

doi:10.6041/j.issn.1000-1298.2021.09.027

日光温室基质培生菜鲜质量无损估算方法

刘林^{1,2} 苑进^{1,2} 张岩¹ 刘雪美^{1,2}

(1. 山东农业大学机械与电子工程学院, 泰安 271018; 2. 山东省农业装备智能化工程实验室, 泰安 271018)

摘要: 设施栽培中作物鲜质量动态变化作为生长发育的指示性特征,是蔬菜长势无损监测的重要指标之一。水培蔬菜通过离水直接称量实现长势无损监测,但是土培或基质培无法通过直接鲜质量称量实现生长过程的无损测量。本文提出了基于表型特征参数和生长过程环境参数融合的鲜质量估算方法,用于日光温室环境下基质培生菜个体和群体的鲜质量无损估算。首先,监测生菜全生命周期的环境参数,采集第1批次生菜生长过程中的多样本图像和部分样本鲜质量,提取样本图像中不同生长期生菜的形状、颜色、纹理等特征,计算环境信息中的累积辐热积等参数。然后,利用高斯过程回归方法建立表型参数和环境参数与生菜鲜质量的关系模型。最后,采集第2批次生菜群体的样本数据,基于上述模型预测生菜3个生长阶段的个体和群体鲜质量,以验证鲜质量估算模型的泛化能力和可靠性。结果表明,与支持向量机、线性回归、岭回归和神经网络相比,高斯过程模型的决定系数 R^2 为0.9493,相对误差的均值和标准差分别为11.50%和11.21%。模型泛化能力试验中,生菜群体鲜质量比个体鲜质量的预测相对误差的平均值小(3个生长阶段分别相差4.44、5.71、5.89个百分点),且随着群体数量增加,群体鲜质量预测的相对误差的均值和标准差逐渐减小。本鲜质量估算方法预测的群体鲜质量数据可为基质培绿叶菜类作物的栽培管理决策提供数据支撑。

关键词: 生菜; 鲜质量估算; 高斯过程回归; 作物表型; 长势无损监测; 日光温室

中图分类号: S636.2 **文献标识码:** A **文章编号:** 1000-1298(2021)09-0230-11

OSID:



Non-destructive Estimation Method of Fresh Weight of Substrate Cultured Lettuce in Solar Greenhouse

LIU Lin^{1,2} YUAN Jin^{1,2} ZHANG Yan¹ LIU Xuemei^{1,2}

(1. College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian 271018, China
2. Shandong Agricultural Equipment Intelligent Engineering Laboratory, Taian 271018, China)

Abstract: As an indicative feature of crop growth and development, the dynamic change of fresh weight of crops in facility cultivation is one of the important indicators for non-destructive monitoring of vegetable growth. Hydroponic vegetables can be directly weighed out of water to achieve non-destructive monitoring of growth, but it is difficult to achieve fresh weight by non-destructive measurement in soil or matrix. To solve this problem, the fresh weight estimation method based on the combination of phenotypic parameters and environmental parameters was proposed to estimate the fresh weight of individuals and groups of lettuce in solar greenhouse. Firstly, the environmental parameters of the whole life cycle of lettuce were monitored. The multi sample images and the fresh weight of some samples were collected during the growth of the first batch of lettuce. The shape, color, texture and other characteristics of lettuce in different growth periods were extracted from the sample images, and the accumulated heat product in the environmental information was calculated. Then, the relationship model between the parameters of phenotype and environment and fresh weight of lettuce was established by Gaussian process regression. Finally, the sample data of the second batch of groups of lettuce were collected to predict the fresh weight of individuals and groups of lettuce at three growth stages based on the above model, so as to verify the generalization ability and reliability of the fresh weight estimation model. The results showed that

收稿日期: 2020-10-09 修回日期: 2021-01-12
基金项目: 国家自然科学基金项目(51675317)和山东省重大科技创新工程项目(2019JZZY010715)
作者简介: 刘林(1987—),男,实验师,博士生,主要从事精确农业与智能装备研究,E-mail: swzcr@163.com
通信作者: 苑进(1972—),男,教授,博士生导师,主要从事智能农业装备技术研究,E-mail: jyuan@sdau.edu.cn

compared with support vector machine, linear regression, ridge regression and neural network, the determination coefficient of Gaussian process model was 0.949 3, and the mean of relative error was 11.50%, while the standard deviation of relative error was 11.21%. In the model generalization ability test, the average value of relative error of prediction of fresh weight of groups of lettuce was smaller than that of individuals of lettuce, and the difference of them were 4.44, 5.71 and 5.89 percentage points at the three growth states. The average value and standard deviation of predicted fresh weight of groups of lettuce was gradually decreased with the increase of groups. The fresh weight data of groups predicted by this method can provide data support for the cultivation and management decision of substrate cultivated green leafy vegetables.

Key words: lettuce; fresh weight estimation; Gaussian process regression; crop phenotype; growth non-destructive monitoring; solar greenhouse

0 引言

设施栽培中作物鲜质量动态变化作为生长发育的指示性特征,是蔬菜长势无损监测的重要指标之一^[1-2]。蔬菜鲜质量作为最重要的生理性状表征之一,可为蔬菜生长状况提供更准确的评估信息^[3],如用于预测蔬菜生长期、确定最佳采收时间和预测蔬菜产量等^[4-5],为栽培数字化管理提供精准数据支撑。

国内外学者对鲜质量估算做了深入研究, REYES-YANES 等^[6]提出了一种深度学习图像分割方法用于在水培生菜鲜质量估算中以冠层图像获取信息,建立作物尺寸与鲜质量间的回归模型。JUNG 等^[5]在环境可控的循环水水培生菜系统中,研究了基于机器视觉的形态学分析方法,将图像处理后的生菜冠层投影图像的像素数与离水称量的生菜实际鲜质量直接关联,建立了生菜冠层投影面积与实际测量鲜质量的关系模型。由于其离水鲜质量称量精度高,因而该模型预测精度高,由此也证明了植株冠层投影面积与其鲜质量间存在显著相关性。基于上述预测模型,JIANG 等^[7]开发了基于在线实时影像处理的水培生菜鲜质量估算系统,在生产环境控制质量较高的状况下,实现了生菜鲜质量的高精度估算。水培蔬菜生长条件下,植株允许暂时脱离营养液而直接称量,且不妨碍其继续生长,因而方便对鲜质量进行无损标定,容易实现无损的高精度鲜质量估算;但土培或基质培蔬菜条件下,虽然植株离开基质直接称量也能得到准确的即时鲜质量,但植株已无法继续生长,即无法对该植株生长过程中的鲜质量进行连续采集^[8-9],难以直接建立该植株样本生长过程中冠层投影面积与其鲜质量变化序列的相关关系,因而使土培或基质培蔬菜在生长过程中的在线无损鲜质量估算成为技术难点。

