

DOI: 10.13652/j.spjx.1003.5788.2024.80468

EfficientNet V2 算法融合 GCN 和 CA-Transformer 的 腐烂草莓分类方法

王 伟¹ 杨世忠¹ 官钰程¹ 高 升^{1,2,3} 邓兆鹏¹

(1. 青岛理工大学信息与控制工程学院, 山东 青岛 266520; 2. 山东荣信水产食品集团股份有限公司, 山东 日照 276800; 3. 青岛科技大学海洋科学与生物工程学院, 山东 青岛 266042)

摘要: [目的] 利用现代计算机视觉技术和深度学习方法, 提升腐烂草莓分类的准确性和效率。[方法] 提出了基于 EfficientNet V2 融合 GCN 和 CA-Transformer 的腐烂草莓分类方法。为基准模型添加了图卷积分支, 通过聚合节点的周围信息来更新特征表示, 更好地捕捉节点在图结构中的上下文信息; 将带有注意力的 Transformer 结构融合到基准模型的主干中, 用该结构替换部分卷积操作, 实现全局和局部特征的融合, 从而更好地识别草莓的腐烂情况; 在传统残差结构的基础上引入学习参数, 以实现特征的动态融合。[结果] GC-EfficientNet V2 模型相比基准模型在准确率上提高了 1.86%, 召回率提升了 1.49%。与 Inception V3、ResNet50、VGGNet、Vision Transformer 和 EfficientNet V2-m 模型相比, 该模型的识别准确率分别提高了 0.93%, 2.08%, 2.79%, 3.26%, 0.47%。[结论] 该模型能够准确地对腐烂草莓进行分类。

关键词: 草莓; 腐烂; 图卷积; CA-Transformer; 可学习残差; EfficientNet V2

Rotten strawberry classification based on EfficientNet V2 algorithm fused with GCN and CA-Transformer

WANG Wei¹ YANG Shizhong¹ GONG Yucheng¹ GAO Sheng^{1,2,3} DENG Zhaopeng¹

(1. School of Information and Control Engineering, Qingdao University of Technology, Qingdao, Shandong 266520, China; 2. Shandong Rongxin Aquatic Food Group Co., Ltd., Rizhao, Shandong 276800, China; 3. School of Marine Science and Bioengineering, Qingdao University of Science and Technology, Qingdao, Shandong 266042, China)

Abstract: [Objective] Improving the accuracy and efficiency of rotting strawberry classification using modern computer vision techniques and deep learning methods. [Methods] A classification method for rotten strawberries based on EfficientNet V2 fusion with Graph Convolutional Network (GCN) and Channel-Attention Transformer (CA-Transformer) has been proposed. Firstly, a graph convolution branch was added to the baseline model, which updated feature representations by aggregating the surrounding information of nodes, better capturing the contextual information of nodes in the graph structure. Secondly, this study integrated the Transformer structure with attention into the backbone of the baseline model, replacing some convolution operations with this structure to achieve the fusion of global and local features, thereby better identifying the rottenness of strawberries. Finally, learning parameters were introduced on the basis of the traditional residual structure to achieve dynamic feature fusion. [Results] The GC-EfficientNet V2 model improved the accuracy by 1.86% and the recall by 1.49% compared to the baseline model. Compared with Inception V3, ResNet50, VGGNet, Vision Transformer, and EfficientNet V2-m, the recognition accuracy of the model was improved by 0.93%, 2.08%, 2.79%, 3.26%, and 0.47%, respectively. [Conclusion] This model can accurately classify rotten strawberries, providing some theoretical support for automatic strawberry sorting.

Keywords: strawberry; rot; GCNs; CA-Transformer; learnable residuals; EfficientNet V2

草莓是一种经济价值和营养价值较高的小浆果^[1], 富含维生素 C、膳食纤维和抗氧化物质, 对免疫系统、消化系
统和心血管健康有益^[2]。但由于草莓具有果皮薄、易损伤
和怕水的特性, 在种植、运输和贮藏等过程中极易发生腐

基金项目: 国家自然科学基金(编号: 62001263); 山东省青年基金(编号: ZR2023QC114)

通信作者: 高升(1988—), 男, 青岛理工大学讲师。E-mail: gaosheng@qut.edu.cn

收稿日期: 2024-05-20 改回日期: 2024-10-01

烂现象^[3]。腐烂草莓上的细菌、真菌或其他病原体可能会通过接触空气传播或共享资源传染给周围的健康草莓,导致更多的草莓腐烂。然而,传统的腐烂草莓检测方法通常依赖于人工视觉和主观判断,存在着识别效率低、主观性强、劳动强度大的问题。

