

doi:10.6041/j.issn.1000-1298.2022.08.038

基于多维度特征和 LightGBM 的大闸蟹质量估算方法

段青玲^{1,2} 陈鑫^{1,2} 许冠华³ 樊宇星^{1,2} 张玉玲^{2,4}

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 中国农业大学国家数字渔业创新中心, 北京 100083;
3. 江苏中农物联网科技有限公司, 宜兴 214200; 4. 中国农业大学烟台研究院, 烟台 261400)

摘要: 大闸蟹是我国特有的名优水产养殖品种,其质量既是确定投喂量的重要依据,亦是评判其生长状况、品质等级的重要指标。为了准确估算蟹体质量,提出一种基于多维度特征和轻量梯度提升机(Light gradient boosting machine, LightGBM)的大闸蟹质量估算方法。首先通过相机获取蟹体图像,其次采用图像处理技术对图像进行分割以获取背甲图像,然后提取背甲二值图像的几何特征构成形状特征(Shape features, SF);提取不同颜色空间背甲图像的各通道分量值构成颜色特征(Color features, CF),并采用标定法计算特征值;最后采用基于 LightGBM 的方法预测大闸蟹质量。本文根据色泽表征其发育状况,提取背甲颜色特征与形状特征构成多维度特征,解决单一形状特征导致预测精度不高的问题;提取背甲轮廓比值作为形状特征,有效降低随机调整相机高度对特征值稳定性的影响;在真实数据集上进行预测,结果表明平均绝对误差(MAE)为 2.751 g,均方根误差(RMSE)为 3.680 g,决定系数 R^2 为 0.949。并与 SF-LightGBM、SF3-LightGBM、area-OLS、MF-BPNN 和 MF-SVM 质量估算方法进行对比,本文方法的各评价指标的性能均有较大幅度提升,能够较准确地估算出大闸蟹蟹体质量。

关键词: 大闸蟹; 水产养殖; 质量估算; 多维度特征; LightGBM

中图分类号: TP391.4; S966.16 文献标识码: A 文章编号: 1000-1298(2022)08-0353-08 **OSID:** 

Chinese Mitten Crab Weight Estimation Method Based on Multi-dimensional Features and LightGBM

DUAN Qingling^{1,2} CHEN Xin^{1,2} XU Guanhua³ FAN Yuxing^{1,2} ZHANG Yuling^{2,4}

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
2. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China
3. Jiangsu Zhongnong Internet of Things Technology Co., Ltd., Yixing 214200, China
4. Yantai Research Institute, China Agricultural University, Yantai 261400, China)

Abstract: Chinese mitten crab is a unique aquaculture species in China. Its weight is not only an important basis for determining the feeding amount, but also an important indicator for judging its growth status and quality. Taking Chinese mitten crab as the research object, a method for estimating its weight based on multi-dimensional features and light gradient boosting machine (LightGBM) was proposed. Firstly, image segmentation were carried out on these collected crab images to obtain the carapace images. Then the geometric features of the carapace binary image was extracted as shape features (SF), extracting each channel component value of carapace images in different color spaces as color features (CF), and feature values were calculated by the calibration method. Finally, the crab weight was estimated by the LightGBM algorithm. The color feature and shape feature were extracted to form multi-dimensional features to solve the problem of low prediction accuracy caused by a single shape feature. The shape feature consisted of different carapace contour ratios, which effectively reduced the impact on the stability of the feature value caused by the random adjustment of the camera height. The proposed Chinese mitten crab weight estimation method was tested on the real dataset with the mean absolute error (MAE) of 2.751 g, the root mean square error (RMSE) of 3.680 g and the coefficient of determination (R^2) of 0.949. Furthermore, when compared with the SF-LightGBM, SF3-LightGBM, area-OLS, MF-BPNN and MF-SVM crab weight estimation methods, the performance of each evaluation metric of

收稿日期: 2021-09-10 修回日期: 2021-11-17
基金项目: 江苏省农业科技自主创新基金项目(CX(19)1003)、宁波市公益科技项目(202002N3034)和烟台市校地融合发展项目(2020XDRHXMXX07)
作者简介: 段青玲(1967—),女,教授,博士生导师,主要从事智能信息处理研究,E-mail: dqing@cau.edu.cn

the proposed method was improved. The experimental results indicated that the proposed method can accurately estimate the Chinese mitten crab weight.

Key words: Chinese mitten crab; aquaculture; weight estimation; multi-dimensional features; LightGBM

0 引言

大闸蟹,学名中华绒螯蟹,又称河蟹,是我国最重要的养殖蟹类之一,其肉质鲜美、营养丰富,年产量超过 80 万 t,养殖面积达 $6.67 \times 10^5 \text{ hm}^2$ 以上^[1]。在大闸蟹养殖过程中,蟹农需要定期称量其质量以确定投饵量、监测其生长状况。大闸蟹成熟后,蟹体质量也是品质分级重要标准之一。传统的蟹体质量测量方法是人工取样并称量,体重计示数容易因蟹体的应激行为而不准确,而且费工费时。因此实现蟹体质量的无损精确测量对大闸蟹科学高效养殖有重要作用。

