



DOI:10.12404/j.issn.1671-1815.2403871

引用格式:张振利,陈源,付豪,等.基于改进 EfficientNetV2 网络的垃圾图像分类算法[J].科学技术与工程,2025,25(10):4229-4238.

Zhang Zhenli, Chen Yuan, Fu Hao, et al. A garbage image classification algorithm based on the improved EfficientNetV2 network[J]. Science Technology and Engineering, 2025, 25(10): 4229-4238.

基于改进 EfficientNetV2 网络的垃圾图像分类算法

张振利,陈源,付豪,曾璐*

(江西理工大学电气工程与自动化学院,赣州 341000)

摘要 目前主流垃圾图像分类算法中存在数据集普适性差、垃圾识别种类少、分类算法局限于特定环境等问题。针对这些问题,结合垃圾图像分类的快速性与准确率的要求,提出了一种基于改进 EfficientNetV2 网络的垃圾图像分类算法。该算法以 EfficientNetV2 网络作为基准模型,通过添加 SK(selective kernel)注意力机制提升分类的快速性,使用迁移学习策略提升分类的准确率。该算法利用深度学习模型框架对垃圾图像进行处理,无需对数据集图像特征进行人工提取,在实现对垃圾图像快速准确分类的同时增加了垃圾识别的种类。实验表明,新的算法在自建数据集上的准确率为 99.71%,相较于 GoogleNet 等其他算法,提升了至少 4.77%。在时间上相较于 VggNet19 算法等,提升了至少 50%。通过改进 EfficientNetV2 网络,实现了更为准确快速的垃圾分类,为日益激增的垃圾问题提供了一种科学高效的解决方案。

关键词 垃圾分类;深度学习;EfficientNetV2;卷积神经网络;SK 注意力机制

中图分类号 TP391.4;

文献标志码 A

A Garbage Imageclassification Algorithm Based on the Improved Efficientnetv2 Network

ZHANG Zhen-li, CHEN Yuan, FU Hao, ZENG Lu*

(School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China)

[Abstract] An improved version of the EfficientNetV2 network is presented for garbage image classification to address the limitations of mainstream algorithms, such as poor dataset universality, limited recognition types, and algorithmic constraints in specific environments. The proposed algorithm emphasized both classification speed and accuracy. The EfficientNetV2 network was utilized as the baseline model, and classification speed was enhanced through the incorporation of the SK (selective kernel) attention mechanism. Transfer learning strategies were employed to improve classification accuracy. By leveraging deep learning model frameworks for garbage image processing, the need for manual feature extraction from dataset images was eliminated, and the scope of garbage recognition was expanded. Experimental results demonstrate that the proposed algorithm achieves an accuracy of 99.71% on a self-built dataset, which is an improvement of at least 4.77% compared to other algorithms, such as GoogleNet. Furthermore, in terms of time efficiency, the proposed algorithm outperforms algorithms like VggNet19 by at least 50%. Through the enhancement of the EfficientNetV2 network, accurate and faster garbage classification is enabled, providing a scientific and efficient solution to the growing challenges posed by garbage issues.

[Keywords] garbage classification; deep learning; EfficientNetV2; convolutional neural network; split-attention mechanism with switchable normalization

随着社会的飞速发展,垃圾的产生越来越多,如何科学有效便捷地处理日益激增的垃圾,已成为亟待解决的问题之一,这对人们的身心健康、幸福生活、环境保护和可持续发展具有非常重要的意义。常见的垃圾涵盖废纸、塑料、玻璃、金属、厨房垃圾等种类,垃圾图像识别的快速性与准确率是现

阶段各种算法关注的重点。现如今的垃圾末端分类处理中,图像处理是最为重要的一环。通过数学模型、特定算法等可以对相关垃圾图像数据集进行训练,得到相对可靠的分类结果。近年来,随着数据处理的进步,卷积神经网络已被证明是处理复杂视觉任务的有效架构^[1]。卷积神经网络是一组相

收稿日期:2024-05-24; 修订日期:2025-01-14

基金项目:国家自然科学基金(62363013)

第一作者:张振利(1976—),男,汉族,河北滦县人,硕士,副教授。研究方向:检测技术与控制。E-mail:47717770@qq.com。

*通信作者:曾璐(1981—),女,汉族,江西赣州人,硕士,副教授。研究方向:智能控制技术。E-mail:zenglu@jxust.edu.cn。

投稿网址:www.stae.com.cn

互连接的神经元,通过处理系统的输入,并根据相应数据的重要程度和其他参数输出的响应程度来为数据处理分配逻辑权重。随着科学技术的不断发展,卷积神经网络已经被广泛地应用到包括垃圾图像分类在内的各种图像分类领域中^[2]。常用的垃圾图像分类方法有基于轻量化的YOLO模型^[3],该方法通过轻量化设计和高效特征提取技术实现了高精度检测和适应实际场景的能力。有一种基于小型卷积神经网络(convolutional neural network, CNN)的垃圾分类方法^[4],该方法通过优化预处理算法和轻量化设计,拥有广泛的实际应用潜力。有一种基于ShuffleNet网络的方法^[5],该方法通过轻量化设计和先进技术应用,在快速响应方面表现出色,适合资源受限设备上的垃圾分类应用。有一种基于DenseNet网络的方法^[6],该方法通过数据增强和遗传算法优化,显著提高了垃圾分类精度和模型的可解释性。有基于MobileNet网络的方法^[7-8],这些方法通过创新显著减少了模型的参数量和计算量,同时提升了分类精度,适合于嵌入式设备和移动设备的实时垃圾分类应用。有基于ResNet网络的方法^[9-10],这些方法实现了对垃圾图像实现高精度和高效能的分类。还有一些研究提出了一些方法,例如利用CapSA优化CNN超参数,并采用ECOC和ANN的混合学习模型进行分类^[11];通过均值漂移滤波、色差灰度模型与改进的OTSU方法^[12];结合像素级空间注意力机制(pixel-level spatial attention, PSATT)、标签平滑正则化损失函数和阶梯形OneCycle学习率控制方法^[13],这些方法各具优势,实现了对垃圾图像进行高效和自动化处理。近些年来,基于深度学习的垃圾图像分类方法虽然取得了很大进展,但同样面临着一些不足,在泛化能力、数据集缺乏普适性、实时性、长期稳定性、硬件兼容性和计算资源需求等方面仍需进一步

研究、验证和优化,以确保在各种实际场景中的有效性和可靠性。

本文研究提出一种新的基于深度学习的垃圾图像分类算法,它使用轻量级的卷积神经网络模型EfficientNetV2作为主干网络,引入SK(selective kernel)注意力机制对移动倒置瓶颈卷积模块(mobile inverted bottleneck convolution, MBConv)进行更新提升垃圾图像分类任务的快速性,使用迁移学习策略提升分类任务的准确性。改进后的算法具备图像特征自动提取的能力,可实现对垃圾图像快速准确的分类。

