

文章编号: 2095-2163(2022)04-0047-07

中图分类号: TP391

文献标志码: A

# 基于 GAICNet 的垃圾识别分类检测网络

张涛

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

**摘要:** 随着社会经济的飞速发展和人们物质生活水平的不断提高, 所产生垃圾的种类和数量也在大幅增加, 为了有效提高垃圾资源的回收利用率 and 垃圾分类智能化水平, 本文提出了基于 GAICNet 的智能垃圾识别分类检测网络。使用华为云人工智能大赛公开的数据集, 对训练集使用 LabelImage 软件进行标注, 然后为了扩大训练数据, 强化网络模型对普通数据集的泛化能力, 设计了一种具有普适性的非线性曲线映射图像增强方法。最终, 构建了一个全局感知特征聚合模块与全新的核心自注意力机制的垃圾识别分类网络 (Garbage identification and classificationNet, GAICNet)。该轻量级网络是以 ResNet18 为骨干网络进行多层次的特征提取, 实验显示该算法在现有复杂实际场景垃圾分类中的检测准确率 (Accuracy) 可达到 97.3%, 具有较高的实用价值和市场前景。

**关键词:** 垃圾分类; 深度学习; 核心自注意力; GAICNet

## Garbage identification, classification and detection method based on GAICNet

ZHANG Tao

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** With the rapid development of social economy and the continuous improvement of human living standards, the types and quantity of waste are also greatly increasing. In order to effectively improve the recycling rate of waste resources and the intelligent level of waste classification, an intelligent waste identification, classification and detection network based on GAICNet is proposed in this paper. Using the data set published by Huawei cloud artificial intelligence competition, the training set is labeled with LabelImage software. Then, in order to expand the training data and strengthen the generalization ability of network model to ordinary data sets, a universal nonlinear curve mapping image enhancement method is designed. Therefore, a garbage identification and classification network (GAICNet) with global perceptual feature aggregation module and a new core self attention mechanism is constructed. The lightweight network takes ResNet18 as the backbone network for multi-level feature extraction. The experiment shows that the detection accuracy of the algorithm in the existing complex actual scene garbage classification can reach 97.3%, which has high practical value and market prospect.

**[Key words]** garbage classification; deep learning; nucleus self-attention; GAICNet

## 0 引言

随着科技革新的不断推进和城市化规模的逐渐扩大, 人们越来越重视周边环境卫生及可持续发展。但是, 有价值垃圾的接管利用和对人类身心健康有威胁的垃圾的利用已经成为一个亟待解决的问题, 所以垃圾的分类处理问题也成为人们越来越关注的热点话题之一<sup>[1]</sup>。许多国家已经出台了垃圾分类的严格政策要求, 这对全球环境保护和各国推行的可持续发展计划有着重要的促进作用。目前的垃圾回收大部分依靠人工挑选那些机器无法准确识别的破碎目标, 这种方式不仅效率低下、劳动强度大、并且可能会对工作人员健康造成伤害<sup>[2-3]</sup>。所以彻底解决这个问题就要从科技创新上来着手, 利

用深度学习方法进行自动化分拣, 实现垃圾的合理分类及回收。

当前, 许多能用于实际的检测模型处于测试阶段, 普遍存在着模型适应性差、精准率低、鲁棒性差等问题。2014年 Lukka 等人成功开发出了第一台可用于垃圾分类的机器人 Zen Robotics, 该机器人问世后, 将其应用到了实际工作中, 但由于其智能化水平低, 故障率高, 只能辅助垃圾分类<sup>[4]</sup>。吴健等人<sup>[5]</sup>针对国内的计算机实验室场景提出了一套实验室垃圾分类的解决方案。然而, 由于垃圾、背景复杂交错, 光照昏暗不明, 手工粗略提取相应数据时, 模型适应性差, 处理相应数据过程十分复杂, 所以无法满足实时检测要求。