计算机视觉技术以其非侵入性、客观、高精度的特点成为无损检测水生生物特征的最重要手段^[2-6],目前已在鱼体质量估算中得到应用^[7-9]。关于蟹体质量估算的研究较少,主要方法分为简单数据拟合方法和机器学习方法两类。基于简单数据拟合的质量估算方法研究中,王静安等^[10]对池塘养殖的大闸蟹质量与甲长、甲宽的关系分别进行研究,发现甲长、甲宽与质量存在很好的线性关系,其中公蟹质量与甲长、甲宽的相关系数分别为 0.889 2、0.813 8,母蟹与甲长、甲宽的相关系数分别为 0.821 4、0.903 3。张超等^[11]采集不同生长阶段的梭子蟹图像并提取其投影面积,利用普通最小二乘法(Ordinary least square,OLS)对蟹体面积和质量分别进行线性拟合、多项式拟合、幂次拟合,其中拟合效果最好的是二次多项式,决定系数为 0.922 0,结果表明梭子蟹体质量与投影面积紧密联系。基于机器学习的质量估算方法研究中,唐杨捷等^[12]采用分水岭分割技术得到梭子蟹头背甲图像,然后提取了投影面积、甲长、甲宽 3 个特征,采用遗传算法(Genetic algorithms,GA)优化的支持向量机(Support vector machine,SVM)算法建立梭子蟹质量三因子回归预测模型,估算蟹体质量,模型预测的均方根误差为 5.80 g,平均绝对百分比误差为 2.23%。唐潮等^[13]在养殖背景下采用模板匹配法分割出梭子蟹背甲,提取投影面积、甲长、甲宽 3 个特征,采用 K-means 聚类算法和支持向量回归(Support vector regression,SVR)建立双模型回归预测结构,实现梭子蟹质量预测,模型预测的均方根误差为 18.90 g,最大相对误差为 17.62%,最小相对误差为 0.67%。

现有研究中,基于机器学习的蟹体质量估算方法主要应用于梭子蟹,且精度有待进一步提高,大闸蟹养殖以池塘养殖为主,实验条件下固定拍摄高度的图像采集方法不适用,需要提供便捷精准的野外质量估算方法。

目前对大闸蟹质量估算的研究主要是从背甲的形状特征着手,但有关生理研究表明大闸蟹质量与其体色密切相关,大闸蟹背甲的色泽体现其发育状况^[14-16]。甲壳动物的色泽与其组织内的类胡萝卜素尤其是虾青素含量密切相关^[17-19],类胡萝卜素是大闸蟹性腺成熟过程中必需的营养元素,在其性腺发育过程中具有重要作用^[20-21],故色泽作为评判大闸蟹肥满和成熟程度的重要依据,被广泛应用到大闸蟹的品质分级中^[21-23]。肥满度为大闸蟹的质量与蟹壳面积的比值,蟹壳大小相同的蟹体越肥满蟹体质量越大,背甲颜色为墨绿色的大闸蟹大都肥满,为黄色的大都瘦弱^[24],因此,形状特征和颜色特征都是大闸蟹质量估算的重要依据,但是把颜色特征应用到大闸蟹质量估算的研究鲜见报道。

本文结合大闸蟹生理特性,将颜色特征应用于大闸蟹质量估算研究,在大闸蟹质量与背甲形状、背甲色泽形成机理的研究基础上,提出一种基于颜色、形状多维度特征和 LightGBM 的大闸蟹质量估算方法。首先采集蟹体图像,采用图像处理技术对图像进行分割以获取蟹体背甲图像,然后提取背甲形状特征和颜色特征,并采用标定法^[25]计算形状特征值,最后基于 LightGBM 方法预测大闸蟹质量。

1 质量估算与方法

1.1 数据采集

为获得大闸蟹质量数据和图像数据,首先把大闸蟹放置在电子秤上进行称量并记录示数,采样电子秤精度为 0.1 g,然后把大闸蟹移至由相机、底板、硬币构成的大闸蟹图像采集装置上拍摄其俯视图,采样相机为 Nikon D7100,拍摄过程中随机调整相机高度^[26],以满足大闸蟹养殖过程中蟹农采集图像时无固定高度的要求。

本文于 2020 年在江苏省宜兴市高腾大闸蟹养殖基地和北京西三旗花鸟鱼虫市场采集个体尺寸不同的大闸蟹 151 只,质量为 58.4 ~ 133.9 g,具体样本质量分布区间见表 1。大闸蟹背部呈墨绿色或灰

黑色^[27],实验中采用白色为放置蟹体的底板颜色。图像尺寸为 4 032 像素 × 3 024 像素,存储格式为 JPG,共采集图像 1 230 幅。

表 1 样本质量分布区间

Tab. 1 Sample quality distribution interval							
质量/g	0 ~ 70	70 ~ 80	80 ~ 90	90 ~ 100	100 ~ 110	110 ~ 120	> 120
图像数量/幅	71	133	200	280	307	175	64

1.2 基于计算机视觉的大闸蟹质量估算方法

本文在大闸蟹背甲的形状、色泽研究的基础上提出了基于多维度特征和 LightGBM 的大闸蟹质量估算方法。将采集到的图像进行预处理,处理后的图像提取投影面积、甲宽、甲长等形态学特征和 RGB 等颜色特征,通过 LightGBM 建立大闸蟹质量预测模型,并在智能终端上实现该方法。本研究方法的整体流程如图 1 所示。

