

基于改进生成对抗网络的甜樱桃数据增强方法

韩翔^{1,2} 李玉强¹ 高昂¹ 马静怡¹ 宫庆福¹ 宋月鹏^{1,3}

(1. 山东农业大学机械与电子工程学院, 泰安 271018; 2. 山东省园艺机械与装备重点实验室, 泰安 271018;

3. 山东省农业装备智能化工程实验室, 泰安 271018)

摘要: 为解决在数据不平衡条件下甜樱桃分类模型出现的长尾类不平衡问题, 提出了一种基于深度卷积生成对抗网络 (Deep convolutional generative adversarial networks, DCGAN) 的缺陷甜樱桃图像增强方法。首先, 在生成器部分引入多尺度残差块 (MSRB) 和 CBAM 注意力机制, 增强了模型特征表达能力和生成图像细节质量, 同时改善了梯度流; 在判别器部分应用谱归一化技术, 并引入 Wasserstein 距离和加梯度惩罚的损失函数, 增强了模型训练稳定性和收敛速度。实验结果表明, 与传统的 GAN 模型相比, 本文模型可以生成更高质量的缺陷甜樱桃图像, 两种缺陷甜樱桃图像的 FID 值 (Fréchet inception distance) 分别为 64.36 和 59.97。本文模型生成的数据增强后, VGG19 和 MobileNetV3 的甜樱桃分类准确率分别提高 16.44 个百分点和 13.94 个百分点。

关键词: 甜樱桃图像; 数据生成; 生成对抗网络; 长尾类不平衡

中图分类号: TP391

文献标识码: A

文章编号: 1000-1298(2024)10-0252-11

OSID:



Data Augmentation Method for Sweet Cherries Based on Improved Generative Adversarial Network

HAN Xiang^{1,2} LI Yuqiang¹ GAO Ang¹ MA Jingyi¹ GONG Qingfu¹ SONG Yuepeng^{1,3}

(1. College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian 271018, China

2. Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment, Taian 271018, China

3. Shandong Provincial Laboratory of Agricultural Equipment Intelligence Engineering, Taian 271018, China)

Abstract: To address the class imbalance in sweet cherry data, a novel image enhancement method based on sweet cherry generative adversarial network, SCGAN was proposed. The generator incorporated multi-scale residual blocks (MSRB) and the convolutional block attention module (CBAM), enhancing the model's feature representation and the quality of generated images. These blocks captured features at various scales, and CBAM focused on channel and spatial information, improving image quality. In the discriminator, spectral normalization and the Wasserstein distance with a gradient penalty loss function were applied. This combination controlled the discriminator's power, prevented overfitting, and boosted training stability and speed. Experimental results showed that SCGAN produced higher quality defective sweet cherry images compared with traditional GANs, with Fréchet inception distance (FID) scores of 64.36 and 59.97 for two types of defects. After data augmentation with SCGAN, classification accuracy for VGG19 and MobileNetV3 was increased by 16.44 percentage points and 13.94 percentage points, respectively. The data augmentation method presented held significant potential in addressing data imbalance issues within the agricultural and food sectors. It not only improved the generalization capability of models but also provided a more reliable data foundation for practical applications. Through this approach, it was possible to more effectively tackle long-tail class imbalance issues, which enhanced the accuracy and efficiency of agricultural and food detection systems.

Key words: sweet cherry images; data generation; generative adversarial network; long-tailed class imbalance

收稿日期: 2024-06-04 修回日期: 2024-07-20

基金项目: 山东省现代农业产业技术体系果品产业创新团队项目 (SDAIT-06-12)

作者简介: 韩翔 (1995—), 男, 博士生, 主要从事图像处理与深度学习算法研究, E-mail: 2021010102@sdau.edu.cn

通信作者: 宋月鹏 (1971—), 男, 教授, 博士生导师, 主要从事果园管理及产后处理生产技术与装备研究, E-mail: uptonsong@sdau.edu.cn

0 引言

甜樱桃是一种具有重要经济和营养价值的水果。然而,在采收过程中,部分甜樱桃果实可能会出现开裂和腐烂等质量问题,严重影响了其市场价值。因此,确保甜樱桃在处理和流通过程中的质量分级准确性显得尤为重要。甜樱桃数据集中存在严重不平衡现象^[1],使得现有质量分级方法往往倾向于关注数据量较大的类别,而忽视了数据量较小的类别,这种偏差可能导致分类误判^[2-4]。

目前,为解决数据集长尾类不平衡的问题,主流方法采用基于过采样的数据增强技术^[5-6]。这种技术通过扩充原始数据集的信息来减少训练集、验证集与未来测试集之间的差异,从而降低模型过拟合风险^[7-9]。然而,传统的数据增强方法(如旋转、裁剪、缩放等)虽能增加数据量^[10-13],但往往只针对单一风格、信噪比低的图像,效率较低,且生成的图像中包含大量冗余信息,给卷积神经网络(CNN)的训练带来了更多不确定性。另一种数据增强技术是利用生成对抗网络(GAN),其在生成保留真实图像特征的假图像方面取得了显著成就,能够从真实图像中挖掘出额外有用信息^[14-17]。但传统 GAN 存在无法精确控制生成图像细节的问题;此外,GAN 模型本身训练困难,容易出现梯度消失和模型崩溃。

为应对甜樱桃数据不平衡问题,本文提出改进的甜樱桃 GAN 模型(Sweet cherry generative adversarial network, SCGAN)。改进模型融入多尺度残差块,增强训练稳定性并加快收敛速度,同时通过特征重用降低计算负荷。此外,集成的 CBAM 注意力机制优化特征提取,强化通道与空间特征的表达。通过谱归一化适度限制判别器能力,维持生成器与判别器间的动态平衡,以期提升模型整体性能。

1 SCGAN 数据增强整体流程

图 1 展示了在数据不平衡的条件下使用生成缺陷甜樱桃的 SCGAN 进行数据增强,以提升甜樱桃分级准确率。

为获取原始数据集,从泰安甜樱桃采摘园(117.36°E,36.00°N)自助采摘甜樱桃样品。甜樱桃品种为布鲁克斯,等级分为健康果、裂果、腐烂果。对采集到的甜樱桃进行图像采集,图像采集系统主要包括手机支架、手机(iPhone 13)、补光灯。图像保存格式为 JPG 格式,像素分辨率为 3 024 像素 × 4 032 像素,手机拍摄高度为 200 mm,光照条件为正常光照。采集到的图像为图 1 所示的原始数据集。

为了降低背景噪声对模型的影响,进一步提高生成图像质量,利用 OpenCV 将采集到的原始 RGB 图像进行颜色空间转换处理为 HSV 图像,选择合适的阈值($0 < \text{Hue} < 180, 0 < \text{Saturation} < 55$ 和 $200 < \text{Value} < 255$),对图像进行初步阈值分割,得到二值图像。然后对二值图像进行高斯模糊,减少噪声。对待分割的图像进行掩膜处理。提取的图像如图 1 所示的预处理后数据集。预处理甜樱桃初始数据集包括 2 000 个健康甜樱桃、146 个腐烂甜樱桃和 47 个裂甜樱桃。将数据集按比例 3:1:1 分为训练集、验证集和测试集。具体划分如表 1 所示。