深度学习是机器学习发展的前沿^[4],在自然语言处理和计算机视觉^[5-6]等方面都取得了优异的成绩。在水果图像分类领域,很多研究利用卷积神经网络对不同水果进行识别分类,并获得了良好的识别性能。Lamb 等^[7]提出了一种基于卷积神经网络的低成本草莓检测系统,并讨论了训练数据和网络结构对模型性能的影响。曾平等^[8]参照 LeNet-5 结构设计了更适合其数据集的 CNN 网络模型,解决了传统水果图像分类识别方法中人工提取特征的缺陷,最终此模型取得了 98.44% 的识别准确率,高于大多数传统水果识别方法。Hossain 等^[9]提出了两种 CNN 模型对水果进行分类,第 1 种由 6 个卷积层构建而成的轻型 CNN 模型,第 2 种是模型微调后的 VGG16 模型,上述两种模型在面对标准、高质量的测试集时都表现出优越的性能,分类准确率分别达到了 99.49%, 99.75%。同时对于网络上随机选取的水果图像,分类准确率达到 85.43%, 96.75%,都有着比较出色的预测效果。Momeny 等^[10]使用改进的卷积神经网络模型来检测樱桃的外观,并与一些传统的方法进行了对比,所提出的方法优于其他方法,精度高达 99.4%。Gao 等^[11]利用改进的 AlexNet 卷积神经网络对草莓成熟度进行分类,其测试集准确率为 98.6%。Ni 等^[12]提出了一种基于深度学习的草莓品质评价方法,该方法经过数据增强后准确率达到 95.75%。Yuan 等^[13]使用 5 种深度学习模型(ShuffleNet、SqueezeNet、EfficientNet、ResNet18 和 MobileNet-V2)在自制数据集上检测水果的新鲜度。Foong 等^[14]使用 ResNet50 模型获得了 98.89% 的准确率,对单个水果图像进行分类仅需 0.2 s。

目前学者对于腐烂草莓分类的研究相对较少,由于腐烂草莓与新鲜草莓之间的视觉差异不够明显,以及腐烂情况和草莓的外观多样性,从图像中精确地区分腐烂和健康的草莓是一项具有挑战性的任务,需要进一步研究并改进分类方法,以提高腐烂草莓分类的准确率。此外,构建复杂场景、不同生长状态腐烂草莓分类数据集也是非常重要的,这有助于训练泛化能力更强的深度学习模型,并将该模型应用于生产实际。为此,研究拟提出 GC-EfficientNet V2 的腐烂草莓分类方法,旨在利用现代计算机视觉技术和深度学习方法,提升腐烂草莓分类的准确性和效率,为及时、有效、在线、智能检测出腐烂草莓提供可靠的方法。

1 材料与方法

1.1 图像数据集

构建数据集所采用的样本来自山东省青岛市黄岛区的红石崖草莓种植基地,该基地拥有 360 多个种植大棚,占地面积超过 500 hm²。草莓试验样本涵盖妙香、雪里香、甜宝、红颜 4 个品种的不同农户的种植棚,以消除个别地理位置和品种造成的数据偏差。

该数据集于 2023 年 5 月在实验室使用手机拍摄,图像尺寸为 1 200 像素×1 600 像素,共收集 2 152 张图像用于腐烂草莓分类。在农户的专业管理下,草莓园极少出现腐烂果实。采摘后,将完好草莓直接带回实验室进行拍摄,构成完好草莓数据集。随后,将草莓进行多天的自然腐烂处理,再次拍摄采集腐烂草莓图片。通过模拟自然腐烂过程,实现现实生活中草莓腐烂状态的复现。

该数据集包括两类:1 031 张完好草莓图片和 1 121 张腐烂草莓图片。在农场主多年种植经验的指导下,对每张图片进行了精确分类,确保了数据集的准确性。最后,该数据集按 8:1:1 的比例被划分^[15]为训练集、验证集和测试集,如表 1 所示。

表 1 划分后的数据集样本数量分布

Table 1 Sample size distribution of the partitioned dataset.

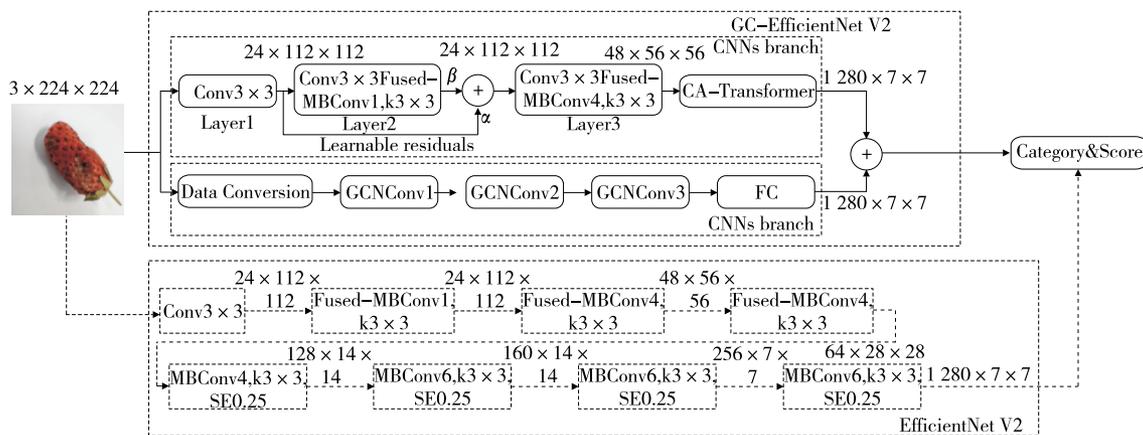
| 数据集 | 完好草莓 | 腐烂草莓 |
|-----|-------|-------|
| 训练集 | 825 | 897 |
| 验证集 | 103 | 112 |
| 测试集 | 103 | 112 |
| 总计 | 1 031 | 1 121 |

