

DOI:10.19651/j.cnki.emt.2518718

基于改进 YOLO11 的生活垃圾检测模型^{*}

任梦晗 赵海燕 宋佳智

(内蒙古民族大学计算机科学与技术学院 通辽 028000)

摘要: 随着城市化进程的加快,生活垃圾量的持续攀升对生态环境形成严峻挑战,因此基于目标检测的智能分拣技术成为关键解决方案。针对现有检测模型在复杂场景下精度不足和部署效率低的问题,提出一种改进的 YOLO11 生活垃圾检测模型。通过引入可变形卷积和自主设计的三分支坐标注意力机制,构建了增强型可变形卷积模块,并用其重构骨干网络中的 C3k2,显著提升了模型对复杂背景中目标的特征提取能力。此外,采用内容感知特征重组算子替代颈部网络中的上采样,增强特征重建效果。引入指数移动平均滑动损失函数,有效提升检测精度并加速模型收敛。在优化后的华为云生活垃圾数据集上进行的实验表明,改进模型在 mAP@0.5 和 mAP@0.5:0.95 指标上分别达到 76.5% 和 64.6%,较基线模型提升 1.8% 和 1.7%。相比其他主流检测算法,改进模型参数量仅为 2.8 M,更适合移动端部署。

关键词: YOLO11; 可变形卷积; 注意力机制; 生活垃圾; 目标检测

中图分类号: TN911.73 文献标识码: A 国家标准学科分类代码: 120.5040

Improved YOLO11-based model for domestic waste detection

Ren Menghan Zhao Haiyan Song Jiazh

(School of Computer Science and Technology, Inner Mongolia Minzu University, Tongliao 028000, China)

Abstract: With the acceleration of urbanization, the continuous increase in domestic waste has posed a severe challenge to the ecological environment. Therefore, intelligent sorting technology based on target detection has become a key solution. Aiming at the problems of insufficient accuracy and low deployment efficiency of existing detection models in complex scenarios, an improved YOLO11 domestic waste detection model is proposed. By introducing deformable convolution and a self-designed three-branch coordinate attention mechanism, an enhanced deformable convolution module is constructed, which is used to reconstruct C3k2 in the backbone network, significantly improving the model's feature extraction capability for targets in complex background. In addition, a content-aware feature recombination operator is adopted to replace the upsampling in the neck network, enhancing the feature reconstruction effect. An exponential moving average sliding loss function is introduced to improve detection accuracy effectively and accelerate model convergence. Experiments on the optimized Huawei Cloud domestic waste dataset show that the improved model achieves 76.5% and 64.6% in mAP@0.5 and mAP@0.5:0.95 metrics, respectively, with an increase of 1.8% and 1.7% compared to the baseline model. Compared with other mainstream detection algorithms, the improved model has a parameter count of only 2.8 M, making it more suitable for mobile deployment.

Keywords: YOLO11; deformable convolution; attention mechanism; domestic waste; object detection

0 引言

随着城市化进程的加快和居民生活水平的提高,生活垃圾产量激增,如何高效、准确地进行垃圾分类成为城市管理面临的重大挑战。传统的人工分类方式难以满足日益增长的垃圾处理需求,近年来,基于计算机视觉和深度学习技

术的自动垃圾检测方法逐渐成为研究热点。

现有的目标检测算法主要分为单阶段和两阶段两类算法,二者在检测精度、计算效率及应用场景上存在显著差异。早期研究多采用两阶段检测器,Chen 等^[1]结合 ResNet-50 与 RPN,提出基于迁移学习的 Faster RCNN 模型,在自建数据集上实现了 84.1% 的 mAP@0.5,但其高计

算成本限制了实际应用;李博威等^[2]通过改进 SSD 的基础特征提取网络并引入 FPN 结构,在减少模型计算量的同时增强了小目标检测能力。

不同于两阶段目标检测算法的高复杂度与计算成本, YOLO 系列的出现为物体检测带来生机。目前生活垃圾检测领域已有一些基于 YOLO 算法的改进颇有成效。范金豪等^[3]替换 YOLOv4 主干为轻量型特征提取网络,并利用 ECA 重构卷积模块,实现了轻量化与检测精度的统一。涂成凤等^[4]在 YOLOv5 的基础上进行改进,引入 ShuffleNetv2 和 GhostNet 实现网络的轻量化,同时引入 SE 注意力提高特征提取能力,从而提升准确率。陈君等^[5]在三类垃圾数据集上测试改进的 YOLOv7,通过添加注意力机制以及更改主干网络模块提升检测能力,并对检测框做出调整,实现了较高的检测精度。陈思羽等^[6]在 YOLOv8 基础模型上加入 GAM 注意力机制,在自行整理的华为云数据集上实现了精度提升。为适应多变的垃圾检测环境,很多生活垃圾检测着力于多目标研究。Xia 等^[7]提出 YOLO-MTG,将 MobileViTv3 作为主干网络,结合 EfficientFormer 和动态卷积优化 MobileViT 块,在自建多目标垃圾数据集上实现了高精度的轻量化检测。在机器人应用方面,林哲等^[8]融合 YOLO 和目标分割算法实现了机器人在遮挡情况下抓取目标,这为垃圾检测的应用落地提供了支持。