1 基于EfficientNetV2网络的垃圾图像分类方法

1.1 EfficientNet网络

传统的神经网络模型通过对图像数据分辨率、网络通道数量和网络深度这3个维度中的任意一个进行常规缩放获得更复杂的网络,如图1所示。Tan等^[14]提出了EfficientNet网络,使用了复合缩放法,通过设置适当的比例扩展输入图像数据分辨率、网络深度和网络通道数3个维度,以获得所需的最佳参数,最大限度地提高识别精度,得到性能更优的模型。EfficientNet网络在使用过程中,出现以下问题:训练图片尺寸很大,训练时间非常长;在浅层网络中使用深度卷积结构会导致训练速度较慢,无法充分利用现有的加速器;在EfficientNetV1中,所有阶段会同时对宽度和深度进行放大,但是每个阶段的参数量和结构是不同的,所以相同放大的方法在使用中不是最优解。为更好地解决以上问题,Tan等^[15]提出了EfficientNetV2网络。

1.2 EfficientNetV2网络

EfficientNetV2网络在训练速度和参数量方面的表现都优于之前的网络,研究验证对于不同的

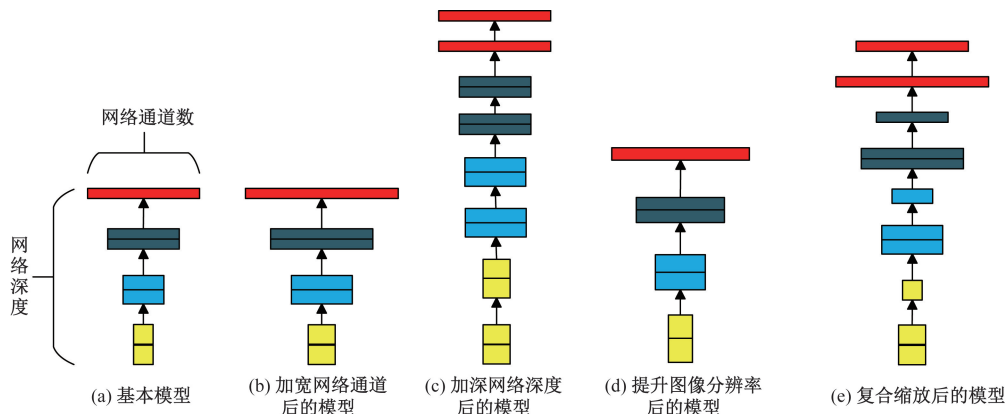


图1 模型缩放示意图

Fig. 1 Model compression schematic

图像尺寸,应该使用不同的正则化方法,新的研究能够自动根据不同图像尺寸使用动态的正则化方法,通过这种改进,模型的预测准确率得到了提高,同时还提升了训练的速度^[16]。因此,选取 EfficientNetV2 网络应用于垃圾分类当中。图 2 为 EfficientNetV2 网络的结构图。

EfficientNetV2 模型的基本结构包括 Fused-MBConv 模块和 MBConv 模块。图 3 展示了这两种结构。这两种模块均引入了 BN 层和 SENet 神经网络中提出的 SE (squeeze-and-excitation networks) 模块^[17],其中 BN 层在深层神经网络中对每个隐层神经元的输入值进行白化操作,SE 模块实现增强网络模型提取关键有用信息的能力。

在 EfficientNetV2 网络中,引入了几项架构改进,在保持性能的同时进一步降低了计算复杂度。

初始卷积层(stem):设计了初始卷积层用来更有效地处理输入图像,减少操作数。反向残差块:这些块使用倒置残差和线性瓶颈,减少了参数和浮点运算数。高效块设计:EfficientNetV2 中每个块的结构经过优化,以最小化冗余操作,例如使用挤压-激励机制和先进的归一化技术。EfficientNetV2 引入了一个尺度系数 ϕ ,同时调整网络的宽度、深度和分辨率。

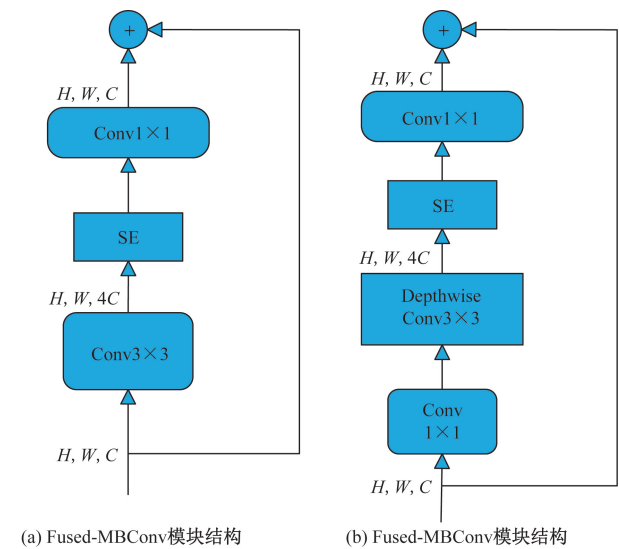
通过这些创新,EfficientNetV2 在减少计算复杂度的同时,保持或提高了准确性和速度等性能指标。

1.3 SK 注意力机制

注意力机制在深度学习领域中起着至关重要的作用,尤其是在处理图像和序列数据时。它可以帮助模型集中在图像中最相关的部分,从而提高分类的准确性和效率。在决定将注意力机制添加到

哪个位置时,需要考虑到模型的整体架构、处理的数据类型以及目标任务的具体需求。

SKNet 使用动态选择机制,该机制能够根据目标物体的尺寸自适应地选择不同的感受野。这个机制主要作用于卷积核,对卷积核的敏感度进行动态的调整,根据不同大小的图像生成相对应大小的卷积核,并尝试找到最适合的卷积核。这种机制的目的是提高模型对不同尺寸目标的感知能力。SKNet 的设计使得它具有目标尺度自适应的特点,从而更好地适应不同尺度的目标检测任务,SKNet 的网络结构如图 4 所示。它的动态选择机制允许神经



H 为图像或特征图的高度,即垂直方向上的像素数量; W 为图像或特征图的宽度,即水平方向上的像素数量; C 为图像或特征图的通道数;Conv (Convolution) 为卷积;Depthwise Conv 为卷积操作的一种变体,是轻量化神经网络中常用的高效卷积操作