当前在深度学习研究中获得迅猛发展的图像检

作者简介: 张涛 (1995-), 男, 硕士研究生, 主要研究方向: 深度学习、智能网联汽车。

通讯作者: 张涛 Email: 2099743507@qq.com

收稿日期: 2021-09-02

测是计算视觉的未来发展趋势,现已运用到各个领域,如自然语言处理、无人驾驶等<sup>[6]</sup>,并取得优异成绩。其中,Song 等人<sup>[7]</sup>提出基于 Inceptionv4<sup>[8]</sup>的垃圾自动分类 DSCR 网络,使得模型在多尺度特征上获得更多的信息,准确率达 94.38%。Dong 等人<sup>[9]</sup>提出一种注意力机制模型,通过完成局部、全局的特征提取和特征融合机制等方法建立了垃圾图像分类模型,该模型有效利用丰富的特征信息进而避免梯度消失的现象。Yang 等人<sup>[10]</sup>设计了一个新的增量学习框架 GarbageNet 以解决垃圾分类中缺乏足够数据、高成本的类别增量等问题,通过 AFM (Attentive Feature Mixup) 消除噪声标签的影响,但这些模型通常只针对通用物体检测,深度学习在垃圾分类领域依旧面临运算复杂、时效性差、鲁棒性欠佳的问题。

综上所述,本文研究了以 ResNet18<sup>[11]</sup>为骨干的浅层特征提取网络,设计了一个全局感知特征聚合模块的垃圾识别分类网络(Garbage identification and classificationNet,GAICNet)。并在整体网络建构中

引入一种具有普适性的非线性曲线映射图像增强方法,有助于算法更有效地对目标进行检测,增强网络模型对于不同拍摄环境、不同视觉效果垃圾的测试鲁棒性。

## 1 算法基础

### 1.1 非线性曲线映射图像增强

对数据做变换,增加基础数据集容量的方法叫做数据增强,避免因为数据集太少导致新对象的细粒度特征就很容易被忽略,也可以让网络能学到某一类别真正的特征,而不是此类别中个别图片的非本质个性特征。

本次研究提出了一种具有普适性的非线性曲线映射图像增强方法,该方法给定输入弱光图像的情况下估计一组最佳拟合的光增强曲线。然后,该框架通过迭代非线性曲线来映射输入的 RGB 通道中所有像素到其增强版本,以获得最终的增强图像。研究得到的模型结构如图 1 所示。

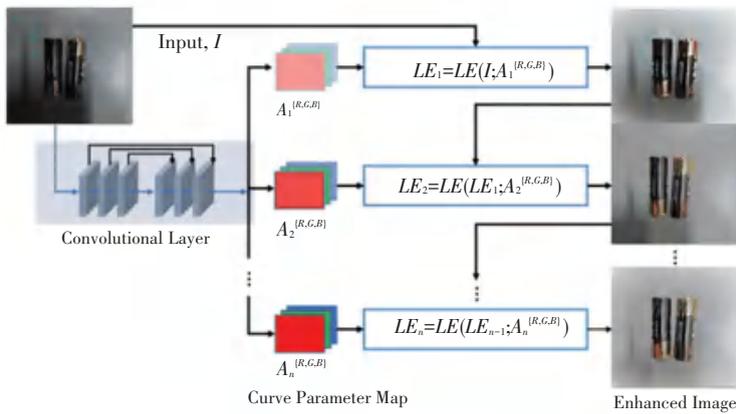


图 1 非线性曲线映射图像增强模型

Fig. 1 Non-linear curve mapping image enhancement model

首先,输入尺寸为  $256 \times 256 \times 3$  的弱光图像  $I$ 。然后使用一个由 7 个卷积层组成的卷积神经网络,这些卷积层是对称连接的。每个卷积层由 32 个卷积核组成,各卷积核的大小为  $3 \times 3$ ,  $step = 1$ ,后跟 Sigmoid 激活函数。此方法丢弃了破坏相邻像素关系的下采样和 BN 层。最后一个卷积层后是 Tanh 激活函数。研究过程中,为了能够在更宽的动态范围内调整图像,本文设计了一种可自动将弱光图像映射到其增强版本的高阶曲线。由此推得的数学定义公式如下:

$$LE_n(x) = \alpha_n LE_{n-1}(x) + A_n(x) LE_{n-1}(x) (1 - LE_{n-1}(x)) \quad (1)$$

