

doi:10.6041/j.issn.1000-1298.2021.11.024

基于注意力机制的葡萄品种多特征分类方法

苏宝峰^{1,2} 沈磊^{1,2} 陈山^{1,2} 米志文^{1,2} 宋育阳³ 陆南⁴

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100; 2. 农业农村部农业物联网重点实验室, 陕西杨凌 712100; 3. 西北农林科技大学葡萄酒学院, 陕西杨凌 712100; 4. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100)

摘要: 针对田间自然背景下葡萄品种鉴别缺乏有效识别方法的问题,提出了一种基于融合注意力机制的残差网络 ResNet50-SE,对自然背景下不同生长时期的葡萄品种进行分类鉴别,分析并验证了网络的识别效果。将 SE 注意力模块引入 ResNet-50 网络,并通过迁移学习实现基于不同时期下葡萄的嫩梢、幼叶及成熟叶片特征的识别;同时为了揭示注意力机制的作用机制,利用 Grad-CAM 可视化方法,对 ResNet50-SE 模型每一层所提取的不同生长阶段下的葡萄特征进行可视化解释;通过 t-SNE 算法对模型提取到的不同葡萄品种的多特征进行聚类分析,进而直观评估模型对多特征提取的性能。结果表明:提出的 ResNet50-SE 网络在田间复杂背景条件下对于葡萄不同时期的多特征识别具有较高的识别率和较强的鲁棒性,模型测试集准确率达到 88.75%,平均召回率达到 89.17%,相比于 AlexNet、GoogLeNet、ResNet-50、VGG-16,测试集准确率分别提高了 13.61、7.64、0.70、6.53 个百分点;注意力机制能明显降低背景影响,强化有效特征;模型对训练集提取的不同生长时期的特征聚类效果较强。可见,SE 模块可明显提升 ResNet-50 模型在特征提取过程的效果,有效降低田间复杂背景对分类结果的影响,为田间复杂背景下葡萄品种的分类识别及田间多特征分类问题提供借鉴。

关键词: 葡萄品种;分类;注意力机制;可视化

中图分类号: S24 文献标识码: A 文章编号: 1000-1298(2021)11-0226-08 OSID: 

Multi-features Identification of Grape Cultivars Based on Attention Mechanism

SU Baofeng^{1,2} SHEN Lei^{1,2} CHEN Shan^{1,2} MI Zhiwen^{1,2} SONG Yuyang³ LU Nan⁴

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
3. College of Enology, Northwest A&F University, Yangling, Shaanxi 712100, China
4. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: In view of the lack of effective identification methods for grape cultivars identification under the field natural background, a residual network ResNet50-SE based on attention fusion mechanism was proposed to classify and identify grape varieties in different growth periods under natural background, and the identification effect of the network was analyzed and verified. The SE attention module was introduced into ResNet-50 network, and the recognition of grape shoots, young leaves and mature leaves in different periods was realized through transfer learning. Besides, in order to reveal the attention mechanism, the grape characteristics of different growth stages extracted from each layer of ResNet50-SE model were visualized and explained by the Grad-CAM visualization method. The t-SNE algorithm was applied to cluster the multi-features of different grape varieties extracted by the model, and then the performance of multi-features extraction of the model was intuitively evaluated. The results indicated that the ResNet50-SE network had a high recognition rate and strong robustness for grape multi-features recognition in different periods under the complex background conditions in the field. The accuracy rate of the model test set reached 88.75%, and the average recall rate reached 89.17%. Compared with AlexNet, GoogLeNet, ResNet-50 and VGG-16, the accuracy of the test set was improved by 13.61,

收稿日期: 2021-07-07 修回日期: 2021-08-15
基金项目: 广西重点研发计划项目(桂科 21076001)、宁夏酿酒葡萄病虫害绿色防控关键技术创新与示范项目(2019BBF02013)和陕西省重点研发计划项目(2021NY-041)
作者简介: 苏宝峰(1981—),男,副教授,博士,主要从事田间植物表型快速获取及应用研究,E-mail: bfs@nwsuaf.edu.cn

7.64, 0.70 and 6.53 percentage points. The attention mechanism can significantly reduce the influence of the background and strengthen the effective features. The model had a strong clustering effect on the features of different growth periods extracted from the training set. Therefore, the SE module can obviously improve the effect of ResNet-50 model in the feature extraction process, and effectively reduce the impact of field complex background on the classification results. The research result can provide a reference for the classification and recognition of grape cultivars multi-features under field complex background.

Key words: grape cultivars; classification; attention mechanism; visualization

0 引言

葡萄品种的有效分类可以帮助葡萄种植者更为方便管理与精准决策^[1]。葡萄在发育成熟过程中,容易受到病害的侵染,了解特定品种对特定疾病的易感性,对其进行有效识别,将有助于更具体和更有针对性地防治葡萄病害^[2]。

传统上,鉴定葡萄品种的方法通常需要对叶片破坏性取样或者手动提取图像特征,其具有一定破坏性且耗时耗力^[3-6]。以卷积神经网络为代表的深度学习^[7-10]方法,在图像的分类中受到广泛应用。文献[11]基于 54 幅榕属植物叶片图像,利用人工神经网络和支持向量机方法,提出了一种自动识别 3 种具有相似叶形态的榕属植物的系统,该系统对榕属植物叶片图像识别准确率达到 83.3%,证明了深度学习对叶片图像的识别能力。

田间环境复杂多变,不同葡萄品种之间的图像具有高度相似性,且易受到自然背景干扰。因此,在田间自然环境背景下识别葡萄品种是一个难点。对于葡萄品种识别的早期工作中,文献[12]提出了一种在夜间环境下针对于彩色图像中的葡萄串检测和定位的系统来区分白葡萄和红葡萄。作物真实的生长环境条件具有高度异质性^[13],并且很难重复。通过夜间受控环境来消除自然光照及背景条件的影响受限于特定的应用场景,难以在实际中推广与应用。