另一方面,因为日光温室环控手段少、调控稳定性差,造成温室内的环境存在高低温、弱光多发频发

等问题,使得蔬菜的生长发育受到影响,难以一直处于健康的生长状态^[10]。温度和辐射是影响生菜光合作用的 2 个重要因子,即显著影响生菜生长过程中的体内含水率,也对叶片干物质积累和产量形成起着重要作用^[11]。因而生菜鲜质量受温度和光照辐射的影响较大,从而最终显著影响生菜产量。与受控环境下的水培生菜不同,日光温室在温度和辐射变化较大,蔬菜生长难以维持在最佳生长状态,因而在叶厚等方面存在显著差异,因此沿用水培生菜的冠层投影图像面积很难准确估算日光温室条件下的生菜鲜质量。

日光温室条件下,不同批次种植的生菜所处的温度和辐射环境不同,存在相同冠层投影面积的生菜,可能存在叶片厚度、颜色、纹理等参数的差别,造成鲜质量差异。这些特征又很难通过冠层投影面积表征,造成日光温室基质培蔬菜难以仅仅依靠冠层投影面积估算鲜质量。本文提出基于生菜表型特征参数和生长过程环境参数融合的鲜质量估算方法,用于日光温室基质培环境下生菜个体和群体的鲜质量无损估算。

1 试验与方法

1.1 试验设计

试验地点(图 1)为山东农业大学科技创新园(园艺实验站)的 6 号日光温室,跨度约 8 m,高度约 4 m,东西长约 50 m(36.16°N,117.16°E)。温室北墙为实体墙,主采光面附有塑料薄膜,形成封闭小气候环境,留有通风口,用于换气除湿降温。冬季早晚依靠卷帘机提供动力完成棉被铺卷任务。试验材料为意大利生菜(河北茂华种业有限公司)。品种主要特性为:耐寒、耐湿、耐热、晚抽苔、株半直立、叶片近圆形、颜色翠绿、散叶不接球。将生菜种子放入穴盘中进行培育,待幼苗生长到“五叶一心”时,将生菜定植到基质槽中,选用蔬菜通用型基质作为生菜生长载体。缓苗成功后,选用山崎配方营养液连续浇灌,保

证生菜正常生长,进行两批次生菜种植试验。

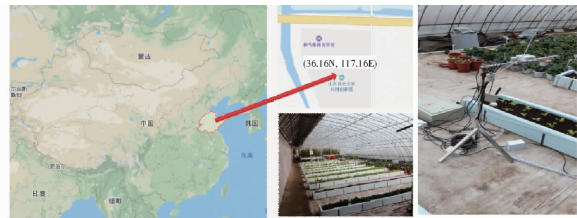


图1 试验地点
Fig.1 Experimental site

1.2 数据采集

生菜鲜质量受多种因素的影响,与生菜形状、颜色、纹理等表型特征因素和环境因素有关^[12]。因此,利用这些特征快速估算生菜的鲜质量具有重要意义。为了能够低成本获取温室中生菜形状、颜色和纹理信息,待生菜缓苗成功后,每隔一段时间,利用可见光成像技术,获取随机样本的俯视图像和侧视图像。利用传感器实时获取温室内温度和光合有效辐射。

如图2所示,采用摄像头获取生菜的两类图像样本。俯视图像是生菜冠层垂直映射到基质表面的投影,侧视图像是生菜映射到竖直平面中的投影。两类图像经过数字图像处理技术得到生菜的形状、颜色和纹理等表型特征信息。形状特征选取冠层投影面积、冠层投影周长和株高。颜色特征选取生菜冠层图像中生菜区域 H 通道和 S 通道图像的特征值。纹理特征选取生菜区域 H 通道和 S 通道图像的熵值。采用温度传感器和光合有效辐射传感器获取温室内环境信息,经过数学运算得到累积辐热积作为环境特征。每次采集完样本图像后,将样本去

除根部,用电子秤(精度0.01 g)称其质量,得到鲜质量。整个生长期中总共获取260株样本用于模型的建立和验证。为了验证模型的泛化能力和可靠性,从第2批次生菜中挑选3个时间(代表不同的生长阶段)采集10株样本数据作为模型的测试集。

1.3 特征提取

利用计算机(CPU:i7-8750,内存16 GB,显卡GTX1060,Windows 10操作系统),对样本图像进行处理得到生菜的形状、颜色和纹理等特征参数。

1.3.1 特征分割

由于获取的图像中生菜与背景有明显的区别,为了能够获取冠层轮廓图像和株高图像,采用超绿色法分割图像。RGB颜色空间是当前使用最广泛的颜色系统之一,几乎包含了人类能感知的所有颜色^[13]。超绿色法作为一种作物识别或者杂草识别最常用的灰度化方法,在RGB颜色空间中经过计算,图像中阴影、草和土壤等均能较明显地被抑制,提取绿色植物图像效果好,使得叶片图像更为突出^[14]。超绿色法计算式为

$$E_{xg} = 2g - r - b \tag{1}$$

式中 E_{xg} ——像素点经过计算后的值
 $r、g、b$ ——像素点的 R 分量、 G 分量和 B 分量
利用超绿色法对生菜初步分割后,提取组合图像像素值分布直方图,如图3所示。从图3中看出,叶片像素值聚集在中间区域,背景像素值聚集在两侧,叶片与背景有明显区别。将两侧背景区域和叶片区域像素值的交界处作为阈值分割点,组合图像经过两次图像阈值处理后得到生菜冠层投影的二值