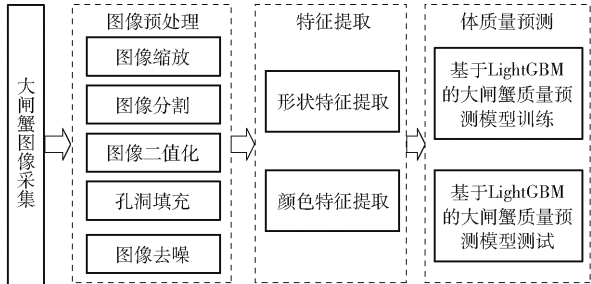


图 1 大闸蟹质量估算方法流程图

Fig. 1 Process of Chinese mitten crab weight estimation

1.3 图像预处理

图像预处理是为了消除图像中的无关信息,提高关键信息的可检测性,从而提升特征提取的效率和效果。本实验预处理包含图像缩放、图像分割、图像二值化、孔洞填充及图像去噪。

为节约图像处理的时间成本,把图像尺寸从 4 032 像素 × 3 024 像素缩放至 672 像素 × 504 像素,缩放前后图像横宽比、大闸蟹形态相同,对实验精度无影响。为获得大闸蟹背甲图像,采用 Grabcut 对胸足和背甲进行分割,如图 2 所示。

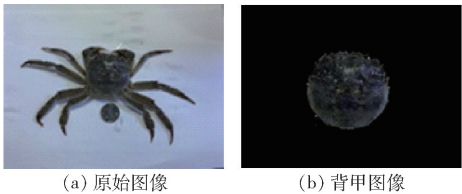


图 2 缩放、分割结果

Fig. 2 Results of image scaling and segmentation

接着采用自定义均值阈值法得到背甲的二值图像,如图 3a 所示,可见二值图像存在孔洞和噪点,这会对中华绒螯蟹形态特征提取结果造成影响,为保

证所提取的特征值能准确反映实际情况,用形态学重建的方法填补二值化图像的孔洞。形态学重建是一种基于集合膨胀、求补和交集的算法,填充的核心公式为

$$H' = [R_{I^C}^D(F)]^C \tag{1}$$

其中 $R_{I^C}^D(F) = F \oplus B' \cap I^C$
式中 H' ——孔洞填充结果图
 I^C ——原始二值图像补集
 B' ——结构元 F ——标记图像
 $\oplus$ ——膨胀操作
 $R_{I^C}^D(F)$ —— F 关于 I^C 的膨胀结果

孔洞填充结果如图 3b 所示。最后,用面积阈值法去噪,提取二值图像中每一个孤立区域轮廓面积信息保存在向量中,设置面积阈值,将面积小于阈值的区域填充背景,结果如图 3c 所示。



图 3 二值化、孔洞填充、阈值去噪结果

Fig. 3 Results of image binarization, hole filling and threshold denoising

1.4 特征变量的提取与计算

图像特征是指每一幅图像能够区别于其他类图像的本质特点或特性,是能够表示整幅图像的具有代表性的信息。图像特征包括颜色、形状、纹理等^[28]。本文主要提取大闸蟹的形状特征和颜色特征。

1.4.1 形状特征提取与计算

形状特征分为单一形状特征和不同形状特征比值两部分。投影面积、甲宽、甲长多次用于大闸蟹质量估算研究中,但是蟹螯收缩状态下的河蟹背甲与地面存在一定的角度,俯拍图像不能反映甲长的实际情况,因此选取投影面积(S')、甲宽(W)两个单一形状特征值为特征变量。

此外,在数据采集过程中随机调整拍摄高度,不同的焦距、物距使得图像的比例尺度不同,数据不具有稳定性,为了减小这一影响,除了 S' 、 W 两个特征外,还选取能反映背甲形态的其他 9 个比例参数:投影面积与外接矩形面积比(R_1)、周长平方与投影面积比(R_2)、圆形度因子(H_{cf})、等效圆直径(E_d)、等效圆直径与甲宽比(R_3)、投影面积与外接圆面积比(R_4)、投影面积与凸包面积比(R_5)、等效椭圆长轴长度与短轴长度比(R_6)、等效椭圆焦点间距与长轴长度比(R_7)。各特征几何含义如图 4 所示,图 4a

中 S' 表示轮廓投影面积, S_b 表示外接矩形面积, W 表示甲宽, C 表示轮廓周长; 图 4b 中 S_c 表示凸包面积, S_e 表示外接圆面积; 图 4c 中 a_{eq} 表示椭圆的长轴长度, b_{eq} 表示椭圆的短轴长度, c_{eq} 表示椭圆的焦距。

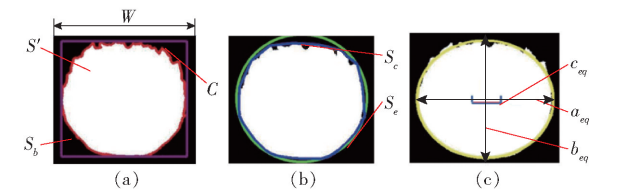


图 4 特征变量几何含义图

实验图像采集时随机调整相机高度以符合水产养殖中实际采样情况, 不同拍摄高度下的同一大闸蟹特征值不同, 本文采用标定法解决上述问题, 实际特征值计算公式为

$$V_s = \frac{V_c}{N_c} \tag{2}$$
$$V_a = V_s N_a \tag{3}$$

式中 V_s ——参照物单个像素的值
 N_c ——参照物像素总数
 V_c ——参照物实际特征值
 V_a ——大闸蟹实际特征值
 N_a ——原始图像像素个数