1.2 GC-EfficientNet V2 整体架构

基于 EfficientNet V2 模型融合 GCN 和 CA-Transformer 深度学习算法,详细结构如图 1 所示,其分为两个部分:一部分利用卷积和 Transformer 操作对图像数据进行处理;另一部分将图像数据转换为图数据,并经过三层图卷积操作。然后,两部分的结果进行融合,输出最终的分类结果和置信度。该模型具有 3 个关键改进。首先,引入了图卷积结构,以捕捉全局图结构信息。通过将 GCN 和 CNN 相结合,能够更好地综合局部和全局信息,特别适用于处理复杂的图像数据。其次,使用具有更强特征提取能力的 CA-Transformer 结构替换了原有的部分主干结构,增强了模型的泛化能力。最后,对传统的残差结构进行了改进,通过添加两个可学习参数来实现动态特征融合。

1.3 改进的 GCN 分支

图卷积网络(graph convolutional networks, GCNs)是



上部虚线框中为原始 EfficientNet V2 模型;下部虚线框中为改进模型

图 1 GC-EfficientNet V2 的网络结构

Figure 1 Network architecture of GC-EfficientNet V2

一种专门设计用于处理图结构数据的深度学习模型。其核心思想是利用邻接矩阵传播信息,使每个节点聚合其邻接节点的特征,并获得更新后的节点表示。通过多层堆叠,GCNs 能够捕捉不同层次的图结构信息,从而使节点的特征可以应用于各种机器学习任务^[16]。计算过程如式(1)所示。

$$H' = \sigma \left(D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} XW \right), \quad (1)$$

式中:

H' ——节点特征;

σ ——ReLU 激活函数;

D ——度矩阵 (degree matrix);

\tilde{A} ——邻接矩阵 A 与单位阵 I 相加;

X ——输入的节点特征。

研究创新地将图卷积结构与传统卷积进行融合,结构如图 1 所示。图像中存在着空间上的关联,而传统卷积操作无法直接捕捉到这种结构信息。图卷积可以通过图形的连接关系,有效地捕捉到图像的结构信息,从而提高分类准确性。

该模型首先将草莓图像转换为图结构,构建图像的邻接图,其中节点代表图像的像素,边代表它们之间的关系,通过建立图数据结构来捕捉周围像素之间的特征关系,详细过程如图 2 所示。首先,找到像素值 > 102 的像素点,作为图数据中的节点,而 < 102 的像素点则被移除。在整个图像中,节点被标号(从 0 到 N , 其中 N 表示节点的总数)。每个节点具有两个特征,即 x 和 y 坐标。接下来,需要获取边的信息。遍历所有节点,并查找每个节点周围的 8 个像素点。如果周围的像素点也是节点(像素值 > 102),则将其与当前节点相连;如果周围的像素点不是节

点(即像素值 < 102),则不进行连通。通过这样的处理,就得到了两个数据:节点和边。节点的形状为 $(N, 2)$, 其中 N 是图中节点的个数,2 表示节点的特征数。边的形状为 $(M, 2)$, 其中 M 是图中边的个数,2 表示边的朝向。

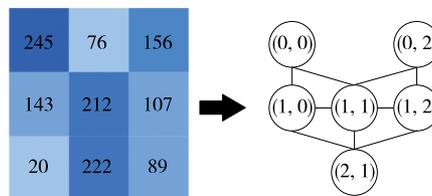


图 2 将图像数据转换为图数据的过程

Figure 2 Process of converting image data into graph data

1.4 提出的 CA-Transformer

CRATE (Coding RATE reduction TransformEr) 是一种新型的白盒 Transformer 结构,旨在通过增量映射学习压缩和稀疏的输入数据表示^[17]。首先,它能够通过逐步的压缩和稀疏化处理更好地捕捉数据的低维结构和稀疏特性。这使得其在处理高维数据和大规模真实世界数据集时表现出色。其次,该结构是可解释的,每个层都可以通过数学解释来理解,提供了更多的灵活性和可控性,可以根据需要设计和定制自己所需的学习表示属性。最后,CRATE 还改进了传统的 Transformer^[18] 模型,通过将自注意力层视为压缩数据表示的梯度下降步骤,以及将多层感知机视为通过构建标记表示的稀疏编码来优化稀疏性。