图 3 Fused-MBConv 模块结构和 MBConv 模块结构

Fig. 3 Structure of Fused-MBConv and MBConv

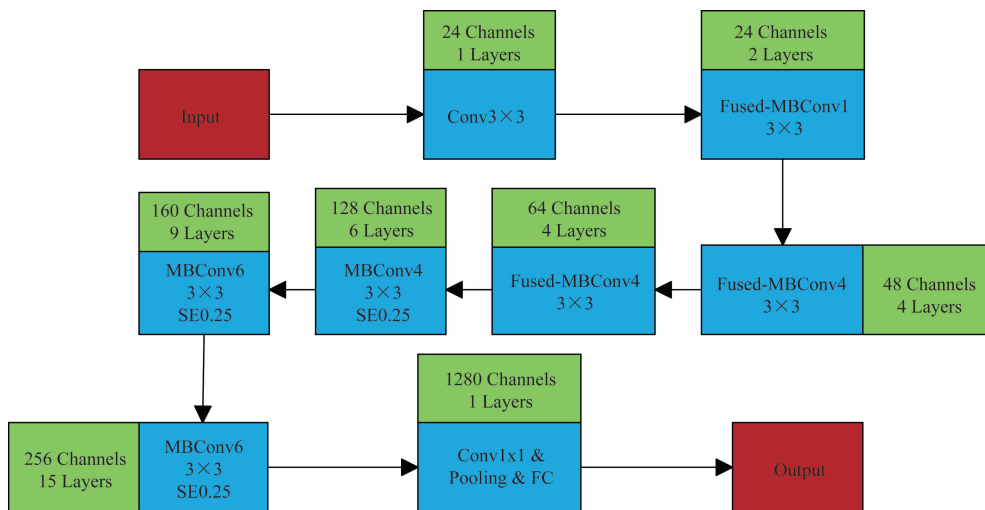
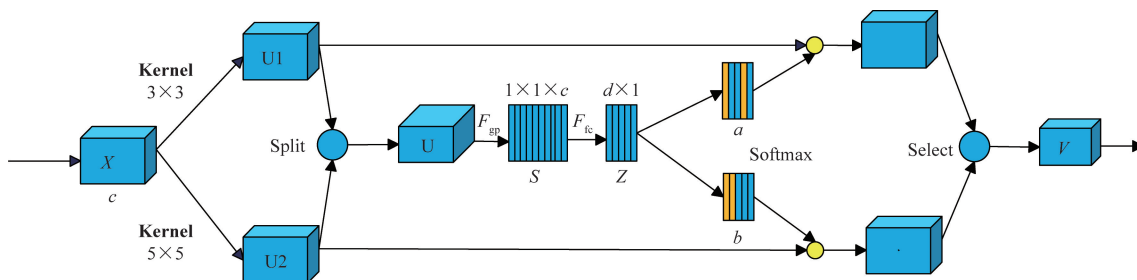


图 2 EfficientNetV2 网络的结构图

Fig. 2 The structure diagram of EfficientNetV2 network



Kernel 为一个小尺寸的权重矩阵,在卷积操作中作为滑动窗口,用于在输入数据上提取局部特征;Split 为处理多尺度特征的重要步骤;Softmax 主要用于生成每个感受野(或每个尺度)的权重;Select 过程是从生成的多尺度特征中选择最相关的特征,并为每个特征分配一个权重,从而增强最终的输出特征

图4 SKNet网络结构图

Fig. 4 SKNet network architecture diagram

元根据不同刺激进行感受野的调节,从而提高了模型的感知能力和准确性。这种自适应性的特点使得 SKNet 在处理多尺度目标时更具优势,从而为目标检测任务带来更好的效果。

SKNet 模块的处理主要包括以下 3 个步骤。

第一步:Split。该操作通过使用两个大小不一样的卷积核对输入的特征图进行 Group Convolution 等操作,如使用 2×2 、 8×8 的卷积核同时进行处理,较小的卷积核(如 2×2)通常更适合捕捉局部细节特征,而较大的卷积核(如 8×8)则更适合捕捉全局上下文信息,会得到两个不同的特征图(feature map),即 U_1 和 U_2 。主要测试不同卷积核对目标的敏感度,以提高检测精度。

第二步:Fuse。对第一步得到的两个不同的特征图 U_1 和 U_2 进行融合,可以使用逐元素相加或逐元素连接的方式将它们合并为一个融合特征图。然后采用全局平均池化操作将该融合特征图转换为一个全局特征向量 S ,以嵌入全局信息,利用全连接获取向量 Z ,从而达到准确自适应选取目的。其中 F_{gp} 表示全局平均池化操作, F_{fc} 表示先降维再升维的两层全连接层。最终,通过门控机制和元素相加的方式将两个分支的信息融合,并通过全局平均池化和全连接层操作嵌入全局信息,得到最终的特征向量 S 。这个特征向量 S 可以在后续任务中使用。

$$S_c = F_{gp}(U_c) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W U_c(i,j) \quad (1)$$

$$Z = F_{fc}(S) = \delta[B(W_s)] \quad (2)$$

式中: $U_c(i,j)$ 为特征图 U_c 中位于位置 (i,j) 的像素值; H 和 W 分别为特征图 U_c 的高和宽; c 为通道数; S_c 为全局平均池化的结果,通常是一个向量,维度为 c 。

通道数的选取会直接影响到模型的性能和计算复杂度。首先要考虑输入特征的维度,通道数不能大于输入特征的维度,否则会导致无法计算或计算异常。其次,通道数的增加会增加计算量,因此

需要根据计算资源的限制来选择合适的通道数,在性能和计算效率之间找到平衡。最后,不同的任务可能对通道数有不同的要求,复杂的任务可能需要更多的通道来捕捉更复杂的特征交互。 W_s 表示一个全连接层的权重矩阵。它对输入 S 进行线性变换,将其映射到另一个特征空间。在门控机制中,使用 ReLU 函数(δ)进行激活, B (批量标准化)用于归一化。特征维度 d 的计算通过取最大值 $\max(c/r, L)$ 来确定,其中 r 为减速率, L 为 d 的最小值。最终得到的特征向量 S 可用于后续任务。

第三步:Select。在向量 Z 上执行 softmax 操作以获得 a_c 、 b_c ,然后将原始特征 U_1 和 U_2 分别与对应的权重乘积 a_c 和 b_c 逐元素相乘,并对结果进行逐像素累加,最终得到输出特征 V 。 V 经过加权和操作后得到的输出特征图。它整合了不同分支特征图 U_1 和 U_2 的信息,且权重由 Softmax 操作动态决定。由于 a_c 与 b_c 之间函数值之和等于 1,从而达到设定每个分支特征图权重的目的。

$$a_c = \frac{e^{A_c Z}}{e^{A_c Z} + e^{B_c Z}} \quad (3)$$

$$b_c = \frac{e^{B_c Z}}{e^{B_c Z} + e^{A_c Z}} \quad (4)$$

$$V_c = a_c U_{1c} + b_c U_{2c} \quad (5)$$

式中: $A, B \in \mathbf{R}^{d \times c}$; $A_c \in \mathbf{R}^{1 \times d}$, 表示 A 的第 c 行; a_c 为 a 中的第 c 个元素; B_c 和 b_c 的含义同上。