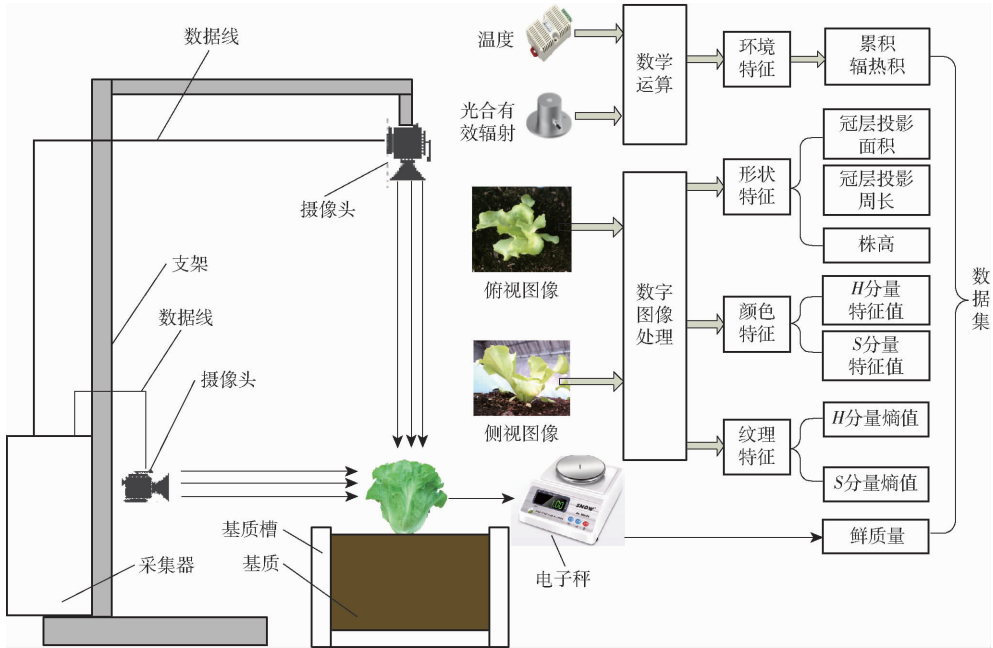


图2 数据集获取示意图
Fig.2 Schematic of data set acquisition

图像。

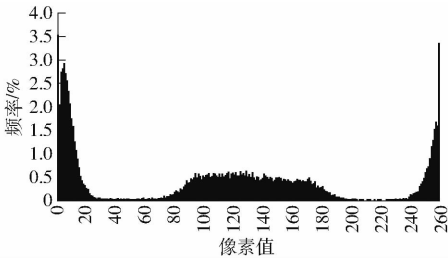


图 3 组合图像像素值分布直方图

Fig.3 Distribution histogram of pixel value of composite image

为了验证分割方法的分割效果,参考 GEORGE 等^[15]建立的图像分割性能指标,应用重合度和识别率作为分割性能评价指标^[16],计算式为

$$e = \frac{\sum_{\substack{i_0=m_0 \\ j_0=n_0}} F_{kq}(i_0,j_0) \cap F_{kp}(i_0,j_0)}{\sum_{\substack{i_0=m_0 \\ j_0=n_0}} F_{kp}(i_0,j_0)} \times 100\% \quad (2)$$

$$f = \frac{\sum_{\substack{i_0=m_0 \\ j_0=n_0}} F_{kq}(i_0,j_0) \cap F_{kp}(i_0,j_0)}{\sum_{\substack{i_0=m_0 \\ j_0=n_0}} F_{kq}(i_0,j_0) \cup F_{kp}(i_0,j_0)} \times 100\% \quad (3)$$

式中 e ——识别区域与人工提取区域识别重合度, %

- F_{kq} ——分割图像的二值图
- k ——图像序号
- q ——分割方法序号
- F_{kp} ——Photoshop 人工提取植物图像二值图
- i_0 ——图像行号 j_0 ——图像列号
- m_0 ——图像行数 n_0 ——图像列数
- f ——分割方法识别率, %

试验期间选择图像 5 幅,以在 Photoshop 软件中采用套绳工具人工提取的生菜区域作为参考依据^[16-17],从表 1 看出,图像分割重合度不小于 98.3%,识别率不小于 97.9%。

表 1 图像分割性能

Tab.1 Performance of segmentation algorithms					%
参数	编号				
	1	2	3	4	5
重合度	99.4	99.6	99.2	98.3	99.0
识别率	98.5	98.7	97.9	98.0	98.2

1.3.2 形状特征提取

生菜形状特征提取包括生菜冠层投影面积、周长和株高特征的提取,其中生菜冠层投影面积和周长特征提取过程如图 4 所示。首先,读取俯视投影原图信息,并提取 RGB 通道图像,将三通道

图像信息按照超绿色法处理后得到组合图像,组合图像经过阈值处理后得到生菜冠层投影的二值图像。利用二值图像提取冠层轮廓并计算面积和周长。

生菜株高特征提取过程如图 5 所示。首先,读取侧视投影原图信息,并提取 RGB 通道图像,将三通道图像信息按照超绿色法处理得到组合图像,组合图像经过阈值处理分割后得到生菜高度二值图像。如果将生菜高度二值图像置于平面直角坐标系中,其中水平向右为 x 轴正方向,竖直向上为 y 轴正方向,则在直角坐标系中,二值图像像素点最低点为生菜基部,横坐标相同纵坐标最大的点形成的区域代表生菜冠层上部区域。将生菜冠层上部区域点的横坐标值不变,其纵坐标值减去生菜基部点的纵坐标值,构造的新点形成生菜高度轮廓曲线。因为生菜散叶呈现的高度也不一样,为了能够更全面反映植株高度,计算高度曲线轮廓上所有点纵坐标的平均值,将计算结果作为生菜植株高度的特征值。

为验证图像方法测量株高的准确性,从生菜高度轮廓曲线中选择 3 个点,用米尺人工测量实际生菜 3 个点相对地高度,图像测量与人工测量株高如表 2 所示,其结果相对误差在 2.5% 以内。

1.3.3 颜色特征提取

由于人的视觉对亮度的敏感程度明显高于对颜色浓度的敏感程度,所以使用 HSI 颜色空间更能反映人类观察彩色的方式。在 HSI 颜色空间中, H 和 S 都不受光强变化影响^[18],因而用于描述生菜图像时能够更好地表征生菜的颜色和纹理情况。 I 表示强度或亮度,极易受到光强变化的影响,不易作为表征生菜的颜色和纹理情况。RGB 转 HSI 通道^[18] 计算式为

$$\left\{ \begin{aligned} i &= \frac{r+g+b}{3} \\ s &= 1 - \frac{\min(r,g,b)}{i} \\ h &= \begin{cases} \arccos \frac{2r-g-b}{[(r-g)^2+(r-b)(g-b)]^{\frac{1}{2}}} & (b \leq g) \\ 2\pi - \arccos \frac{2r-g-b}{[(r-g)^2+(r-b)(g-b)]^{\frac{1}{2}}} & (b > g) \end{cases} \end{aligned} \right. \quad (4)$$

式中 h,s,i ——像素点 H 分量、 S 分量和 I 分量

在提取生菜的颜色特征和纹理特征时,减少因为采集过程中光强对成像图像质量的影响,选取用于描述生菜图像时能够更好地表征生菜颜色情况的

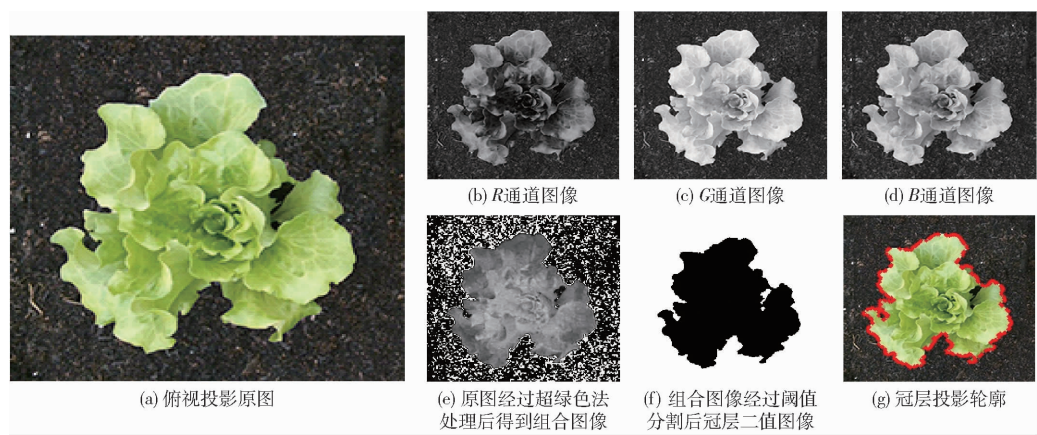


图 4 生菜冠层投影面积和周长特征提取示意图

Fig. 4 Schematics of extraction of projection area and perimeter features of lettuce canopy

