

基于改进对抗擦除的食品图像识别方法

李艳灵*, 梁庆琪, 李佳漫, 赵添宇, 杨志鹏

(信阳师范大学 计算机与信息技术学院, 河南 信阳 464000)

摘要: 提出一种基于改进对抗擦除技术的食品识别方法, 渐进式获取判别性区域。该方法通过Otsu算法和形态学操作获得判别性区域, 降低噪声干扰。为验证所提食品识别方法的有效性, 在Sushi-50、ETH Food-101和Vireo-172数据集上, 与其他文献的方法进行对比。实验结果表明, 所提方法更有效地降低食品图像复杂背景的干扰, 提升食品识别性能。在ETH Food-101数据集上, 相较于ResNet-50, 该方法在Top-1和Top-5准确率上分别提升2.6和0.8个百分点。

关键词: 卷积神经网络; 复杂背景; 对抗擦除; Otsu算法

中图分类号: TP391.4

文献标志码: A

开放科学(资源服务)标识码(OSID):



Food image recognition based on improved adversarial erasing

LI Yanling*, LIANG Qingqi, LI Jiaman, ZHAO Tianyu, YANG Zhipeng

(College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China)

Abstract: A food recognition method was proposed based on the improved adversarial erasing technique that incrementally obtain discriminative regions. The method could identify each discriminative region using Otsu algorithm and morphological operations, thereby reducing noise interference. To validate the effectiveness of the proposed food recognition method, the comparative experiments were conducted on the Sushi-50, ETH Food-101 and Vireo-172 datasets with the methods presented in other literatures. The experimental results demonstrated that the proposed method can effectively mitigate interference from complex backgrounds in food images, thereby enhancing food recognition performance. Compared to ResNet-50, the method improved the Top-1 and Top-5 accuracy on the ETH Food-101 dataset by 2.6 and 0.8 percentage point, respectively.

Key words: convolutional neural network; complex backgrounds; adversarial erasing; Otsu algorithm

0 引言

食品不仅是人类生存的基本需求,也是文化、价值观和生活方式的重要组成部分^[1],在促进健康和预防疾病方面起着关键作用。随着肥胖问题日益严重,糖尿病和心血管疾病等多种慢性病的风险上升,科学的饮食结构愈发重要^[2]。近年来,研究者针对食品类别和成分的识别展开相关研究。食品识别作为众多健康应用的重要环节,特别是

在食品日志^[3]和健康导向的食品推荐^[4]中,发挥着不可或缺的作用。同时,食品识别作为细粒度视觉分类的重要分支,具有重要的研究意义。

与通用物体识别类似,食品识别的关键在于提取判别性视觉特征。早期的食品识别技术主要依赖人工提取的特征。例如,ZHU等^[5]提取并融合颜色和纹理特征,实现食品图像识别。随着计算机视觉技术的发展,基于卷积神经网络(convolutional neural network, CNN)的方法逐渐

收稿日期:2024-10-17;修回日期:2024-11-11;*. 通信联系人. E-mail: ly175@163.com

基金项目:河南省科技攻关项目(252102211025);河南省研究生教育优质课程项目(YJS2022KC34);信阳师范大学研究生科研创新基金项目(2024KYJJ087);河南省研究生精品教材项目立项名单(YJS2025JC30)

作者简介:李艳灵(1975—),女,河南新乡人,教授,博士,硕士生导师,主要从事模式识别及图像处理研究。

引用格式:李艳灵,梁庆琪,李佳漫,等.基于改进对抗擦除的食品图像识别方法[J].信阳师范大学学报(自然科学版),2025,38(4):408-413.

LI Yanling, LIANG Qingqi, LI Jiaman, et al. Food image recognition based on improved adversarial erasing[J]. Journal of Xinyang Normal University (Natural Science Edition), 2025, 38(4):408-413.