1.4 SK 注意力机制的添加

EfficientNetV2 网络被分成 8 个阶段,每个阶段都有不同的结构和功能。EfficientNetV2 模型的总体结构如表 1 所示。阶段 1:这是一个普通卷积层,使用 3×3 大小的卷积核和步距为 2,同时包含批量归一化(BN)层和激活函数 Swish。阶段 2~阶段 4:这些阶段都由 Fused-MBConv 模块进行重复累加所构成,表中的层数表示该阶段中 Fused-MBConv 结构的重复次数。Fused-MBConv 结构的设计旨在减少

网络的参数量和计算量,以提高网络的效率。阶段5~阶段7:这些阶段都由 MBConv 结构组成,同样是重复累加得到的。表中的层数表示该阶段 MBConv 结构的重复次数。阶段8由 1×1 的普通卷积层、池化层和全连接层构成。这些层的组合用于进一步处理和提取特征,并完成最终的特征处理和分类。通过使用 Fused-MBConv 和 MBConv 结构, EfficientNetV2 网络能够在保持较低的参数和计算量的同时,提供强大的特征提取能力。

在 EfficientNetV2 中, MBConv^[18] 结构是对 MobileNetV2 网络模型中的倒残差结构进行微调的一种改进。相比于传统的残差结构,倒残差结构先升维后降维,从而减小信息损失的可能性,形成更稀疏的特征表示。主要进行了两个方面的调整:第一,使用了不同的激活函数, EfficientNetV2 的 MBConv 用 Swish 激活函数替换 Relu6 激活函数, Swish 激活函数能够提供更好的非线性激活能力;第二,在每个 MBConv 中内置通道注意力机制,在深度卷积层之后增加 SE 结构,计算通道权重,从而增强模型的表达能力。通过这些微调, EfficientNetV2 中的 MBConv 结构在保持倒残差结构优点的同时,增强了特征提取和表示能力。MBConv 结构如图 5 所示。

MBConv 结构由以下几个关键组件组成。 1×1 普通卷积层:这是 MBConv 结构的起始部分,使用 1×1 的卷积核、步距为 1,同时包含批量归一化

表 1 EfficientNetV2 模型的总体结构

Table 1 Overall architecture of EfficientNetV2-S model

阶段	操作	步长	通道数	层数
1	Conv 3×3	2	24	1
2	Fused-MBConv1, $k3 \times 3$	1	24	2
3	Fused-MBConv4, $k3 \times 3$	2	48	4
4	Fused-MBConv4, $k3 \times 3$	2	64	4
5	MBConv4, $k3 \times 3$, SE0.25	2	128	6
6	MBConv6, $k3 \times 3$, SE0.25	1	160	9
7	MBConv6, $k3 \times 3$, SE0.25	2	256	15
8	Conv 1×1 & Pooling & FC	—	1 280	1

(BN) 和 SiLU (Swish) 激活函数,它主要起到升维的作用,增加特征的维度; $k \times k$ 深度卷积层:在 1×1 卷积层之后,接入一个 $k \times k$ 的深度卷积层,其中 k 的值通常为 3 或 5,较小的 k 可以保留更多的局部特征,适合于捕捉图像中的细节信息;较大的 k 则有助于捕捉更大的上下文信息。选择合适的 k 取决于任务的复杂性和资源限制。深度卷积层的步距可以是 1 或 2,步距为 1 时,输出特征图大小与输入相似,适合于保留更多的空间信息;步距为 2 时,特征图尺寸减半,适合于减少计算量和参数数量,同时增大感受野。选择步距时需要考虑输入特征图的大小和网络的深度。这一步骤大大减少计算量和参数数量,提高了网络的效率;SE 模块:在深度卷积层之后,接入 SE 模块。SE 模块通过引入全局信息池化操作和一层全连接层,计算每个通道的权重,以增强关键信息的表达能力,SE 模块可以帮助网络更好地捕捉重要的特征; 1×1 普通卷积层:在 SE 模块之后,使用 1×1 的卷积核和批量归一化 (BN) 层,起到降维的作用,减少特征的维度;Dropout 层:MBConv 结构的最后一步是 Dropout 层,用于随机深度。随机深度是指根据倒残差层数 l 的线性函数,以一定的概率丢弃整个 MBConv 模块,从而减小网络的深度。这种随机深度并不使用传统的随机失活方式,而是以一定的概率使整个模块消失;MBConv 模块结构的设计旨在提高网络的效率和表达能力。通过升维、深度卷积、SE 模块、降维和随机深度等步骤,它能够同时减少计算量和参数数量,提高特征的表达能力,并增强网络的泛化能力。

1.5 改进后的 EfficientNetV2 网络

将 SK 注意力机制集成到 MBConv 类中是一种创新的方法,旨在增强垃圾图像分类模型的性能。这种集成方式充分利用了 SK 注意力机制的特点,即通过动态调整特征映射来适应图像中不同尺度的特征。改进后的 MBConv 命名为 SK-MBConv 模块,其结构图如图 6 所示。