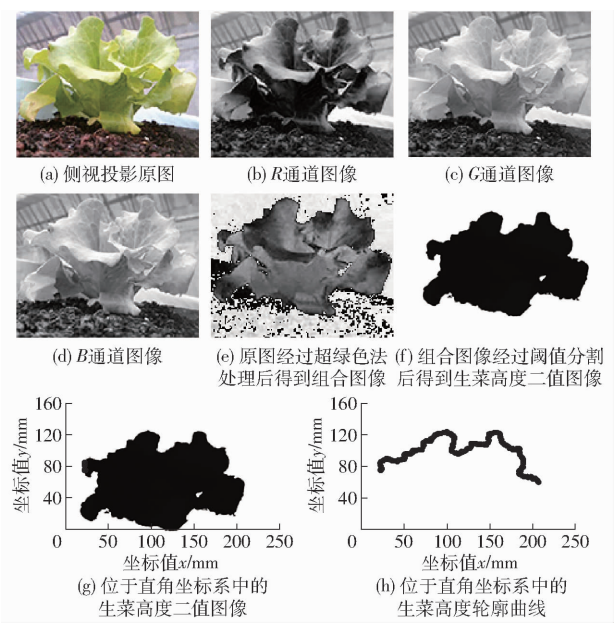


图 5 生菜植株高度特征值提取示意图

Fig. 5 Schematics of extraction of plant height characteristic value of lettuce

表 2 图像测量与人工测量株高比较

Tab.2 Comparison of image measurement and artificial measurement of plant height

	高度 1/mm	高度 2/mm	高度 3/mm	平均值/mm
图像测量	122	96	123	113. 7
人工测量	124	98	120	114
相对误差/%	1. 6	2. 0	2. 5	0. 3

H 通道和 S 通道图像作为计算图形^[18]。生菜颜色特征提取过程如图 6 所示。首先,将 RGB 通道图像(图 4b~4d),利用式(4)得到 H 通道和 S 通道分量图像,如图 6a、6d 所示。由于背景在整个图像中所占的比例比较大,在实际计算中,必须将背景部分去除,只提取生菜区域的 H 通道和 S 通道图像。将图 6a、6d 分别做掩膜运算,得到背景为黑色的图 6b、6e。

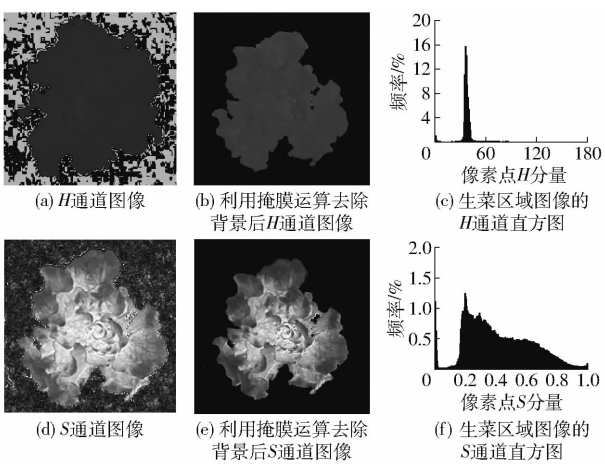


图 6 生菜颜色特征提取示意图

Fig. 6 Schematics of lettuce color feature extraction

掩膜运算是用选定的物体、图形或图像,遮挡待处理的全部或局部图像,来控制图像处理区域,其中用于覆盖的特定图像称为掩膜图像^[19]。如图 7 所示,以 3×3 的图像与 3×3 的掩膜图像进行运算为例:原图中的每个像素和掩膜图像中的对应像素进行运算,如果掩膜图像中的像素点为 1,则原图中的像素点保留;如果掩膜图像中的像素点为 0,则原图中的像素点被设置为 0。

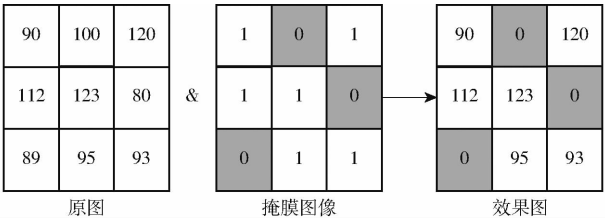


图 7 掩膜运算原理图

Fig. 7 Mask operation principle

利用掩膜运算原理,将图 4f 中的冠层二值图像进行处理得到掩膜图像(将白色区域的像素点设置为 0,黑色区域的像素点设置为 1),在图 6a 和图 6d 做掩膜运算时,保留的像素点为生菜区域图像,其像

素的频率分布直方图如图 6c、6f 所示。为了兼顾运算效率,选取生菜区域 H 通道和 S 通道像素值的众数作为颜色的特征值。

1.3.4 纹理特征提取

纹理特征是一种区域性特征,用于表征像素的分布规律和变化规律,并能代表像素与其邻域的空间关系,在反映物体的表面粗糙度、颗粒度、光滑度、规范程度和随机性等方面较为客观,可以很好地兼容图像宏观性质与微观结构,可用于描述物体表面的结构组织排列信息及其与环境的关系^[20]。在众多纹理特征中,熵值能够反映图像中纹理特征的非均匀程度,其值越大表示图像纹理越复杂^[21],生菜的纹理特征选取生菜区域图像的 H 通道和 S 通道图像的熵值作为特征值。计算式为

$$E_{NT} = - \sum_{i=1}^L \sum_{j=1}^L P(i,j) \lg P(i,j) \tag{5}$$

式中 E_{NT} ——熵值 $P(i,j)$ ——图像像素

1.3.5 生长环境特征提取

温度和辐射是影响生菜叶片生长的两个重要因子,对植物的光合作用、植株生长和产量形成起着重要作用。鲜质量作为衡量作物产量的重要指标,受温度和辐射的累积影响,温度和辐射对生菜鲜质量的影响可用累积辐热积来度量,具体计算公式为^[22-23]

$$R_{TE} = \begin{cases} 0 & (T < T_b \text{ 或 } T > T_m) \\ (T - T_b) / (T_{ob} - T_b) & (T_b \leq T < T_{ob}) \\ 1 & (T_{ob} \leq T \leq T_{ou}) \\ (T_m - T) / (T_m - T_{ou}) & (T_{ou} < T \leq T_m) \end{cases} \tag{6}$$

表 3 部分生菜样本表型和环境信息

Tab.3 Phenotypic and environmental information of some lettuce samples

序号	冠层投影 面积/mm ²	冠层投影 周长/mm	株高/ mm	H 分量 特征值	S 分量 特征值	H 分量 熵值	S 分量 熵值	累积辐热积/ (MJ·m ⁻²)	鲜质量/g
1	10 559	550	81.43	33	0.75	3.15	5.50	14.44	9.30
2	11 110	568	101.72	33	0.78	3.79	5.58	14.59	14.84
3	11 916	520	81.14	35	0.76	3.29	6.05	14.61	15.33
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
258	47 939	1 071	172.28	40	0.68	3.67	7.11	21.31	275.51
259	48 018	1 274	152.09	39	0.69	4.16	7.17	18.92	281.15
260	55 005	1 029	244.06	40	0.71	4.38	7.25	14.78	293.49