文献[2]提出了一种基于 AlexNet 体系结构和迁移学习^[14-16]的方法,在田间环境下通过葡萄叶片图像对 6 个葡萄品种进行识别分类,测试准确率仅为 77.30%。该研究利用葡萄叶片图像鉴别葡萄品种的方法,仅限于葡萄特定的生长时期。然而,不同生长阶段下葡萄叶片的尺寸、颜色、形态与纹理结构具有明显的差异^[17],只针对某一时期的葡萄叶片识别在实际生产应用上存在一定的局限性。

葡萄品种的分类属于典型的图像多特征分类问题。然而以上研究多针对单一时期葡萄特征开展研究,模型难以泛化到更为异构的数据集,且在特征学习中因自然环境的影响而受限于特定场景^[14,18]。因此,为了能够解决田间复杂背景条件下对不同生

长时期的葡萄品种鉴别问题,本文在田间自然环境下采集 12 类不同品种不同时期葡萄特征图像,通过引入注意力机制降低自然环境影响,改进 ResNet-50^[19],提出一种在田间环境下对葡萄多特征图像鉴别的方法,并对深度学习可解释性差的问题,通过可视化对网络作用机制进行解释,以期对不同生长时期的葡萄品种进行可靠的识别分类。

1 材料与方法

1.1 实验数据集制备

本实验图像采集地点为陕西省杨陵区曹新庄实验农场酿酒葡萄生产示范基地,收集了 12 类酿酒葡萄品种在不同生长时期的嫩梢、幼叶以及成熟叶片特征图像,如图 1 所示。采集时间集中在 2020 年 6—8 月,都在晴天以及阴天的自然光线条件下进行。相机型号为 SONY ILCE-5100L,分辨率为 3 008 像素×1 668 像素,光圈值为 f/3.2,曝光时间为 1/60 s。每类品种每种特征各 200 幅,合计 7 200 幅图像。

将收集到的图像按 7:2:1 的比例随机划分为训练集、验证集以及测试集。为了保持每类样本特征图像数量均衡、消除随机性,每类品种中的 3 种特征图像的数量均相等。

1.2 实验环境与条件

实验在 Linux 系统环境下进行 (Ubuntu 20.04),计算机显卡为 NVIDIA Titan RTX,显存 24 GB,CPU 型号为 Intel(R) Xeon(R) Silver 4214@2.20 GHz,深度学习框架为 Pytorch (Python 3.6, Pytorch 1.2, torchvision 0.4)。

1.3 模型构建与训练参数设计

针对田间自然背景环境下对葡萄多特征识别分类存在的问题,本研究基于注意力机制,通过改进 ResNet-50,提出一种新的网络模型 ResNet50-SE,并利用迁移学习方法对模型进行训练,以实现不同生长时期下葡萄品种鉴别。

1.3.1 SE 模块

SENet^[20] (Squeeze-and-excitation networks) 核心思想在于通过网络学习特征权重,放大有效的特征

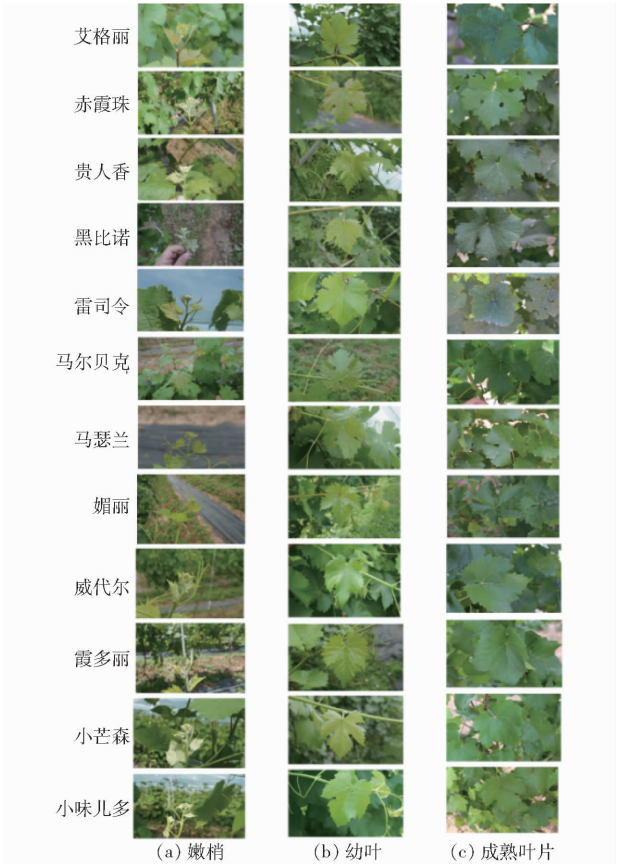


图 1 不同生长阶段下葡萄特征图像

Fig. 1 Grape characteristic images in different growth stages

通道,抑制无效或效果小的特征通道,从而使训练模型达到更好的结果。SE 模块为其核心结构,是一种注意力机制,如图 2 所示。SE 模块主要由挤压 (Squeeze, F_{sq}) 和激发 (Excitation, F_{ex}) 两步操作组成,Squeeze 操作先将输入尺寸为 $H \times W \times C$ 的特征图 (其中 H 、 W 、 C 分别为输入特征图长、宽和特征通道数) 采用全局平均池化 (Global average pooling) 计算,得到 $1 \times 1 \times C$ 的特征向量,使其具有全局感受野,然后进入两层全连接层,全连接层之间采用 ReLU 激活函数,并且输入与输出的特征通道维度相同。Excitation 操作通过最后的 Sigmoid 激活函数计算生成每个特征通道的权重。最后 Scale 操作 (F_{scale}) 将输出的特征通道权重向量与原输入特征图相乘,完成在通道维度上的原始特征标定,使提取的特征具有更强的指向性,从而提高分类性能^[21]。