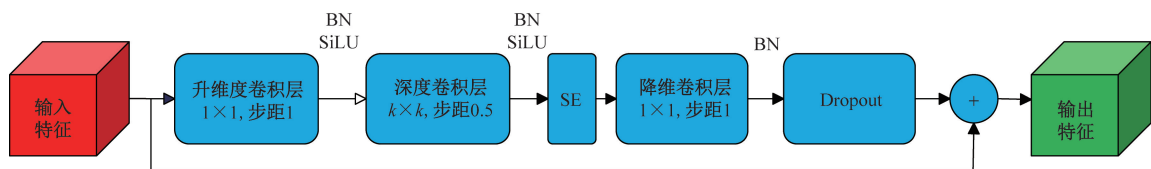


图 5 MBConv 结构原理

Fig. 5 Principle of MBConv structure

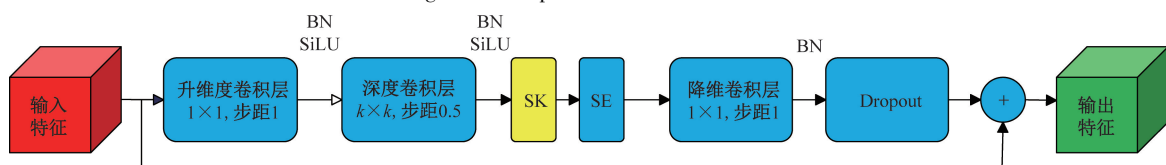


图 6 SK-MBConv 结构图

Fig. 6 SK-MBConv structure diagram

集成到 MBCConv 类中的好处。在 MBCConv 中集成 SK 注意力机制意味着在深度卷积和 SE(squeeze-and-excitation)模块之间添加一个能动态调整特征映射重要性的层。这样的集成使得 MBCConv 不仅可以通过 SE 模块处理通道间的关系,而且能够通过 SK 层处理不同尺寸特征间的关系。这种集成有助于网络更好地适应图像中的不同尺度特征,从而提高识别和分类的能力。相比于传统的 MBCConv 结构,这种改进可以提升模型在处理多尺度特征时的灵活性和准确性。改进后的 MBCConv 结构的影响。改进后的结构通过引入 SK 层,使得网络能够更加有效地处理和解析不同尺度的图像特征。这种动态选择机制使得网络能够更好地适应图像中的不同尺度特征,无论是大尺度的结构性特征还是小尺度的细节特征。这种方法特别适合于垃圾分类这样的应用中,因为垃圾图像往往包含多种尺寸和类型的对象。

新的算法通过 Stem 模块开始,执行 3×3 卷积操作以提取初始特征,并在之后应用批归一化和激活函数。接着,利用一系列 MBCConv 模块,每个模块包含以下步骤:首先是 1×1 卷积扩展输入通道数,然后是深度可分离卷积进行空间特征提取,可选地应用 SE 模块以增强通道关注度,最后通过 1×1 卷

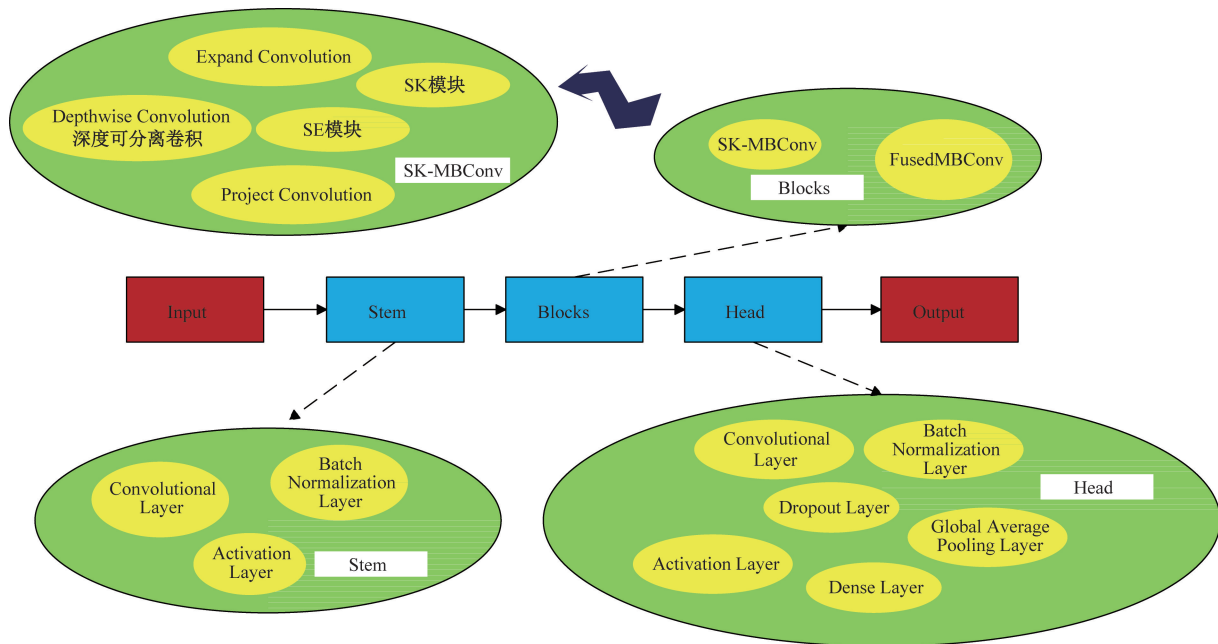
积将特征映射到较低维度空间。整个过程通过 Head 模块完成,其中包括全局平均池化和全连接层,用于生成最终的类别预测。图 7 为新算法的算法流程图。

综上所述,将 SK 注意力机制集成到 MBCConv 类中显著增强了 EfficientNetv2 网络在垃圾图像分类任务上的性能。通过此方法,模型不仅能够更加精确地处理不同尺寸的特征,而且还能提高整体的分类准确率。通过改进, EfficientNetv2 网络在处理复杂和多变的图像数据时将表现出更高的精确度和鲁棒性。这种结构上的改进为深度学习在复杂图像处理任务中提供了新的可能性。

1.6 迁移学习

迁移学习是一种机器学习技术,通过利用在一个任务上学到的知识来改善在另一个相关任务上的性能。在深度学习中,通常会使用预训练的模型作为初始模型,并通过调整模型的参数来适应新的任务。它的核心思想就是将预训练模型的特征学习能力和泛化能力转移到新的任务中。通过共享和迁移模型在源任务上学到的知识,可以加速模型收敛、降低对大量标注数据的需求,提高在新任务上的性能。

具体而言,使用迁移学习需要经过以下步骤:选择适合的预训练模型,冻结初始模型的参数;定



Input 为经过预处理后的图像数据;Stem 用于处理输入图像并生成初始的特征表示;Convolutional Layer 为卷积层;Batch Normalization Layer 用于加速训练速度、提高模型稳定性;Activation Layer 用于引入非线性特性;Blocks 是构建神经网络的基础模块,用于表示网络中各个层级的特定运算单元;Head 为网络的最后部分,用于对提取的高级特征进行处理,输出最终的分类或回归结果;Dropout Layer 为正则化技术,用于减少神经网络的过拟合现象;Dense Layer 连接上一层的所有神经元到当前层的每个神经元;Global Average Pooling Layer 为池化层;

Output 为模型的最终输出层,直接生成模型的预测结果

图 7 改进后新算法的算法流程图

Fig. 7 Improved algorithm flowchart

义新任务, 调整模型结构以适应新任务的需求; 使用新任务的数据进行微调 and 训练。

通过迁移学习, 可以利用已有的知识和经验, 快速构建和训练适应新任务的模型, 全面提高模型的性能和泛化能力, 尤其在数据有限或相似任务的情况下具有显著的优势。

2 实验与性能分析

2.1 实验参数设置

本次实验运行环境配置如下。操作系统: Windows11 家庭中文版; CPU: 12th Gen Intel (R) Core (TM) i7-12700H 2.30 GHz; GPU: NVIDIA GeForce RTX 3060 Laptop GPU; CUDA Version: 12.2; RAM: 16 384 MB; VRAM: 6 009 MB; Env: PyCharm2020.1 × 64。