1.4.2 高斯过程回归

机器学习的高斯过程是一个通用的监督学习方法,高斯过程回归是使用高斯过程先验对数据进行回归分析的非参数模型。高斯过程回归中预测是对观察值的插值且带有概率,可以用来计算经验置信区间和超越概率,以便重新拟合预测,在小样本上能

$$T_{EP} = \sum R_{TEP} \tag{7}$$

其中 $R_{TEP} = \sum_{i=1}^{24} (3\ 600 R_{TEi} P_{ARi} / 10^6) \tag{8}$

式中 T_b ——生长下限温度,℃
 T_m ——生长上限温度,℃
 T_{ob} ——生长最适温度下限,℃
 T_{ou} ——生长最适温度上限,℃
 T ——环境温度,℃
 R_{TE} ——相对热效应
 R_{TEP} ——每日相对辐热积,MJ/(m²·d)
 R_{TEi} ——第 i 小时内的相对热效应
 P_{ARi} ——第 i 小时平均光合有效辐射,
MJ/(m²·d)
 T_{EP} ——累积辐热积,MJ/m²

在栽培阶段,根据生菜生长发育期知识^[11],本文设定 T_b 、 T_{ob} 、 T_{ou} 和 T_m 分别为 5、15、30、40℃。

为了提高鲜质量估算精度和减小计算复杂度,将样本采集时刻前 5 d 的累积辐热积作为估算生菜鲜质量的生菜生长环境特征。

生菜样本部分表型中的形状、颜色、纹理特征和环境中累积辐热积如表 3 所示。

1.4 模型建立

1.4.1 数据集划分

数据集划分是模型实用性的关键,选择合适的训练集,能使模型预测能力增强^[24]。一般训练集包括总样本 50% ~ 75%,试验共采集到样本 260 个,将数据按照 7:3 的比例分为训练集和验证集。其中训练集数为 182,验证集数为 78。

够获得泛化能力强的模型^[25]。

1.4.3 支持向量机回归

支持向量机回归是一种通过引入一个损失函数,使经验风险与置信风险之和最小化的回归分析方法,其遵循结构风险最小化原则,可以巧妙解决小样本、高维度、非线性问题,模型对未来数据具有较

好的泛化能力^[26]。

1.4.4 线性回归

线性回归是利用线性回归方程的最小平方差函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析,数据使用线性预测函数来建模,并且未知的模型参数也是通过数据来估计^[27]。

1.4.5 岭回归

岭回归是一种旨在消除最小二乘估计因多重共线性而引起的不良后果的回归估计方法,是一种有偏估计,由 HOERL 和 KENNARD 提出的一套改进方法^[28]。

1.4.6 神经网络回归

神经网络回归是模拟人脑神经元进行预测的一种经典机器学习算法,其特点是预测精度高,应用广泛,具备非常良好的非线性收敛能力,并且在训练过程中具有简单高效的特点,可以模拟系统的非线性特征^[29-32]。

1.5 模型评价指标

1.5.1 决定系数

决定系数用于考察数据集样本预测值和真实值之间的相关程度,取值范围为 0~1,越接近 1,表明模型对数据拟合越好。

1.5.2 误差

因为生菜在不同的生长阶段,样本鲜质量差别很大,所以不适合采用平方根误差、均方根误差作为模型的评价指标。相对误差 RE 是测量所造成的绝对误差与被测量真值的比值,更能反映生菜鲜质量样本点测量的可信程度。相对误差平均值 MRE 是

所有样本点的相对误差的平均值,所以采用相对误差的平均值更适合描述生菜鲜质量估算模型,其值越小,表明模型对样本预测能力越高。相对误差标准差 σ 是所有样本点相对误差的标准差,采用相对误差标准差能够反映模型预测样本相对误差的波动程度,其值越小,表明模型对样本预测能力越高。

2 结果与讨论

2.1 生菜鲜质量预测最优模型

选用机器学习中的高斯过程回归、支持向量机回归、线性回归、岭回归和神经网络回归模型对训练集数据进行训练,建立训练模型。将验证集数据导入训练的模型中,用决定系数、相对误差平均值和标准差评价 5 种模型,选出最优模型。

从图 8 可以看出,日光温室生菜鲜质量估算的最优模型为高斯过程回归建立的模型。模型的决定系数为 0.949 3,高于支持向量机、线性回归、岭回归和神经网络回归模型。模型相对误差的平均值和标准差分别为 11.50% 和 11.21%,明显小于支持向量机回归、线性回归、岭回归和神经网络模型。这主要因为高斯过程回归模型中预测值是对观察值的插值,预测是带有概率,可以用来计算经验置信区间和超越概率,以便重新拟合预测,特别适用于小数据集预测模型^[26-27]。其中线性回归模型和岭回归模型当生菜鲜质量比较小时,预测出的鲜质量有负值出现;神经网络回归当生菜鲜质量低于 50 g 时,预测出的鲜质量误差比较大。所以,利用高斯过程回归方法预测的生菜个体鲜质量更准确。

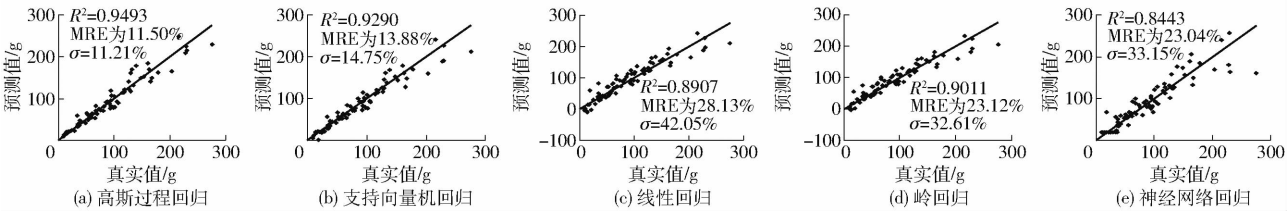


图 8 5 种模型下生菜鲜质量真实值与预测值对比

Fig. 8 Comparison of real value and predicted value of fresh weight of lettuce under five models

2.2 不同输入特征下模型比较

将模型输入分别去除颜色、纹理和环境特征后,利用剩余特征数据训练得到相关模型,分别定义为模型 a、模型 b 和模型 c,其中将模型输入原特征(带有颜色、纹理和环境特征)数据训练得到的模型定义为模型 d。用决定系数、相对误差平均值和标准差,评价颜色、纹理和环境参数在生菜鲜质量估算中的作用。

如表 4 所示,对比模型 d,模型 a 的 R^2 减小 0.020 9、MRE 增加 5.01 个百分点、 σ 增加 8.37 个

表 4 不同输入特征下回归模型性能参数比较

Tab.4 Comparison of performance parameters of regression models under different input characteristics				
指标	模型 a	模型 b	模型 c	模型 d
R^2	0.928 4	0.936 5	0.934 3	0.949 3
MRE/%	16.51	13.47	12.48	11.50
σ /%	19.58	13.22	11.31	11.21

百分点,模型 b 的 R^2 减小 0.012 8、MRE 增加 1.97 个百分点、 σ 增加 2.01 个百分点,模型 c 的 R^2 减小 0.015 0、MRE 增加 0.98 个百分点、 σ 增加 0.10 个百分点,说明颜色、纹理和环境特征在生菜鲜质量估