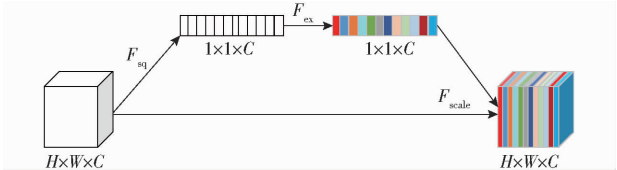


图 2 SE 模块结构示意图

Fig. 2 Structure diagram of SE module

1.3.2 ResNet 模型

随着网络层数的加深,网络性能逐渐出现退化问题^[19],ResNet 模型通过堆叠残差结构,可有效解决这一问题^[22]。残差结构如图 3 所示, X 为网络输入,它使用了一种“Shortcut connection”连接方式,将网络原本要学习的特征 X 从捷径分支与从经过权重层学到的特征 $F(X)$ 直接相加,并通过 ReLU 函数进行激活。ResNet-50 作为其经典结构,是目前主流的图像分类方法之一。

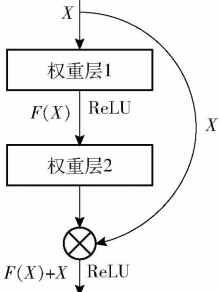


图 3 残差模块

Fig. 3 Residual module

1.3.3 改进的 ResNet-50 模型

通过对 ResNet-50 模型多特征识别分类精度以及可视化结果分析,在田间复杂背景条件下,容易受到背景影响,使模型提取到一些无效信息,从而降低分类性能。基于注意力机制的方法,使得模型关注并加强有效的分类信息,而忽略掉一些无用的特征信息,从而提高模型的鲁棒性。鉴于 SE 模块对特征提取的校正性作用,本研究提出一种将 SE 模块嵌入 ResNet-50 模型结构,为区别于其它 SE 模块嵌入 ResNet-50 的网络模型,将其命名为 ResNet50-SE,模型结构如图 4 所示。

ResNet50-SE 模型通过在每个残差结构中嵌入 SE 模块以及在模型第 1 层 (Conv1) 和输出层中引入 SE 模块对网络提取到的特征进行校正。网络主要的组成结构为 Bottleneck-SE 结构 (图 4),主要有 4 种不同的类型,其数量分别为 3、4、6、3。Bottleneck-SE 结构所使用的卷积核尺寸相同,主要区别在于卷积核的数量 (即卷积核维度) 以及 SE 模块中 2 个全连接层的输出维度不同,这是由于上一层残差结构输出通过与下一层的输出直接相加时维度需匹配。详细的网络结构参数如表 1 所示。

1.3.4 模型训练与参数设计

由于大部分数据或任务都存在相关性,故本研究通过加载在 ImageNet 数据集^[23]上预训练网络 ResNet-50 的部分权重信息对 ResNet50-SE 模型进行迁移学习,以加快模型的收敛速度。在模型训练过程中,将图像缩放至 224 像素 $\times$ 224 像素 $\times$ 3 作

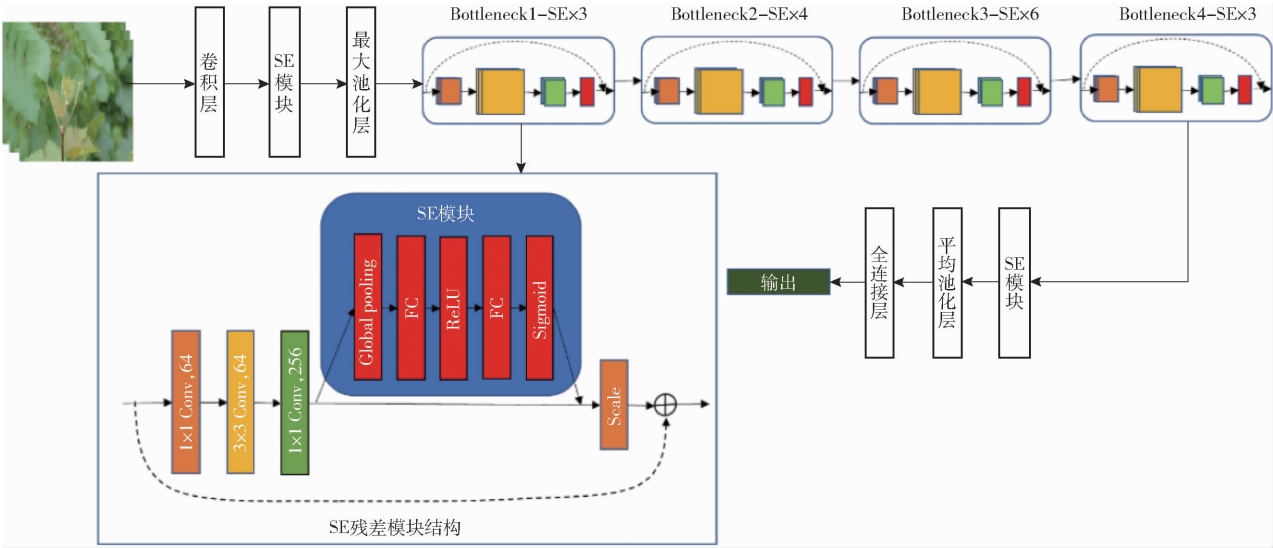


图 4 ResNet50 – SE 网络结构

Fig. 4 ResNet50 – SE network structure

表 1 ResNet50 – SE 网络结构参数

Tab. 1 ResNet50 – SE network structure parameters

层数	输出尺寸/ (像素 × 像素)	网络结构
Conv1	112 × 112	$\begin{pmatrix} 7 \times 7, 64 \\ SE(4, 64) \\ \text{最大池化层} \end{pmatrix}$
Conv2_x	56 × 56	$\begin{pmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \\ SE(16, 256) \end{pmatrix} \times 3$
Conv3_x	28 × 28	$\begin{pmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \\ SE(32, 512) \end{pmatrix} \times 4$
Conv4_x	14 × 14	$\begin{pmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1\,024 \\ SE(64, 1\,024) \end{pmatrix} \times 6$
Conv5_x	7 × 7	$\begin{pmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2\,048 \\ SE(128, 2\,048) \end{pmatrix} \times 3$
输出层	1 × 1	SE(128, 2 048) 平均池化层 全连接层 SoftMax (12 维)