1.4.2 颜色特征提取与计算

提取蟹壳图像 RGB 颜色空间和 HIS 颜色空间各通道分量为颜色特征变量^[29]。RGB 颜色空间采用物理三基色表示, 以 R (红)、 G (绿)、 B (蓝) 三基色按不同比例相加混合而成; HIS 颜色空间从人类视觉系统出发, H (Hue) 表示不同的颜色, I (Intensity) 表示颜色的明暗程度, S (Saturation) 表示颜色深度。 I 分量与彩色信息无关, 故本文不予考虑, 最终提取 R 、 G 、 B 、 H 、 S 共 5 个通道分量为颜色特征变量。

对于特征变量的计算, 由于同一大闸蟹背甲的不同区域颜色差距较大, 针对单个或者局部像素点进行颜色特征值量化都不具有代表性, 本研究以像素为单位, 取整个背甲图像颜色信息平均值为特征变量, 平均化公式为

$$\begin{bmatrix} \overline{R}_{ij} \\ \overline{G}_{ij} \\ \overline{B}_{ij} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=i_0}^{i_m} \sum_{j=j_0}^{j_m} R_{ij}}{ij} \\ \frac{\sum_{i=i_0}^{i_m} \sum_{j=j_0}^{j_m} G_{ij}}{ij} \\ \frac{\sum_{i=i_0}^{i_m} \sum_{j=j_0}^{j_m} B_{ij}}{ij} \end{bmatrix} \tag{4}$$

式中 R_{ij} ——第 i 行第 j 列的 R 值
 G_{ij} ——第 i 行第 j 列的 G 值
 B_{ij} ——第 i 行第 j 列的 B 值
 $\overline{R}_{ij}$ ——背甲图像 R 通道分量均值
 $\overline{G}_{ij}$ ——背甲图像 G 通道分量均值
 $\overline{B}_{ij}$ ——背甲图像 B 通道分量均值

由 R 、 G 、 B 计算出特征变量 H 、 S , 计算公式为

$$H = \begin{cases} \theta & (B \leq G) \\ 360 - \theta & (B > G) \end{cases} \tag{5}$$

$$S = 1 - \frac{3}{R + G + B} \min(R, G, B) \tag{6}$$

其中 $\theta = \arccos \frac{\frac{1}{2} [(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{1/2}}$

综上所述, 颜色特征变量的提取流程如图 5 所示。遍历目标图像得到各个像素 R 、 G 、 B 值, 通过式(4)求得 RGB 相应的平均值 $\overline{R}_{ij}$ 、 $\overline{G}_{ij}$ 、 $\overline{B}_{ij}$ 即求出 R 、 G 、 B 值, 再由式(5)、(6)求出 H 、 S 值。

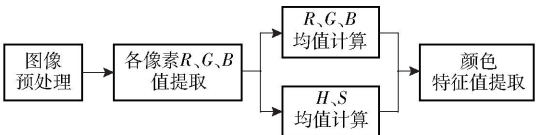


图 5 颜色特征值计算流程图

Fig. 5 Flow chart of color feature value calculation

1.5 基于 LightGBM 的大闸蟹质量预测模型

通过预测算法构建质量估算模型, 实现大闸蟹的质量估算。LightGBM 引入了基于梯度的单边采样 (Gradient-based one-side sampling, GOSS) 和互斥特征绑定 (Exclusive feature bundling, EFB) 实现对 GBDT 的优化, 取得训练的高效率和高精度^[30], 在多个预测领域均展示出了良好的效果^[31-33], 所以本文采用 LightGBM 算法对大闸蟹的质量进行预测。

初始算法模型为

$$F_M(x) = \sum_{m=1}^M T(x, \Theta_m) \tag{7}$$

式中 F_M ——提升树函数
 M ——决策树的个数
 T ——决策树函数
 x ——大闸蟹各特征变量
 Θ_m ——决策树的参数
 m ——迭代次数

将大闸蟹特征值样本放入弱学习器进行训练, 并通过最小化损失函数来确定参数 Θ_m 。

$$\Theta_m = \arg \min_{\Theta_m} \sum_{i=1}^N L(y_i, f_{m-1}(x_i, \Theta_m)) \tag{8}$$

式中 y_i ——大闸蟹质量
 x_i ——与大闸蟹质量相关的特征变量

f ——损失函数 L ——最小化损失函数
 N ——样本数据集序列长度
其中损失函数初始化为

$$f_0 = \arg \min_{\theta_m} \sum_{i=1}^N L(y_i, c) \tag{9}$$

式中 c ——损失函数极小化常数
训练过程中每次迭代所得残差与 x_i 拟合成回归树,最小化当前损失函数求出各个叶子结点的输出值 $C_{m,j}$ 为

$$C_{m,j} = \arg \min_c \sum_{x \in R_{m,j}} L(y, f_{m-1}(x_i) + c) \tag{10}$$

式中 j ——结点区域序号
 y ——大闸蟹质量
 $R_{m,j}$ ——第 m 棵树的叶结点区域
 $C_{m,j}$ 更新回归树后对每个弱学习器加权,得到最终的大闸蟹质量估算模型为

$$\hat{f}_M(x) = \sum_{m=1}^M \sum_{j=1}^J C_{m,j} I' \quad (x \in R_{m,j}) \tag{11}$$