2.2 数据集选择与处理

本文所使用的垃圾分类数据集为 Garbage Classification 公共数据集, 该数据集选择日常生活中常见的垃圾, 共 15 515 张图片, 分成 12 类。根据生活中的实际情况, 所获得的图像来源并非十分理想的问题, 例如每张图像的清晰度不同、图像发生偏转、关键部分发生缺失等问题, 对准备好的数据集进行数据预处理^[18], 使其更加符合实际状况, 并且将所有图像的尺寸都进行了大小为 150 × 150 的统一处理。将所有图像的尺寸进行统一处理, 有以下几点原因: ①统一输入尺寸: 深度学习模型要求输入尺寸固定, 简化设计和训练; ②计算效率: 固定尺寸有助于并行计算, 提高整体性能; ③避免尺寸偏差: 尺寸差异大可能影响模型泛化能力; ④数据一致性: 相同尺寸提高数据一致性, 便于模型学习关键特征。通过增强图像清晰度、随机旋转图像、上下翻转图像、高斯模糊等方法对数据集进行增强, 最终获得扩充后的数据集共 77 575 张图片, 在原有的 15 515 张图片的基础上增加了 62 060 张图片, 增强了数据集的普适性, 更加符合生活中的实际情况。将准备好的数据集划分为训练集与测试集, 其中训练集占 90%, 测试集占 10%。综上所述, 通过统一图像尺寸和图像增强技术可提升数据集的质量和一致性, 减少图像质量差异对模型的负面影响, 增加数据的多样性, 确保了数据集的可靠性。通过扩充数据集的规模, 可以增强数据集的覆盖范围, 提高模型的泛化能力。各类垃圾图片数量如表 2 所示。

2.3 不同模型性能对比

采用 AlexNet、GoogleNet、ResNet50、MobileNet-V2、VggNet19、Swin Transform、ConvNeXt 和改进之后的新算法(命名为 Ours) 8 个模型进行对比实验。每

一个模型都在 GPU 上进行了 batch_size = 8, epoch = 30 的训练。评价指标包括 3 项, 训练集识别准确率(以下简称识别准确率)、模型参数和运行总时间。其中, 识别准确率取 30 轮中的最优值, 实验结果如表 3 所示。不同的模型在 30 个 epoch 中的识别准确率如图 8 所示, 训练损失值变化曲线如图 9 所示。

表 2 实验所用数据集中的具体数据

Table 2 Detailed data in the experimental dataset

类别	训练集 (Train)	测试集 (Val)	数量
Battery	4 253	472	4 725
Biological	4 433	492	4 925
Brown glass	2 732	303	3 035
Cardboard	4 010	445	4 455
Clothes	23 963	2 662	26 625
Green glass	2 831	314	3 145
Metal	3 461	384	3 845
Paper	4 725	525	5 250
Plastic	3 893	432	4 325
Shoes	8 897	988	9 885
Trash	3 137	348	3 485
White glass	3 488	387	3 875

表 3 不同的模型实验结果对比

Table 3 Comparison of experimental results from different models

模型	识别准确率/%	模型参数数量	运行总时间/s
AlexNet	88.88	14 606 028	1 629
GoogleNet	94.94	10 340 180	9 197
ResNet50	94.44	25 671 628	4 893
MobileNet-V2	93.83	2 273 356	2 623
VggNet19	93.89	75 627 596	28 673
SwinTransform	74.20	27 557 394	273 294
ConvNeXt	89.72	27 829 356	268 279
本文模型	99.71	20 346 732	10 856

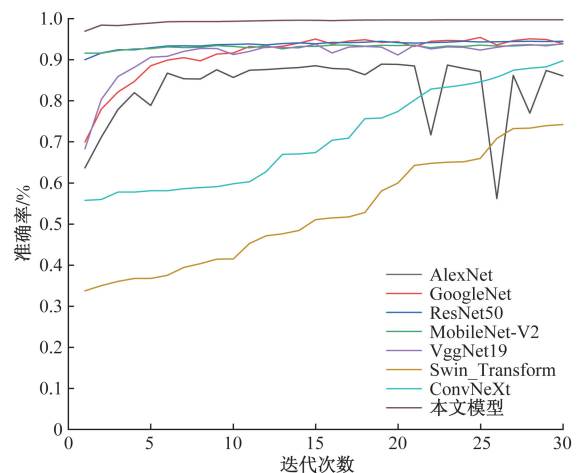


图 8 识别准确率

Fig. 8 Recognition Accuracy

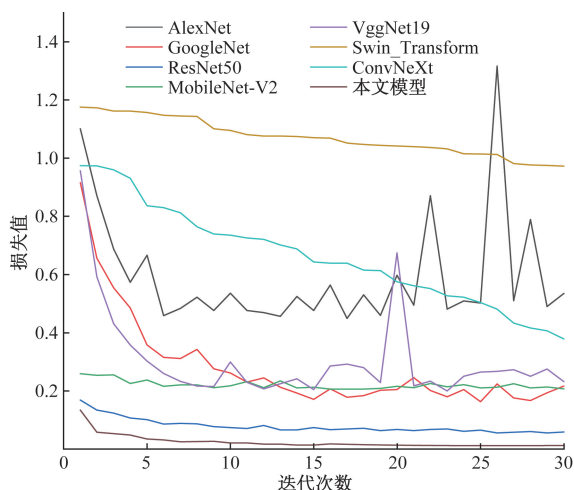


图9 训练损失值

Fig. 9 Training loss values

由表3可以看到,新模型的准确率要远远高于其他5个算法。并且在6个模型中,新模型对于内存的需求不是最大,MobileNet-V2模型所需参数虽然为2 273 356个,但是它的训练集识别准确率和测试集识别准确率均不如新模型。从运行时间来看,新模型所需时间不是最长。因此综合模型参数和运行总时间来看,新模型针对垃圾分类问题具有一定的实际应用价值。