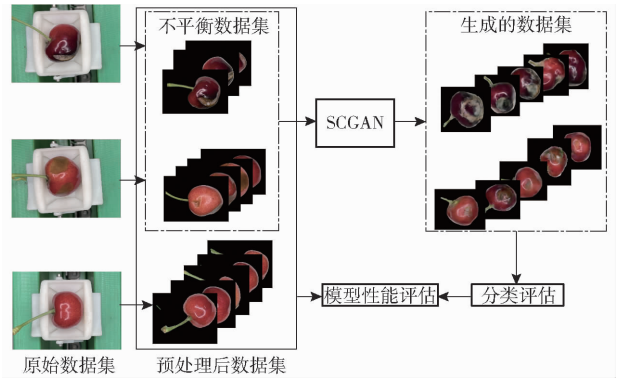


图 1 SCGAN 甜樱桃分级数据增强流程图

Fig. 1 Data augmentation process for sweet cherry quality grading using SCGAN

表 1 甜樱桃数据集划分

Tab. 1 Division of sweet cherry data set

种类	全部	训练集	验证集	测试集
腐烂果	146	88	29	29
裂果	47	28	10	9
健康果	2 000	1 200	400	400

由表 1 可以发现,原始预处理数据集存在不平衡问题,尤其是腐烂果和裂果存在着采集难、数据量少的问题。于是使用 SCGAN 模型在不平衡训练集上进行训练,生成各类别的缺陷甜樱桃图像,以平衡训练数据。分类模型初次以原始数据集进行训练,然后使用由 SCGAN 增强后的数据集进行再训练,全面评估数据增强对模型性能的影响。这一过程中,验证集和测试集保持不变以确保评估的一致性和公正性。分类任务采用卷积神经网络(CNN),特别是 VGG19 和 MobileNetV3,因为 VGG19 擅长捕捉图像纹理和细节,而 MobileNetV3 则适用于资源受限环境和实时应用,结合两种模型各自优势。VGG19 和 MobileNetV3 模型在这些经过平衡的训练数据上训练,以进一步验证数据增强在实际应用中的有效性。全面评估图像质量与实用性。性能评估随后仅在测试集上进行,确保结果公正性和可比性。

2 GAN 和 DCGAN

受零和博弈概念 (Zero - Sum Game) 的启发, GOODFELLOW 等^[18] 提出了生成对抗网络 (Generative adversarial networks, GAN), 其整体结构如图 2 所示, 是由生成模型和判别模型两部分构成。GAN 网络目标函数可表示为

$$\minmax_{G,D} V(D,G) = E_{x \sim p_{data}(x)} [\lg D(x)] + E_{z \sim p_z(z)} [\lg (1 - D(G(x)))]$$

(1)

式中 D ——判别器 G ——生成器
 $V(D,G)$ ——优化目标函数
 E ——分布函数数学期望
 $p_{data}(x)$ ——原始数据分布
 $E_{x \sim p_{data}(x)} [\lg D(x)]$ ——判别真实数据为真实样本的概率
 z ——随机噪声
 $p_z(z)$ ——随机噪声数据分布
 $E_{z \sim p_z(z)} [\lg (1 - D(G(z)))]$ ——判别生成的样本 $G(z)$ 为生成样本的概率

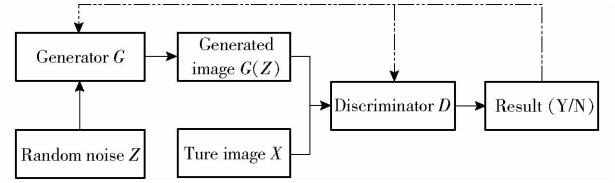


图 2 生成对抗网络
Fig. 2 Generative adversarial network

生成器 G 的目标是学习真实样本数据的分布, 并生成逼真的新样本。其输入随机噪声, 试图最大化判别模型将其生成的样本判别为真实样本的可能性。而判别器 D 的任务则是区分真实样本和生成器产生的假样本, 旨在最小化将生成样本误判为真实样本的可能性。在经过多次对抗训练后, 生成器和判别器达到纳什平衡, 这意味着判别器对真假样本的判断概率接近 0.5, 表明生成器产生的图像质量高, 难以被区分^[19]。

传统的 GAN 模型在训练过程中常遇到模式崩溃等稳定性问题, 限制了样本多样性。早期 GAN 使用的较浅网络结构难以学习复杂的数据分布, 且整个训练过程中需精细调整参数以保持生成器与判别器之间的学习平衡。为改善这些问题, DCGAN 采用全卷积网络结构, 去除全连接层, 并广泛使用卷积层和反卷积层优化图像处理能力。此外, DCGAN 在生成器和判别器中引入批归一化和 LeakyReLU 激活函数, 旨在减少内部协变量偏移、提供非零梯度, 从而有效地提高图像生成质量并增强训练稳定

性^[20-21]。

3 甜樱桃生成对抗网络

在处理分选网络过程中, 发现甜樱桃数据集存在着严重的长尾类不平衡, 即在采收过程中某些类别的甜樱桃数量远多于其他类别。这种不平衡对于训练有效的机器学习模型构成了挑战, 因为模型往往偏向于识别较多样本的类别, 而忽视了少数类别, 从而影响了整体分类性能。为应对这一问题, 引入用于生成缺陷甜樱桃的生成对抗网络 (Sweet cherry generative adversarial network, SCGAN)。

基于 DCGAN 的底层架构提出的 SCGAN 结构如图 3 所示, 主要包含 2 个对立网络: 生成器 G 提取输入样本中的特征, 获取样本数据分布, 并输出生成样本; 将真实样本与生成样本作为一组数据, 成对输入到判别器 D 中。然后, D 判断输入图像是来自训练数据集中的真实图像还是 G 生成的假图像, 并对输入图像进行正确分类。由于 G 和 D 的目标相互矛盾, 所以 SCGAN 的训练过程是寻求并实现纳什均衡。此外, 在 G 中嵌入多尺度残差块与 CBAM 模块, 用于增强特征通道和空间注意力。在 D 中嵌入谱归一化模块, 提升了判别器训练稳定性以及生成图像质量。

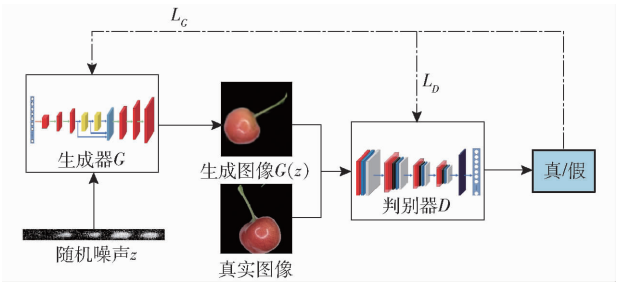


图 3 SCGAN 网络结构
Fig. 3 Structure of SCGAN

3.1 SCGAN 生成器

如图 4 所示, SCGAN 网络生成器由多层卷积结构构成, 结合上采样层、全连接层、多层卷积、多尺度残差块和 CBAM 模块, 旨在生成特征图维度 (Feature map dimensions, CHW) 为 $3 \times 128 \times 128$ 的甜樱桃图像。生成过程始于一个全连接层, 将输入的潜在空间向量 (随机噪声) 转换为一个具有足够信息量的扁平向量。这个向量被重塑为 512 个通道、8 像素 $\times$ 8 像素特征图。随后, 通过批量归一化、上采样和卷积层来逐步增大特征图的空间维度 (从 8×8 到 16×16 , 再到 32×32 , 然后到 64×64 , 最终到 128×128), 同时逐步减少特征图的深度 (从 512 到 256, 再到 128, 然后到 64, 最后到 32)。在这个过程中, 应用多尺度残差块进一步丰富特征图表示能