算中具有一定作用。一般来说,生菜外层叶子外轮廓表现为整株生菜冠层轮廓特征(冠层面积和冠层周长),中层或内层叶子很难通过整株生菜冠层轮廓特征表征。当生菜冠层轮廓特征比较接近时,如果生菜缺水,呈现的纹理和颜色特征会发生变化。另一方面,假设生菜历经 5 d 辐射和温度累积的影响后,如果生菜健康生长,叶片表现较厚,如果生菜不健康生长,叶片表现较薄。同时,由于叶片存在相互遮挡,叶片厚度将很难通过主视或俯视二维图像呈现。上述情况下若仅依靠冠层投影轮廓将很难准确估算生菜鲜质量,因此,颜色、纹理和环境特征参数在生菜鲜质量估算中有不可替代的作用。

2.3 生菜鲜质量模型分段预测能力

将验证集中生菜样本按照鲜质量分成 3 个区间,代表生菜生长不同阶段样本。评估利用高斯过程回归方法建立的模型对生菜不同生长阶段个体鲜质量的预测能力。

将验证集数据按照鲜质量真实值划分为 0 ~ 55 g、55 ~ 102 g、102 ~ 300 g 3 个鲜质量区间,每个区间 26 个样本数据,分别编号,得到样本区间的相对误差的平均值和标准差。从图 9 可以看出,0 ~ 55 g 区间的预测生菜鲜质量相对误差的平均值和标准差明显大于 55 ~ 102 g、102 ~ 300 g 2 个区间,而且出现一个相对误差为 74.06% 的数据,表明当生菜鲜质量比较小时,鲜质量估算相对误差比较大,主要是因为当生菜比较小时,叶片薄且数量少,形状特征参数很容易受到外界环境(例如温度、光照、风速等)变化的影响;当生菜比较大时,生菜的叶片相对较厚、数量多且紧凑,形状特征参数不易受到外接环境变化的影响,由图 9b 生菜预测鲜质量的相对误差的平均值和标准差居中可以得到验证;在生菜生长后期,由于生菜的叶面积较大,根部吸收的水分和叶片蒸散的水分比较大,对生菜鲜质量的影响较前一阶段影响大,由图 9c 预测生菜鲜质量的相对误差的平均值和标准差比 55 ~ 102 g 区间大可以得到验证。

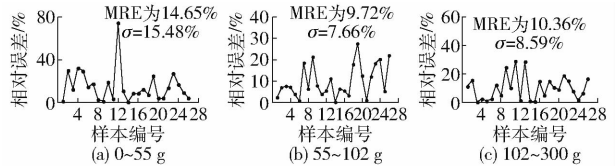


图 9 模型对生菜不同生长阶段预测能力对比
Fig. 9 Comparison of prediction ability of models for different growth stages of lettuce

2.4 验证模型的泛化能力

在第 2 批次生菜的生长期内选择 3 个阶段,每个阶段选取 10 个样本数据,预测生菜群体的鲜质

量,验证模型的泛化能力和可靠性。个体可以看作样本数量为 1 的群体,为了能够利用有限的样本获得不同数量的群体样本,将每阶段生菜样本做组合处理,具体步骤如下:

- (1)从第 1 阶段 10 个样本中随机选择 $l(l = 1, 2, \dots, 10)$ 个样本,总共有 C_{10}^l 个样本数为 l 的群体样本。
- (2)计算第 1 阶段预测生菜群体鲜质量样本数为 l 的相对误差的平均值和标准差。
- (3)依次选取第 2、3 阶段的样本,分别循环步骤(1)~(2),得到第 2 阶段和第 3 阶段的预测生菜群体鲜质量样本数为 $l(l = 1, 2, \dots, 10)$ 的相对误差的平均值和标准差。

如图 10 所示,可以看出模型预测个体($l = 1$)生菜鲜质量相对误差的平均值和标准差较大,预测生菜群体鲜质量的相对误差的平均值和标准差较小,预测生菜群体($l = 10$)鲜质量相对误差的平均值与预测的生菜个体($l = 1$)鲜质量相对误差的平均值在 3 个生长阶段分别相差 4.44、5.71、5.89 个百分点;随着群体样本数的增多,预测生菜群体鲜质量相对误差的平均值和标准差逐渐减小,其中相对误差的平均值最终趋于稳定,相对误差的标准差逐渐减小。主要因为生菜样本个体差异,预测的生菜个体鲜质量有正偏离,有负偏离,经过求和,抵消了一部分鲜质量差异,使得预测群体鲜质量的相对误差明显减小。

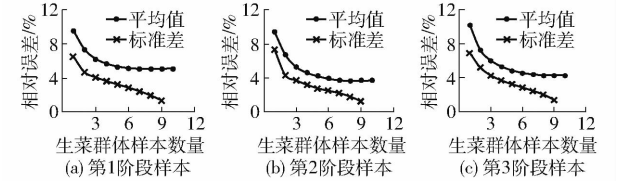


图 10 测试集生菜群体估算鲜质量相对误差对比
Fig. 10 Comparison of real value and predicted value of fresh weight of groups of lettuce

在实际的应用过程中,比较容易获取群体生菜的冠层投影图像信息,群体生菜的株高参数却不易获取^[33]。所以在群体鲜质量估算时,研究去除株高特征参数的模型具有重要的意义。因此,将数据集 中的株高特征参数去除,仅利用冠层投影图像中的形状特征(冠层投影面积和周长)、颜色特征、纹理特征和环境信息,采用高斯过程回归方法建立模型,验证模型对生菜群体鲜质量的预测能力,并将两种模型(有株高(模型 1)、无株高(模型 2))的预测能力做比较。生菜群体样本数量为 10。

从图 11 可看出,模型 1 和模型 2 预测生菜群体鲜质量的相对误差比较接近,最大不超过 1.54 个百分点,说明采用模型 2 能够达到模型 1 估算生菜群

体鲜质量的目的。模型 2 的输入数据集中去除了生菜株高这一维度的特征数据,导致模型 2 预测的个体鲜质量比模型 1 预测的相对误差波动大,表明模型 2 预测生菜个体鲜质量能力有所下降。然而,对于生菜群体鲜质量来说,较大的个体鲜质量误差,最终对群体的预测能力有可能增加,从图 11b 中模型 2 比模型 1 预测的生菜群体鲜质量的相对误差小得到验证。所以,模型 1 和模型 2 都能够达到准确预

测生菜群体鲜质量的目的。对于实时监测生菜群体鲜质量的场所,采用模型 2 更具有优势。因为不需要获得生菜的株高信息,而仅需要获得生菜群体冠层图像就可以准确预测出群体鲜质量。因此,在无损监测生菜群体鲜质量方面,模型 2 比模型 1 更具有优势。利用上述方法预测的群体鲜质量,可为基质培绿叶菜类作物的栽培管理决策提供数据支撑。