2.4 新算法的稳定性

将EfficientNetV2网络和SK注意力机制结合起来时,新算法的稳定性相较于结合之前,得到了一些提升,可以从以下几个方面详细说明。

(1)特征提取的稳定性。EfficientNetV2网络通过缩放网络的深度、宽度和分辨率,在不同任务和数据集上实现了更好的性能平衡。这种策略能够有效地控制特征表达的复杂度,避免特征提取过程中的不稳定性和过拟合问题。SK注意力机制进一步增强了特征表示能力,通过动态选择卷积核尺寸来适应不同的输入特征,这种选择能力增强了模型对于不同场景和输入数据分布的适应能力。这种自适应性能有效地提高特征提取的稳定性,使得模型能够更准确地捕捉和表达关键特征。

(2)模型的泛化能力。EfficientNetV2通过网络结构的优化,确保模型在不同数据集和任务上都能够保持较好的泛化能力。SK注意力机制的引入进一步增强了模型对于输入数据分布的适应能力,使得模型面对不同的视觉任务时能够灵活调整,从而提高泛化能力。

(3)对抗性和鲁棒性。SK注意力机制的引入不仅提升了特征表达能力,还加强了模型的对抗性和鲁棒性。SK注意力机制能够动态地选择卷积核

尺寸,将更多的注意力集中在重要的特征上,抑制无关信息的影响。这种机制有助于模型在处理复杂场景时,更加关注关键的视觉特征,提高了模型对于复杂输入的鲁棒性。通过更精确地选择和表达特征,能够帮助模型避免过拟合问题,提高模型在训练数据外的泛化能力。这种稳定的特征提取能力是保证模型在实际应用中稳定性的关键因素之一。

综上所述,EfficientNetV2网络和SK注意力机制的结合不仅在理论上有助于提高模型的稳定性,同时也在实际的垃圾分类任务中展示了良好的表现,使得模型能够更加稳定、可靠地应对各种复杂的输入和挑战。

2.5 消融实验

2.5.1 消融实验1

针对基准模型中MBConv模块中,添加SK注意力机制是否为最佳策略设计了对比实验,即不同注意力机制对模型性能的影响,实验结果如表4所示。由表4可知,在都使用迁移学习策略的前提下,无论对比识别准确率或者运行总时间,SK注意力机制与MBConv模块的结合效果最好。

表4 消融实验1

Table 4 Causal experiment 1

方案	是否添加迁移学习策略	识别准确率/%	运行总时间/s
MBConv	是	98.81	14 294
MBConv + SGE	是	34.31	15 885
MBConv + ECA	是	61.11	16 050
MBConv + CBAM	是	50.51	17 744
MBConv + SK	是	99.71	10 856

2.5.2 消融实验2

为验证改进方法中不同的模块和方法对模型性能的贡献,以EfficientNetV2作为基准模型设计了对比实验,实验结果如表5所示。可以看出,单独将SK注意力机制引入MBConv模块之后,识别准确率相比于基准模型提升了1.83%。单独采用迁移学习策略时,识别准确率为98.81%。无论是否使用迁移学习策略,添加SK注意力机制之后,运行总时间都有所减少。而将SK注意力机制和迁移学习策略都添加到基准网络之后,由实验结果可以看出,相比于基准模型的识别准确率与运行总时间,改进之后的网络识别准确率提升了53.20%,运行总时间减少了26.87%。

测试结果表明,SK注意力机制和迁移学习策略使模型性能均有所提升,二者同时添加的效果最好。因此本文模型能更好地胜任垃圾识别任务。

表 5 消融实验 2

Table 5 Causal experiment 2

基准模型	SK 注意力机制	迁移学习	识别准确率/%	运行总时间/s
✓			46.51	14 844
✓	✓		48.34	11 337
✓		✓	98.81	14 294
✓	✓	✓	99.71	10 856

注: ✓表示包含此模块和方法。

2.6 算法可视化结果

为了更加直观地分析算法改进后的分类效果, 将本文算法预测结果进行可视化测试, 使用一组真

实场景中常见的垃圾图像, 将图像输入模型进行识别, 结果如图 10 所示。

3 结论

为提高垃圾分类的准确率和减少垃圾分类的时间, 提出了一种基于改进 EfficientNetV2 网络的垃圾图像分类算法。使用相同的数据集分别在其余几个传统算法模型上进行了实验, 实验结果表明, 新算法在垃圾分类任务中表现出色, 达到了 99.71% 的识别准确率, 远高于其他传统算法模型如 AlexNet (88.88%)、GoogleNet (94.94%)、ResNet50