其中,  $x$  表示像素坐标;  $LE_n(x)$  是给定输入  $I$  的最终增强迭代版本;  $\alpha \in (-1, 1)$  是可训练的曲线参数且调整 LE 曲线的幅度,也控制曝光水平;  $n$  是迭代次数,控制曲率;  $A$  是与给定图像大小相同的参数图。该框架将像素归一化为  $[0, 1]$ ,以避免溢出截断引起的信息丢失。再通过迭代应用曲线来映射输入的 RGB 通道的所有像素,并降低过饱和的风险,以获得最终的增强图像。受益于简单的曲线映

射形式和轻量级的网络结构,则能够改善高级视觉任务,计算效率高。图像增强结果见图1。

## 1.2 总体架构

基于传统的深度学习图像检测方法需要大量手工标注的边界框注释用于训练,这对于获得高分辨且易分类的注释数据来说是十分困难的。在这项工作中,为了充分利用带注释的新对象的特征并捕获查询对象丰富的细粒度特征,文中提出了带有上下文感知聚合的密集关系蒸馏来解决带有弱光和复杂背景图像的识别分类问题。首先,将原始图像输入基于 ResNet18<sup>[11]</sup> 构建的主干网共享特征提取器用

于提取低分辨率特征。这里,本文的简化 ResNet18 将输出通道切割为“16、32、64、128”以避免过拟合,并将下采样因子设置为8以减少通道结构细节的损失。同时将获得的特征信息通过专用的深度编码器编码成 *Key* 和 *Value*, 接着馈送到 DRD 模块测量 *Key* 和 *Value* 值图的权重之间像素的相似性。然后使用新设计的 CAFA 模块将 DRD 模块中所映射的有效特征聚合起来。最后执行一种核心自注意力关注机制,以解决尺度变化大和获取详细的局部上下文信息的问题,最终输出是聚集特征之后的3个特征的加权和。整体框架如图2所示。

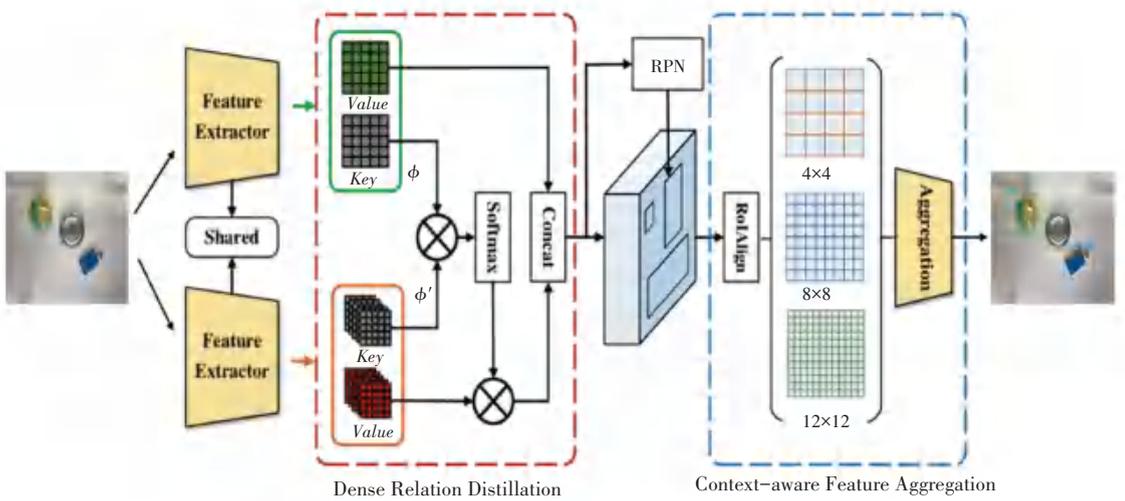


图2 整体模型架构

Fig. 2 Overall model architecture

## 1.3 密集关系蒸馏(DRD)模块

因为在某些垃圾分类数据集中个别同类别垃圾之间大小不一、复杂背景环境,颜色五花八门,这会影响模型的检测精度。针对这些问题,本文构建了密集关系提取模块,支持从特征中提取像素相关视觉语义信息,从像素级区分目标对象。由图2可知,深度编码器以一个或多个特征作为输入,查询特征来自前一个解码器层,并为每个输入特征输出2个特征映射: *Key* 和 *Value*, 这2个值图是并行的3×3卷积层,用于压缩输入特征的通道维数,以节省计算成本。*Key* 图用于支持特征之间的相似性,这不仅能够确定在哪里检索相关支持值,而且也有助于匹配编码视觉语义信息。*Value* 用于存储识别的详细信息。之后所生成的 *Key* 值和 *Value* 值被进一步馈送到关系提取部分。

