境压力能力等问题，已成为许多国家的研究热点[3-4]。考种技术对农作物的改良起着重要作用[5]，在各种粮食作物和果实的育种中得到了广泛的应用。在利用不同技术获取的西瓜表型和内部特征中，横径、纵径、皮厚以及糖度分布是西瓜考种过程中考察的核心参数[6]，这些指标对西瓜品质的评价和育种选择具有重要意义，能够有效指导考种人员进行品种优化。

西瓜的横径和纵径是衡量其大小和形状的关键指标，考种中对西瓜形状有特定要求，理想的形状（如较圆或椭圆）有助于提升市场吸引力。瓜皮厚度则影响果实的运输与贮存性能，较厚的瓜皮可以减少运输中的损坏风险，但过厚的瓜皮会降低果肉的食用比例，因此皮厚是评价果实耐贮运性和商品性的关键因素。代表糖度的可溶性固形物含量（Soluble solid content，SSC）反映了西瓜的甜度，较高的 SSC 通常预示着更好的口感和品质，是评估西瓜品质的重要指标[7-9]。

西瓜的 SSC 测量方法主要包括无损检测和破坏性检测，这两种方法各具优势，为品质评估提供了可靠依据。高光谱成像技术因其信息量大、光谱分辨率高、操作方便等特点，已广泛应用于苹果、猕猴桃、梨和柑橘等果蔬内部品质参数的检测。许丽佳等[10]利用高光谱成像技术对猕猴桃糖度进行无损检测，其最优预测模型的预测集相关系数 R_p 为 0.839 0。李斌等[11]通过高光谱成像技术确定了涌泉蜜桔糖度检测的最优位置为赤道部位，其预测集相关系数 R_p 为 0.955 0。邵园园等[12]采用高光谱成像技术对山东聊城大桃进行可溶性固形物和硬度可视化分析检测，糖度和硬度预测相关系数分别为 0.843 9 和 0.877 2。冯树南等[13]采集枇杷高光谱图像并以整个样本为感兴趣区域建立的最好多元线性回归决定系数 $R_p^2 = 0.822$，并基于 SPA − MLR 模型计算枇杷每个像素点的糖度，生成糖度可视化分布图。

破坏性 SSC 测量方法通常采用折光率法，该方法足够准确和稳定。然而，从西瓜果肉中榨取西瓜汁进行 SSC 测量是一个耗时费力的过程，在处理大量样本时，由于工作量大，容易出现人工测量误差。文献[14]设计了一种西瓜育种中 SSC 自动测量系统，该系统包括自动切割和 SSC 测量两部分；通过一个三轴机械臂驱动折光仪到达西瓜切面中心点位置，后将折光仪插入西瓜果肉中获取西瓜 SSC。该系统只对西瓜切面的中心糖度做出测量，应用于西瓜考种工作中不具有代表性，且折光仪在对上一个西瓜 SSC 测量时传感器探头上会保留西瓜的汁液，再次测量会出现误差。考虑到西瓜考种工作需要将西瓜切开以后测量西瓜的表型特征和西瓜内部果肉品质，因此本文为提高西瓜考种效率、降低育种成本，设计一种基于机器视觉和高光谱成像的西瓜考种系统，以实现对西瓜表型特征的测量和切面各个位置 SSC 的可视化检测。

1　材料与方法

1.1　试验材料

本研究使用的西瓜为小型早熟吊秧西瓜 L600，西瓜样本均产于北京市大兴区庞各庄。共计采摘 30 个成熟的西瓜，采摘当天运送到实验室分别编号。在 25℃下，静置 24 h，使西瓜内部温度保持一致，以减小温度差异对高光谱采集和糖度测定的影响。每个西瓜取 10 个不同的 ROI 作为独立的区域样本，共计 300 个区域样本。

1.2　试验设备

西瓜视觉信息采集装置由 4 部分组成，主要包括 1 台分辨率为 1 920 像素 ×1 200 像素的 GigE 相机、1 台焦距为 16 mm 的镜头、1 个直径为 500 mm 亮度可调节的圆顶光源、1 个称量平台，以上组件放置在暗室光箱中，如图 1 所示。为满足对不同尺寸和品种的西瓜都能够获取到高质量的图像，将相机安装在距西瓜切面 550 ~ 600 mm 正上方的可调支

架上,用于获取西瓜的整果图像和切面图像。

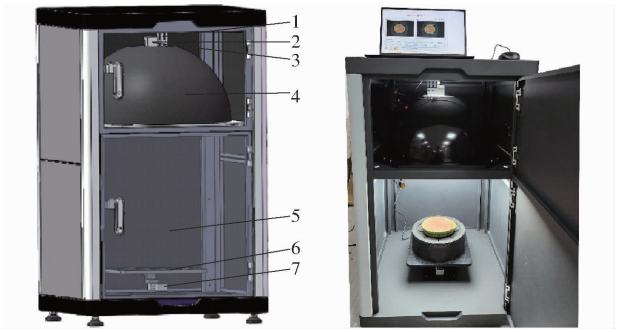


图 1 西瓜视觉信息采集装置设计图和实物图

Fig. 1 Watermelon visual information acquisition device

1. 相机支架 2. GigE 相机 3. 可调镜头 4. 圆顶光源 5. 光箱 6. 称量平台 7. 称量传感器

高光谱信息采集系统由 6 部分组成,主要包括 1 台分辨率为 1 004 像素 × 1 000 像素的面阵 EMCCD 相机,1 个焦距为 23 mm、光圈为 F/2.4 的镜头,1 台成像光谱仪,2 个 150 W 输出功率可调卤素灯源,1 套电动传送装置,以上组件放置在暗室中,如图 2 所示。暗室外连接了 1 台包含光谱采集软件的计算机。为保证能够采集到西瓜整个切面的高光谱信息,设置面阵相机距西瓜切面的高度为 550 mm。