式中 $\hat{f}_M$ ——最终得到的大闸蟹质量估算函数
 I' ——更新回归树时,参与迭代的参数
 J ——结点区域个数

1.6 模型性能评估方法
将本文方法与其他蟹体质量估算方法(本文提取的 11 个形状特征和 LightGBM (SF - LightGBM)、3 因子形状特征和 LightGBM (SF3 - LightGBM)、area -OLS^[11]、MF - SVM^[12]、MF - BPNN^[26])进行了对比。
在智能终端上实现本文方法,用户可在微信小程序上上传大闸蟹图像,从而获取大闸蟹质量。

实验开发环境为:算法编程语言为 Python 3.7 (64 位),开发环境为 Anaconda 3。图像预处理、特征提取与计算基于 OpenCV 包实现,LightGBM、SVM 模型基于 sklearn 包实现,BPNN 基于 Keras 包实现。本文采用平均绝对误差 (Mean absolute error, MAE)、均方根误差 (Root mean square error, RMSE)、决定系数 R^2 评估模型的准确性。

2 结果与讨论

2.1 图像预处理结果

图 6 为 5 只大小、姿态不同的大闸蟹图像的预处理结果。为了更好呈现预处理效果,背甲图像面积均扩大为原图的 1.5 倍。

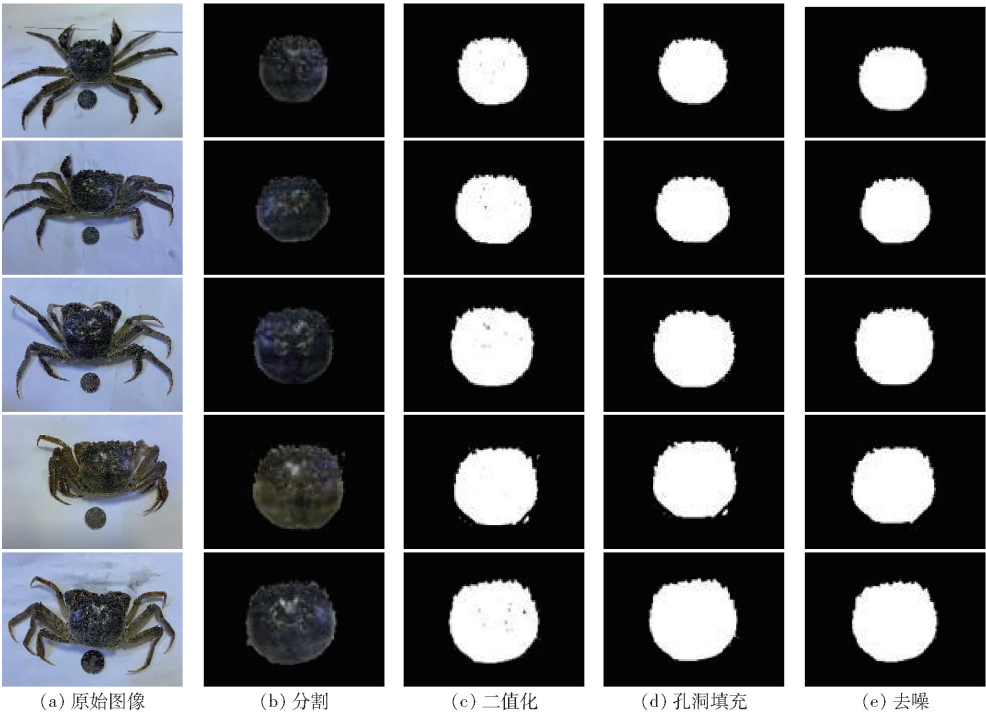


图 6 大闸蟹图像预处理结果

Fig.6 Results of Chinese mitten crab image preprocessing

2.2 特征提取结果与分析

采用定标法计算形状特征值,以同一图像在不同拍摄高度下的投影面积为例,由参照物求出单个像素面积,背甲投影面积与单个像素面积的乘积即为实际特征值,实际投影面积如表 2 所示,不同拍摄

高度下的图像经过计算后与实际特征值相近。其他形状特征计算方法同上。
对背壳图像颜色信息平均化计算颜色特征值,遍历大闸蟹背甲 RGB 图像,求得各通道分量平均值 $\bar{R}_y$ 、 $\bar{G}_y$ 、 $\bar{B}_y$,既求得了颜色特征变量 R 、 G 、 B 值,进而

表 2 实际投影面积
Tab.2 Actual projected area

序号	单个像素面积/mm ²	投影面积/像素	实际特征值/像素
1	0.469	6 047	2 836.043
2	0.385	7 441	2 846.785
3	0.321	8 910	2 860.110
4	0.301	9 455	2 845.955
5	0.277	10 248	2 838.696