通过在 CRATE 结构中引入注意力机制,包括通道注意力和空间注意力,命名为 CA-Transformer (结构如图 3 所示)。通道注意力主要关注特征的通道维度,通过学习

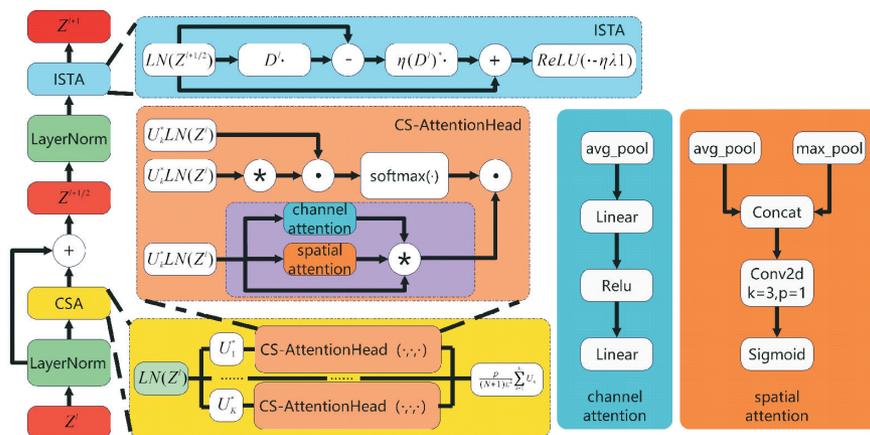


图 3 CA-Transformer 网络结构

Figure 3 CA-Transformer network architecture

每个通道的重要性权重,模型能够自适应有选择性地聚焦于不同通道的特征。另一方面,空间注意力机制主要关注特征点的空间维度,通过学习每个空间位置的权重,能够提升模型对局部细节的感知能力。CA-Transformer 结构是一种基于注意力机制的深度学习模型,与传统的 Transformer 结构相比,能更有效地捕捉图像特征。多个 CS-AttentionHead 允许模型同时关注不同特征通道,通过计算不同通道之间的相关性来提取更丰富的信息,计算过程如式(2)所示。将 CA-Transformer 结构集成到 GC-EfficientNet V2 模型的卷积分支最后一层,有效地增强了模型对草莓腐烂情况的识别能力。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)(\mathbf{V} \times C_A(\mathbf{V}) \times S_A(\mathbf{V})), \quad (2)$$

式中:

- \mathbf{Q} ——查询向量;
- \mathbf{K} ——键向量;
- \mathbf{V} ——值向量;
- d_k ——向量维度;
- C_A ——通道注意力;
- S_A ——空间注意力。

1.5 改进的残差

通过在传统的残差^[19]结构中添加两个可学习参数,实现了动态特征融合,如图4所示。传统的残差结构在深度学习中取得了显著成功,通过引入跳跃连接来解决梯度消失问题,有效促进模型的收敛。然而,传统的残差结构在每个跳跃连接中使用简单的加法操作,缺乏对特征之间复杂关系的建模能力。在该结构中引入可学习参

数,并将其与两个分支分别相乘,增强模型对特征之间关系的建模能力。

使用 softmax 激活函数对学习参数进行归一化处理,确保其总和为 1,计算过程如式(3)所示。这种设计允许学习参数动态分配不同特征通道的重要性,并能根据任务需求进行自适应调整。

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}. \quad (3)$$

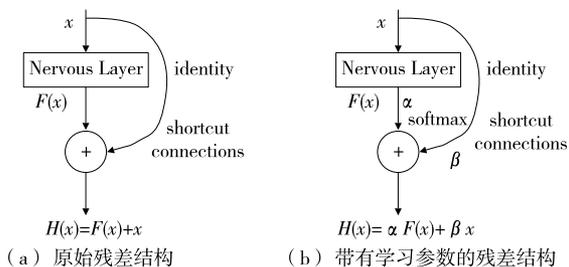


图 4 改进后的残差结构

Figure 4 Improved residual structure

2 结果与讨论

2.1 试验平台

操作系统: Windows 10, CPU: Intel I5-8300H, GPU: NVIDIA GeForce GTX 1050Ti, 深度学习框架: Deep Learning Framework: NVIDIA GeForce GTX 1050Ti, 深度学习框架: PyTorch。

2.2 评价指标

(1) 准确率 (accuracy): 准确率是最常用的图像分类评价指标之一,表示分类正确的样本数量与总样本数量之间的比率^[20],其计算式如式(4)所示。

$$A = \frac{T_p + T_N}{T_p + T_N + F_N + F_p} \times 100\%, \quad (4)$$

式中:

A ——准确率, %;

T_p ——预测为完好草莓, 实际是完好草莓的样本数;

F_p ——预测为完好草莓, 实际是腐烂草莓的样本数;

T_N ——预测为腐烂草莓, 实际是腐烂草莓的样本数;