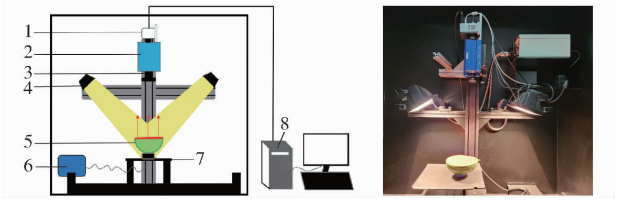


图 2 高光谱信息采集系统原理图和实物图

Fig. 2 Hyperspectral information acquisition system

1. EMCCD 相机 2. 成像光谱仪 3. 镜头 4. 面光源 5. 西瓜样本 6. 电机 7. 载物台 8. 计算机

本研究基于 Solidworks、LabVIEW 和 OpenCV 设计开发了视觉信息采集装置和西瓜自动化考种软件。通过 LabVIEW 软件开发的西瓜自动化考种系统,调用图像采集与图像处理算法,以获取西瓜的整果图像和切面图像,并基于采集的切面图像计算西瓜的横径、纵径和瓜皮厚度。通过 HyperSpectral Image 软件控制光谱采集装置载物平台的移动速度、相机曝光时间等参数来实现西瓜切面高光谱信息的采集。对采集到的高光谱信息在经过黑白校正后,通过 Matlab 2021a 软件进行西瓜多点位的糖度建模。

1.3 感兴趣区域的选取与数据采集

1.3.1 感兴趣区域选取

每个瓜取 10 个半径为 80 像素的圆形区域作为 ROI。将西瓜沿瓜蒂至瓜脐方向从中间位置切开,分为左右两半。在左半部分,取瓜蒂处、中心处和瓜脐处分别为 1、2 和 3 号 ROI,上、下边缘处分别为 4 号和 5 号 ROI。在右半部分,取 1 号和 4 号 ROI 中间位置处为 6 号 ROI,取 1 号和 5 号 ROI 中间位置处为 7 号 ROI,取 2 号和 7 号 ROI 中间位置处为 8 号 ROI,取 3 号和 4 号 ROI 中间位置处为 9 号 ROI,取 3 号和 5 号中间位置处为 10 号 ROI,如图 3 所示。设置 8 号 ROI 作为边糖与中心糖之间的过渡区域,利用这 10 个 ROI 可以全面覆盖西瓜切面不同位置的糖度分布。

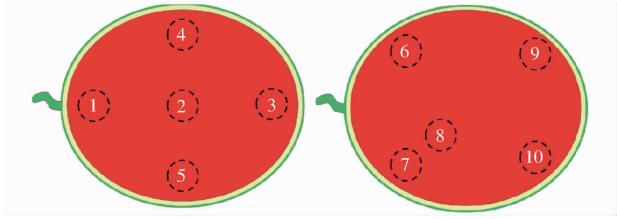


图 3 感兴趣区域选取

Fig. 3 Selection of region of interest

1.3.2 西瓜表型和糖度数据采集

在切面图像采集完成后,人工使用数显游标卡尺 3 次测量西瓜的横径和纵径,取平均值作为真实值(精度 0.01 mm)。用直尺测量西瓜的皮厚,取 4 个位置测量的平均值作为真实值(精度 0.1 mm)。

在高光谱信息采集完成后,开始对西瓜进行糖度测量。使用挖球勺挖取对应 ROI 内的西瓜果肉,使用手动榨汁机将果汁挤压到对应标签的干净烧杯中。用玻璃棒搅拌均匀后,胶头滴管吸取果汁滴在 PAL-1 型数字折光仪(ATAGO 公司,日本,分辨率 0.1%,精度 ±0.2%)上测量糖度。为提高糖度测量值的精确性,每次测量完先使用蒸馏水进行清洗,再用待测样品冲洗数字折光仪,取测量 3 次的平均值作为西瓜每个部位糖度的最终值。

2 西瓜表型特征测量与内部糖度建模方法

2.1 西瓜表型特征测量

2.1.1 西瓜图像预处理

对采集到的西瓜切面图像进行预处理,图像预处理可以优化图像数据,去除和过滤掉图像中无用的图像噪声和背景特征。使用 OpenCV 的 Python 接口调用 OpenCV 函数库,对采集的切面图像进行灰度处理来减少和消除图像的噪声信息以增强图像特征[15-16]。应用 OpenCV 函数库中的均值滤波、中值滤波、高斯滤波和双边滤波对图像进行降噪。通过对比 4 种滤波处理方式和降噪结果,如图 4 所示,得到最优降噪方式为高斯滤波。

对降噪后的图像分别进行全局阈值分割和迭代法阈值分割。根据图像的全局阈值直方图,如图 5a

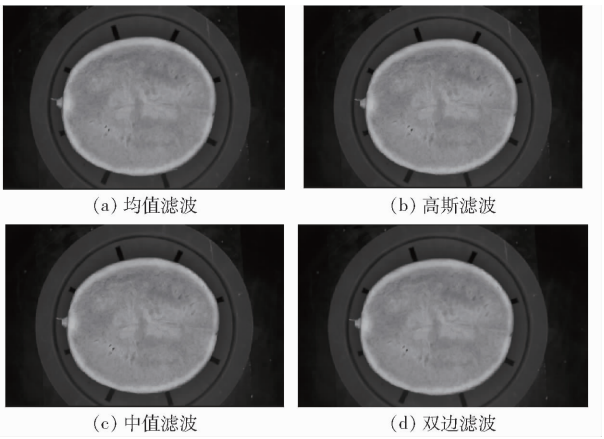


图 4　不同滤波效果对比图

Fig. 4　Comparison of different filtering effects

所示，得到双峰的中间阈值为 90，二值化结果如图 5b 所示。通过迭代法阈值分割，得到最佳阈值为 85，二值化图像结果如图 5c 所示。对比两种阈值分割算法的过程和结果[17]，采用迭代法阈值分割进行二值化操作。

西瓜切面图像经过降噪与二值化处理后，图像中仍存在一部分噪声信息，运用图像形态学开运算对图像进行二次降噪[18]，完全将西瓜切面的前景分离出来。经运算，结构元 K 尺寸设为 4 ×4 时进行两次开运算效果最好，如图 5d 所示。