注:SE 括号内数字为 SE 模块 2 个全连接层的输出维度。

为网络输入尺寸。为了提高模型的鲁棒性,进行数据集增强,随机对图像进行水平、垂直翻转,同时对图像进行归一化处理。

通过多次组合实验,综合考虑到模型准确率和收敛时间。在训练时批次量 (Batch size) 设置为 16 个样本,初始学习率 (Learning rate) 设置为 0.001,迭代周期 (Epoch) 设置为 40。为防止模型过拟合,

设置权重衰减 (Weight decay) (也称为 L2 正则化) 为 10^{-4} 。同时,在每个 Epoch 下进行训练和验证前,数据都进行了随机打乱。

本实验采用带动量的随机梯度下降算法^[14,24] (Stochastic gradient descent with momentum, SGDM) 更新参数和优化训练过程,参数更新函数表达式为

$$\theta_{i+1} = \theta_i - \alpha \Delta L_R(\theta_i) + m(\theta_i - \theta_{i-1}) \quad (1)$$

式中 i ——迭代次数

θ ——网络参数 (权重矩阵 W 和偏置向量 b)

$\Delta L_R(\theta_i)$ ——损失函数梯度

m ——动量因子 α ——学习率

训练过程中 m 设置为 0.9,学习率 α 调整采用等间隔调整策略,每训练 20 个 Epoch,学习率调整为 $\alpha\gamma$,其中 γ 为学习率调整倍数,设置为 0.1。

1.4 模型评估指标

在评估一个模型的性能时通常选取准确率 (Accuracy)、精确度 (Precision)、召回率 (Recall) 及特异度 (Specificity) 作为评价指标^[14]。其中特异度计算公式为

$$S = \frac{T_n}{F_p + T_n} \times 100\% \quad (2)$$

式中 S ——特异度

F_p ——假阳性样本数量

T_n ——真阴性样本数量

1.5 特征可视化

对于深度学习网络可解释性差的问题,本研究利用 Grad-CAM^[25] 的方法对模型在多特征提取中做出可视化解释,并通过 T-SNE^[26] 算法在二维空间中可视化高维特征,对训练集提取到的多特征进行聚类分析进而直观评估模型的分类性能。

1.5.1 Grad-CAM 可视化

通过 Grad-CAM 对模型每一层提取的特征进行了可视化解释,以热力图的形式呈现模型用作分类最为关注的特征。它主要通过采用梯度的全局平均来计算特征的权重,首先计算 c 类别的模型得分对于某个卷积层的梯度,同时对上述过程得到的梯度信息,在每个通道维度上对各像素值取平均(类似全局平均池化),得到神经元重要性权重,第 c 类的分类得分为 S_c ,特征图尺寸为 $c_1 \times c_2$,则

α_i^c = 1/Z ∑_{k=1}^{c₁} ∑_{j=1}^{c₂} ∂S_c / ∂A_{kj}ⁱ (3)

式中 Z——特征图的像素数
A_{kj}ⁱ——第 i 个特征图的第 k 行、第 j 列像素值
α_i^c—— c 类相对于最后一层卷积层输出的特征图的第 i 个通道权重

通过加权平均,再取 ReLU 激活函数得到 Grad-CAM 特征图,计算公式为

L_{Grad-CAM}^c = ReLU (∑_i α_i^c A_i) (4)

式中 L_{Grad-CAM}^c—— c 类的激活热力图
A_i——第 i 个特征图

1.5.2 t-SNE 聚类

t-SNE 是一种表示特征空间分布的方法,它通过在二维或三维图中给出每个数据点的位置来可视化高维数据,可直观地评估模型分类性能。t-SNE 算法主要依据具有对称性的联合概率分布 P_{ij} 来表示高维空间中样本点之间的距离^[27]。本研究针对训练集中的 12 类葡萄品种的 5 040 幅葡萄特征图像,利用 t-SNE 算法观察每幅图像的高维特征分布。实验从葡萄分类模型 (ResNet50-SE) 的全连接层提取每幅图像的特征,对每个品种的 10 幅图像的特征求取平均值,并在二维空间中可视化高维特征,对特征进行层次聚类分析。Perplexity 参数设置为 30,迭代次数为 1 000 次。

2 结果与分析

2.1 3 个时期内模型测试结果

为了测试改进网络模型 ResNet50-SE 的性能,在未参与模型训练以及验证过程的测试集中进行评估,生成的混淆矩阵如图 5 所示,通过计算得到模型准确率为 88.75%,结果如表 2 所示。每类预测精确度均在 79.4% 以上,其中小芒森和马瑟兰品种的预测精确度达到了 100%,但小芒森的召回率仅有 75%。根据混淆矩阵图可知模型将小芒森误分为贵人香、黑比诺、雷司令、马尔贝克、霞多丽以及小味儿多,可能由于小芒森品种在某些纹理或颜色形状特