F_N ——预测为腐烂草莓, 实际是完好草莓的样本数。

(2) 精确度 (precision): 预测正确的正类样本占有所有预测为正类样本的百分比, 用来衡量模型在预测为正类的样本的准确率, 其计算式如式(5)所示。

$$P = \frac{T_p}{T_p + F_p} \times 100\%, \quad (5)$$

式中:

P ——精确度, %;

T_p ——预测为完好草莓, 实际是完好草莓的样本数;

F_p ——预测为完好草莓, 实际是腐烂草莓的样本数。

(3) 召回率 (recall): 正确预测的正类样本占有所有实际正类样本的百分比^[21]。在医学图像分类问题中, 召回率也被称为灵敏度, 高灵敏度表示对病灶区域更敏感。召回率的计算式如式(6)所示。

$$R = \frac{T_p}{T_p + F_N} \times 100\%, \quad (6)$$

式中:

R ——召回率, %;

T_p ——预测为完好草莓, 实际是完好草莓的样本数;

F_N ——预测为腐烂草莓, 实际是完好草莓的样本数。

(4) FLOPs (floating point operations per second): 表示模型在推理或训练过程中所需要执行的浮点运算次数。FLOPs 越高, 意味着模型计算量越大, 所需的硬件资源和计算时间也会相应增加。

(5) FPS (frames per second): 在计算机视觉任务中, 表示模型每秒可以处理的图像帧数。FPS 越高, 模型的实时处理能力越强。

2.3 Grad-CAM 特征可视化

Grad-CAM 融合了梯度类激活图和类别感兴趣区域的加权组合, 能够可视化模型对草莓图像中不同区域的关注程度^[22], 从而直观地观察模型在腐烂草莓分类任务中的决策依据^[23]。

完好草莓通常呈现出鲜艳的红色或粉红色外皮, 表面光滑且均匀, 无明显的凹陷或软化的部分。而腐烂草莓往往呈现出明显的褐色、黑色或发霉的斑点, 外皮可能出现凹陷、皱缩、软化或变得黏稠。热力图显示模型能够较好地关注到草莓的腐烂区域, 对损坏区域高度敏感(如

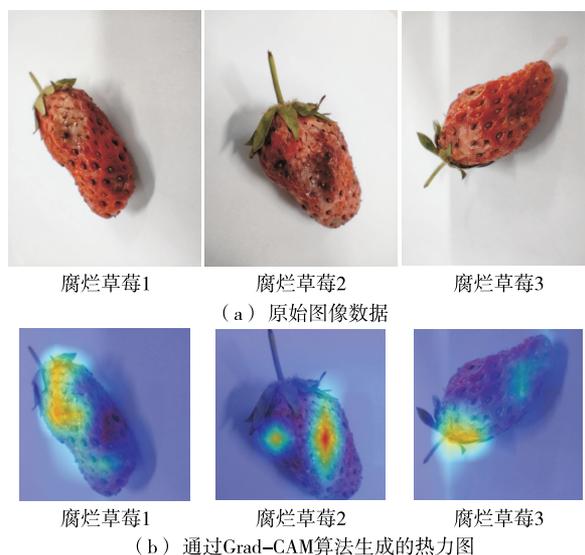


图5 部分数据的热力图
Figure 5 Heatmaps of a subset of the data

图5所示), 表明模型成功学习到了腐烂草莓与完好草莓之间的差异。通过原始图片和模型热力图进行对比能够看到, 模型成功关注到草莓的腐烂部位, 为接下来的分类提供重要且准确的特征信息。

对腐烂草莓2样本进行卷积特征可视化处理, 结果(图6)发现, 卷积层提取了输入图像的边缘、轮廓和线条等特征, 这些特征在特征图中清晰可见。随着网络层数的增加, 深层卷积相较于浅层卷积提取的特征更加复杂。这一现象可以归因于深度神经网络具有的分层表示能力。浅层卷积层主要关注图像的低级特征, 例如边缘和纹理。随着信息在网络中向深层卷积层传播, 神经网络通过多个非线性变换层逐渐提取更高级的特征表示。

图像熵在一定程度上反映了信息的混乱程度, 熵值越大则图像所包含的信息越多。为了代表当前卷积层提取的特征信息量, 选择当前卷积层中图像熵最大的特征图。通过累加不同灰度值的熵值来计算图像的熵, 计算为:

$$S = - \sum_{i=0}^{255} p_i \log_2(p_i), \quad (7)$$

式中:

S ——图像的熵值, bits;

p_i ——灰度值 i 在该图像中出现的概率(可由灰度直方图获取)。

随着深度神经网络层数的增加, 最大熵也逐渐增加, 表明模型能够学到腐烂草莓分类数据集中更多的信息, 能够更好地捕捉到草莓之间的复杂结构。与浅层模型相比, 深层神经网络能够通过多个非线性变换层级来逐渐提取更高级别的特征表示, 更好地适应分拣过程的复杂情况。