2.1.2　西瓜切面轮廓提取与横纵径计算

经过图像预处理后，图像中的噪声信息已经被消除，将上方经过形态学处理后的二值图像作为

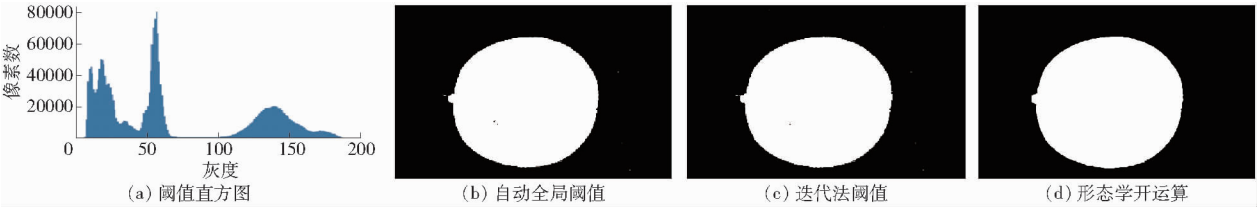


图 5　二值化结果对比图

Fig. 5　Comparison of binarization results

mask 掩膜图像，对原始图像掩膜操作，实现西瓜切面轮廓的提取，如图 6a 所示。为了验证提取效果，用提取出来的切面图像与原始切面图像进行加权拟合，权重 α 设置为 0.6，拟合结果如图 6b 所示。

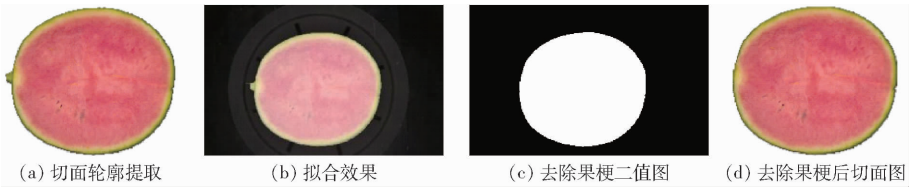


图 6　切面轮廓提取

Fig. 6　Cross-sectional contour extraction

在计算西瓜横纵径前，需要去除西瓜的果梗。对多幅西瓜图像果梗区域进行标记，得到在果梗与西瓜表皮接触处果梗的列像素值达到最大值。剔除二值图像中果梗区域列像素和小于最大值的列，得到去除果梗后的西瓜二值化图像，如图 6c 所示。将去除果梗后的二值图像作为 mask 掩膜图像，提取的西瓜切面图像作为目标图像，进行掩膜得到去除果梗后的西瓜切面图像，如图 6d 所示。

因西瓜的形状是一个类椭圆体的不规则球体，本研究为提高横纵径测量的准确性，选择西瓜横径和纵径的测量最大值作为测量结果。采用最小外接矩形法来计算西瓜的横纵径[19]，通过 cv2. minAreaRect 函数找到去除果梗后西瓜二值图像的最小外接矩形，如图 7 所示，再调用 cv2. boxPoints 函数得到返回矩阵 4 个角点的坐标 $(x,y+h)$、$(x+w,y+h)$、$(x+w,y)$、(x,y)。计算最小外接矩形的长 w 和宽 h 并除以像素精度作为西瓜横径和纵径的真实值。

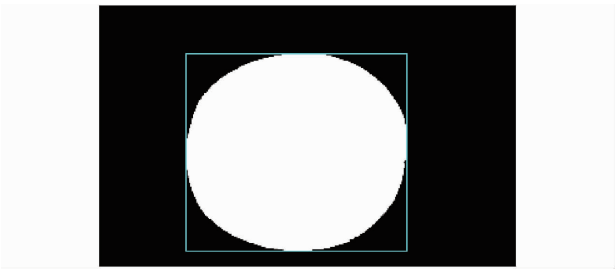


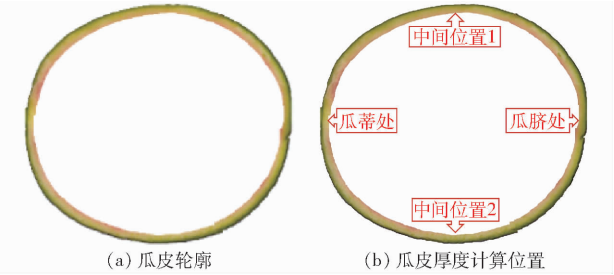
图 7　最小外接矩形

Fig. 7　Minimum bounding rectangle

2.1.3　瓜皮轮廓提取与尺寸计算

本研究应用 Canny 边缘检测算法实现对瓜皮轮廓的提取[20-22]。应用 cv2. split 函数对去除果梗后的西瓜图像做单通道数组分离，得到 3 幅分别为 B、G、R 单通道数组的灰度图像。因为瓜皮与果肉连接处存在着多个 R 和 G 混合的单通道数组，通过 Canny 边缘检测算法中的高斯滤波对分离后单通道

数组 R 的灰度图像做平滑处理，采用 Sobel 滤波器来计算梯度和方向。利用非极大值抑制技术细化边缘，并保留边缘方向上具有极大值的像素点，利用双阈值的技术判定可能潜在的边界，突出显示瓜皮轮廓边界。用提取出的瓜皮轮廓二值图像对原始图像进行掩膜操作，从而得到瓜皮轮廓，如图 8a 所示。



(a) 瓜皮轮廓 (b) 瓜皮厚度计算位置

图 8 瓜皮轮廓提取

Fig. 8 Rind contour extraction

计算瓜皮轮廓中间行和中间列的轮廓像素个数，得到西瓜瓜蒂、瓜脐和两个中间位置的瓜皮厚度，如图 8b 所示，取 4 个位置厚度的平均值作为瓜皮的真实厚度。

2.2 西瓜内部糖度建模

2.2.1 高光谱图像采集与校正

在采集西瓜高光谱图像前，设置相机曝光时间 10 ms，平台移动速度 1 mm/s，西瓜切面距离镜头高度约 550 mm，并对光谱仪预热 30 min。将西瓜样本放置于移动平台上，通过线扫描方式获取样本高光谱图像。在采集过程中，因光源在各波段下强度分布的差异和相机暗电流噪声的影响，会夹杂部分噪声信息[23]。为保证高光谱图像定性或定量分析模型的精度和稳定性，有必要对高光谱图像进行黑白校正，以消除噪声影响。