成为食品图像识别的主流^[6]。现有的食品识别方法大多使用CNN技术从食品图像中直接提取卷积特征进行识别。TANNO等^[7]采用深度CNN提取视觉特征进行食品识别。文献[8]也通过微调Inception V3网络模型进行识别。但这些方法未能充分考虑图像背景的复杂性,识别性能仍有待提升。

近年来,一些研究者致力于设计面向食品图像识别任务的网络模型。例如,文献[9]提出一种新的网络框架,采用弱监督学习方法充分挖掘食品图像中的显著性区域,进行食品图像识别。文献[10]提出一种区域增强网络用于食品识别,该网络由渐进的局部特征学习和区域特征增强两个模块组成。前者通过改进的渐进式训练策略学习多样化且互补的局部特征,后者利用注意力机制将局部特征与上下文信息结合,进一步提升食品识别的效果。

食品图像包含大量的背景元素,如其他物体或多样图案,这些元素易引入视觉噪声,干扰目标识别的性能。针对此问题,本文提出一种基于改进对抗擦除的食品图像识别方法。该方法采用新型网络框架OER-Net,能够学习全局特征和互补的局部特征,从而提高食品识别的准确性。

OER-Net由4部分组成:1)原始图像分类网络(original image classification network, O-Net),负责对原始图像进行分类,获得全局表示;2)擦除图像分类网络(erased image classification network, E-Net),对擦除图像进行分类,并通过大津法(Otsu算法)^[11]和形态学操作获得判别性区域;3)区域分类网络(region classification network, R-Net),负责对所有的判别性区域进行分类,从而获得互补的局部表示;4)额外的全连接层将原始图像与挖掘区域的表示进行拼接,形成稳健且全面的表示,进行最终预测。选取PARNet^[9]和ResNet^[12]模型与所提模型OER-Net进行对比实验。结果表明本文方法更有效地消除食品图像复杂背景干扰,提升识别性能。

本文的贡献有以下两个方面:

(1)提出一种基于改进对抗擦除的食品图像识别方法,该方法采用一种新型网络框架OER-Net,渐进式获取判别性区域。通过大津法(Otsu算法)减少复杂背景对目标食品识别的干扰,并通过形态学操作进一步降低噪声干扰。

(2)在3个食品数据集Sushi-50^[9]、ETH Food-101^[13]和Vireo Food-172^[14]上评估所提出方法,与PAR-Net和ResNet模型相比,本文所提方法在3个数据集上取得了最优的识别性能。

1 基于改进对抗擦除的食品识别

对抗擦除技术从一个小目标区域开始,通过对已挖掘区域进行对抗擦除,促使分类网络逐步发现新的互补目标区域,逐步定位并扩展目标区域。该方法可以有效获得判别性区域,降低复杂背景对食品识别的干扰。基于此,本文提出一种改进对抗擦除技术的食品识别方法,逐步挖掘判别性区域,采用ORE-Net网络框架进行食品识别。

1.1 判别性区域生成模块

通过应用大津法(Otsu算法),形态学操作和连通性标记方法,自适应获取判别性区域。生成判别性区域的过程如下:

首先,生成类激活图(class activation map, CAM)。CAM是一种可视化技术,用于突出显示分类网络识别目标的区域,由CNN的最后一个卷积层生成,将一幅训练图像输入模型,经过全局平均池化层(Global Average Pooling, GAP),获得特征向量,并将其连接至模型的输出层,生成CAM。计算公式如下:

$$\text{CAM}(O, c) = \sum_{k=1}^N \omega_{k,c} \cdot F_k, \quad (1)$$

式中: c 表示图像 O 的目标类别; $F_k \in \mathbf{R}^{H \times W}$ 表示第 k 个特征图,总数为 N ; $\omega_{k,c}$ 表示全连接层的权重。

然后,对生成的CAM进行上采样处理获得热力图,采用大津法(Otsu算法)进行自适应分割。大津法(Otsu算法)的具体分割过程和对应的计算公式为:

1)根据每个灰度级的频率 $P(i)$ 计算前景权重 $W_1(T)$ 和背景权重 $W_2(T)$,其中:

$$W_1(T) = \sum_{i=0}^T P(i), \quad (2)$$

$$W_2(T) = 1 - W_1(T). \quad (3)$$

2)计算前景均值 $\mu_1(T)$ 、背景均值 $\mu_2(T)$ 和类间方差 $\sigma_B^2(T)$ 。

$$\mu_1(T) = \sum_{i=0}^T (i \cdot P(i)) / W_1(T), \quad (4)$$

$$\mu_2(T) = \sum_{i=T+1}^L (i \cdot P(i)) / W_2(T), \quad (5)$$

$$\sigma_B^2(T) = W_1(T) \cdot W_2(T) \cdot (\mu_1(T) - \mu_2(T))^2 \quad (6)$$

3) 计算最大类间方差对应的最佳阈值 T^* ,

$$T^* = \arg \max \sigma_B^2(T), \quad (7)$$

式中: i 表示灰度级, T 表示当前阈值, L 表示灰度级的总数。

接着,为去除食品图像中包含的微小噪声,采用形态学操作,使用结构元素对图像进行腐蚀以去除微小的突出部分,然后进行膨胀操作以恢复原始物体的大小,确保食品轮廓完整。此过程不仅有效地平滑图像噪声,还保持了食品的主要特征。

最后,使用连通域标记技术,计算各连通区域

的像素强度之和,将识别强度最大的区域作为判别区域。

1.2 OER-Net 整体结构

OER-Net 网络结构如图 1 所示。OER-Net 由 4 个部分组成:1) O-Net 网络对原始图像进行分类获得全局表示,然后通过判别性区域模块获得首个判别区域;2) E-Net 对擦除图像进行分类,并通过判别性区域生成模块逐步挖掘判别性区域;3) R-Net 则对判别性区域进行分类,获得互补的局部表示;4) 额外的全连接层将原始图像与挖掘的区域表示拼接,实现最终预测。其中,O-Net 基于 ResNet-101, E-Net 基于 ResNet-34, R-Net 基于 ResNet-50。