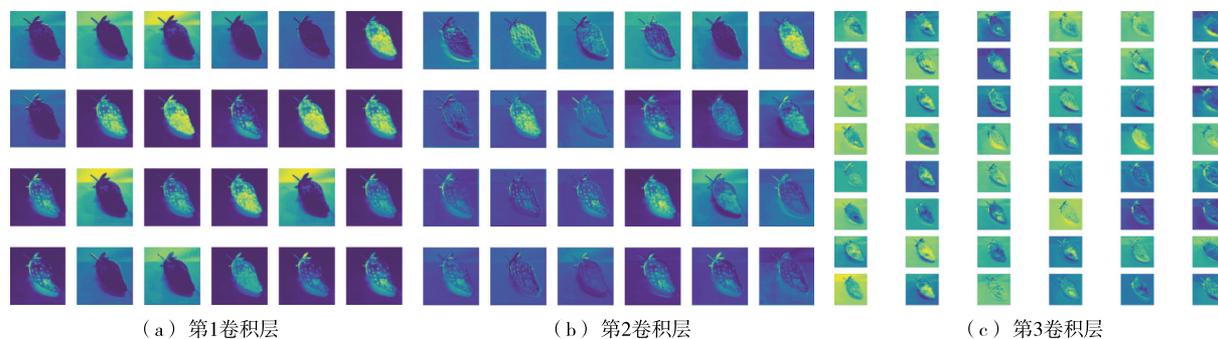


图 6 卷积特征图可视化

Figure 6 Convolutional feature map visualization

2.4 对比试验

为了评估深度学习模型的性能,进行了对比试验,将改进模型 GC-EfficientNet V2 与其他主流模型进行了比较,结果见表 2。

由表 2 可知,与传统的卷积神经网络模型(VGG、InceptionV3 和 ResNet)以及基于 Transformer 的 Vit 模型相

比,在自建的腐烂草莓分类数据集上,GC-EfficientNet V2 模型取得了更高的分类准确率。GC-EfficientNet V2 准确率相较于 EfficientNet V2-s 得到了 1.86% 的提升,同时召回率提升了 1.49%。GC-EfficientNet V2 在仅使用 EfficientNet V2-m 模型 9.7% 参数量的情况下,取得了更优秀的效果。

表 2 不同方法的腐烂草莓分类性能对比

Table 2 Comparison of classification performance for rotten strawberries using different methods

| 模型 | Top-1 准确率/% | 精确度/% | 召回率/% | 参数量/M | FLOPS/G | FPS |
|-----------------------------------|-------------|-------|-------|-------|---------|-------|
| EfficientNet V2-s ^[23] | 96.74 | 96.74 | 97.58 | 96.73 | 19.28 | 2.68 |
| Inception V3 ^[24] | 97.67 | 97.67 | 97.54 | 96.91 | 21.79 | 5.74 |
| ResNet 50 ^[25] | 96.52 | 96.52 | 28.97 | 28.89 | 23.50 | 4.13 |
| VggNet | 95.81 | 95.81 | 95.81 | 94.74 | 132.86 | 7.60 |
| Vision Transformer | 95.34 | 95.34 | 95.34 | 93.56 | 28.97 | 5.70 |
| Efficientnet V2-m | 98.13 | 97.88 | 97.47 | 51.50 | 5.19 | 17.97 |
| GC-EfficientNet V2 | 98.60 | 99.09 | 98.22 | 5.01 | 1.09 | 47.76 |

如图 7 所示,GC-EfficientNet V2 在训练初期就展现出了极高的准确率,并且随着训练的进行,准确率保持了较快的提升速度。此外,GC-EfficientNet V2 相较于其他模型具有更高的准确率,可有效分拣腐烂草莓。

2.5 消融试验

为了评估所添加的图卷积、CA-Transformer 和改进的残差结构对 EfficientNet V2 分类性能的影响,进行了消融试验,结果见表 3。在原始模型基础上添加图卷积之后准确率得到 0.93% 的提升,在引入 CA-Transformer 结构后准确率得到了 1.31% 的提升。当使用普通残差时模型准确率为 98.13%,改进残差后得到了 0.47% 的提升,并且召回率也有相应的改善,各模型的准确率随着训练轮数的变化曲线如图 8 所示。由图 9 可知,GC-EfficientNet V2 对完好草莓的分类准确率达到 98.22%,对腐烂草莓的分类准确率高达 98.98%,满足实际生产环境的使用要求。

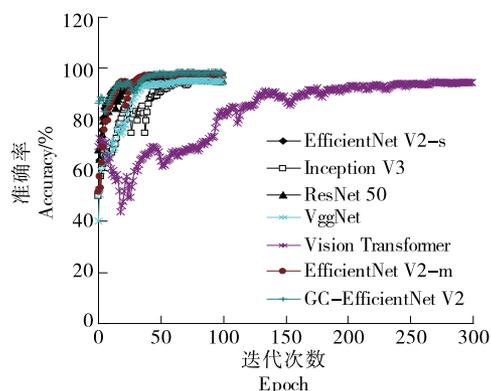


图 7 不同模型的准确率随着训练轮数的变化曲线
Figure 7 Accuracy curves of different models with increasing training epochs