征的水平上与上述的品种特征出现了混淆造成,也可能是由于拍摄角度和光照的影响导致。媚丽的召回率为 100%,没有出现误分类现象,相较于其他品种,媚丽的特征更为显著。从表 2 中还可以得出每类品种的特异度均高于 97.7%,说明模型对于误分类率很小。通过以上分析,证明了本研究所提出的改进模型在田间复杂背景条件下对葡萄品种的多特征识别分类具有较强的鲁棒性。

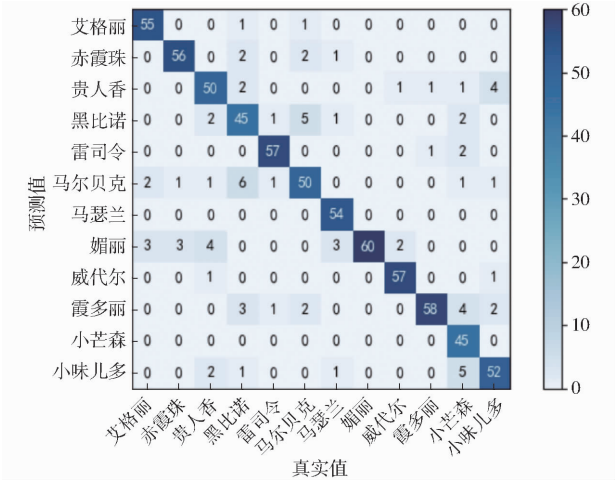


图 5 测试集混淆矩阵结果

Fig. 5 Test set confusion matrix results

表 2 测试集各品种预测精确度、召回率及特异度结果

Tab. 2 Precision, recall and specificity results of each variety in test set %			
品种	预测精确度	召回率	特异度
艾格丽	96.50	91.70	99.70
赤霞珠	91.80	93.30	99.20
贵人香	84.70	83.30	98.60
黑比诺	80.04	75.00	98.30
雷司令	95.00	95.00	99.50
马尔贝克	79.40	88.30	98.00
马瑟兰	100	90.00	100
媚丽	80.00	100	97.70
威代尔	96.60	95.00	99.70
霞多丽	82.90	96.70	98.20
小芒森	100	75.00	100
小味儿多	85.20	86.70	98.60

2.2 单个时期内模型性能对比

为测试 ResNet50-SE 模型在单个时期内对葡萄品种的分类性能,在相同训练条件下对模型在单个时期下的嫩梢、幼叶以及成熟叶片数据集分别训练,得到 ResNet50-SE-T1、ResNet50-SE-T2、ResNet50-SE-T3 模型。分别对模型在只包含单个时期图像测试集 A 和包含 3 个时期图像测试集 B 下进行测试,其对比结果如表 3 所示。

由 3 种不同模型在不同测试集下的测试结果可得出,单一时期特征图像训练的 ResNet50-SE-T1、

表 3 单一时期模型对不同测试集测试结果

Tab.3 Test results of single period model for

模型	different test sets		%
	测试集 A 准确率	测试集 B 准确率	
ResNet50 - SE - T1	91.25	40.28	
ResNet50 - SE - T2	85.83	36.53	
ResNet50 - SE - T3	92.62	41.39	

ResNet50 - SE - T2、ResNet50 - SE - T3 模型虽然在其当前单个时期模型测试集准确率较高,但对于包含 3 种不同时期特征图像的识别性能较差,从而也说明单一时期特征图像训练的模型对葡萄品种分类的鲁棒性较差,难以泛化到其他生长时期内。

2.3 分类模型性能对比

在相同训练条件下,对 4 种典型的卷积神经网络 (AlexNet^[28]、GoogLeNet^[29]、ResNet - 50、VGG - 16^[30]) 进行迁移学习和比较,训练集的准确率结果如图 6 所示。从图 6 中可看出在第 20 个时期 (Epoch) 后,除 AlexNet 模型外,其他模型的训练集准确率维持在 99.9% 左右,虽然 ResNet - 50 的准确率上升较其他模型快,但从图 7 中可知 ResNet - 50 验证集损失值曲线随着训练的 Epoch 增加最终保持在 0.299。然而 ResNet50 - SE 模型最终达到收敛时损失值为 0.216,相比于其他模型损失值较低。AlexNet 模型对于葡萄识别分类效果最差,其原因在于 AlexNet 相比于其他模型网络深度不够,网络提取到的特征信息量较少。模型训练在 20 个 Epoch 前,VGG - 16、AlexNet 和 ResNet - 50 验证集损失值波动较大,模型稳定性较差。GoogLeNet 模型收敛过程虽较为平稳,但准确率与 ResNet50 - SE 相比较低,从而说明 ResNet50 - SE 模型对于葡萄品种的多特征图像识别分类方面具有显著优势。