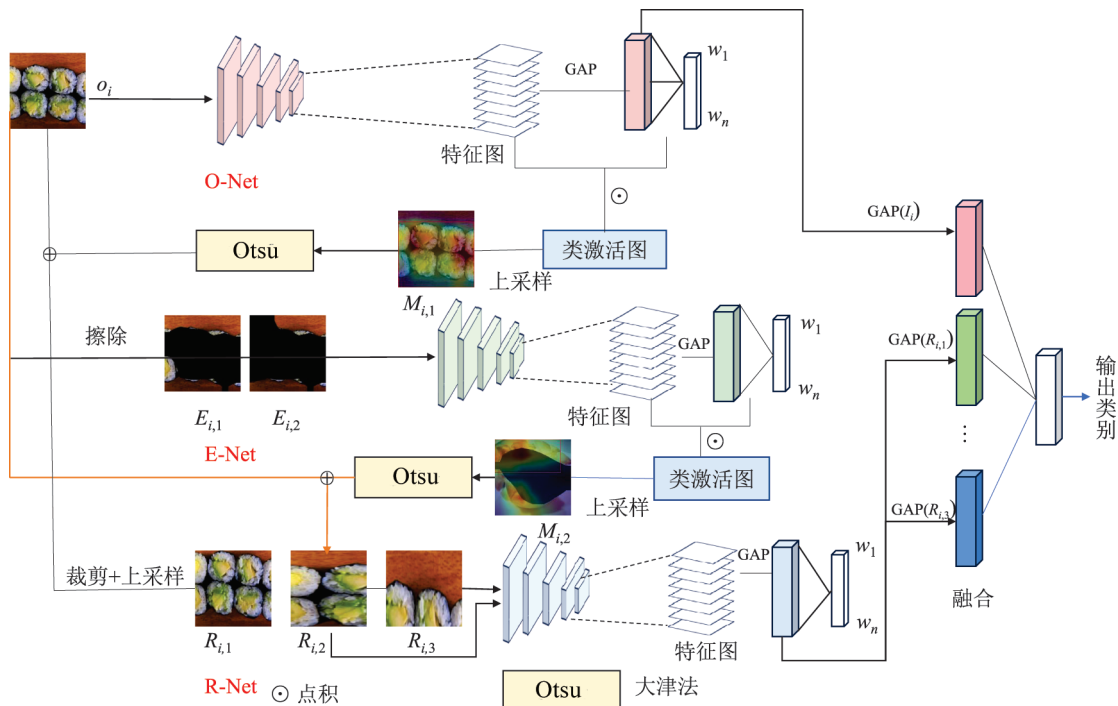


图 1 OER-Net 网络结构

Fig. 1 The network structure of OER-Net

具体来说,给定图像和真实标签对 $\{O_i, y_i\}$ 作为输入以及区域挖掘次数 N 。将原始图像 O_i 输入到 O-Net 进行分类,然后根据式(1)计算的 $CAM(O_i, y_i)$ 上采样,以热力图的形式将目标食品区域叠加在原图上,直观地展示图中的食品区域。 $M_{i,n}$ 、 $R_{i,n}$ 和 $E_{i,n}$ 分别表示基于 CAM 生成的热力图、第 n ($n \leq N$) 次挖掘步骤获得的挖掘区域和擦除图像。所有的 $R_{i,n}$ 和 $E_{i,n}$ 都使用原始图像 O_i 的真实标签 y_i 。在上采样后,第一张热力图 $M_{i,1}$ 的大小与 O_i 相同。对 $M_{i,1}$ 应用 Otsu 方法,并通过形态学操作去除食品图像中的小噪声,使用连通性标

记技术得到判别性区域 $R_{i,1}$ 。接着,使用紧凑边界框裁剪该区域,通过双线性插值将裁剪后的图像块上采样到与 O_i 相同尺寸,并输入到 R-Net 进行识别。然后进行第一次擦除操作,将 O_i 中包含 $R_{i,1}$ 内的像素置零,得到第一张擦除图像 $E_{i,1}$,将 $E_{i,1}$ 输入到 E-Net 进行分类,继续通过 CAM 得到判别性区域,而 R-Net 则继续识别由 E-Net 提供的判别性区域。从 O-Net 和 R-Net 的 GAP 提取原始图像 O_i 的全局表示和区域 $R_{i,n}$ 的局部表示,将这些表示拼接构成一个全面的特征,最后将全面的特征表示输入到额外的全连接层进行分类。随着

挖掘次数的增加, $E_{i,n}$ 将包含更少的判别区域。因此, 设置挖掘次数 $N=3$, 得到判别性区域图像 ($R_{i,1}, R_{i,2}, R_{i,3}$) 和擦除图像 ($E_{i,1}, E_{i,2}$)。

1.3 损失函数

OER-Net 涉及的所有分类任务应用交叉熵损失。交叉熵损失函数的计算公式为:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \ln(\hat{y}_{i,c}), \quad (8)$$

式中: L_{CE} 为交叉熵损失, N 为样本数, C 为类别数, $y_{i,c}$ 为实际标签, $\hat{y}_{i,c}$ 为第 i 个样本属于类别 c 的预测概率。

OER-Net 的损失函数由 4 部分组成: O-Net 分类原始图像 O_i 得到的损失 $L_{o,n}$, E-Net 分类 $E_{i,n}$ 得到的损失 $L_{e,n}$, R-Net 分类判别性区域 $R_{i,n}$ 得到的损失 $L_{r,n}$, 以及额外的全连接层分类图像 O_i 和判别性区域 $R_{i,n}$ 的拼接表示得到的损失 L_{concat} 。

当 $N=0$ 时, 输入图像 O_i 仅通过 O-Net 进行识别, 此时模型总损失 L 的值等于 L_o 。当 $N=1$ 时, 计算 E-Net 的损失 $L_{e,1}$ 和额外的全连接层损失 L_{concat} 。E-Net 的损失仅在成功挖掘判别性区域时计算。对于 $N=3$, OER-Net 的总损失 L 定义为:

$$L = L_o + \sum_{n=1}^3 L_{r,n} + L_{concat} + \sum_{n=1}^2 L_{e,n}, \quad (9)$$