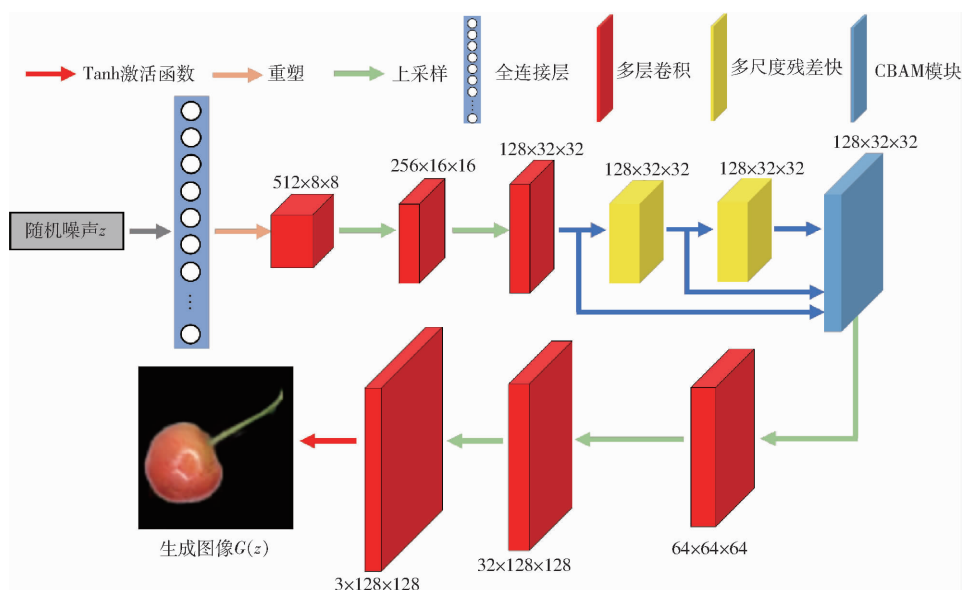


图 4 生成器结构图

Fig. 4 Structure of generative network

力, CBAM 模块对特征图进行特征优化。一系列的上采样和卷积操作依次进行, 逐步放大特征图尺寸, 并最终生成目标图像。

为了改进生成对抗网络生成图像质量或分辨率, 一般会采用增加模型参数, 即通过使用更深层次网络架构, 进而实现捕捉更多的图像细节。但这种方法会导致计算成本增加, 并且存在着难以调参、过拟合风险大的问题。为了进一步提升 SCGAN 对细节的关注度和特征显著性, 在特征图尺寸为 $128 \times 32 \times 32$ 的位置设置一个 CBAM 模块, 对通道和空间两个维度进行“动态加权”, 增强中间特征表达能力, 使模型能够更好地捕捉重要的局部和全局信息; 改善特征图传递效果, 在中间位置引入注意力机制可以避免过早使用而导致丢失有用信息或过晚应用导致信息的冗余, 从而提升生成图像质量; 确保在适当的阶段优化特征图; 平衡计算成本和模型性能, 在提升性能同时控制计算资源消耗; 提高生成器对复杂特征的建模能力, 生成更高质量和多样性的图像。

CBAM 是一个结合了通道和空间注意力机制的轻量级模块, 其通过增强特征表达和优化特征提取效率, 显著提升了网络对关键特征的关注度和图像质量, 同时并未显著增加计算负担。这一模块能够与卷积网络无缝集成, 实现端到端训练。CBAM 作为一个高效且轻量级的注意力模块, 可以灵活融入各种卷积网络架构并支持端到端学习。如图 5 所示, CBAM 模块通过两阶段注意力机制——通道注意力和空间注意力, 分别增强了对特征通道和图像关键区域的关注, 进一步提升了网络表现力。

CBAM 在通道注意力阶段利用全局平均池化和最大池化将输入特征图 F 缩减为两个描述符 F_{avg}^C 和

F_{max}^C , 并通过含两层全连接层的多层感知器 (MLP) 生成通道注意力图 $M_c(F)$, 从而强调每个通道的重要性, 即

$$M_c(F) = \sigma(\text{MLP}(F_{avg}^C) + \text{MLP}(F_{max}^C)) \quad (2)$$

式中 σ ——Sigmoid 函数

F_{avg}^C ——特征图 F 的每个通道的全局平均池化结果, 每个通道平均激活值

F_{max}^C ——特征图 F 的每个通道的全局最大池化结果, 每个通道最大激活值

空间注意力则计算描述符的平均值和最大值, 通过 7×7 卷积和 Sigmoid 函数形成的空间注意力图 $M_s(F')$ 为

$$M_s(F') = \sigma(\text{Conv}_{7 \times 7}(\text{Concat}(F_{avg}^{S'}, F_{max}^{S'}))) \quad (3)$$

式中 $\text{Conv}_{7 \times 7}$ —— 7×7 的卷积操作

$F_{avg}^{S'}$ ——特征图 F' 的每个通道的全局平均池化结果, 每个位置平均激活值

$F_{max}^{S'}$ ——特征图 F' 的每个通道的全局最大池化结果, 每个位置最大激活值

CBAM 通过结合通道和空间注意力图来增强网络对图像关键特征的捕捉, 提升模型性能和识别精度, 其定义为

$$M(F'') = M_c(F) \otimes M_s(F') \quad (4)$$

式中 $\otimes$ ——元素级乘法

这一机制优化特征提取, 细化到局部特征, 无需大量计算资源即可提高图像质量。在 SCGAN 网络中, CBAM 模块被应用在输入特征图 F 形状为 $(N, 128, 32, 32)$ 的位置。通道注意力模块将输入特征图通过平均池化和最大池化转换为 $(N, 128, 1, 1)$, 然后通过两层全连接层和 Sigmoid 函数生成通道注意力