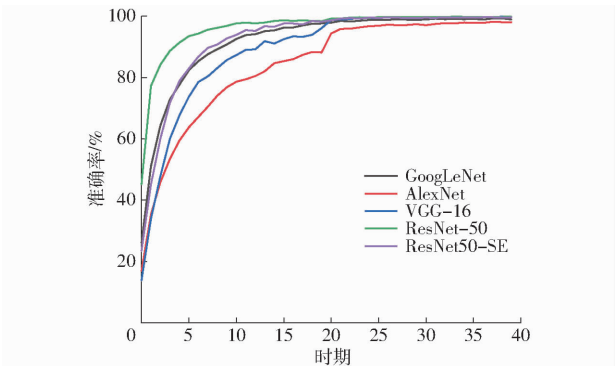


图 6 各网络模型训练集准确率结果对比

Fig.6 Comparison of accuracy results of each network model training set

同时本研究对所提出的模型在验证集准确率以及测试集的准确率和召回率进行对比分析,结果如表 4 所示。结果表明 ResNet50 - SE 在验证集上准

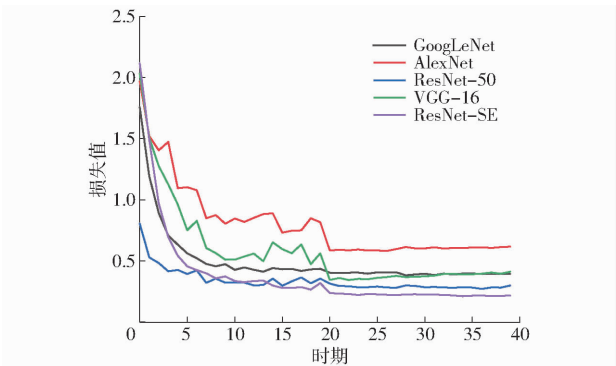


图 7 各网络模型验证集损失值对比

Fig.7 Comparison of loss values of each network model validation set

表 4 各网络模型分类性能对比

模型	Comparison of classification performance of each network model			%
	验证集准确率	测试集准确率	测试集召回率	
AlexNet	82.64	75.14	75.13	
GoogLeNet	88.68	81.11	81.13	
VGG - 16	90.00	82.22	82.23	
ResNet - 50	93.06	88.05	88.05	
ResNet50 - SE	94.10	88.75	89.17	

确率为 94.10%, 在测试集上准确率为 88.75%, 平均召回率为 89.17%。ResNet50 - SE 相比于 AlexNet、GoogLeNet、ResNet - 50、VGG - 16, 验证集准确率分别提高了 11.46、5.42、1.04、4.10 个百分点, 测试集准确率分别提高了 13.61、7.64、0.70、6.53 个百分点, 平均召回率分别提高了 14.04、8.04、1.12、6.94 个百分点。

2.4 特征可视化结果

2.4.1 Grad - CAM 可视化结果

随机挑选测试集的葡萄不同时期下的 3 幅特征图像, 并且根据 Grad - CAM 方法生成模型每一层的可视化热力图, 如图 8 所示。由图 8 可知, 随着网络层数的加深, 模型学习到的特征逐渐增加, 并且由于 SE 模块注意力机制作用, 降低了背景对于模型特征提取的影响。对于葡萄嫩梢, 模型更多是提取到梢尖以及副梢部分的特征信息。对于幼叶和成熟叶片, 模型更为关注叶片边缘的一些形状信息以及叶片的纹理特征信息。

同时, 提取 4 种卷积神经网络的最后卷积层所提取特征的可视化热力图与 Resnet50 - SE 进行比较, 结果如图 9 所示。由图 9 可看出, AlexNet 模型对于特征提取的效果最差, 对于嫩梢和幼叶特征的提取易受到自然背景因素的影响, 难以提取有效特征, 对于幼叶特征关注范围较小, 这也是在对 2.3 节所提到 AlexNet 对于其模型精度相较于其他模型较低的合理

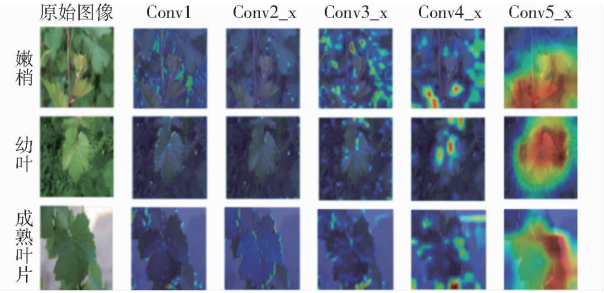


图 8 模型提取各层特征可视化热力图

Fig. 8 Visualization thermodynamic diagram of each layer feature extracted by mode

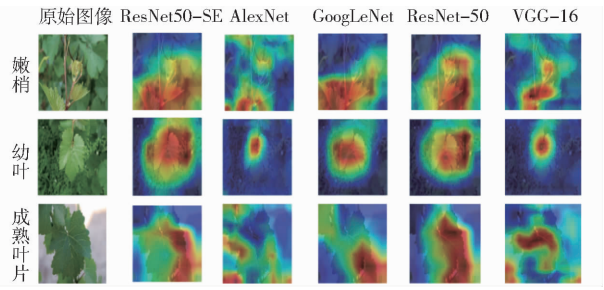


图 9 各网络提取特征可视化对比结果

Fig. 9 Visual comparison results of extracted features of each network

解释。ResNet50-SE 对于嫩梢和幼叶特征提取效果均优于 ResNet-50,而对于成熟叶片特征提取效果较差,这也是导致 ResNet50-SE 测试结果相较于 ResNet-50 优势不明显的重要原因,但是相较于其他 3 种模型(AlexNet、GoogLeNet、VGG-16)却有显著优势,同时也印证了 2.3 节模型性能对比结果。

为了证明 SE 模块在模型特征提取时的校正性,同时还探究了在 ResNet-50 模型 Conv1 层后引入 SE 模块所提取到的葡萄嫩梢、幼叶以及成熟叶片的特征,并且与 Conv1 层所提取的特征(图 8)进行了比较,结果如图 10 所示。在网络 Conv1 层之后加入 SE 模块,将会强化有效特征,弱化一些无效或者低效的特征,使得模型在复杂田间背景条件下能捕获到用于分类的有效特征。

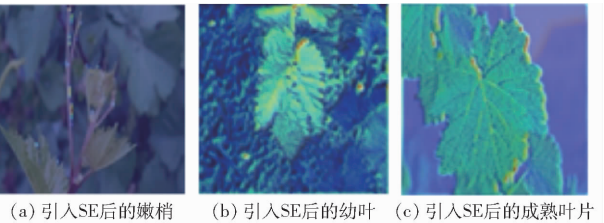


图 10 SE 模块对特征提取的影响结果

Fig. 10 Effect of SE module on feature extraction

2.4.2 t-SNE 聚类结果

利用 t-SNE 算法在训练集上由 ResNet50-SE 模型所获取的 2048 维特征在二维空间(Dim1、Dim2)可视化特征聚类结果如图 11 所示,每种颜色