3 结论

基于 EfficientNet V2 算法融合 GCN 和 CA-

表 3 消融试验

Table 3 Ablation experiment

| 模型 | Top-1 准确率/% | 精确度/% | 召回率/% | 参数量/M | FLOPS/G | FPS |
|--|--------------|-------|-------|-------|---------|-------|
| EfficientNet V2-s | 96.74 | 97.58 | 96.73 | 19.28 | 2.68 | 76.97 |
| EfficientNet V2-s+ GCNs | 97.67(+0.93) | 97.88 | 97.77 | 19.28 | 2.68 | 7.65 |
| EfficientNet V2-s+ GCNs+ CA-Transformer | 98.05(+1.31) | 98.38 | 97.47 | 5.01 | 1.09 | 49.70 |
| EfficientNet V2-s+ GCNs+ CA-Transformer+Residual | 98.13(+1.39) | 98.52 | 97.47 | 5.01 | 1.09 | 48.38 |
| EfficientNet V2-s+GCN+CA-Transformer+Learnable residuals | 98.60(+1.86) | 99.09 | 98.22 | 5.01 | 1.09 | 47.76 |

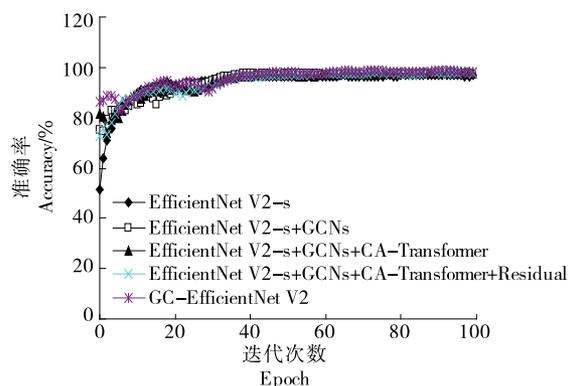


图 8 各模型的准确率随着训练轮数的变化曲线

Figure 8 Accuracy curves of different models with increasing training epochs

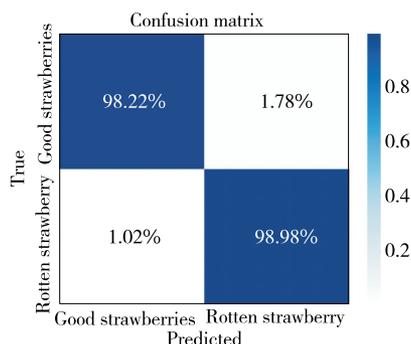


图 9 混淆矩阵

Figure 9 Confusion matrix

Transformer的腐烂草莓分类方法,通过图卷积网络增强了对图像结构信息的理解,并将CA-Transformer融入主干网络,实现全局与局部特征的有效融合,同时改进了残差结构,引入了学习参数,提升了特征提取能力。试验结果显示,GC-EfficientNet V2模型在准确率、精确率和召回率方面均显著优于基准模型,参数量减少至原始模型的25.98%。该方法在与Inception V3、ResNet50等模型的对比中表现出色,具备扩展应用于其他水果和蔬菜分类的潜力。

参考文献

[1] 刘小刚,范诚,李加念,等.基于卷积神经网络的草莓识别方

法[J].农业机械学报,2020,51(2):237-244.
 LIU X G, FAN C, LI J N, et al. Identification method of strawberry based on convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(2): 237-244.
 [2] 黄家才,赵雪迪,高芳征,等.基于改进YOLOv5s的草莓多阶段识别检测轻量化算法[J].农业工程学报,2023,39(21):181-187.
 HUANG J C, ZHAO X D, GAO F Z, et al. Recognizing and detecting the strawberry at multi-stages using improved lightweight YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(21): 181-187.
 [3] WU F, GUAN Z F, WHITAKER V. Optimizing yield distribution under biological and economic constraints: florida strawberries as a model for perishable commodities[J]. Agricultural Systems, 2015, 141: 113-120.
 [4] CHEN Y, HUANG Y Q, ZHANG Z Z, et al. Plant image recognition with deep learning: a review[J]. Computers and Electronics in Agriculture, 2023, 212: 108072.
 [5] 焦方圆,申金媛,郝同盟.一种基于卷积神经网络的烟叶等级识别方法[J].食品与机械,2022,38(2):222-227.
 JIAO F Y, SHEN J Y, HAO T M. A method of tobacco leaf grade recognition based on convolutional neural network[J]. Food & Machinery, 2022, 38(2): 222-227.
 [6] 周胜安,黄耿生,张译匀,等.基于深度学习的水果缺陷实时检测方法[J].食品与机械,2021,37(11):123-129.
 ZHOU S A, HUANG G S, ZHANG Y J, et al. Real time detection method of fruit defects based on deep learning[J]. Food & Machinery, 2021, 37(11): 123-129.
 [7] LAMB N, CHUAH M C. A strawberry detection system using convolutional neural networks[C]// 2018 IEEE International Conference on Big Data (Big Data). Seattle, WA: IEEE, 2018: 2515-2520.
 [8] 曾平平,李林升.基于卷积神经网络的水果图像分类识别研究[J].机械设计与研究,2019,35(1):23-26,34.
 ZENG P P, LI L S. Classification and recognition of common fruit images based on convolutional neural network[J]. Machine Design & Research, 2019, 35(1): 23-26, 34.
 [9] HOSSAIN M S, AL-HAMMADI M, MUHAMMAD G. Automatic fruit classification using deep learning for industrial