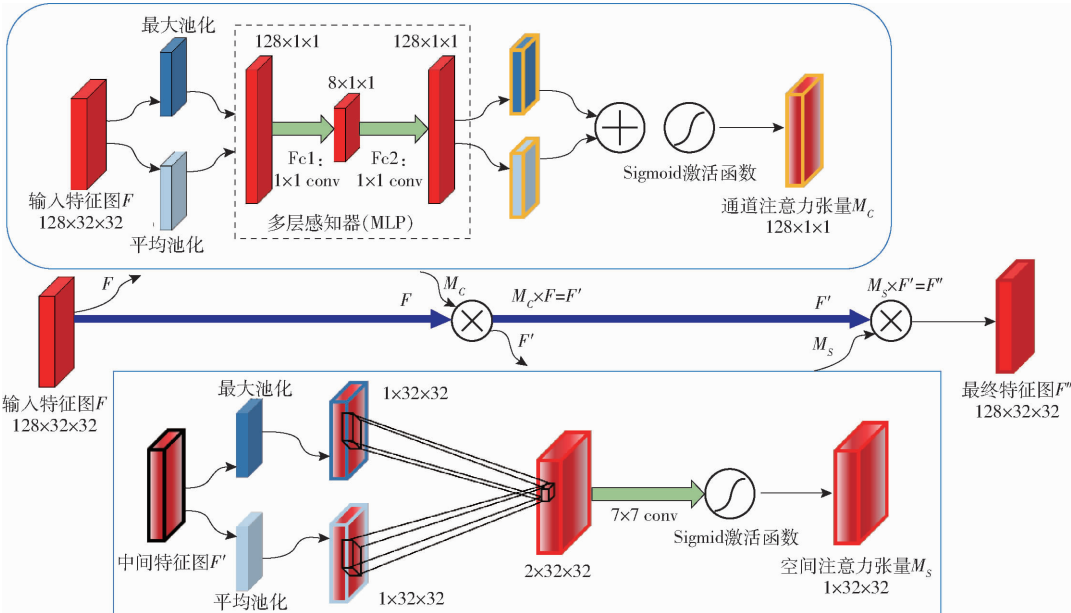


图 5 CBAM 结构图

Fig. 5 Structure of CBAM

图 M_c 。空间注意力模块对中间特征图进行通道平均池化和最大池化,生成两个 $(N, 1, 32, 32)$ 的特征图,并通过 7×7 卷积层和 Sigmoid 函数生成空间注意力图 M_s ,形状为 $(N, 1, 32, 32)$ 。最终输出的特征图经过 CBAM 模块的优化后,保持原有的形状 $(N, 128, 32, 32)$ 。

CBAM 模块进行最后的上采样和卷积步骤,将特征图尺寸从 32×32 上采样到最终的 128×128 ,并逐步减少特征深度至图像通道数 3,使用 Tanh 作为激活函数输出归一化后的图像数据。

为了克服深层网络训练中常见的梯度消失问题,并精确捕捉甜樱桃从细微裂纹到显著腐烂区域的多变外表和纹理特征,在 SCGAN 生成器的第 3 个卷积层(即 $C \times H \times W = 128 \times 32 \times 32$ 的卷积层)后面、CBAM 模块的前面连续内嵌入 2 个多尺度残差块,其结构如图 6 所示。

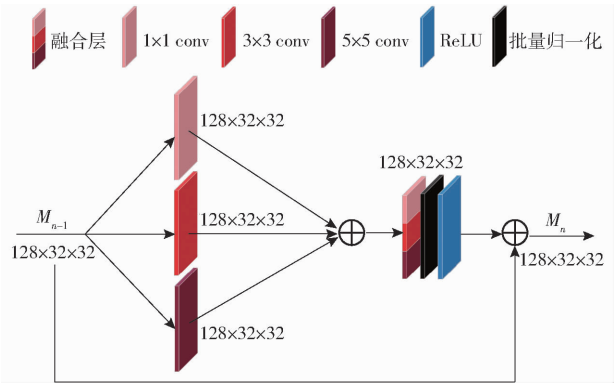


图 6 多尺度残差块结构图

Fig. 6 Structure of multi-scale residual block (MSRB)

在多尺度残差块中,输入张量形状为 $[128 \times 32 \times 32]$ 。输入张量首先被分成 3 个并行卷积层,

分别使用不同卷积核: 1×1 、 3×3 和 5×5 。 1×1 卷积层卷积核大小为 1×1 ,步幅为 1,填充为 0; 3×3 卷积层卷积核大小为 3×3 ,步幅为 1,填充为 1; 5×5 卷积层卷积核大小为 5×5 ,步幅为 1,填充为 2。以上 3 个卷积层的输出形状均保持为 $[128 \times 32 \times 32]$ 。然后将这 3 个卷积层的输出相加并经过批量归一化和 ReLU 激活函数处理。最后,通过残差连接将处理后输出与初始输入相加,最终输出的张量形状保持为 $[128 \times 32 \times 32]$ 。这种结构通过多尺度特征提取和残差连接,能够在不同空间尺度上提取关键信息,从而显著提升图像逼真度和模型泛化性能。残差连接设计确保了信息和梯度在深层网络中的有效传递,显著增强了训练稳定性,并加速了模型收敛速度^[22]。此外,这种多尺度处理策略不仅优化了梯度流动,降低了训练难度,还通过特征重用减轻了计算负担,提高了整体训练效率。

整个生成器结构优点在于其对特征提取和细节重建强大能力,以及通过注意力机制和残差连接增强学习稳定性,这些都是生成高质量图像所必需的。通过这种设计,生成器能够有效地从复杂数据分布中学习,并生成具有高度真实性图像,为各种图像生成任务提供了强大支持。

3.2 SCGAN 判别器

判别器网络 D 展现了一个精心设计的架构,结构如图 7 所示。该网络由一系列的谱归一化层、卷积层和随机丢弃层构成。它接收一个尺寸为 $128 \times 128 \times 3$ 的图像作为输入,并通过多个采用 3×3 大小卷积核卷积层以及全连接层进行深度处理。每一个卷积层后,都应用了 LeakyReLU 非线性激活函数

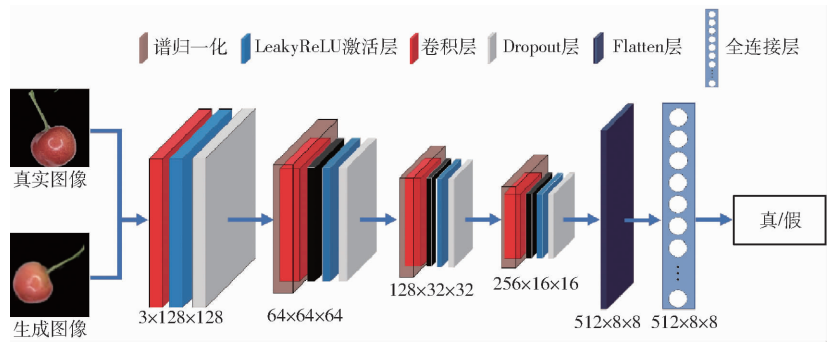


图 7 判别器结构图

Fig. 7 Structure of discriminative network

来增强模型表达能力。初始层使用 4×4 卷积核,步长为 2,并采用丢弃率 50% 进行初步特征提取,LeakyReLU 激活函数负斜率设为 0.2,以增强模型非线性表达。在深层卷积层中,为了增强 SCGAN 的结构稳定性,除了第一个卷积层未设置谱归一化以保证输入信息的完整性之外,之后每层卷积均配备了谱归一化。此外,还添加了批归一化、LeakyReLU 激活函数和 Dropout 层以防止过拟合。通过这一系列层次逐步增加通道数并减少空间尺寸。最终,Flatten 层和一个全连接层将卷积输出转化为输出向量,该向量评估图像真实性,为 GAN 模型提供了有效的训练反馈,确保模型对生成图像真伪的鉴别能力。