求出颜色特征变量 H 、 S 值,颜色特征值结果如表 3 所示。

表 3 颜色特征值
Tab.3 Color feature value

序号	R	G	B	H	S
1	31.055 68	36.348 56	37.291 00	187.973 90	0.110 01
2	36.550 73	34.308 39	31.862 90	31.449 14	0.069 44
3	37.394 51	34.580 94	31.977 01	28.734 66	0.077 16
4	36.592 91	32.981 89	30.256 67	25.398 28	0.090 76
5	34.779 26	30.451 33	27.704 66	22.658 4	0.105 67

2.3 质量估算结果与分析

本实验共有 1 230 个样本数据,将样本数据按 7:3 划分为训练集和验证集,将 LightGBM 的采样率设置为 0.8,学习率设置为 0.1,算法输入层的回归树数目设置为 200。将回归树叶子结点数设置为 70,最大深度设置为 7,防止过拟合的同时能够获得较好的精度。通过多次实验结果表明,LightGBM 模型使用 LGBMRegressor 回归器并选取 Lasso 作为评估函数时,对大闸蟹的质量预测结果误差较小。基于 Goss 和 EFB 的 LihgtGBM 算法在几乎不损失模型精度的同时极大地加速了模型训练速度,最终实验结果 MAE 为 2.751 g, RMSE 为 3.680 g, R^2 为 0.949, 预测效果较好。图 7 为样本的预测结果散点图。

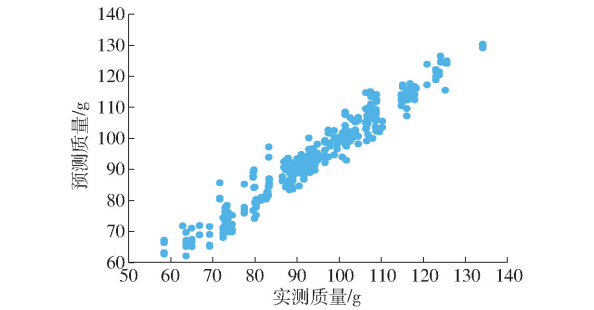


图 7 大闸蟹质量预测结果

Fig.7 Results of Chinese mitten crab weight prediction

实验共有 369 个样本预测结果,随机选取 20 个样本的预测结果组成误差分析表,如表 4 所示,所选样本实测质量为 73.3 ~ 123.9 g,预测质量 76.1 ~ 122.1 g,表内数据取样具有随机性,质量区间与全体数据集相近,故表 4 对于实验的整体结果有较好

的代表性。大部分预测结果的相对误差小于 3%,对于较大的相对误差,如样本编号 270 为 5.910%,原因可能是蟹螯和蟹壳颜色相似。

表 4 预测结果误差分析
Tab.4 Error analysis of prediction results

样本编号	实测质量/g	预测质量/g	绝对误差/g	相对误差/%
18	106.7	106.5	0.2	0.220
36	106.2	101.4	4.8	4.514
54	93.5	90.3	3.2	3.402
72	73.3	76.1	2.8	3.841
90	81.4	81.5	0.1	0.123
108	86.4	85.1	1.3	1.541
126	115.9	115.5	0.4	0.319
144	86.4	86.7	0.3	0.312
162	88.2	90.1	1.9	2.108
180	77.4	76.6	0.8	1.087
198	86.4	85.7	0.7	0.847
216	123.9	122.1	1.8	1.442
234	92.6	91.7	0.9	0.930
252	99.4	101.7	2.3	2.351
270	72.9	68.6	4.3	5.910
288	101.6	103.2	1.6	1.580
306	90.4	91.8	1.4	1.603
324	90.1	88.5	1.6	1.802
342	115.8	117.1	1.3	1.127
360	106.7	107.5	0.8	0.759

2.4 不同模型对比结果与分析

为了进一步验证本文方法的性能,将其与本文提取的 11 个形状特征和 LightGBM (SF - LightGBM)、3 因子形状特征和 LightGBM (SF3 - LightGBM)、area - OLS^[11]、MF - SVM^[12]、MF - BPNN^[26]进行对比,SVM 的参数设定为惩罚因子 $C' = 0.988\ 3$,核函数参数 $p = 2.424\ 2$;BPNN 模型的拓扑结构为 $16 \times 33 \times 1$,即输入层节点数为 16,隐含层节点数为 33,输出层节点数为 1,学习率设置为 0.01,隐含层激活函数为 ReLU 函数,输出层激活函数为 linear 函数,结果如表 5 所示。

由表 5 可知,本文方法的性能优于其他模型。与模型 SF - LightGBM 相比,平均绝对误差降低了 49.92%,均方根误差降低了 46.33%,决定系数提高了 20.89%;与 SF3 - LightGBM 相比,平均绝对误差降低了 59.00%,均方根误差降低了 54.91%,决定系数提高了 36.35%;与模型 area - OLS 相比,平均绝对误差降低了 61.90%,均方根误差降低了 58.22%,决定系数提高了 37.74%;与模型 MF - BPNN 相比,平均绝对误差降低了 44.76%,均方根误差降低了 42.27%,决定系数提高了 16.44%;与模型 MF - SVM 相比,平均绝对误差降低了 51.12%,均方根误差降低了 51.65%,决定系数提