2.2.2 光谱预处理

由于外界环境的干扰，导致在采集原始光谱信息时会出现随机噪声和表面散射等现象，从而使提取的光谱数据中存在许多干扰信息，对模型的精度和稳定性有显著影响[24]。光谱预处理能够消除这些不利影响，从而提高其信噪比。本研究采用 SGS，分别结合 MSC、SNV、UVN 共 3 种预处理方式，针对西瓜切面 ROI 的光谱信息挑选出合适的预处理方式。

全光谱波长数据量很大，存在着大量的无信息波长，导致数据处理速度缓慢。通过算法提取出最具代表性的波长，以构建更具稳定性和鲁棒性的糖度含量模型，并且能够简化建模过程[25-26]。本研究分别采用一次 CARS 和 SPA、一次 CARS + SPA 进行特征波长筛选。

2.2.3 回归预测模型的建立及性能评价

PLSR 可以同时分解光谱矩阵 $\boldsymbol{X}$ 和浓度矩阵 $\boldsymbol{Y}$，并建立和强化两者之间的关系。使用 PLSR 建立模型时，通过选择潜在变量(LVs)的最佳数量，来避免出现过拟合和欠拟合现象。本研究采用交叉验证方法，依据残差平方和(PRESS)达到最小值时对应的 LVs 为最佳潜在变量数。

本文主要通过相关系数 R、均方根误差(RMSE)和残差预测偏差(RPD)来评价预测模型性能[27]。

3 结果与分析

3.1 西瓜表型测量结果与分析

本文通过测量结果准确率来评价系统的检测性能，将西瓜样本分为 5 组，表 1 为 5 组西瓜 3 项表型特征测量结果的准确率。从表 1 可以看出系统对西瓜横径和纵径测量值的准确率最高分别为 98.68% 和 98.82%，对皮厚测量值的准确率最高为 93.81%。

表 1 西瓜表型测量准确率

Tab. 1 Accuracy of phenotypic measurement in watermelon %

表型特征参数	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组
横径测量值	98.04	97.95	98.59	98.68	98.47
纵径测量值	98.35	98.63	98.82	98.78	97.94
皮厚测量值	92.80	92.24	93.81	93.39	93.67

图 9 为 3 项表型特征系统测量结果与人工测量结果拟合图。3 项测量结果与真实值相比决定系数 R^2 分别为 0.94、0.93、0.90，均方根误差分别为 2.43、2.08、1.63 mm。

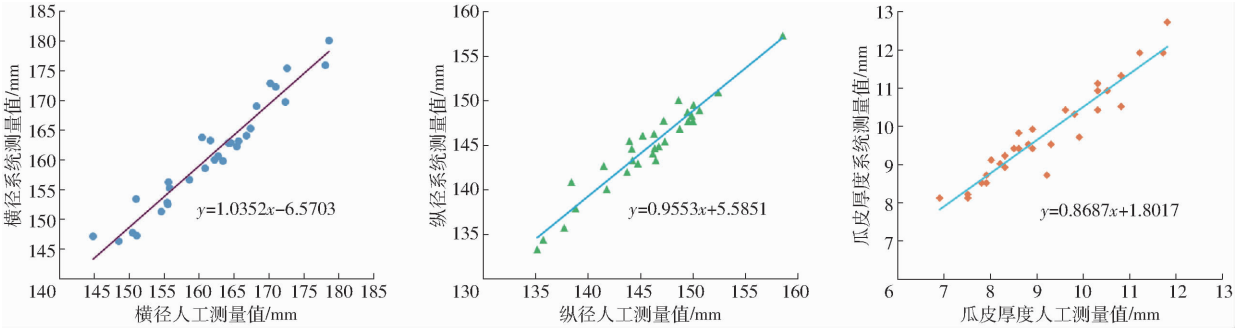


图 9 表型特征系统测量值与人工测量值线性拟合分析

Fig. 9 Linear fitting analysis of phenotypic feature system measurements and manual measurements

表型特征测量产生误差原因如下：在采集视觉信息的过程中，人工放置西瓜切面时存在不同的倾斜角度，导致横纵径和瓜皮厚度测量出现误差。此外，由于西瓜表皮含水率高，表皮较脆，在切割时容易造成瓜皮边缘崩碎使得切面不完整，从而进一步导致测量误差。

3.2　糖度模型结果与分析

3.2.1　样本数据集划分

使用 SPXY 算法将所有 300 个区域样本按照比例 3∶1划分校正集和预测集[28]，其中校正集 225 个用于建立预测模型，预测集共 75 个用于评价模型性能。西瓜切面区域样本糖度为 8.2 ~ 13.3°Brix，平均值为 11.0°Brix，标准偏差为 0.97°Brix，样本集覆盖了足够大的糖度范围，能够反映出不同感兴趣区域的糖度差异，有利于提高所建模型的适用性。

3.2.2　光谱特征分析及预处理结果

在原始光谱采集过程中，由于仪器设备误差和样本形状不规则等内外因素的影响，会携带一些与样本化学性质无关的成分。这会导致所采集的光谱中出现光谱平移、重叠等噪声，特别是光谱曲线在首尾两端波动较大。因此，本文选取 400 ~ 1 000 nm 范围内的光谱作为有效光谱。图 10a 为西瓜切面 ROI 的原始平均反射光谱。在波长 705 nm 附近观察到有一个明显的光谱曲线吸收峰，吸收峰形成原因是西瓜果肉中含有多种有机化合物和大量水分，为不同物质官能团的特征谱带重叠而形成。其中，C—H 基团的 3 级倍频伸缩振动出现在 670 ~ 720 nm，O—H 基团的 2 级倍频伸缩振动出现在 730 ~ 760 nm，因此这个吸收峰可能由多种化合物的官能团形成。