DRD 在获得 *Key* 值和 *Value* 值的映射之后执行

关系提炼,以非局部方式执行像素相似性,公式为:

$$F(k_{q_i}, k_{s_j}) = \phi(k_{q_i})^T \phi'(k_{s_j}) \quad (2)$$

其中,  $i$  和  $j$  是支持位置的索引;  $\phi$  和  $\phi'$  表示2个不同的线性变换;  $k_{q_i}$  和  $k_{s_j}$  分别表示查询特征的 *Key* 值和 *Value* 值的输出映射。计算像素特征的相似度后,输出最终权重  $w_{ij}$ :

$$w_{ij} = \frac{\exp(F(k_{q_i}, k_{s_j}))}{\sum_j \exp(F(k_{q_i}, k_{s_j}))} \quad (3)$$

在此基础上,通过产生的权的加权求和来检索支持特征的值,此后将其与查询特征的值图连接。

## 1.4 上下文感知特征聚合(CAFA)模块

执行密集关系蒸馏后,DRD 模块完成了指定任务。细化的查询特征随后被馈送到区域建议输出的区域规划网。RoIAlign 模块以建议和特征为输入,进行特征提取,用于最终的类预测和包围盒回归。此外,由于尺寸、颜色差异被放大,模型趋向于失去

对新类的泛化能力,而新类对不同尺度具有足够的鲁棒性。为此,研究提出了上下文感知特征聚合模块。根据经验选择4、8和12三种分辨率,继而执行并行池化操作,以获得更全面的特征表示。较大的分辨率倾向于专注局部详细的上下文信息,而较小

的分辨率目标是捕捉整体的信息。因为每个生成的特征包含不同级别的语义信息,因此这种简单灵活的方式可用来解决了尺度变化问题。为了有效地聚合特征,研究中进一步提出了一种核心自注意力关注机制。如图3所示。

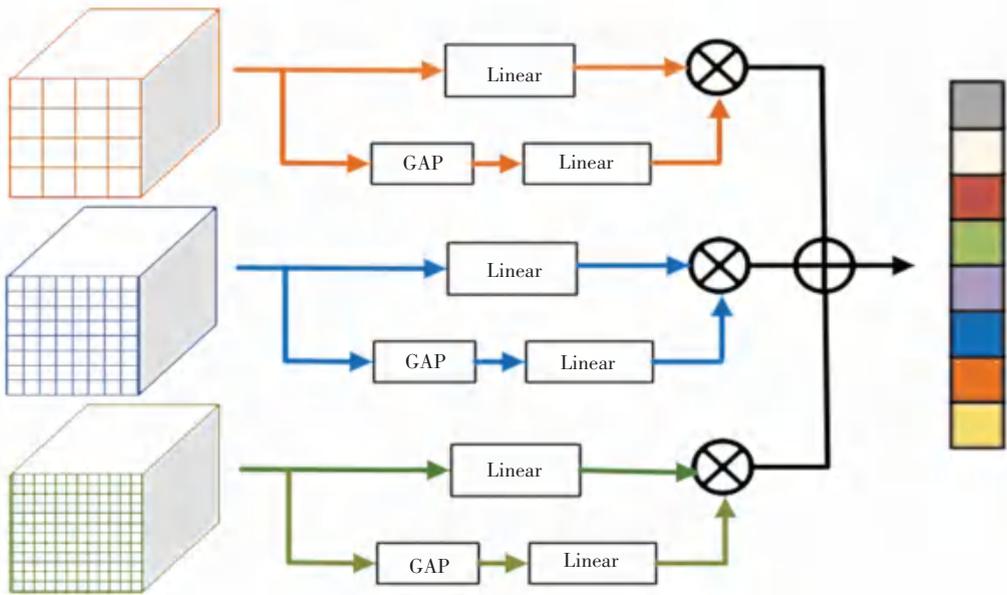


图3 核心自注意力关注机制

Fig. 3 Focused self-attention mechanism

研究中,为每2个块组成的特征添加一个注意分支。第一个块包含一个全局平均池,第二个包含2个连续的FC层。研究中,向生成的权重添加最大值归一化,以平衡每个特征的贡献。此后可得到,聚合特征的最终输出是3个特征的加权总和。最终输出公式为:

$$y = (v_q, w * v_s) \quad (4)$$

其中,“\*”表示矩阵内积, $v_q$ 和 $v_s$ 分别表示支持特征的Key值和Value值的输出映射。

至此,模型通过ResNet18融合密集关系蒸馏(DRD)模块和上下文感知特征聚合(CAFA)模块获取了丰富的、不同级别的特征信息,同时再利用获取到的不同层次特征生成密集的和范围框的预测,最后对范围框采用NMS输出最终结果。