表 5 大闸蟹质量估算方法实验对比结果

Tab.5 Experimental comparison results and analysis of weight estimation methods

模型	预测算法	提取特征	MAE/g	RMSE/g	R ²
本文方法	LightGBM	11 个形状特征 + 5 颜色特征	2. 751	3. 680	0. 949
SF - LightGBM	LightGBM	11 个形状特征	5. 493	6. 857	0. 785
SF3 - LightGBM	LightGBM	面积、甲宽、甲长	6. 710	8. 161	0. 696
area - OLS	OLS	面积	7. 220	8. 808	0. 689
MF - BPNN	BPNN	11 个形状特征 + 5 颜色特征	4. 980	6. 375	0. 815
MF - SVM	SVM	11 个形状特征 + 5 颜色特征	5. 628	7. 611	0. 755

高了 25. 70% 。

提取了颜色、形状特征的模型与仅提取形状特征的模型相比,性能得到显著提升,说明大闸蟹背甲色泽是衡量发育状况的重要指标,是估算大闸蟹质量的重要依据。当模型输入特征为面积或面积、甲宽、甲长 3 个特征时性能最差,可见形状特征提取中选取的反映背甲形态的其他 9 个比例参数,能在很大程度上解决随机调整相机高度导致的特征值数据不稳定的问题。

综上所述,本文方法具有更好的准确性,MAE、RMSE、R²均优于其他模型,预测值最接近实际测量值。

本文通过开发的小程序在智能终端上实现大闸蟹质量预测,小程序基于 Fastapi 框架,前端采用 WXML、WXSS、JS,后端采用 Python 3. 7,适用于普通智能手机。上传大闸蟹图像进行测试,大闸蟹实际质量为 107. 4 g,预测结果为 107. 8 g,预测绝对误差为 0. 4 g,相对误差为 0. 372%,本文方法在智能终端

上预测精度较高,可应用于实际养殖。

3 结论

(1) 提出了一种基于多维度特征和 LightGBM 的大闸蟹质量估算方法。提取背甲轮廓比值作为形状特征,解决了随机调整相机高度导致的特征值数据不稳定的问题,提取大闸蟹背甲颜色特征,把大闸蟹品质分级的重要依据应用到质量估算中,与仅提取形状特征模型对比,平均绝对误差降低 49. 92%,均方根误差降低 46. 33%,决定系数提高 20. 89%,模型准确率有效提高。

(2) 将本文方法与 SF - LightGBM、SF3 - LightGBM、area - OLS、MF - BPNN 和 MF - SVM 进行对比,结果表明,本文方法可以快速准确地实现大闸蟹的质量估算,MAE 为 2. 751 g, RMSE 为 3. 680 g, R² 为 0. 949,并在智能终端实现了大闸蟹质量的精准预测,对于实现不同生长阶段大闸蟹的精准投喂、生长监测、商品蟹品质分级具有重要意义。

参 考 文 献

[1] 国家统计局. 中国渔业年鉴[M]. 北京:中国农业出版社, 2019—2020.

[2] 段延娥,李道亮,李振波,等. 基于计算机视觉的水产动物视觉特征测量研究综述[J]. 农业工程学报, 2015, 31(15): 1 - 11.

DUAN Yan'e, LI Daoliang, LI Zhenbo, et al. Review on visual characteristic measurement research of aquatic animals based on computer vision[J]. Transactions of the CSAE, 2015, 31(15):1 - 11. (in Chinese)

[3] ZION B. The use of computer vision technologies in aquaculture—a review[J]. Computers and Electronics in Agriculture, 2012, 88: 125 - 132.

[4] 孙龙清,刘婷,陈帅华,等. 多目标鱼体对象提议检测算法研究[J]. 农业机械学报, 2019, 50(12): 260 - 267.

SUN Longqing, LIU Ting, CHEN Shuaihua, et al. Multi-target fish detection algorithm based on object proposals[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 260 - 267. (in Chinese)

[5] 李鑫星,董保平,杨铭松,等. 基于 SVM 核机器学习的三文鱼新鲜度检测系统[J]. 农业机械学报, 2019, 50(5): 376 - 384.

LI Xinxing, DONG Baoping, YANG Mingsong, et al. Detection system of salmon freshness based on SVM kernel-based machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 376 - 384. (in Chinese)

[6] 赵德安,曹硕,孙月平,等. 基于联动扩展神经网络的水下自由活蟹检测器研究[J]. 农业机械学报, 2020, 51(9): 163 - 174.

ZHAO Dean, CAO Shuo, SUN Yueping, et al. Small-sized efficient detector for underwater freely live crabs based on compound scaling neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 163 - 174. (in Chinese)

[7] SABERIOON M, CISAR P. Automated within tank fish mass estimation using infrared reflection system[J]. Computers and Electronics in Agriculture, 2018, 150: 484 - 492.

[8] VIAZZI S, VAN HOESTENBERGHE S, GODDEERIS B M, et al. Automatic mass estimation of jade perch *Scortum barcoo* by computer vision[J]. Aquacultural Engineering, 2015, 64: 42 - 48.