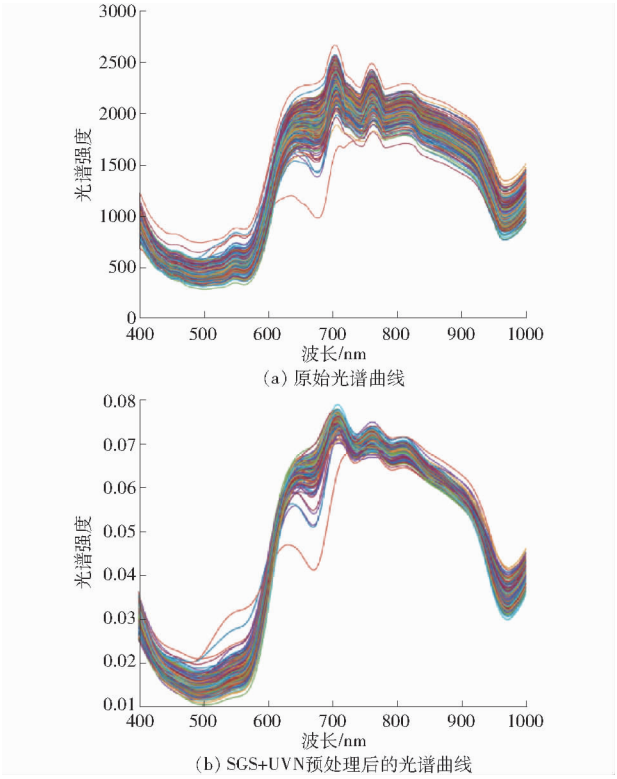


图 10　切面感兴趣区域的原始光谱曲线和预处理后的光谱曲线

Fig. 10　Original spectral curves and preprocessed spectral curves of region of interest in cross-section

对经过不同预处理后的西瓜光谱数据进行 PLSR 建模分析（建模结果如表 2 所示），发现使用 SGS（拟合二次多项式，窗口宽度为 7）结合 UVN 预处理的效果较好，其校正集相关系数 R_c 和均方根误差分别为 0.897 5 和 0.424 3°Brix，预测集相关系数 R_p 和均方根误差分别为 0.854 4 和 0.531 6°Brix，RPD 为 1.887 5，预处理后的平均反射光谱如图 10b 所示。该预处理后的预测效果优于 SGS 分别结合 MSC、SNV，故后续均基于 SGS 结合 UVN 预处理方法进行。

表 2　不同预处理西瓜切面糖度预测 PLSR 模型结果

Tab. 2　PLSR model results for predicting sugar content in watermelon cross-sections under different treatments

预处理方法	LVs	校正集（n = 225）		预测集（n = 75）		
		R_c	RMSEC/°Brix	R_p	RMSEP/°Brix	RPD
RAW	8	0.919 0	0.380 5	0.769 4	0.699 0	1.446 0
SGS + MSC	8	0.857 3	0.495 6	0.821 5	0.572 3	1.760 9
SGS + SNV	7	0.901 6	0.396 0	0.840 1	0.636 3	1.832 3
SGS + UVN	6	0.897 5	0.424 3	0.854 4	0.531 6	1.887 5

3.2.3　特征光谱变量提取

高光谱数据量大，且存在一些冗余信息波长之间的相互干扰，本文分别采用一次 CARS 和 SPA 降维以及一次 CARS + SPA 组合降维从全波段光谱中提取特征光谱，实现光谱数据的降维，以达到提高模型运算效率和预测精度的目的。

（1）CARS 特征光谱变量提取

对预处理后的光谱用 CARS 算法提取特征光谱变量时，设定蒙托卡罗采样次数为 100，采用 5 折交叉验证法。图 11 为特征波长的选择过程，图 11a 呈现出指数衰减函数的选择过程，特征光谱变量的数量随着采样次数的增加先迅速下降然后平缓减少，具有粗选和精选两个特征。由图 11b 可知，随着蒙托卡罗采样次数的增加，内部交叉验证均方根误差（RMSECV）呈现出缓慢减小后陡然增大的趋势，当采样次数为 21 次时，RMSECV 达到最低值，此时的变量子集为 CARS 筛选的最优波段。图 11c 是特征光谱变量随着采样次数变化的回归系数路径图，当 RMSECV 达到最小值时，各特征光谱变量的回归系

数位于图 11c 中的“ * ”所在的垂直线位置,此时采样次数为 21 次,提取出 45 个特征光谱变量,占全波段的 11.6% 。

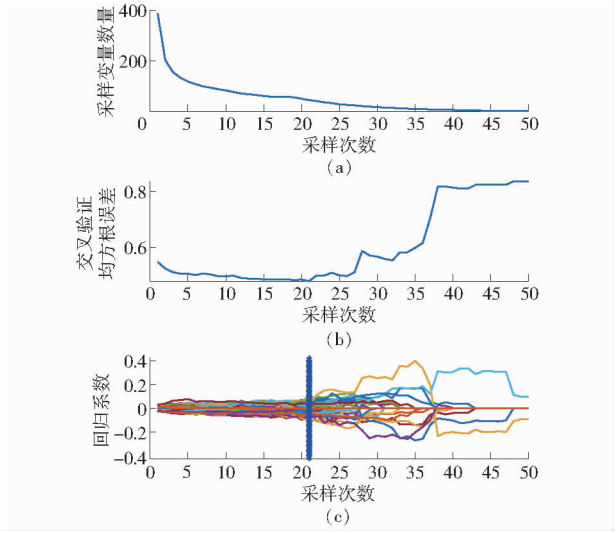


图 11 CARS 特征波长筛选过程

Fig. 11 CARS feature wavelength selection process

(2)SPA 特征光谱变量提取

利用 SPA 算法选取特征光谱变量,为了确定最佳的特征光谱数量,通常会计算不同数量特征光谱变量下的校正均方根误差(RMSEC)。一般情况下,选择 RMSEC 最小时对应的特征光谱数量作为最佳选择。设置特征光谱变量数的变化范围为 1 ~ 50,当 RMSEC 达到最小值 0.587 01 时,此时提取了 27 个特征光谱变量,占全波段的 6.8% ,特征光谱变量在全波段中的分布如图 12 所示。