## 2 实验结果与分析

### 2.1 实验平台

目标检测模型训练对计算机的配置和算力都有着较高的要求,本实验模型训练采用ubuntu18.04,由于ubuntu占用了大量计算机内存,故选择

minconda搭建TensorFlow-GPU环境,并下载安装相应版本的cuda,cudnn来支持GPU计算,最后是使用一块型号为Tesla P100-PCIE-16 GB的GPU进行模型训练。

### 2.2 样本数据集

本文采用的基础实验数据是华为云人工智能大赛提供的生活中常见的40种垃圾图片数据集,图片数量合计14 802张。该数据集均为真实、高质量的数据资源。将生活中的垃圾分为4大类,分别为:厨余垃圾、有害垃圾、可回收垃圾,除以上垃圾以外的归为“其他垃圾”。具体分类及每种垃圾数量见图4。其中,训练样本数量占比70%,有10 361张;测试样本数量占比20%,有2 960张;检测样本数量占比10%,有1 480张。

另外,为了减少数据集中因为拍摄设备、拍摄环境造成的噪声与异常值带来的干扰,系统使用滤波器去噪对数据集图像进行处理。线性均值滤波去噪是在图像上对目标像素的中心以 $m \times n$ 的像素为一个模板,通过均值计算以全体像素的平均值赋值给原有的模板中心像素,其数学表达式为:

$$f(x,y) = \frac{1}{hw} \sum_{(x,y) \in S} g(S,t) \quad (5)$$

其中,  $f(x,y)$  为滤波后的图像;  $g(S,t)$  为原始图像;  $S$  为以  $(x,y)$  为中心, 以  $h \times x \times w$  为尺寸的模