- [9] DE VERDAL H, VANDEPUTTE M, PEPEY E, et al. Individual growth monitoring of European sea bass larvae by image analysis and microsatellite genotyping[J]. *Aquaculture*, 2014, 434: 470–475.
- [10] 王静安,陈开健,刘小燕,等. 大通湖水域河蟹体长、体重及头胸甲关系研究[J]. *湖南文理学院学报(自然科学版)*, 2017, 29(1): 35–39.
- [11] 张超,徐建瑜,王文静. 基于机器视觉的梭子蟹质量估计方法研究[J]. *宁波大学学报(理工版)*, 2014, 27(2): 49–51.
- [12] 唐杨捷,胡海刚,张刚,等. 基于计算机视觉和 GA-SVM 的梭子蟹体重预测[J]. *宁波大学学报(理工版)*, 2019, 32(1): 32–37.
- [13] 唐潮,胡海刚,史策,等. 基于改进 Mask RCNN 和 SVR 的无接触梭子蟹体质量预测研究[J]. *宁波大学学报(理工版)*, 2021, 34(2): 31–41.
- [14] LONG X, WU X, ZHAO L, et al. Effects of dietary supplementation with *Haematococcus pluvialis* cell powder on coloration, ovarian development and antioxidation capacity of adult female Chinese mitten crab, *Eriocheir sinensis*[J]. *Aquaculture*, 2017, 473: 545–553.
- [15] 李清清. 中华绒螯蟹色泽遗传参数评估及色泽形成机制研究[D]. 上海:上海海洋大学, 2019.
- [16] PARISENTI J, BEIRAO L H, TRAMONTE V, et al. Preference ranking of colour in raw and cooked shrimps[J]. *International Journal of Food Science & Technology*, 2011, 46(12): 2558–2561.
- [17] 龚志,伊淑平,孔丽,等. 3 种类胡萝卜素源对大闸蟹体色改善效果的比较研究[J]. *动物营养学报*, 2014, 26(8): 2408–2413.
- GONG Zhi, YI Shuping, KONG Li, et al. Comparison of the improvement effect on body colour of Chinese mitten crab caused by three carotenoid sources[J]. *Chinise Journal of Animal Nutrition*, 2014, 26(8): 2408–2413. (in Chinese)
- [18] NIU J, WEN H, LI C H, et al. Comparison effect of dietary astaxanthin and β -carotene in the presence and absence of cholesterol supplementation on growth performance, antioxidant capacity and gene expression of *Penaeus monodon* under normoxia and hypoxia condition[J]. *Aquaculture*, 2014, 422–423: 8–17.
- [19] SUPAMATTAYA K, KIRIRATNIKOM S, BOONYARATPALIN M, et al. Effect of a *Dunaliella* extract on growth performance, health condition, immune response and disease resistance in black tiger shrimp (*Penaeus monodon*) [J]. *Aquaculture*, 2005, 248(1–4): 207–216.
- [20] DALL W, SMITH D M, MOORE L E. Carotenoids in the tiger prawn *Penaeus esculentus* during ovarian maturation[J]. *Marine Biology*, 1995, 123(3): 435–441.
- [21] LINAN-CABELLO M A, MEDINA-ZENDEJAS R, SANCHEZ-BARAJAS M, et al. Effects of carotenoids and retinol in oocyte maturation of crayfish *Cherax quadricarinatus*[J]. *Aquaculture Research*, 2004, 35(9): 905–911.
- [22] 石海军. 基于光谱图像技术的中华绒螯蟹品质鉴别研究[D]. 镇江:江苏大学, 2019.
- SHI Haijun. Quality evaluation of *Eriocheir sineusis* based on spectrum and imaging technology[D]. Zhenjiang: Jiangsu University, 2019. (in Chinese)
- [23] 韩洋. 基于机器视觉的典型产品特征提取与分级算法研究[D]. 扬州:扬州大学, 2018.
- HAN Yang. Research on typical product feature extraction and classification algorithm based on machine vision [D]. Yangzhou: Yangzhou University, 2018. (in Chinese)
- [24] 成春到. 螃蟹的挑选与食用[J]. *海洋世界*, 2000(2): 35.
- [25] WANG W, ZHANG F, WANG W. Image area measurement based on horizontal scan line[J]. *International Journal of Intelligent Information Technolog Application*, 2009, 2(1): 8–11.
- [26] ZHANG L, WANG J, DUAN Q. Estimation for fish mass using image analysis and neural network[J]. *Computers and Electronics in Agriculture*, 2020, 173: 105439.
- [27] 王武,李应森. 河蟹生态养殖[M]. 北京:中国农业出版社, 2010.
- [28] PATEL J M, GAMIT N C. A review on feature extraction techniques in content based image retrieval[C]//IEEE, 2016.
- [29] 张云,杨宏伟,罗克勇. 中华绒螯蟹蟹壳颜色的识别量化研究[J]. *江苏农业科学*, 2005(6): 115–117.
- [30] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Long Beach, CA, 2017.
- [31] LI D, PENG J, HE D. Aero-engine exhaust gas temperature prediction based on LightGBM optimized by improved bat algorithm[J]. *Thermal Science*, 2021, 25(2): 845–858.
- [32] WANG Y, WANG T. Application of improved LightGBM model in blood glucose prediction[J]. *Applied Sciences*, 2020, 10(9): 3227.
- [33] GUO Y, LI Y, XU Y, et al. Study on the application of LSTM-LightGBM model in stock rise and fall prediction[J]. *MATEC Web of Conferences*, 2021, 336: 5011.