判别器太强,它可能会轻易地辨别出所有生成的图像,导致生成器无法有效地学习和改进。通过谱归一化,判别器能力被适度限制,有助于维持 GAN 训练中生成器和判别器之间的动态平衡。具体来说,谱归一化首先对网络层权重矩阵 W 进行奇异值分解 SVD。在数学上,奇异值分解将矩阵 W 分解为 3 个矩阵的乘积,即

$$W = U \Sigma V^T \quad (5)$$

其中 U 和 V 是正交矩阵,而 Σ 是对角矩阵,其对角线上的元素是 W 的奇异值。光谱范数 $\|W\|_{\text{spectral}}$,即 W 的最大奇异值,用来衡量矩阵 W 。其计算公式为

$$\|W\|_{\text{spectral}} = \max(\Sigma) \quad (6)$$

其中 $\max(\Sigma)$ 表示 Σ 中的最大值,即 W 最大奇异值。谱归一化通过将原始权重矩阵 W 除以其谱范数来实现归一化。归一化后的权重矩阵 $W_{\text{normalized}}$ 表示为

$$W_{\text{normalized}} = \frac{W}{\|W\|_{\text{spectral}}} \quad (7)$$

保证了权重矩阵不会变得过大,避免了梯度爆炸问题,提高了模型训练稳定性^[23]。

3.3 损失函数

在传统的生成对抗网络(GAN)训练过程中,损

失函数可能引发训练不稳定,导致网络难以有效收敛;此外,还可能出现模式崩溃现象,即生成器开始输出高度相似的图像,极大地减少了输出多样性。为了解决这些问题,引入 Wasserstein 距离作为一种新的损失函数,它已被证明可以显著提升训练稳定性和输出质量^[24-27]。Wasserstein 距离量化了将一个概率分布转换为另一个概率分布所需的“最小运输成本”,提供了一种更加平滑和连续的方式来优化生成器。使用 Wasserstein 距离公式为

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (8)$$

式中 $\inf$ ——所有可能情况中的最小值

P_r ——真实数据分布

P_g ——生成数据分布

$\Pi(P_r, P_g)$ —— P_r 和 P_g 组合得到的所有可能的联合分布集合

$\mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|]$ ——样本 x 和 y 的距离

Wasserstein 距离利用这种度量来优化生成器,使得训练过程中梯度更加平滑,减少了模式崩溃风险,并提供了更连续和有意义的损失,改进了模型收敛行为,使得 SCGAN 通常比传统 GAN 更加稳健。

4 实验

4.1 实验设置

实验平台硬件环境的中央处理器(CPU)为 Intel Xeon Silver 4314 型 CPU@2.40 GHz,图形处理器(GPU)为 NVIDIA Quadro RTX 4000。软件环境使用 Win 10 操作系统,Anaconda 版本为 4.14.0,Python 版本为 3.8.8,Pytorch 版本为 1.7.1,图像处理库为 OpenCV 4.6.0。

4.2 网络训练

生成器和判别器采用交替训练的方式,通过对抗过程持续地增强各自功能。这种方式在实施中经常遇到协调难题,可能引起训练不稳定性。为了解决这些挑战,制定了一套详尽的训练流程。SCGAN

训练步骤为:

Input: training samples $X_{real} = \{ (X_{real}) \}_{i=1}^n$, number of training epochs $E = 4\ 000$, batch size $n = 32$. Save the weight file every 100 iterations, with the learning rates for the generator and discriminator set to $lr_g = 0.000\ 3$ and $lr_d = 0.000\ 2$, respectively.

Initialization: Initialize the weights of the generator G and discriminator D.

Training:

for E training iterations do:

 Configure the real images X_{real} .

 Generate n d-dimensional random noise samples

$Z = \{ z_j^d \}_{j=1}^n$.

 Input the noise Z into the generator G.

 The generator G generates n fake samples $X_{fake} = \{ (X_{fake}) \}_{i=1}^n$.

 Input the n fake samples X_{fake} into the discriminator D.

 Calculate the loss of the discriminator D (L_D^{fake}), and take the negative of this loss as the loss of the generator G (L_G).

 Optimize G by adjusting its weights to minimize L_G .

 Evaluate D's ability to distinguish real images by calculating the loss for real images (L_D^{real}).

 Derive the total loss of D (L_D) by subtracting L_D^{real} from L_D^{fake} .

 Update D's weights to minimize L_D .

 If necessary, perform weight clipping to ensure D's weights remain within the desired range.

End.

Output: generating fake defective sweet cherry images $X_{fake} = \{ (X_{fake}) \}_{i=1}^n$.

4.3 生成图像质量评估

生成图像质量与多样性是用来衡量生成对抗网络性能的关键指标。质量保证生成图像具有可识别性,并与训练数据保持一致性,而多样性则开启了更广泛创新可能性,助力挖掘训练集中潜在的额外价值和信息。

弗雷歇初始距离 (Fréchet inception distance, FID) 是一个衡量真实图像与生成图像之间特征距离的指标,用于评估生成图像质量。FID 得分低意味着生成的图像与真实图像更为接近,显示出模型高性能和图像高质量。此外,FID 值变化趋势也能反映对抗网络收敛性。随着 GAN 的训练,FID 值应该逐渐减小并趋于稳定,这表明生成图像越来越接

近真实图像,模型逐渐收敛。通过定期计算和监控 FID 值,可以直观地观察模型在训练过程中的进展。较低且稳定的 FID 值不仅表明生成图像质量提升,也证明了模型架构和训练策略的有效性。因此,FID 值不仅是评估生成图像质量重要指标,也是衡量对抗网络收敛性和稳定性的有力工具。FID 计算式为

$$FID(real, fake) = \| \mu_{real} - \mu_{fake} \|^2_2 + \frac{\text{Tr}(\Sigma_{real} + \Sigma_{fake} - 2 \sqrt{\Sigma_{real} \Sigma_{fake}})}{2}$$
 (9)

式中 μ ——图像特征的均值向量

Tr——迹运算,用于测量协方差矩阵的规模和形状

μ_{real} 、 μ_{fake} ——真实数据集和生成数据集的特征的均值向量

Σ_{real} 、 Σ_{fake} ——真实数据集和生成数据集特征协方差矩阵

采用 T-SNE 技术和熵值分析,通过将生成的图像和真实图像进行降维处理,直观评估图像多样性。比较高维空间的概率分布 P 与低维映射后的分布 Q ,其中 P 基于高斯分布计算点对之间的相似度概率,而 Q 则基于 t 分布,利用 Kullback-Leibler divergence 表达这两个分布间的差异,即

$$C = KL(P \parallel Q) = \sum_i \sum_j p_{ij} \lg \frac{p_{ij}}{q_{ij}}$$
 (10)

式中 p_{ij} ——高维空间中数据点 i 和 j 共同出现的概率

q_{ij} ——低维空间中数据点 i 和 j 共同出现的概率

$\lg \frac{p_{ij}}{q_{ij}}$ —— p_{ij} 和 q_{ij} 比率对数,即两个概率之间相对熵

该方法通过将数据映射到低维空间并保持原始数据点间的局部关系,可视化生成图像与真实图像的分布差异。生成图像与真实图像的分布在低维空间中高度重叠,表明生成模型能够有效捕捉数据多样性;否则,显示出需要进一步优化的空间。

为进一步评价模型生成图像多样性,通过计算概率 P 分布的熵来量化数据不确定性和复杂性,熵 $H(P)$ 计算式为

$$H(P) = - \sum_i p_i \lg(p_i)$$
 (11)

式中 p_i ——概率分布中第 i 个事件的概率

高熵值表明图像类别内部展示出较大的复杂性和变异性。结合 T-SNE 的可视化结果,这不仅能有效地揭示生成图像多样性的潜在缺陷,还能验证生成图像与真实图像分布的相似性。通过这种方

法,可以全面评估和优化生成模型,以确保其生成的图像质量和多样性。

图 8 清晰地展示了在整个生成过程中各种 GAN 模型所记录的 FID 得分变化。从图可以看出,大多数 GAN 模型在训练初期的 FID 得分均呈明显的下降趋势,表明模型在学习过程中初期能够有效改进生成图像质量。然而,DCGAN 模型在训练后期显示出较大的得分波动,甚至呈上升趋势,表明模型可能遇到了训练不稳定,例如梯度消失或模式崩溃的技术难题。相比之下,WGAN 虽也有波动,但整体趋于平稳,显示其在训练过程中更为稳定。SCGAN 则表现出较低波动性和更稳定下降趋势,尤其是在处理单独腐烂果和裂果数据集时,其 FID 得分显著低于其他模型,分别为 64.36 和 59.97(表 2)。