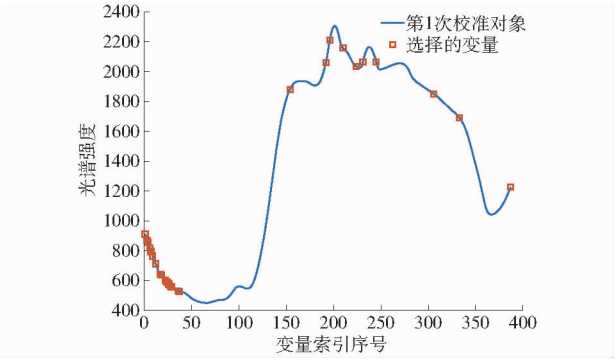


图 12 SPA 特征光谱变量分布图

Fig. 12 SPA feature spectral variable distribution diagram

(3)CARS + SPA 组合降维

考虑到经过 CARS 算法处理后提取出的 45 个特征光谱个数较多,且相邻波段间可能存在一定的共线性问题,因此通过 SPA 算法进行组合降维以消除可能存在的共线性问题。设置特征光谱数量变化范围为 1 ~ 30,SPA 算法再次降维提取出 21 个特征光谱变量,如图 13a 所示;图 13b 表示所选的 21 个特征变量在原始有效光谱中对应的波长位置,占全光谱波段的 5.4% 。

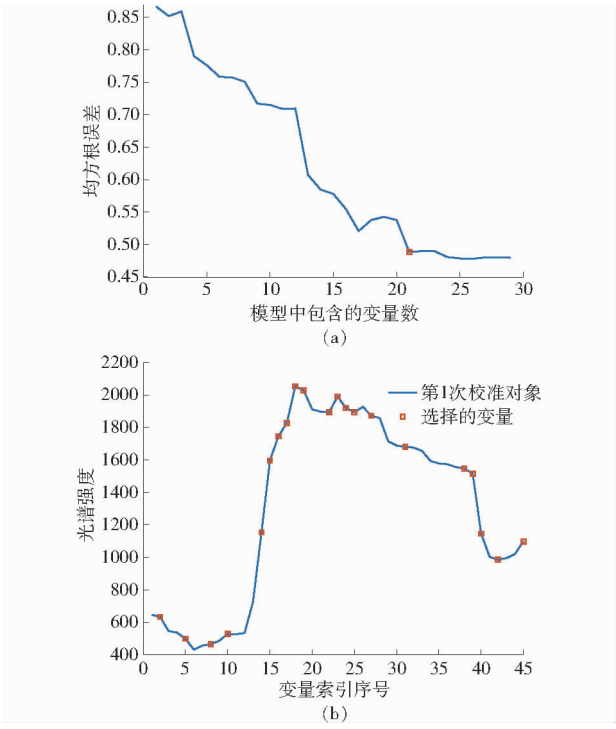


图 13 CARS + SPA 特征光谱变量分布图

Fig. 13 CARS + SPA feature spectral variable distribution diagram

3.2.4 预测模型的结果与分析

根据特征波长筛选结果,建立西瓜切面的 PLSR 糖度预测模型,预测结果如表 3 所示。由表 3 可知,经过 CARS 算法筛选特征波长后模型的校正集和预测集效果都得到了显著提升,这说明 CARS 算法剔除了光谱数据中的干扰信息;经过 SPA 算法处理后,模型校正集的相关系数和 RMSEC 效果均变差,预测集的相关系数、预测均方根误差(RMSEP)以及

表 3 不同特征提取方法建立的 PLSR 模型预测结果

Tab. 3 PLSR model prediction results by using different feature extraction methods

特征筛选算法	变量数	LVs	校正集(n = 225)		预测集(n = 75)		
			R_c	RMSEC/°Brix	R_p	RMSEP/°Brix	RPD
CARS	45	8	0.920 4	0.376 0	0.917 2	0.466 8	2.490 0
SPA	27	8	0.885 2	0.466 8	0.868 8	0.504 6	1.988 4
CARS + SPA	21	8	0.863 6	0.484 8	0.897 7	0.442 6	2.267 0

RPD 的提升效果不显著。这表明，SPA 算法在处理过程中剔除了西瓜切面 ROI 光谱数据中部分与糖度相关的重要信息。在 CARS 算法的基础上进行一次 SPA 组合降维后，模型校正集效果明显下降，预测集的效果和 RPD 有所提升，但不及单独使用 CARS 算法的效果显著。在 CARS 挑选出的波长中，多重共线性已经较低，进一步使用 SPA 进行降维不一定能提升效果，反而会剔除一些对模型拟合有帮助的细节信息。

因此，综合分析确定西瓜切面糖度最佳预测方法为(SGS + UVN) - CARS - PLSR，其 R_c 和 R_p 分别为 0.920 4 和 0.917 2，RMSEC 和 RMSEP 分别为 0.376 0° Brix 和 0.466 8° Brix，RPD 为 2.490 0。图 14 为 PLSR模型对西瓜切面 SSC 的预测值与真实值之间的散点图。

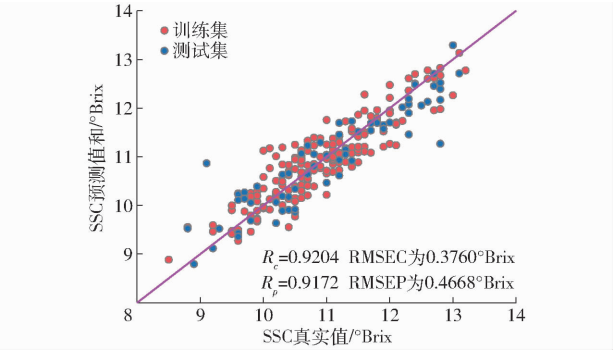


图 14　西瓜糖度含量 PLSR 模型预测散点图

Fig. 14　Scatter plot of PLSR model predictions for watermelon sugar content

3.2.5　西瓜切面糖度分布可视化预测

高光谱成像技术的显著特点在于其能够获取测试样本中每个像素点的详细光谱信息。本文利用这些光谱信息生成糖度分布的可视化预测图，从而使得西瓜样本中 SSC 在不同区域的分布情况得以直观地呈现。