式中: n 为挖掘次数。

2 实验结果与分析

2.1 食品图像数据集

实验使用 3 个公开的通用食品图像数据集: Sushi-50^[9]、ETH Food-101^[13] 和 Vireo-172^[14]。Sushi-50 数据集包含 3963 张食品图像, 共 50 个类别, 训练集和测试集的比例约为 7:3。ETH Food-101 数据集包含 101 个类别, 每个类别含 1000 张图像, 随机选取 75% 的图像作为训练集, 25% 的图像作为测试集。Vireo-172 是一个大规模的中国食品图像数据集, 包含 110 241 张图像, 分为 172 个类别, 其中 60% 的图像用于训练, 10% 用于验证, 30% 用于测试。

2.2 实验设置

实验平台硬件配置如下: GPU 为 NVIDIA A100 80 GB PCIe, 内存为 80 GB。实验平台软件配置如下: 操作系统为 Linux, Python 版本为 3.7.16, 深度学习框架 PyTorch 的版本为 1.13.1, CUDA 版本为 11.7, cuDNN 版本为 9.4.0。

OER-Net 的每个子网络均基于 ResNet^[12]。在

训练过程中, 从原始图像中随机采样单个裁剪, 并通过缩放和纵横比增强将其调整为 224×224 的输入大小。在测试阶段, 使用中心裁剪从调整至 256×256 的原始图像中提取 224×224 大小的图像。

2.3 评价指标

Top-1 准确率是指模型预测的第一类别与真实结果一致的准确率, 而 Top-5 准确率则是指模型预测的前五个类别中包含真实结果的准确率。这两个指标是食品图像识别任务常用指标, 因此, 本文选择这两个指标评估 OER-Net 性能。

2.4 消融实验

评估 OER-Net 各组件间的有效性, 在 3 个数据集上对各组件识别性能试验。表 1 和表 2 展示了原始图像 O , 通过挖掘步骤 n 获得的判别性区域 R_n 、擦除图像 E_n 以及拼接表示 concat 的识别性能。

表 1 O, R_n, E_n 和 concat 的 Top-1 准确率

Tab. 1 Top-1 accuracies on O, R_n, E_n and concat

图像区域	Top-1 准确率/%		
	Sushi-50	ETH Food-101	Vireo-172
O-Net(O)	89.6	88.0	88.3
R-Net(R_1)	89.8	88.2	88.8
R-Net(R_2)	86.0	83.3	85.1
R-Net(R_3)	82.7	72.2	74.9
E-Net(E_1)	34.5	71.9	70.2
E-Net(E_2)	17.3	45.8	38.9
concat	92.1	89.8	90.2

表 2 O, R_n, E_n 和 concat 的 Top-5 准确率

Tab. 2 Top-5 accuracies on O, R_n, E_n and concat

图像区域	Top-5 准确率/%		
	Sushi-50	ETH Food-101	Vireo-172
O-Net(O)	98.5	97.6	97.7
R-Net(R_1)	98.3	97.0	97.9
R-Net(R_3)	97.7	95.7	96.7
R-Net(R_3)	96.2	90.3	92.6
E-Net(E_1)	67.7	91.7	90.9
E-Net(E_2)	47.2	74.7	68.2
concat	98.8	97.9	98.3

由表 1 和表 2 可以看出, 与单一模型相比, OER-Net 能够提升食品图像的识别性能。例如, 在 ETH Food-101 数据集上, 与 O-Net 分类性能相比, OER-Net 在 Top-1 准确率上提升了 1.8 个百分

点, OER-Net 在 Top-5 准确率上提升了 0.3 个百分点。实验结果表明, 随着挖掘次数 n 的增加, 挖掘区域分类准确率逐渐下降, 因为后挖掘区域比前挖掘区域的判别性更弱。同样, 对于擦除判别性区域的图像, 由于后续图像与先前图像相比剩余判别性区域减少, 因此分类性能会下降。在 3 个数据集上的实验结果表明, 拼接表示准确率最优, 验证了全局和局部表示相结合的有效性。

2.5 对比性分析

表 3、表 4 和表 5 展示了 ResNet-50、ResNet-101 和 PAR-Net 与本文提出的 OER-Net 在 3 个数据集上的识别性能。

表 3 数据集 Sushi-50 上的实验结果对比 %
Tab. 3 The comparison of experimental results on the Sushi-50 dataset %