代表不同葡萄品种类别,共 12 种类别。同一品种的特征聚类的分布位置有所偏差,这是由于不同时期下同一品种葡萄的特征(如嫩梢、幼叶和成熟叶片)具有一定的差异,并且同一生长时期下不同葡萄品种之间的特征具有高度的相似性,但是从整体来看不同的品种在不同生长时期下所反映的特征呈现出较好的聚类效果,这是区分不同葡萄品种的关键。由 2.1 节模型测试结果可知,本文所提出的模型在田间复杂背景条件下对于葡萄高度相似的特征的识别具有较强的鲁棒性,也说明模型在训练集提取到了有利于葡萄识别分类的特征,进而证明了 ResNet50-SE 模型对不同葡萄品种在不同生长阶段下的多特征识别分类的有效性。

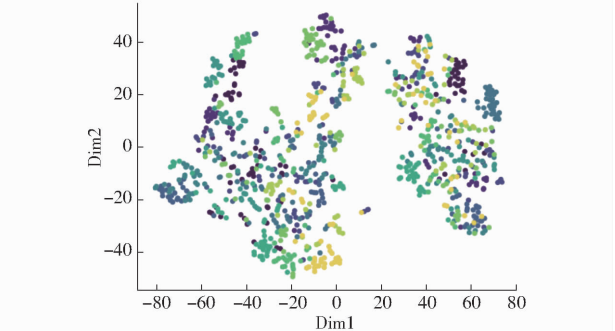


图 11 训练集特征聚类结果

Fig. 11 Feature clustering results of training set

3 结论

(1)提出的模型 ResNet50-SE 在田间复杂背景下对于葡萄多特征图像的识别具有较强的分类性能和鲁棒性,测试集准确率(Accuracy)达到了 88.75%,平均召回率(Recall)为 89.17%,相比于 AlexNet、GoogLeNet、ResNet-50、VGG-16, ResNet50-SE 验证集准确率分别提高了 11.46、5.42、1.04、4.10 个百分点,测试集准确率提高了 13.61、7.64、0.70、6.53 个百分点,平均召回率提高了 14.04、8.04、1.12、6.94 个百分点,分类效果均优于其他模型。

(2)利用 Grad-CAM 可视化方法,对于 ResNet50-SE 模型每一层所提取到的不同生长阶段下的葡萄特征通过激活热力图进行了可视化解释,并且对比了其他 4 种典型卷积神经网络提取特征可视化图。同时,对于引入 SE 模块前后 3 种不同时期下葡萄特征热力图进行对比,证明本研究所提出的 ResNet50-SE 模型对复杂背景条件下特征的提取具有较强的校正性。

(3)通过 t-SNE 算法对不同葡萄品种提取到的不同时期下的特征进行聚类分析,验证了 ResNet50-SE 模型在复杂田间背景环境下对葡萄多特征图像提取特征的有效性以及对葡萄品种鉴别的可靠性。

参 考 文 献

- [1] CECOTTI H, RIVERA A, FARHADLOO M, et al. Grape detection with convolutional neural networks[J]. *Expert Systems with Applications*, 2020,159:113588.
- [2] PEREIRA C S, MORAIS R, REIS M. Deep learning techniques for grape plant species identification in natural images[J]. *Sensors (Basel, Switzerland)*, 2019,19(22):4850.
- [3] TIAN K, LI J, ZENG J, et al. Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm[J]. *Computers and Electronics in Agriculture*, 2019,165:104962.
- [4] ZABAWA L, KICHERER A, KLINGBEIL L, et al. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020,164:73–83.
- [5] ALERCIA A, BECHER R, BOURSQUOT J, et al. OIV descriptor list for grape varieties and vitis species [C] // The International Organisation of Vine and Wine, Paris, 2009.
- [6] LORENZ D, EICHHORN K, BLEIHOLDER H, et al. Growth stages of the grapevine: phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. *vinifera*)—codes and descriptions according to the extended BBCH scale[J]. *Australian Journal of Grape and Wine Research*, 1995,1(2):103–133.
- [7] BRESILLA K, PERULLI G D, BOINI A, et al. Single-shot convolution neural networks for real-time fruit detection within the tree[J]. *Frontiers in Plant Science*, 2019,10:611.
- [8] JIANG F, LU Y, CHEN Y, et al. Image recognition of four rice leaf diseases based on deep learning and support vector machine[J]. *Computers and Electronics in Agriculture*, 2020,179:105824.
- [9] 翟肇裕, 曹益飞, 徐焕良, 等. 农作物病虫害识别关键技术研究综述[J/OL]. *农业机械学报*, 2021,52(7):1–18.
ZHAI Zhaoyu, CAO Yifei, XU Huanliang, et al. Review of key techniques for crop disease and pest detection [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(7):1–18. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20210701&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2021.07.001. (in Chinese)
- [10] KAMILARIS A, PRENAFETA-BOLDU F X. Deep learning in agriculture: a survey [J]. *Computers and Electronics in Agriculture*, 2018,147:70–90.
- [11] KHO S J, MANICKAM S, MALEK S, et al. Automated plant identification using artificial neural network and support vector machine[J]. *Frontiers in Life Science*, 2017,10(1):98–107.
- [12] REIS M, MORAIS R, PERES E, et al. Automatic detection of bunches of grapes in natural environment from color image[J]. *Journal of Applied Logic*, 2012,10(4):285–290.
- [13] MILELLA A, MARANI R, PETITTI A, et al. In-field high throughput grapevine phenotyping with a consumer-grade depth camera[J]. *Computers and Electronics in Agriculture*, 2019,156:293–306.
- [14] 张瑞青, 李张威, 郝建军, 等. 基于迁移学习的卷积神经网络花生荚果等级图像识别[J]. *农业工程学报*, 2020,36(23):171–180.
ZHANG Ruiqing, LI Zhangwei, HAO Jianjun, et al. Image recognition of peanut pod grades based on transfer learning with convolutional neural network[J]. *Transactions of the CSAE*, 2020,36(23):171–180. (in Chinese)
- [15] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. *软件学报*, 2015,26(1):26–39.
ZHUANG Fuzhen, LUO Ping, HE Qing, et al. Survey on transfer learning research[J]. *Journal of Software*, 2015, 26(1):26–39. (in Chinese)
- [16] WEISS K, KHOSHGOFTAAR T M, WANG D D. A survey of transfer learning[J]. *Journal of Big Data*, 2016,3(1):1–40.
- [17] JG A, KH B, FR B, et al. An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding[J]. *Biosystems Engineering*, 2019,183:170–183.
- [18] 孙红, 李松, 李民赞, 等. 农业信息成像感知与深度学习应用研究进展[J/OL]. *农业机械学报*, 2020,51(5):1–17.
SUN Hong, LI Song, LI Minzan, et al. Research progress of image sensing and deep learning in agriculture [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020,51(5):1–17. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200501&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.05.001. (in Chinese)
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C] // IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, USA, 2016.
- [20] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[EB/OL]. [2017-09-05]. <https://arxiv.org/abs/1709.01507>.
- [21] 刘天真, 滕桂法, 苑迎春, 等. 基于改进 YOLOv3 的自然场景下冬枣果实识别研究[J/OL]. *农业机械学报*, 2021,52(5):17–25.
LIU Tianzhen, TENG Guifa, YUAN Yingchun, et al. Winter jujube fruit recognition based on improved YOLOv3 under natural scene[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021,52(5):17–25. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20210502&flag=1. DOI:10.6041/j.issn.1000-1298.2021.05.002. (in Chinese)