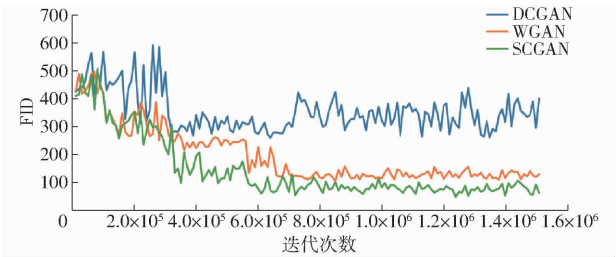


图 8 DCGAN、WGAN 和 SCGAN 模型的 FID 曲线

Fig. 8 FID scores of DCGAN, WGAN and SCGAN during generation

表 2 DCGAN、WGAN 和 SCGAN 模型 FID 值			
Tab. 2 FID scores of DCGAN, WGAN and SCGAN			
模型	DCGAN	WGAN	SCGAN
腐烂果	257.53	103.22	64.36
裂果	260.59	79.78	59.97

图 9 为 DCGAN、WGAN、SCGAN 3 种模型分别生成的两类缺陷甜樱桃图像在三维空间中的分布。从图 9 可知,与 SCGAN 相比,DCGAN 和 WGAN 所生成图像的空间分布与原始甜樱桃图像相比显示出更大的差异。从图中还可发现 SCGAN 所生成图像的空间分布与原始甜樱桃图像相比也存在着一定的差异,这是因为 DCGAN、WGAN、SCGAN 3 种模型不仅复制了训练数据,还在不同程度上内化了数据的潜在结构,增强了泛化能力,使其在未见过的数据上表现出更高的鲁棒性。通过在特征空间中创造不同的新样本,从而实现丰富数据集目标。为了更客观地分析生成图像与原始图像分布之间的差异,进而更精确地描述生成图像特征分布差异,可以通过计算每个类别在特征空间中的熵值。表 3 为原图与生成图像的熵值。

由表 3 可知,虽然 DCGAN 和 WGAN 在模拟原始图像多样性方面表现了局限性(DCGAN生成图

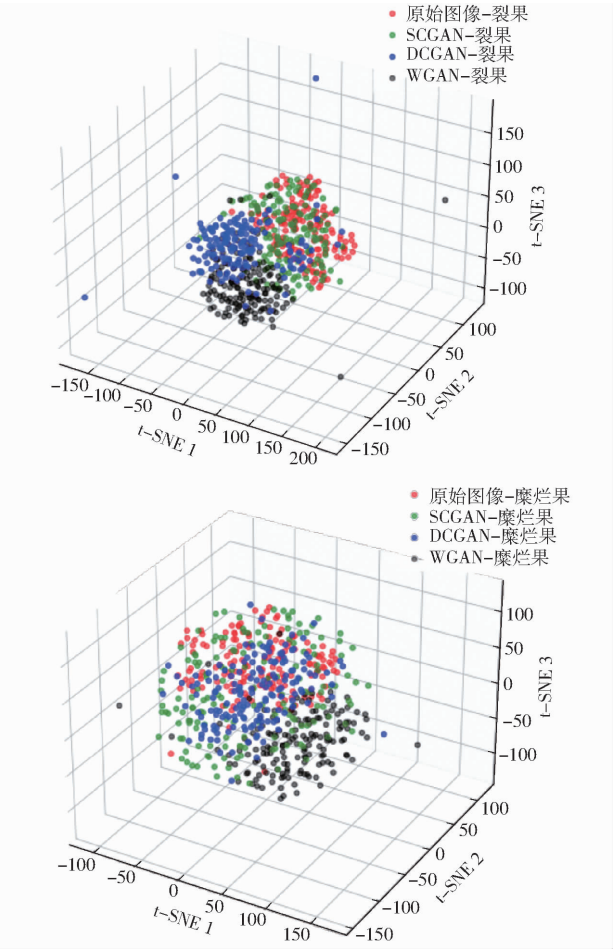


图 9 DCGAN、WGAN 和 SCGAN 生成的甜樱桃图像的 3D 分布可视化图

Fig. 9 3D distribution visualizations of sweet cherry images generated by DCGAN, WGAN and SCGAN

表 3 原图与生成图像的熵值				
Tab. 3 Entropy of original and generated images				
模型	DCGAN	WGAN	SCGAN	原图
腐烂果	2.685	2.667	2.949	3.269
裂果	2.589	2.730	3.207	3.247

信息熵为 2.589,WGAN 生成图信息熵为 2.667),但 SCGAN 在这方面做得更好。SCGAN 不仅保持了与原始甜樱桃图像较高的相似性(原始图信息熵为 3.269),其生成图像信息熵达到 3.207,显著高于其他模型。这不仅直接反映 SCGAN 在特征空间中的丰富多样性,也表明其在生成具有真实感图像的同时,能够更有效地捕捉到原始数据中的变异性,生成视觉上丰富多变的新图像。这种差异反映了生成模型在理解和复制数据本质属性方面的复杂性,即使是最优模型也难以完全达到原始图像多样性和复杂性。

4.4 数据集平衡前后分类结果

为了评估甜樱桃分类模型性能,选用总体准确率(Overall accuracy, OA)、精确率(Precision)、召回

率 (Recall) 以及 F1 值作为分类评估指标。

图 10 为使用不同生成式对抗网络模型进行数据增强前后的 VGG19 和 MobileNetV3 两个模型分类结果的混淆矩阵。使用原始数据的分类结果受分布不平衡的影响较大,缺陷甜樱桃分类准确率分别为 78.95%、57.14% 和 78.12%、65.22%。另外,可以看出,不同 GAN 模型的数据增广都可以在一定程度上缓解数据不平衡的问题。其中,通过 DCGAN 模型生成的数据增强后的总体准确率提高 1.25 个

百分点,对分类结果的改善甚微。尽管 WGAN 生成的樱桃图像质量相对较差,但它仍然学习了樱桃主要特征,因此生成的图像改善了分类模型效果。然而,由表 4 可知,通过 SCGAN 模型生成的数据增强后分类模型,OA 和 F1 值分别提高 3.15、16.44 个百分点和 4.92、13.94 个百分点。这是由于 SCGAN 生成的图像在相似性和多样性上的质量较高,可以有效提高不同分类模型在不平衡数据下的泛化能力和分类能力。