通过提取西瓜切面样本中每个像素点的光谱信息，用已建立的西瓜糖度预测模型可以有效地预测各像素点的糖度，以像素点空间位置及对应糖度绘制了糖度分布伪彩图，如图 15 所示，可以清晰直观地看到西瓜切面各个位置的糖度分布情况。分布图上西瓜的背景为黑色，糖度显示为零；颜色越红表示糖度越高。西瓜是一个复杂的生物体，具有丰富的物质组成成分，无论是径向还是轴向均呈不均匀分布，从图中可以观察到西瓜果肉糖度从边缘到中心位置逐渐增大，这也符合文献[29]对西瓜的糖度分布的研究。图中瓜皮糖度高的原因是因为瓜皮与果肉的物质组成成分存在差异，并且在切瓜时西瓜果肉的汁液会随着切刀流向瓜皮边缘，造成瓜皮 SSC 偏高。

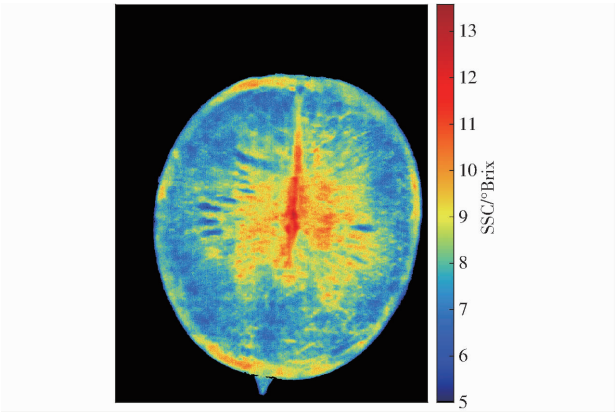


图 15　西瓜糖度分布可视化

Fig. 15　Visualization of watermelon sugar content distribution

4　结论

(1)通过最小外接矩形法计算西瓜的横径、纵径，得到 5 组西瓜测量结果的最高准确率分别为 98.68%、98.82%，测量结果与人工测量值的决定系数 R^2 分别为 0.94、0.93，均方根误差分别为 2.43、2.08 mm；使用 Canny 边缘检测算法测得瓜皮厚度的最高准确率为 93.81%，决定系数 R^2 为 0.90，均方根误差为 1.63 mm。

(2)对西瓜切面原始光谱采用 SGS 分别结合 UVN、MSC、SNV 进行预处理，通过建模结果显示 SGS(拟合二次多项式，窗口宽度为 7)结合 UVN 预处理效果最好。对预处理后的光谱分别进行一次 CARS、SPA 降维和一次 CARS + SPA 组合降维，提取到的特征变量数分别为 45、27 和 21，占全光谱波段的 11.6%、6.8% 和 5.4%。

(3)基于 3 种特征提取方法，分别建立 PLSR 预测模型。综合对比后使用 CARS - PLSR 模型的预测性能最好，其相关系数 R_c 和 R_p 分别为 0.920 4 和 0.912 7，均方根误差 RMSEC 和 RMSEP 分别为 0.376 0° Brix 和 0.466 8° Brix，预测相对分析误差 RPD 为 2.490 0。将西瓜样本各个像素点的光谱信息输入所建预测模型中，得到西瓜糖度分布的可视化图，可直观地观察到西瓜各个位置的糖度分布。

参　考　文　献

[1]　小西瓜日益受欢迎[N]. 经济日报，2024 - 06 - 17(8).

[2]　李干琼，王志丹. 我国西瓜产业发展现状及趋势分析[J]. 中国瓜菜，2019，32(12)：79 - 83.