Electronics in Agriculture, 2019, 166: 104982.

- [21] 马晓虹, 尹向雷. 基于相关滤波器的目标跟踪方法综述[J]. 电子技术应用, 2018, 44(6): 3-7, 14.
MA Xiaohong, YIN Xianglei. Method of object tracking based on correlation filters: a survey [J]. Application of Electronic Technology, 2018, 44(6): 3-7, 14. (in Chinese)
- [22] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C] // The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, 2010, 6(1): 2544-2550.
- [23] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [24] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[J]. ECCV Workshop, 2014 (8926): 254-265.
- [25] DANELLJAN M, KHAN F S, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 143(1): 1090-1097.
- [26] LEI Q, LIU K, YAO B, et al. Real-time visual tracking with ELM augmented adaptive correlation filter[J]. Pattern Recognition Letters, 2019, 127(NOV): 138-145.
- [27] 李亿杨, 孙龙清, 邹远炳, 等. 基于混合高斯模型的生猪个体检测算法[J]. 中国农业文摘——农业工程, 2018, 30(4): 8-12.
LI Yiyang, SUN Longqing, ZOU Yuanbing, et al. Individual pig object detection algorithm based on Gaussian mixture model [J]. China Agricultural Abstracts—Agricultural Engineering, 2018, 30(4): 8-12. (in Chinese)
- [28] WANG C, WANG T, WANG E, et al. Flying small target detection for Anti-UAV based on a Gaussian mixture model in a compressive sensing domain [J]. Sensors, 2019, 19(9): 2168.
- [29] 张冬梅, 武杰, 李丕丁. 基于机器视觉的运动目标检测算法综述[J]. 智能计算机与应用, 2020, 10(3): 192-195, 201.
ZHANG Dongmei, WU Jie, LI Piding. A summary of moving target detection algorithm based on machine vision [J]. Intelligent Computers and Applications, 2020, 10(3): 192-195, 201. (in Chinese)
- [30] 范良忠, 刘鹰, 余心杰, 等. 基于计算机视觉技术的运动鱼检测算法[J]. 农业工程学报, 2011, 27(7): 226-230, 394.
FAN Liangzhong, LIU Ying, YU Xinjie, et al. Fish motion detecting algorithms based on computer vision technologies [J]. Transactions of the CASE, 2011, 27(7): 226-230, 394. (in Chinese)
- [31] 赵婷, 郑紫微. 基于改进的帧间差分运动目标提取算法[J]. 无线通信技术, 2016, 25(2): 46-49, 53.
ZHAO Ting, ZHENG Ziwei. Motion detection based on improved frame difference [J]. Wireless Communication Technology, 2016, 25(2): 46-49, 53. (in Chinese)

(上接第 233 页)

- [22] 陈娟, 陈良勇, 王生生, 等. 基于改进残差网络的园林害虫图像识别[J/OL]. 农业机械学报, 2019, 50(5): 187-95.
CHEN Juan, CHEN Liangyong, WANG Shengsheng, et al. Pest image recognition of garden based on improved residual network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 187-195. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190522&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.05.022. (in Chinese)
- [23] JIA D, WEI D, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
- [24] BOTTOU L, CURTIS F E, NOCEDAL J. Optimization methods for large-scale machine learning[J]. Siam Review, 2018, 60(2): 223-311.
- [25] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.
- [26] LAURENS V D M, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605): 2579-2605.
- [27] LIU Z, WANG J, TIAN Y, et al. Deep learning for image-based large-flowered chrysanthemum cultivar recognition[J]. Plant Methods, 2019, 15(1): 1-11.
- [28] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C] // Proceeding of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012.
- [29] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C] // IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015.
- [30] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C] // International Conference on Learning Representations, San Diego, USA, 2015.