方法	Top-1 准确率	Top-5 准确率
ResNet-50 ^[12]	88.7	98.3
ResNet-101 ^[12]	89.6	98.5
PAR-Net ^[9]	91.9	98.6
OER-Net(本文方法)	92.1	98.8

从表 3、表 4 和表 5 可以看出, 本文方法的识别性能均优于 ResNet-50、ResNet-101 和 PAR-Net。在 Sushi-50 数据集上, 与 ResNet-50 相比, OER-Net 的 Top-1 准确率和 Top-5 准确率分别提升了 3.4 个百分点和 0.5 个百分点; 与 ResNet-101 相比, OER-Net 的 Top-1 和 Top-5 准确率分别提升了 2.5 个百分点和 0.3 个百分点; 与 PAR-Net 相比, OER-

表 4 数据集 ETH Food-101 上的实验结果对比 %
Tab. 4 The comparison of experimental results on the ETH Food-101 dataset %

方法	Top-1 准确率	Top-5 准确率
ResNet-50 ^[12]	87.2	97.1
ResNet-101 ^[12]	88.0	97.6
PAR-Net ^[9]	89.6	97.7
OER-Net(本文方法)	89.8	97.9

表 5 数据集 Vireo-172 上的实验结果对比 %
Tab. 5 The comparison of experimental results on the Vireo-172 dataset %

方法	Top-1 准确率	Top-5 准确率
ResNet-50 ^[12]	88.1	97.3
ResNet-101 ^[12]	88.3	97.7
PAR-Net ^[9]	89.9	97.9
OER-Net(本文方法)	90.4	98.3

Net 的 Top-1 和 Top-5 准确率均提升了 0.2 个百分点。在其他数据集上的实验结果同样表明, 所提方法性能均优于其他方法。与其他方法相比, 本文的改进方法能更有效地提升识别性能。

2.6 可视化分析

为进一步验证方法的有效性, 对 OER-Net 的实验结果进行可视化分析。图 2 展示 OER-Net 在一些食品图像上的可视化分析结果。其中区域 1、区域 2 和区域 3 分别表示 OER-Net 获得的判别性区域 $R_{i,1}$ 、 $R_{i,2}$ 和 $R_{i,3}$ 。热力图 1、热力图 2 和热力图 3 分别表示通过 O_1 、 $E_{i,1}$ 和 $E_{i,2}$ 生成的热力图。