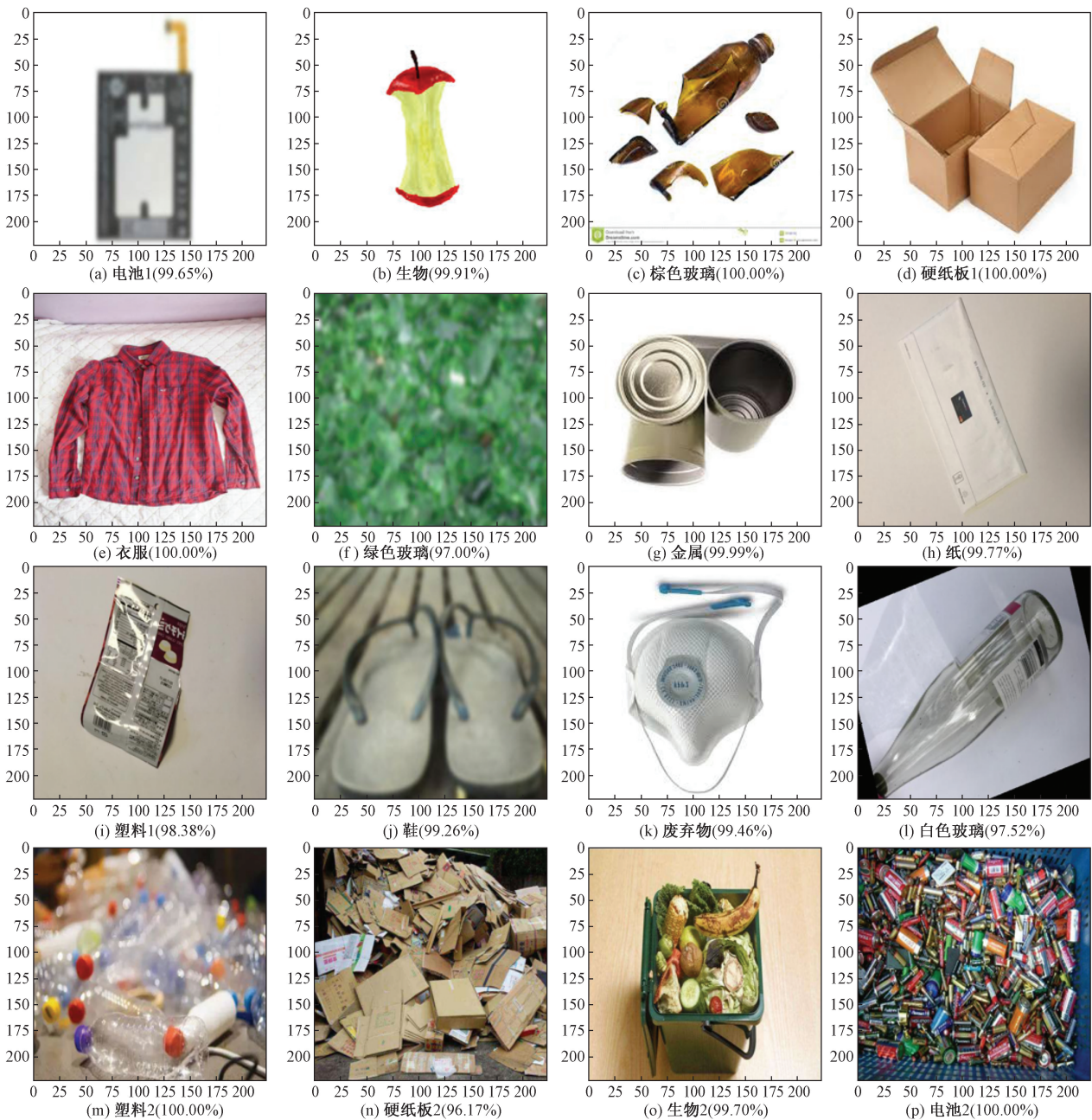


图 10 垃圾图像识别结果

Fig. 10 Results of garbage image recognition

(94.44%)、MobileNet-V2 (93.83%)、VggNet19 (93.89%)、Swin Transform (74.20%) 和 ConvNeXt (89.72%)。消融实验对比了是否添加 SK 注意力机制,是否使用迁移学习策略;在使用迁移学习策略的前提下在 MBCConv 模块中使用不同注意力机制对模型性能的影响。新的算法具有高效的特征提取能力和良好的泛化能力与鲁棒性。然而,新的算法也存在挑战和不足。首先,如同其他深度学习模型,其性能高度依赖于大规模和高质量的训练数据,这些训练数据对垃圾分类任务而言尤为重要,需要确保数据集覆盖各种垃圾类别和场景,以避免模型的偏差。其次,深度学习模型的复杂性限制了其在解释性方面的表现,这在需要对分类结果进行解释或依据结果采取行动的应用中尤为显著,提高模型的解释性是未来研究的重要方向之一。接下来的研究将集中在提升模型解释性的技术探索,如结合可视化方法或生成对抗网络生成解释,以增强模型结果的可解释性和理解性。还将探索在数据稀缺或无监督学习情况下如何优化模型性能,以及如何通过环境感知和自适应学习来提高模型在不同环境条件下的鲁棒性。

参 考 文 献

- [1] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
Xu Bingbing, Cen Keting, Huang Junjie, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5): 755-780.
- [2] 陈亚宇, 孙骥晟, 李建龙, 等. 基于深度学习与图像处理的废弃物分类与定位方法[J]. 科学技术与工程, 2021, 21(21): 8970-8975.
Chen Yayu, Sun Jisheng, Li Jianlong, et al. Research on waste classification and location method based on deep learning and image processing[J]. Science Technology and Engineering, 2021, 21(21): 8970-8975.
- [3] Xia Z X, Zhou H, Yu H, et al. YOLO-MTG: a lightweight YOLO model for multi-target garbage detection[J]. SIVIP, 2024, 18: 5121-5136.
- [4] Yang Z, Xia Z, Yang G, et al. A garbage classification method based on a small convolution neural network[J]. Sustainability, 2022, 14(22). DOI: 10.3390/su142214735.
- [5] Chen Y, Luo A, Cheng M, et al. Classification and recycling of recyclable garbage based on deep learning[J]. Journal of Cleaner Production, 2023, 414. DOI: 10.1016/j.jclepro.2023.137558.
- [6] Mao W L, Chen W C, Wang C T, et al. Recycling waste classification using optimized convolutional neural network[J]. Resources, Conservation and Recycling, 2021, 164. DOI: 10.1016/j.resconrec.2020.105132.
- [7] Tian X, Shi L, Luo Y, et al. Garbage classification algorithm based on improved MobileNetV3 [J]. IEEE Access, 2024, 12: 44799-44807.
- [8] 袁建野, 南新元, 蔡鑫, 等. 基于轻量级残差网络的垃圾图片分类方法[J]. 环境工程, 2021, 39(2): 110-115.
Yuan Jianye, Nan Xinyuan, Cai Xin, et al. Garbage image classification by lightweight residual network[J]. Environmental Engineering, 2021, 39(2): 110-115.
- [9] Kang Z, Yang J, Li G, et al. An automatic garbage classification system based on deep learning [J]. IEEE Access, 2020, 8: 140019-140029.
- [10] Gu Y, Ge B. Research on lightweight convolutional neural network in garbage classification [J]. IOP Conference Series: Earth and Environmental Science, 2021, 781(3). DOI: 10.1088/1755-1315/781/3/032011.
- [11] Wang J. Application research of image classification algorithm based on deep learning in household garbage sorting [J]. Heliyon, 2024, 10(9): e29966.
- [12] 汤伟, 刘思洋, 高涵, 等. 基于视觉的水面垃圾清理机器人目标检测算法[J]. 科学技术与工程, 2019, 19(3): 136-141.
Tang Wei, Liu Siyang, Gao Han, et al. A target detection algorithm for surface cleaning robot based on machine vision[J]. Science Technology and Engineering, 2019, 19(3): 136-141.
- [13] 高明, 陈玉涵, 张泽慧, 等. 基于新型空间注意力机制和迁移学习的垃圾图像分类算法[J]. 系统工程理论与实践, 2021, 41(2): 498-512.
Gao Ming, Chen Yuhan, Zhang Zehui, et al. Classification algorithm of garbage images based on novel spatial attention mechanism and transfer learning[J]. Systems Engineering Theory and Practice, 2021, 41(2): 498-512.
- [14] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks [C]//International Conference on Machine Learning, PMLR. Long Beach: PMLR, 2019: 6105-6114.
- [15] Tan M, Le Q. EfficientNetV2: smaller models and faster training [C]//International Conference on Machine Learning, PMLR, Virtual Event. Long Beach: PMLR, 2021: 10096-10106.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
- [17] Qu H, Yang J, Shen M, et al. Fault diagnosis of rolling bearing under time-varying speed conditions based on EfficientNetV2 [J]. Measurement Science and Technology, 2022, 33(6): 065023.
- [18] Zhang Y, Li D, Law K L, et al. Idr: self-supervised image denoising via iterative data refinement [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 2098-2107.