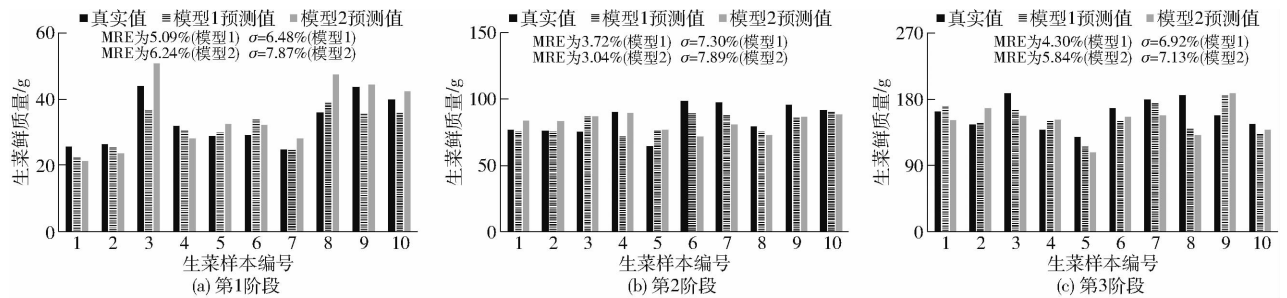


图 11 2 种模型下测试集生菜群体鲜质量真实值与预测值对比

Fig. 11 Comparison between real value and predicted value of fresh weight of groups of lettuce tested under two models

3 结论

(1)提出了基于表型特征参数和生长过程环境参数融合的鲜质量估算方法,并用于日光温室基质培环境下生菜个体和群体的鲜质量无损估算。利用具有少样本建模优势的高斯过程回归方法,以生菜生长过程为对象,建立了表型参数和环境参数与生菜鲜质量的回归关系模型。

(2)通过采集同批次生菜样本个体在不同生长阶段的表型特征参数及其鲜质量,用于表征该生长点的作物表型参数与鲜质量的回归关系,通过在生菜全生育期上密集取样,获得了整个生长期上表型参数与鲜质量的完备表达。

(3)日光温室基质培生菜鲜质量估算,最优模型为高斯过程回归建立的模型,模型的决定系数 R^2

为 0.949 3,相对误差平均值为 11.50%,相对误差标准差为 11.21%。与支持向量机回归、线性回归、岭回归和神经网络回归模型相比,高斯过程回归模型更适用于小样本数据集建立的鲜质量估算模型,预测精度更高,相对误差更小。

(4)本文生菜鲜质量估算模型,可以预测出生菜在生长过程中个体鲜质量和群体鲜质量。虽然生菜的个体鲜质量预测相对误差比较大,但是生菜群体鲜质量预测相对误差较小,预测生菜群体鲜质量相对误差的平均值与预测的生菜个体鲜质量相对误差的平均值在 3 个生长阶段分别相差 4.44、5.71、5.89 个百分点;随群体数量的增加,预测生菜群体鲜质量相对误差的平均值和标准差逐渐减小。利用预测的群体鲜质量数据可为基质培绿叶菜类作物的栽培管理决策提供数据支撑。

参 考 文 献

[1] KONG Yun, KONG Xiangyue, ZHENG Youbin. Estimating fresh weight of individual pea shoots using measurable morphological characteristics[J]. Hortscience, 2020,55(7): 1111 - 1118.

[2] 向云,刘秀珍,李静波,等.新型有机无机复混肥对油菜产量和品质的影响[J].中国农学通报,2011,27(7): 198 - 201. XIANG Yun, LIU Xiuzhen, LI Jingbo, et al. Effect of new organic-inorganic fertilizer on rape yield and quality[J]. Chinese Agricultural Science Bulletin,2011,27(7):198 - 201. (in Chinese)

[3] BUMGARNER N R, MILLER W S, KLEINHENZ M D. Digital image analysis to supplement direct measures of lettuce biomass [J]. Hort Technology,2012,22(4):547 - 555.

[4] 李永秀,罗卫红. 温室蔬菜生长发育模型研究进展[J].农业工程学报,2008,24(1):307 - 312 LI Yongxiu, LUO Weihong. Review on research progress of greenhouse vegetable growth and development models [J]. Transactions of the CSAE, 2008, 24(1): 307 - 312. (in Chinese)

[5] JUNG D H, PARK S H, HAN X Z, et al. Image processing methods for measurement of lettuce fresh weight[J]. Journal of Biosystems Engineering, 2015, 40(1): 89 - 93.

[6] REYES-YANES A, MARTINEZ P, AHMAD R. Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds[J]. Computers and Electronics in Agriculture, 2020, 179: 105827.