[3] 刘文革."十三五"我国西瓜遗传育种研究进展[J].中国瓜菜,2021,34(12):1-9.
LIU Wen'ge. Research progress on watermelon genetic breeding during "the 13th Five-Year Plan" in China[J]. China Cucurbits and Vegetables, 2021,34(12):1-9. (in Chinese)
[4] 苏利军. 常见西瓜品种性状、产量结构对比试验及种植效益分析[J].现代农业, 2022(3): 42-45.
[5] 张伟进,王福顺,孙小华,等.传统图像分割算法在农作物籽粒考种应用中的研究进展[J].中国农机化学报,2024,45(2):280-287.
ZHANG Weijin,WANG Fushun,SUN Xiaohua, et al. Research progress of traditional image segmentation algorithm in seed testing of crops[J]. Journal of Chinese Agricultural Mechanization, 2024,45(2):280-287. (in Chinese)
[6] 梁朴,田淑霞,杨秀秀,等. 宁夏麒麟瓜品质检测分析与评价[J].园艺与种苗,2024, 44(1): 40-44.
LIANG Pu, TIAN Shuxia, YANG Xiuxiu, et al. Analysis and evaluation of quality of Ningxia Kylin melon[J]. Horticulture & Seed, 2024, 44(1): 40-44. (in Chinese)
[7] 严忠伟,田喜,张艺飞,等.基于全透射近红外光谱的西瓜不同部位可溶性固形物含量在线检测研究[J].光谱学与光谱分析,2023,43(6):1800-1808.
YAN Zhongwei, TIAN Xi, ZHANG Yifei, et al. Online detection of soluble solids content in different parts of watermelons based on full transmission near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2023,43(6):1800-1808. (in Chinese)
[8] LI X, LI C, SUN J, et al. Dynamic changes of enzymes involved in sugar and organic cid level modification during blueberry fruit maturation[J]. Food Chemistry, 2020, 309: 125617.
[9] JAWAD U, GAO L, GEBREMESKEL H, et al. Expression pattern of sugars and organic acids regulatory genes during watermelon fruit development[J]. Scientia Horticulturae, 2020, 265: 109102.
[10] 许丽佳,陈铭,王玉超,等.高光谱成像的猕猴桃糖度无损检测方法[J].光谱学与光谱分析,2021,41(7):2188-2195.
XU Lijia, CHEN Ming, WANG Yuchao, et al. Study on non-destructive detection method of kiwifruit sugar content based on hyperspectral imaging technology[J]. Spectroscopy and Spectral Analysis, 2021,41(7):2188-2195. (in Chinese)
[11] 李斌,万霞,刘爱伦,等.基于高光谱成像技术的涌泉蜜桔糖度最优检测位置[J].中国光学(中英文),2024,17(1):128-139.
LI Bin, WAN Xia, LIU Ailun, et al. Optimal position for suger content detection of Yongquan honey oranges based on hyperspectral imaging technology[J]. Chinese Optics, 2024,17(1):128-139. (in Chinese)
[12] 邵园园,王永贤,玄冠涛,等.基于高光谱成像的肥城桃品质可视化分析与成熟度检测[J].农业机械学报,2020,51(8):344-350.
SHAO Yuanyuan, WANG Yongxian, XUAN Guantao, et al. Visual detection of SSC and firmness and maturity prediction for feicheng peach by using hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8):344-350. (in Chinese)
[13] 冯树南,谭涛,尚静,等.枇杷糖度无损检测及可视化研究[J].中国农机化学报,2024,45(2):157-163.
FENG Shu'nan, TAN Tao, SHANG Jing, et al. Non-destructive detection and visualization of sugar content in loquat[J]. Journal of Chinese Agricultural Mechanization, 2024,45(2):157-163. (in Chinese)
[14] HUANG J, ZOU T, HU H, et al. Automatic Brix measurement for watermelon breeding[J]. Applied Sciences,2022,12(23): 12227.
[15] 刘星星,王烁烁,徐丽明,等. 基于 OpenCV 的动态葡萄干色泽实时识别[J].农业工程学报, 2019, 35(23): 177-184.
LIU Xingxing, WANG Shuoshuo, XU Liming, et al. Real time color recognition of moving raisin based on OpenCV[J]. Transactions of the CSAE, 2019, 35(23): 177-184. (in Chinese)
[16] 邹宏伟. 基于 OpenCV 的数字图像处理技术研究与实现[J].无线互联科技,2019, 16(22): 118-119.
ZOU Hongwei. Research and implementation of digital image processing technology based on OpenCV[J]. Wireless Internet Science and Technology, 2019, 16(22): 118-119. (in Chinese)
[17] 汤朋文, 陶华敏, 肖山竹, 等. 几种常用图像分割算法自适应性的分析比较[J]. 数字技术与应用, 2016(5): 140-143.
[18] 刘慧力,贾洪雷,王刚,等. 基于深度学习与图像处理的玉米茎秆识别方法与试验[J].农业机械学报, 2020, 51(4): 207-215.
LIU Huili,JIA Honglei, WANG Gang, et al. Method and experiment of maize(*Zea Mays* L.) stems recognition based on deep learning and image processing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 207-215. (in Chinese)
[19] 杨长辉,王卓,熊龙烨,等. 基于 Mask R-CNN 的复杂背景下柑橘树枝干识别与重建[J].农业机械学报, 2019,50(8): 22-30,69.
YANG Changhui, WANG Zhuo, XIONG Longye, et al. Identification and reconstruction of citrus branches under complex background based on Mask R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 22-30,69. (in Chinese)
[20] 陈利利,李斌,梁晓晴. 基于 Canny 边缘检测的红枣识别算法研究[J].信息技术与信息化, 2024(3): 25-28.

[21] 于新善,孟祥印,金腾飞,等. 基于改进 Canny 算法的物体边缘检测算法[J]. 激光与光电子学进展, 2023,60(22): 221 – 230.
YU Xinshan, MENG Xiangyin, JIN Tengfei, et al. Object edge detection algorithm based on improved Canny algorithm[J]. Laser & Optoelectronics Progress, 2023, 60(22): 221 – 230. (in Chinese)

[22] 刘浩,任宏,赵丁选,等. 基于亚像素定位的图像边缘检测策略研究[J]. 农业机械学报, 2024, 55(2): 242 – 248,294.
LIU Hao, REN Hong, ZHAO Dingxuan, et al. Image edge detection strategy based on sub-pixel location[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 242 – 248,294. (in Chinese)

[23] 郭志明, 赵春江, 黄文倩, 等. 苹果糖度高光谱图像可视化预测的光强度校正方法[J]. 农业机械学报, 2015, 46(7): 227 – 232.
GUO Zhiming, ZHAO Chunjiang, HUANG Wenqian, et al. Intensity correction of visualized prediction for sugar content in apple using hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(7): 227 – 232. (in Chinese)

[24] ZHANG D ,XU L ,WANG Q , et al. The optimal local model selection for robust and fast evaluation of soluble solid content in melon with thick peel and large size by vis-nir spectroscopy[J]. Food Analytical Methods, 2019, 12(1): 136 – 147.

[25] 孙博康, 刘贵珊. 基于高光谱技术检测香水梨硬度的研究[J]. 食品安全导刊, 2021(19): 118 – 120.

[26] ZHANG F, LI B, YIN H, et al. Study on the quantitative assessment of impact damage of yellow peaches using the combined hyperspectral technology and mechanical parameters[J]. Journal of Spectroscopy, 2022, 2022: 7526826.

[27] 廉小亲,汤燊淼,吴静珠,等. 基于近红外的兰州百合品质定量建模方法研究[J].食品科技, 2020, 45(7): 298 – 302.
LIAN Xiaoqin, TANG Shenmiao,WU Jingzhu, et al. Research on quantitative model of lilium Lanzhou quality by near infrared spectroscopy[J]. Food Science and Technology, 2020, 45(7): 298 – 302. (in Chinese)

[28] 赵娟,沈懋生,浦育歌,等. 基于近红外光谱与多品质指标的苹果出库评价模型研究[J]. 农业机械学报, 2023, 54(2): 386 – 395.
ZHAO Juan, SHEN Maosheng, PU Yuge, et al. Out-of-warehouse evaluation and prediction model of apple based on near-infrared spectroscopy combined with multiple quality indexes[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 386 – 395. (in Chinese)

[29] 朱倩倩, 管劼昊, 李旷开, 等. 小型西瓜果实内部糖度分布及其影响因素研究[J]. 食品安全质量检测学报, 2012, 3(6): 661 – 666.
ZHU Qianqian, GUAN Jiehao, LI Kuangkai, et al. Internal Brix distribution and its influencing factors of small watermelon [J]. Journal of Food Safety & Quality, 2012, 3(6): 661 – 666. (in Chinese)