- applications[J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(2): 1 027-1 034.
- [10] MOMENY M, JAHANBAKHSI A, JAFARNEZHAD K, et al. Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach[J]. *Postharvest Biology and Technology*, 2020, 166: 111204.
- [11] GAO Z M, SHAO Y Y, XUAN G T, et al. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning[J]. *Artificial Intelligence in Agriculture*, 2020, 4: 31-38.
- [12] NI J G, GAO J Y, LI J, et al. E-AlexNet: quality evaluation of strawberry based on machine learning[J]. *Journal of Food Measurement and Characterization*, 2021, 15(5): 4 530-4 541.
- [13] YUAN Y, CHEN X L. Vegetable and fruit freshness detection based on deep features and principal component analysis[J]. *Current Research in Food Science*, 2024, 8: 100656.
- [14] FOONG C C, MENG G K, TZE L L. Convolutional Neural Network based Rotten Fruit Detection using ResNet50[C]// 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC). Shah Alam, Malaysia: IEEE, 2021: 75-80.
- [15] WANG Q Y, WU D H, SUN Z Z, et al. Design, integration, and evaluation of a robotic peach packaging system based on deep learning[J]. *Computers and Electronics in Agriculture*, 2023, 211: 108013.
- [16] 孙杰, 杨静, 丁书杰, 等. 基于多注意力机制与编译图神经网络的高光谱图像分类[J]. *农业机械学报*, 2024, 55(3): 183-192, 212.
- SUN J, YANG J, DING S J, et al. Hyperspectral image classification based on multi-attention mechanism and compiled graph neural networks[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(3): 183-192, 212.
- [17] YU Y D, CHU T Z, TONG S B, et al. Emergence of segmentation with minimalistic white-box transformers[J/OL]. *arXiv*. (2023-08-30) [2024-08-13]. <https://arxiv.org/abs/2308.16271>.
- [18] 沈宇麒, 崔衍. 基于双层路由注意力及特征融合的细粒度图像分类[J]. *计算机技术与发展*, 2024, 34(6): 23-28.
- SHEN Y Q, CUI Y. Fine-grained image classification based on bi-level routing attention and feature fusion[J]. *Computer Technology and Development*, 2024, 34(6): 23-28.
- [19] 陈辉, 张甜, 陈润斌. 基于残差卷积与多头自注意力的 CXR 图像分类[J]. *工程科学与技术*, 2024, 56(3): 219-227.
- CHEN H, ZHANG T, CHEN R B. CXR image classification based on residual convolution and multi-headed self-attentions [J]. *Advanced Engineering Sciences*, 2024, 56(3): 219-227.
- [20] 刘玉良, 白玉洁. 基于 CSNet 网络的肺炎图像分类研究[J]. *天津科技大学学报*, 2024, 39(3): 49-55.
- LIU Y L, BAI Y J. Research on pneumonia image classification based on CSNet network[J]. *Journal of Tianjin University of Science & Technology*, 2024, 39(3): 49-55.
- [21] 逮登科, 罗亦冰, 张紫怡, 等. 基于 ResNet50 与通道注意力的遥感图像场景分类[J]. *江西科学*, 2024, 42(2): 396-404.
- LU D K, LUO Y Y, ZHANG Z Y, et al. Remote sensing image scene classification based on ResNet50 and channel attention [J]. *Jiangxi Science*, 2024, 42(2): 396-404.
- [22] 郑世杰, 王高才. 基于 ConvNeXt 热图定位和对比学习的细粒度图像分类研究[J]. *计算机科学*, 2023, 50(10): 119-125.
- ZHENG S J, WANG G C. Study on fine-grained image classification based on ConvNeXt heatmap localization and contrastive learning[J]. *Computer Science*, 2023, 50(10): 119-125.
- [23] QIAN C Y, SUN S G, DONG C H, et al. A study on phenotypic micro-variation of stored melon based on weight loss rate[J]. *Postharvest Biology and Technology*, 2023, 204: 112464.
- [24] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016: 2 818-2 826.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016: 770-778.