- [7] JIANG J, KIM H J, CHO W J. On-the-go image processing system for spatial mapping of lettuce fresh weight in plant factory[J]. IFAC Papers OnLine, 2018, 51(17): 130–134.
- [8] 刘志刚, 王纪章, 徐云峰, 等. 基质配方和灌溉方式对生菜根系和产量的影响[J/OL]. 农业机械学报, 2014, 45(2): 156–160.
LIU Zhigang, WANG Jizhang, XU Yunfeng, et al. Lettuce yield and root distribution in substrates under drip irrigation and micro-sprinkler irrigation[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(2): 156–160. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20140226&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2014.02.026. (in Chinese)
- [9] 陈晓丽, 杨其长, 马太光, 等. 不同频率LED红蓝光交替照射对生菜生长与品质的影响[J/OL]. 农业机械学报, 2017, 48(6): 257–262.
CHEN Xiaoli, YANG Qichang, MA Taiguang, et al. Effects of red and blue LED irradiation in different alternating frequencies on growth and quality of lettuce[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6): 257–262. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170633&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.06.033. (in Chinese)
- [10] 刘文科. 温光控制工程技术推进日光温室生产提质增效[J]. 农业工程, 2015, 5(增刊): 15–17, 42.
LIU Wenke. Engineering technology development of temperature and lighting control promote productivity of Chinese solar greenhouse[J]. Agricultural Engineering, 2015, 5(Supp.): 15–17, 42. (in Chinese)
- [11] 刘志刚, 徐勤超. 基于辐热积的温室微灌基质栽培生菜生长模拟[J]. 江苏农业学报, 2016, 32(6): 1315–1319.
LIU Zhigang, XU Qinchao. A simulation for lettuce growth and development in greenhouse based on product of thermal effectiveness and photosynthetically active radiation[J]. Jiangsu Journal of Agricultural Sciences, 2016, 32(6): 1315–1319. (in Chinese)
- [12] 胡玥, 崔雯, 金敏凤, 等. 不同栽培方式对生菜生长和营养品质的影响[J]. 上海师范大学学报(自然科学版), 2019, 48(5): 566–573.
HU Yue, CUI Wen, JIN Minfeng, et al. Effects of different cultivation methods on the growth and nutritional qualities of lettuce[J]. Journal of Shanghai Normal University (Natural Sciences), 2019, 48(5): 566–573. (in Chinese)
- [13] 杨璟, 朱雷. 基于RGB颜色空间的彩色图像分割方法[J]. 计算机与现代化, 2010(8): 147–149.
YANG Jing, ZHU lei. Color image segmentation method based on RGB color space[J]. Computer and Modernization, 2010(8): 147–149. (in Chinese)
- [14] 毛文华, 王一鸣, 张小超, 等. 基于机器视觉的苗期杂草实时分割算法[J]. 农业机械学报, 2005, 36(1): 83–86.
MAO Wenhua, WANG Yiming, ZHANG Xiaochao, et al. Real-time algorithm based on machine vision to segment weeds at seedling[J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(1): 83–86. (in Chinese)
- [15] GEORGE E M, JOAO C N. Verification of color vegetation indices for automated crop imaging applications[J]. Computers and Electronics in Agriculture, 2008, 63(2): 282–293.
- [16] 孙国祥, 汪小昆, 闫婷婷, 等. 基于机器视觉的植物群体生长参数反演方法[J]. 农业工程学报, 2014, 30(20): 187–195.
SUN Guoxiang, WANG Xiaochan, YAN Tingting, et al. Inversion method of flora growth parameters based on machine vision[J]. Transactions of the CSAE, 2014, 30(20): 187–195. (in Chinese)
- [17] ZHENG L, ZHANG J, WANG Q. Mean-shift-based color segmentation of images containing green vegetation[J]. Computers and Electronics in Agriculture, 2009, 65(1): 93–98.
- [18] 李大华, 赵相飞, 许亮, 等. 基于纹理特征与HSI空间的苹果识别与标定[J]. 图学学报, 2016, 37(5): 688–693.
LI Dahua, ZHAO Xiangfei, XU Liang, et al. Apple identification and calibration based on the texture characteristics and HSI space[J]. Journal of Graphics, 2016, 37(5): 688–693. (in Chinese)
- [19] 王婷, 上官宏, 刘伟, 等. 基于顶帽变换的反锐化掩膜算法[J]. 计算机工程与设计, 2014, 35(1): 198–202.
WANG Ting, SHANGGUAN Hong, LIU Yi, et al. Unsharp masking algorithm based on top-hat transform[J]. Computer Engineering and Design, 2014, 35(1): 198–202. (in Chinese)
- [20] 宋荣杰, 宁纪锋, 刘秀英, 等. 基于纹理特征和SVM的QuickBird影像苹果园提取[J/OL]. 农业机械学报, 2017, 48(3): 188–197.
SONG Rongjie, NING Jifeng, LIU Xiuying, et al. Apple orchard extraction with QuickBird imagery based on texture features and support vector machine[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 188–197. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170324&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.03.024. (in Chinese)
- [21] 徐贵力, 毛罕平, 李萍萍. 差分百分率直方图法提取缺素叶片纹理特征[J]. 农业机械学报, 2003, 34(2): 76–79.
XU Guili, MAO Hanping, LI Pingping. Research on extraction leaf texture features as sample of nutrient shortage by percent histogram of differentiation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2003, 34(2): 76–79. (in Chinese)
- [22] HANG T, LU N, TAKAGAKI M, et al. Leaf area model based on thermal effectiveness and photosynthetically active radiation in lettuce grown in mini-plant factories under different light cycles[J]. Scientia Horticulturae, 2019, 252(6): 113–120.

- [23] ZHOU Tianmei, WU Zhen, WANG Yachen, et al. Modelling seedling development using thermal effectiveness and photosynthetically active radiation[J]. Journal of Integrative Agriculture, 2019, 18(11): 2521–2533.
- [24] 詹雪艳, 赵娜, 林兆洲, 等. 校正集选择方法对于积雪草总苷中积雪草苷 NIR 定量模型的影响[J]. 光谱学与光谱分析, 2014, 34(12): 3267–3272.
ZHAN Xueyan, ZHAO Na, LIN Zhaozhou, et al. Effect of algorithms for calibration set selection on quantitatively determining asiaticoside content in centella total glucosides by near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2014, 34(12): 3267–3272. (in Chinese)
- [25] 苑进, 胡敏, Kesheng Wang, 等. 基于高斯过程建模的物联网数据不确定性度量与预测[J/OL]. 农业机械学报, 2015, 46(5): 265–272.
YUAN Jin, HU Min, WANG Kesheng, et al. Uncertainty measurement and prediction of IOT data based on Gaussian process modeling[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(5): 265–272. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20150538&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2015.05.038. (in Chinese)
- [26] 郭霖, 白丹, 王新端, 等. 基于支持向量机的滴灌灌水器流量预测模型建立与验证[J]. 农业工程学报, 2018, 34(2): 74–82.
GUO Lin, BAI Dan, WANG Xinduan, et al. Establishment and validation of flow rate prediction model for drip irrigation emitter based on support vector machine[J]. Transactions of the CSAE, 2018, 34(2): 74–82. (in Chinese)
- [27] 江力, 徐海川, 张磊, 等. NLOS 环境中的线性回归最小二乘三维定位算法[J]. 湖南工业大学学报, 2019, 33(6): 23–28.
JIANG Li, XU Haichuan, ZHANG Lei, et al. 3D localization algorithm based on linear regression and least square in NLOS environment[J]. Journal of Hunan University of Technology, 2019, 33(6): 23–28. (in Chinese)
- [28] 肖玲玲, 郑华, 林烁烁, 等. 基于岭回归的四带图像偏色校正算法[J]. 计算机系统应用, 2019, 28(8): 129–135.
XIAO Lingling, ZHENG Hua, LIN Shuoshuo, et al. Four-band image color cast correction algorithm based on ridge regression[J]. Computer Systems & Applications, 2019, 28(8): 129–135. (in Chinese)
- [29] 齐亮, 赵茂程, 赵婕, 等. 基于太赫兹光谱分析的猪肉 K 值无损检测方法[J/OL]. 农业机械学报, 2017, 48(12): 346–353.
QI Liang, ZHAO Maocheng, ZHAO Jie, et al. Non-destructive detection method of K value in pork with Terahertz spectroscopy[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 346–353. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20171243&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.12.043. (in Chinese)
- [30] 谢为俊, 丁冷春, 王凤贺, 等. 基于卷积神经网络的油茶籽完整性识别方法[J/OL]. 农业机械学报, 2020, 51(7): 13–21.
XIE Weijun, DING Yechun, WANG Fenghe, et al. Integrity recognition of *Camellia oleifera* seeds based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 13–21. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200702&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2020.07.002. (in Chinese)
- [31] 李旭青, 张秦雪, 安志远, 等. 基于可变形全卷积神经网络的冬小麦自动解译研究[J/OL]. 农业机械学报, 2020, 51(9): 144–151.
LI Xuqing, ZHANG Qinxue, AN Zhiyuan, et al. Automatic interpretation of winter wheat based on deformable full convolution neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 144–151. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200917&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2020.09.017. (in Chinese)
- [32] 刘忠超, 何东健. 基于卷积神经网络的奶牛发情行为识别方法[J/OL]. 农业机械学报, 2019, 50(7): 186–193.
LIU Zhongchao, HE Dongjian. Recognition method of cow estrus behavior based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 186–193. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190719&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.07.019. (in Chinese)
- [33] 李 Xiaobin, WANG Yushun, FU Lihong. Monitoring lettuce growth using K-means color image segmentation and principal component analysis method[J]. Transactions of the CSAE, 2016, 32(12): 179–186.