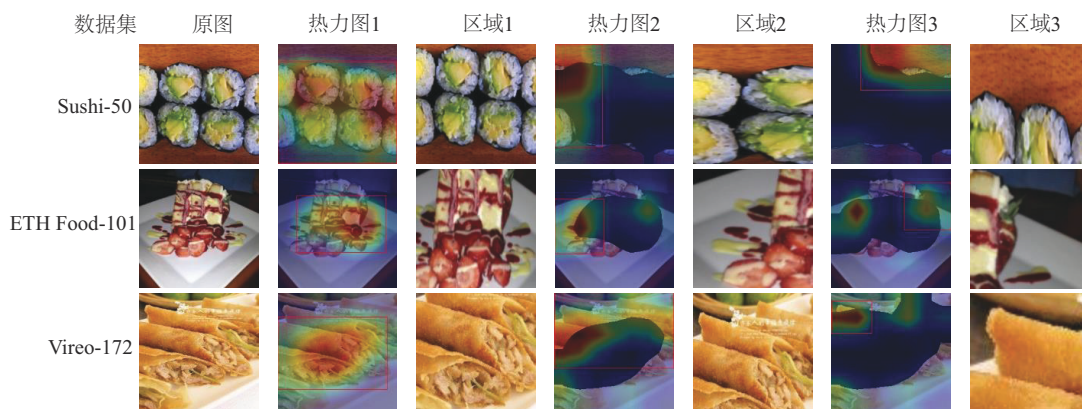


图 2 OER-Net 的可视化分析结果

Fig. 2 The visualization results of OER-Net

对于具有复杂背景的食品图像, 仅使用单一 CNN 难以提取关键特征, 从而更易受背景干扰。在可视化结果的每一行中, 使用基于 ResNet-101

的 O-Net 对原图像进行预测的结果均错误, 而通过整个 OER-Net 模型的预测结果均正确。本文的改进方案能够针对性地解决食品图像因背景复杂而

导致的识别率较低的问题。可视化结果也验证了所提方法的有效性,并表明该方法在多个食品数据集上表现出一致的性能。

3 结论

为解决食品图像复杂背景对食品目标识别的干扰问题,提出了一种基于改进对抗擦除的食品

图像识别方法。该方法使用判别性区域生成模块,自适应生成判别性区域,从而降低背景噪声干扰。通过将原始图像的全局特征与判别性区域的局部特征相结合,构建更稳健全面的特征表示,实现最终预测。实验结果表明,该方法能有效抑制复杂背景干扰,突出判别性区域,进一步提升模型的识别性能。

参考文献:

- [1] ENRIQUEZ J P, ARCHILA-GODINEZ J C. Social and cultural influences on food choices: A review[J]. *Critical Reviews in Food Science and Nutrition*, 2022, 62(13): 3698-3704.
- [2] 衡双平,崔梦迪,李晓琳,等. 芥菜紫色叶片相关分子标记的开发和应用[J]. *信阳师范学院学报(自然科学版)*, 2024, 37(3): 360-365.
HENG Shuangping, CUI Mengdi, LI Xiaolin, et al. Development and application of molecular markers related to purple leaf in *Brassica juncea* [J]. *Journal of Xinyang Normal University (Natural Science Edition)*, 2024, 37(3): 360-365.
- [3] 杨文梅,元成伟,赵纪国,等. 基于卷积神经网络的食品图像识别与营养分析[J]. *中国食品工业*, 2024(24): 118-120.
YANG Wenmei, YUAN Chengwei, ZHAO Jiguo, et al. Food image recognition and nutrition analysis based on convolutional neural network[J]. *China Food Industry*, 2024(24): 118-120.
- [4] CUNHA C, REBELO J, DUARTE R. Unveiling neural networks for personalized diet recommendations [J]. *Procedia Computer Science*, 2024, 241: 606-611.
- [5] ZHU Fengqing, BOSCH M, SCHAP T R, et al. Segmentation assisted food classification for dietary assessment [C]//*International Society for Optics and Photonics. Computational Imaging IX*. Bellingham, WA: SPIE, 2011, 7873: 77-84.
- [6] SHAH B, BHAVSAR H. Depth-restricted convolutional neural network-a model for Gujarati food image classification [J]. *The Visual Computer*, 2024, 40(3): 1931-1946.
- [7] TANNO R, OKAMOTO K, YANAI K. DeepFoodCam: A DCNN-based real-time mobile food recognition system [C]//*Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. New York: ACM, 2016: 89-89.
- [8] HASSANNEJAD H, MATRELLA G, CIAMPOLINI P, et al. Food image recognition using very deep convolutional networks [C]// *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. New York: ACM, 2016: 41-49.
- [9] QIU J, LO F P W, SUN Y, et al. Mining discriminative food regions for accurate food recognition[J]. *arXiv Preprint arXiv:2207.03692*, 2022.
- [10] MIN Weiqing, WANG Zhiling, LIU Yuxin, et al. Large scale visual food recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 9932-9949.
- [11] 刘勍,黄金,张亚亚,等. 基于灰狼优化算法的PCNN中药材显微图像分割[J]. *信阳师范学院学报(自然科学版)*, 2024, 37(1): 120-126.
LIU Qing, HUANG Jin, ZHANG Yaya, et al. Microscopic image segmentation of Chinese herbal medicine based on gray wolf optimization PCNN algorithm[J]. *Journal of Xinyang Normal University (Natural Science Edition)*, 2024, 37(1): 120-126.
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [13] BOSSARD L, GUILLAUMIN M, VAN GOOL L. Food-101-mining discriminative components with random forests [C]//*Computer Vision-ECCV 2014*. Zurich: Springer, 2014: 446-461.
- [14] CHEN Jingjing, NGO C W. Deep-based ingredient recognition for cooking recipe retrieval[C]// *Proceedings of the 24th ACM International Conference on Multimedia*. New York: ACM, 2016: 32-41.

责任编辑:郭红建 陈松楠