板。图像通过均值滤波器能够过滤掉由于受到角度、光源以及分辨率产生的影响, 避免产生噪声, 使图像平滑, 突出图片局部以及全局丰富的细粒度特征。

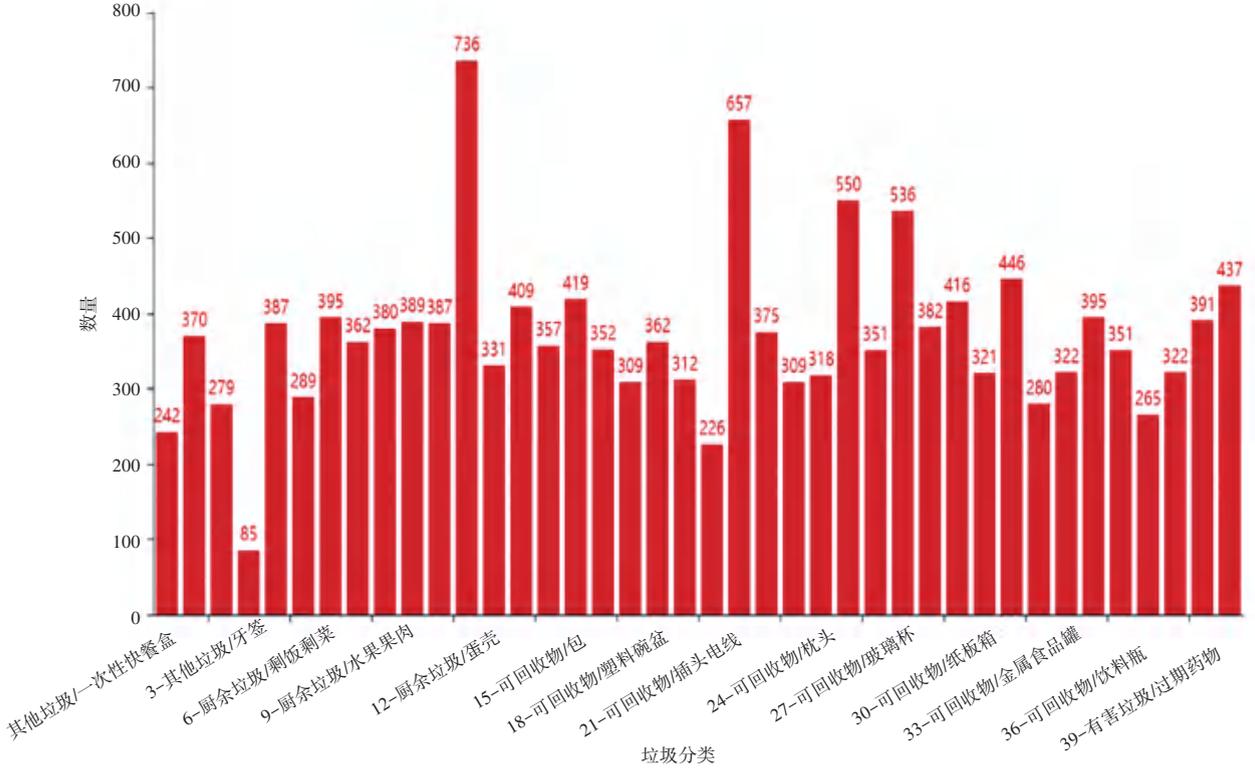


图 4 不同类别数据分布

Fig. 4 Data distribution of different categories

### 2.3 实验设置

本实验参数方面设置为: 使用 SGD 优化算法进行模型练习,  $batch = 12$ , 动量系数为 0.99, 共设置 50 个迭代周期, 基础学习率为 0.01, 设置  $step = 10^5$  并且为了加快模型收敛速度, 将 COCO 数据及上部分已经预训练好的参数迁移至模型训练。由于加入迁移学习和聚类分组归一化模型训练速度加快, 将模型在  $step = 9\ 000$  次时停止, 训练结束时保存正确率值最高和效果最好的模型参数。

### 2.4 结果分析

对 GAICNet 模型训练过程中的准确率 ( $accuracy$ ) 和损失 ( $loss$ ) 进行分析与记录, 并与当前主流的几种方法进行结果对比。图 5 展示了此算法所构建的 GAICNet 网络模型以及目前多种主流的图像识别分类模型在华为云数据集上进行  $10^5$  次迭代练习之后的  $accuracy$  结果对比曲线, 包括 DenseNet、YOLOv3、LSSD 等。图 6 显示了此算法所

构建的 GAICNet 网络模型以及目前多种主流的图像识别分类模型在阿里云天池进行的垃圾分类大赛中的数据集进行  $10^5$  次迭代训练之后的损失值结果对比曲线。

从图 5、图 6 中可以得出结论, GAICNet 的准确率比 DenseNet、YOLOv3、LSSD 高, 并且收敛速度快。  $loss$  值则要大大低于其他 3 种方法。当迭代 18 000 次左右时, GAICNet 的精度已经达到了 97.3%, 说明本文构建的密集关系蒸馏 (DRD) 模块和上下文感知特征聚合 (CAFA) 模块均能够有效提升算法的  $accuracy$ 。通过  $accuracy$  和  $loss$  图的详细比较, 可以得出结论是 GAICNet 鲁棒性好且分类准确度很高。

在与优秀的图像检测算法进行比较后, 图 7 给出了 GAICNet 在测试集上的部分识别结果的详细展示。仔细分析后发现, GAICNet 在各个差异较大类别的图像样例检测中, 都取得了令人满意的结果。