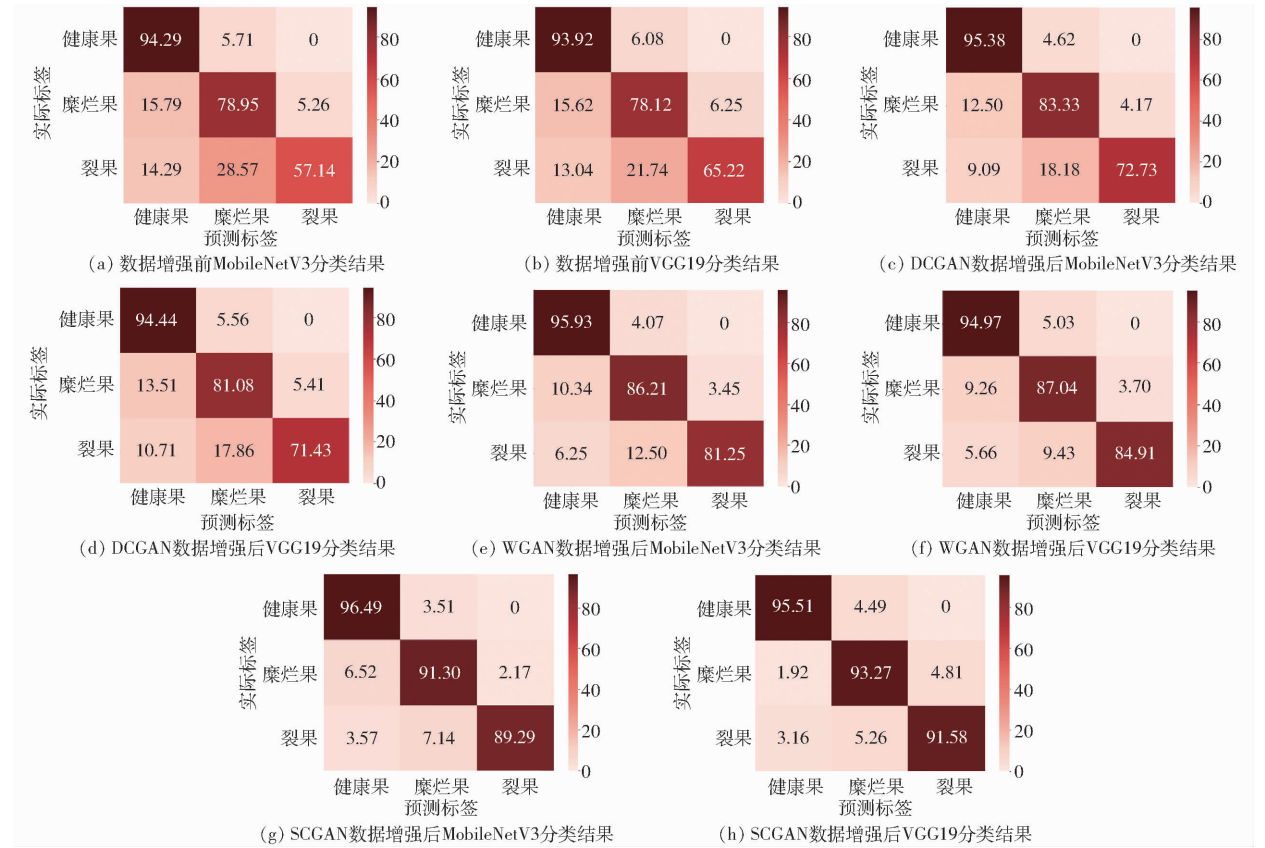


图 10 分类混淆矩阵

Fig. 10 Classification confusion matrices

表 4 用 DCGAN、WGAN 和 SCGAN 进行数据增强后的分类结果					
Tab.4 Classification results after data augmentation by DCGAN, WGAN and SCGAN %					
类型	指标	原数数据集	DCGAN	WGAN	SCGAN
MobileNetV3	总体准确率	91.54	92.79	93.55	94.69
	精确率	77.73	84.40	88.01	92.60
	召回率	76.79	83.81	87.80	92.36
	F1 值	75.94	83.52	87.60	92.38
VGG19	总体准确率	88.98	89.80	91.61	93.90
	精确率	81.57	84.36	89.43	93.30
	召回率	79.09	82.32	88.97	93.45
	F1 值	79.40	82.71	88.99	93.34

4.5 消融实验

为了评估提出的 SCGAN 中 MSRB、CBAM、谱归

一化以及 Wasserstein 距离损失函数的有效性,设计了相应的消融实验以验证其性能。图 11 为生成过程中的 FID 值。结果显示,随着迭代次数的增加,与消融模型相比,SCGAN 表现出更稳定的收敛性,并在后期实现了更低的 FID 值,这进一步证实了模型改进的有效性。

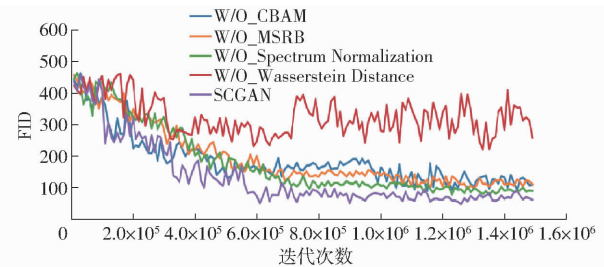


图 11 移除改进措施后的 FID 值

Fig. 11 FID scores after removing improvement measures

从图 11 可以看出,不同模型的 FID 值在初始迭代次数时都较高,但随着迭代次数的增加,大部分模型的 FID 值显著下降。SCGAN 模型下降趋势最为明显,最终表现最佳。相比之下,W/O_Wasserstein Distance 模型表现最差,其 FID 值在整个过程中波动较大且始终较高,无法稳定下降。而 W/O_CBAM 模型和 W/O_MSRB 模型在迭代过程中也逐渐下降并趋于稳定,但最终效果仍不及 SCGAN 模型。

表 5 移除改进措施后 FID 值
Tab.5 FID scores after removing improvement measures

模型	W/O_MSRB	W/O_CBAM	W/O_Spectrum Normalization	W/O_Wasserstein Distance	SCGAN
腐烂果	87.15	92.32	72.86	219.94	64.36
裂果	84.52	90.84	69.54	204.53	59.97

表 6 移除改进措施后分类结果
Tab.6 Classification result after removing improvement measures

模型	指标	W/O_MSRB	W/O_CBAM	W/O_Spectrum Normalization	W/O_Wasserstein Distance	SCGAN
MobileNetV3	总体准确率	93.95	93.90	94.09	91.21	94.69
	精确率	90.41	90.06	90.91	86.45	92.60
	召回率	89.50	89.28	90.71	85.26	92.36
	F1 值	89.77	89.48	90.66	85.60	92.38
VGG19	总体准确率	93.18	92.74	93.48	91.24	93.90
	精确率	91.92	90.94	92.54	86.78	93.30
	召回率	92.17	91.25	92.75	86.85	93.45
	F1 值	91.99	91.03	92.61	86.57	93.34

5 结论

(1)证实了通过 SCGAN 进行数据增强可有效提升甜樱桃分类模型在数据不平衡条件下的性能。经过一系列改进和优化后,SCGAN 在数据增强方面表现出色,将 VGG19 和 MobileNetV3 的 OA 和 F1 值分别提高 3.15、16.44、4.92、13.94 个百分点,展示了其在生成数据集的相似性和多样性方面的优势。

(2)在模型结构方面,以 Wasserstein 距离作为损失函数不仅增强了训练稳定性并加快了收敛速

表 5 为不同模型生成的缺陷甜樱桃图像的 FID 值。相比消融模型,由 SCGAN 生成的缺陷甜樱桃图像质量最高,FID 值分别降低 155.58 和 144.56,这从定量角度证明了改进措施有效提升了合成图像质量。表 6 为数据增强后分类结果。可以看出,使用 SCGAN 进行数据增强后,分类结果显著优于其他消融模型,F1 值分别提高 6.78、6.77 个百分点。表明了改进措施的有效性。

度。集成的多尺度残差块(MSRB)能够在不同空间尺度上提取关键信息,显著提升了图像逼真度。此外,集成的 CBAM 注意力机制优化了特征提取过程,增强了通道与空间特征的表达。谱归一化技术通过适度限制判别器的能力,维持了生成器与判别器之间的动态平衡,不仅提高了生成性能,也进一步提升了甜樱桃质量分级模型性能。

(3)SCGAN 的数据增强显著提高了甜樱桃分类模型性能,证明了其在多角度优化甜樱桃分级方面的有效性。

参 考 文 献

[1] 彭明霞. 基于图像处理的棉花幼苗期杂草识别方法研究[D]. 武汉: 华中农业大学, 2023.
PENG Mingxia. Research on the method of identifying weeds in cotton seedling stage based on image processing[D]. Wuhan: Huazhong Agricultural University, 2023. (in Chinese)

[2] CENGGORO T W. Deep learning for imbalance data classification using class expert generative adversarial network[J]. Procedia Computer Science, 2018, 135: 60–67.