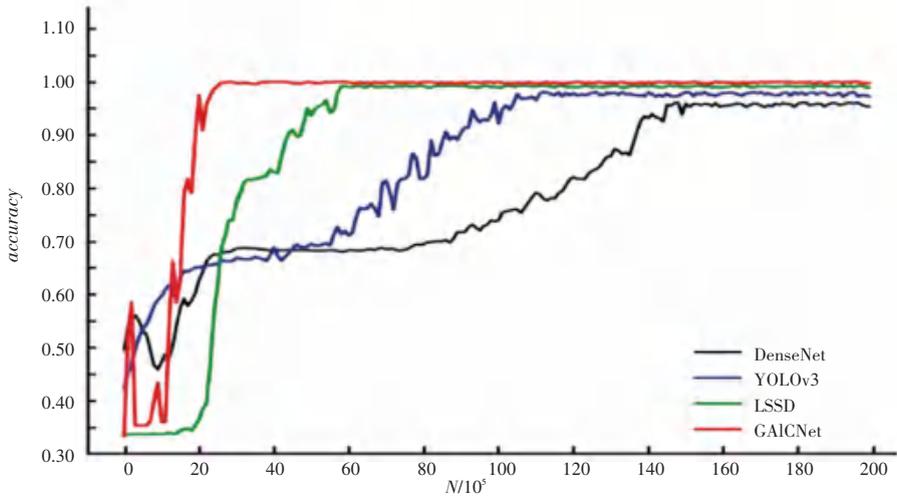


图5 GAICNet的精确度对比图

Fig. 5 Accuracy comparison chart of GAICNet

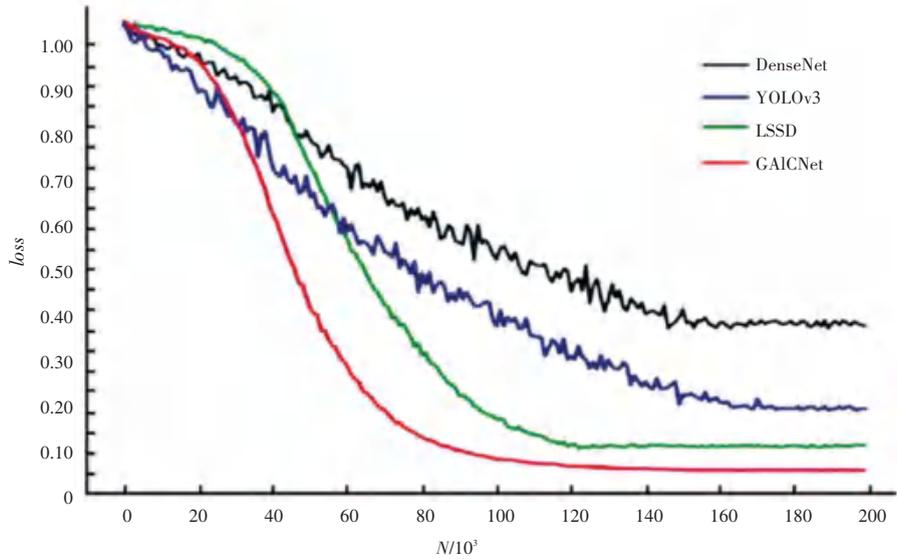


图6 GAICNet的loss对比图

Fig. 6 GAICNet's loss comparison chart

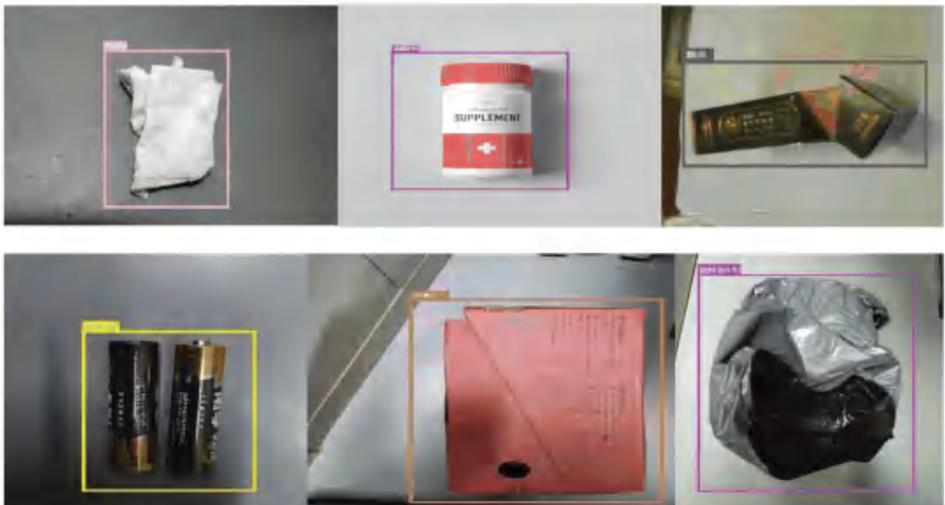


图7 GAICNet数据测试结果

Fig. 7 GAICNet data test results

### 3 结束语

本文主要对人工智能算法在垃圾智能识别分类应用上展开研究,构建了一种基于 GAICNet 的智能垃圾分类模型。该算法创新性地设计了利用非线性曲线映射的变光图像增强方法,提高了模型的泛化能力,有效地扩充了样本数据集,提升了算法的鲁棒性。同时,添加的密集关系蒸馏(DRD)模块采用查询和支持特征的稠密匹配策略,充分挖掘支持信息,可以显著提高性能。上下文感知特征聚合(CAFA)模块,使模型能够自适应地聚合来自不同尺度的图像特征,以获得更全面的特征表示。本文所构建的基于深度学习的垃圾识别分类算法能够减少同类垃圾所带来的差异性的影响,使模型在复杂场景上取得了97.3%的准确率,符合现有场景垃圾分类的实际使用需求。

### 参考文献

- [1] 鞠海翔,樊东燕. 公共场所智能垃圾桶系统的设计[J]. 山西电子技术,2017(06):28-31.
- [2] 赵双双,左琳娜. “互联网+”背景下职业教育教学资源库的建设与应用研究[J]. 南方农机,2019,50(24):81.

(上接第46页)

- [7] LIANG Zhenning, ZHANG Jinxiu, QIU Longhui, et al. Studies on deformation measurement with non-fixed camera using digital image correlation method[J]. Measurement, 2021, 167: 108139.
- [8] SONG Junfang, SONG Huansheng, WANG Shuyu. PTZ camera calibration based on improved DLT transformation model and vanishing point constraints[J]. Optik, 2021, 225: 165875.
- [9] 刘军, 后士浩, 张凯, 等. 基于单目视觉车辆姿态角估计和逆透视变换的车距测量[J]. 农业工程学报, 2018, 34(13): 70-76.
- [10] 黄同愿, 杨雪姣, 向国徽, 等. 基于单目视觉的小目标行人检测与测距研究[J]. 计算机科学, 2020, 47(11): 205-211.
- [11] 吴骏, 李文杰, 耿磊, 等. 基于单目视觉的前方车辆检测与测距[J]. 计算机工程, 2017, 43(02): 26-32.

- [3] 马娟. 互联网+背景下网络教学平台资源库建设的思考[J]. 数字化用户, 2019, 25(21): 279.