[3] CHEN H, LI C, YANG W, et al. Deep balanced cascade forest: an novel fault diagnosis method for data imbalance[J]. ISA Transactions, 2022, 126: 428–439.

[4] SUN C, CUI H, ZHOU W, et al. Epileptic seizure detection with EEG textural features and imbalanced classification based on EasyEnsemble learning[J]. International Journal of Neural Systems, 2019, 29(10): 1950021.

[5] LIU J, XU K, CAI B, et al. Fault prediction of on-board train control equipment using a CGAN – Enhanced XGBoost method with unbalanced samples[J]. Machines, 2023, 11(1): 114.

[6] WANG H, HU T, XU Z, et al. A radar waveform recognition method based on ambiguity function generative adversarial network

data enhancement under the condition of small samples[J]. IET Radar, Sonar & Navigation, 2023, 17(1): 86 – 98.

[7] 王军, 高放, 省海先, 等. 基于 Swin Transformer 生成对抗网络的图像生成算法[J]. 小型微型计算机系统, 2024, 45(1): 241 – 248.
WANG Jun, GAO Fang, SHENG Haixian, et al. Image generation algorithm based on Swin Transformer to generate adversarial networks[J]. Journal of Chinese Computer Systems, 2024, 45(1): 241 – 248. (in Chinese)

[8] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1 – 48.

[9] 韦婷婷, 葛晓月, 熊俊涛. 基于层级多标签的农业病虫害问句分类方法[J]. 农业机械学报, 2024, 55(1): 263 – 269, 435.
WEI Tingting, GE Xiaoyue, XIONG Juntao. Hierarchical multi-label classification of agricultural pest and disease interrogative questions[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 263 – 269, 435. (in Chinese)

[10] GEETHARAMANI G, PANDIAN A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network [J]. Computers & Electrical Engineering, 2019, 76: 323 – 338.

[11] OFUSA K, MIYAZAKI T, SUGAYA Y, et al. Glyph-based data augmentation for accurate kanji character recognition[C]// 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2017, 1: 597 – 602.

[12] ZHANG Y D, DONG Z, CHEN X, et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation[J]. Multimedia Tools Applications, 2019, 78: 3613 – 3632.

[13] ZHOU G, CHEN Y, CHIEN C. On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks[J]. BMC Medical Informatics Decision Making, 2022, 22(1): 226.

[14] 陈佛计, 朱枫, 吴清潇, 等. 生成对抗网络及其在图像生成中的应用研究综述[J]. 计算机学报, 2021, 44(2): 347 – 369.
CHEN Foji, ZHU Feng, WU Qingxiao, et al. A survey about image generation with generative adversarial nets[J]. Chinese Journal of Computers, 2021, 44(2): 347 – 369. (in Chinese)

[15] KHOSLA C, SAINI B S. Enhancing performance of deep learning models with different data augmentation techniques: a survey [C]// 2020 International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2020: 79 – 85.

[16] PEREZ L, WANG J. The effectiveness of data augmentation in image classification using deep learning[J]. arXiv preprint arXiv, 2017: 1712.04621.

[17] 李云红, 张蕾涛, 李丽敏, 等. 基于 CycleGAN – IA 方法和 M – ConvNext 网络的苹果叶片病害图像识别[J]. 农业机械学报, 2024, 55(4): 204 – 212.
LI Yunhong, ZHANG Leitao, LI Limin, et al. Image recognition of apple leaf disease based on CycleGAN – IA method and M – ConvNext network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 204 – 212. (in Chinese)

[18] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems, 2014: 2672 – 2680.

[19] 张林鎔, 巴音塔娜, 曾庆松. 基于 StyleGAN2 – ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. 农业机械学报, 2024, 55(1): 241 – 252.
ZHANG Linxuan, BA Yintana, ZENG Qingsong. Grape disease detection method based on StyleGAN2 – ADA and improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 241 – 252. (in Chinese)

[20] 刘羿漩, 葛广英, 齐振岭, 等. 基于改进深度卷积生成对抗网络的刺绣图像修复[J]. 激光与光电子学进展, 2023, 60(20): 68 – 78.
LIU Yixuan, GE Guangying, QI Zhenling, et al. Research on embroidery image restoration based on improved deep convolutional generative adversarial network[J]. Laser & Optoelectronics Progress, 2023, 60(20): 68 – 78. (in Chinese)

[21] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv, 2015: 1511.06434.

[22] 陶兆胜, 宫保国, 李庆萍, 等. 基于残差网络和特征融合的小麦图像修复模型[J]. 农业机械学报, 2023, 54(3): 318 – 327.
TAO Zhaosheng, GONG Baoguo, LI Qingping, et al. Wheat image inpainting based on residual networks and feature fusion [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 318 – 327. (in Chinese)

[23] 李洁松, 刘韬, 伍星. 谱归一化 CycleGAN 的轴承故障迁移诊断研究[J]. 振动与冲击, 2023, 42(24): 282 – 289.
LI Jiesong, LIU Tao, WU Xing. Spectral normalization CycleGAN for bearing fault transfer diagnosis[J]. Journal of Vibration and Shock, 2023, 42(24): 282 – 289. (in Chinese)

[24] 李庆旭, 王巧华, 马美湖. 基于生成对抗网络的禽蛋图像数据生成研究[J]. 农业机械学报, 2021, 52(2): 236 – 245.
LI Qingxu, WANG Qiaohua, MA Meihu. Poultry egg image data generating based on generative adversarial network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2): 236 – 245. (in Chinese)

[25] 詹燕, 胡蝶, 汤洪涛, 等. 基于改进生成对抗网络的图像数据增强方法[J]. 浙江大学学报(工学版), 2023, 57(10): 1998 – 2010.
ZHAN Yan, HU Die, TANG Hongtao, et al. Image data augmentation method based on improved generative adversarial network[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(10): 1998 – 2010. (in Chinese)

[26] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]// International Conference on Machine Learning. PMLR, 2017: 214 – 223.

[27] 王志强, 于雪莹, 杨晓婧, 等. 基于 WGAN 和 MCA – MobileNet 的番茄叶片病害识别[J]. 农业机械学报, 2023, 54(5): 244 – 252.
WANG Zhiqiang, YU Xueying, YANG Xiaojing, et al. Tomato leaf diseases recognition based on WGAN and MCA – MobileNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(5): 244 – 252. (in Chinese)