- [4] ANONYMOUS. A robot that recycles[J]. Machine Design, 2011, 83(11): 25.
- [5] 吴健, 陈豪, 方武. 基于计算机视觉的废物垃圾分析与识别研究[J]. 信息技术与信息化, 2016(10): 81-83.
- [6] 陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述[J]. 计算机科学, 2019, 46(03): 63-73.
- [7] SONG Fei, ZHANG Ying, ZHANG Jing. Optimization of CNN-based garbage classification model[C]// Proceedings of the 4<sup>th</sup> International Conference on Computer Science and Application Engineering. Sanya, China: ACM, 2020: 1-5.
- [8] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI, 2017: 4278-4284.
- [9] DONG Ziyuan, HAN Weiguang. Garbage image classification algorithm based on convolutional neural network [J]. Computer Systems & Applications, 2020, 29(8): 199-204.
- [10] YANG Jianfei, ZENG Zhaoyang, WANG Kai, et al. GarbageNet: A unified learning framework for robust garbage classification [J]. IEEE Transactions on Artificial Intelligence, 2021, 2(4): 372-380.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

- [12] 高敏, 曹欢. 单目图像中姿态角估计的坦克目标测距方法[J]. 光子学报, 2015, 44(05): 140-147.
- [13] STEIN G P, MANO O, SHASHUA A. vision-based ACC with a single camera: Bounds on range and range rate accuracy [C]// 2003 IEEE Intelligent Vehicles Symposium (IV'03). Ohio: IEEE, 2003: 120-125.
- [14] LIU Changhui, Shuai Kao, Yang Weirong. Design of ranging system for embedded vision[J]. Journal of Wuhan University of Technology, 2015, 37(04): 65-68.
- [15] 张云飞, 王桂丽, 周旭廷, 等. 基于单目视觉的测距算法[J]. 计算机与数字工程, 2020, 48(02): 356-360.