现有生活垃圾检测研究存在两个主要局限:一是现有研究多集中于特定垃圾子类的识别,且部分数据集背景单一,缺乏多样化的真实场景;二是现有检测模型参数量较大,难以兼顾轻量化与检测精度。YOLO11 作为 YOLO 系列较新的轻量化模型,具有参数量少的优势。然而经文献调研发现,目前尚未见以 YOLO11 为基线模型应用于生活垃圾检测场景的相关研究。为此,本研究融合多个垃圾数据集创建了一个以华为云垃圾数据集为基础的生活垃圾数据集,并以 YOLO11 为基础模型进行改进,旨在实现生活垃圾种类的多样化与高精度检测。

1 YOLO11 算法

YOLO11 总体网络结构主要包括骨干网络(backbone)、颈部网络(neck)和检测头(head)。相较于 YOLOv8 模型, YOLO11 利用 C3k2 模块替换 C2f 模块,增强了模型的特征提取能力,提升了不同尺度特征的聚合能力;在 SPPF 模块后添加了一个基于多头注意力机制的 C2PSA 模块;将 YOLOv10 的思想引入到检测头中,使用深度可分离卷积来减少冗余计算,提高了物体检测的精确度以及效率。其网络结构如图 1 所示。

YOLO11 骨干网络通过 5 次 Conv 生成不同尺度的特征图,C3k2 根据 CSP 原理,将输入特征图分为两部分,一部分直接参与后续融合,另一部分经过系列卷积提取更深层的特征。颈部采用双向特征金字塔结构,将主干提取的

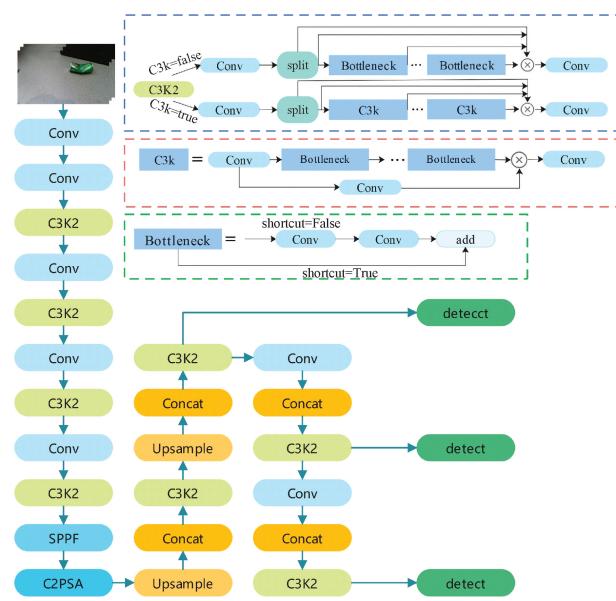


图 1 YOLO11 网络结构

Fig. 1 YOLO11 network structure

不同尺度、不同层次的特征进行整合和融合,使得后续检测头能够利用来自低层和高层的信息,增强模型对复杂目标的检测能力。检测头接收来自颈部的特征融合信息,采用解耦结构,将定位和分类任务分离成独立的分支,使每个分支专注于各自的任务,进而提升整体检测精度。

2 改进的 YOLO11 算法

生活垃圾种类繁多,且因堆叠和形变导致外观差异大,易引发误检和漏检,因此需要高精度检测模型。同时,考虑到实际部署需求,模型必须轻量化。所以,本研究选择在轻量级 YOLO11n 模型基础上进行改进,以兼顾检测精度和计算效率。

首先,提出了一个增强型可变形卷积模块(enhanced deformable convolutional networks v2, EDCNv2)模块对骨干网络中的 C3k2 模块进行重构,形成 C3k2_EDCNv2。其中可变形卷积模块(deformable convolutional networks v2, DCNv2)^[9]的普通局部卷积无法捕捉详细上下文信息,故本文设计了三分支坐标注意力机制(three-branch coordinate attention, TBCA)改进 DCNv2 结构,使其能够动态地提取特征。其次,针对颈部上采样的特征重建能力不足问题,引入内容感知特征重组算子(content-aware reassembly of features, CARAFE)^[10],以增强上采样对特征图的感知能力。最后,采用指数移动平均滑动损失函数(exponential moving average slide loss, EMASlideLoss)^[11],优化模型对正负样本的处理,提升模型的收敛速度。改进的模型结构如图 2 所示。

2.1 骨干网络中 C3k2 的重构

生活垃圾形态各异,在形状、状态方面具有不稳定性。

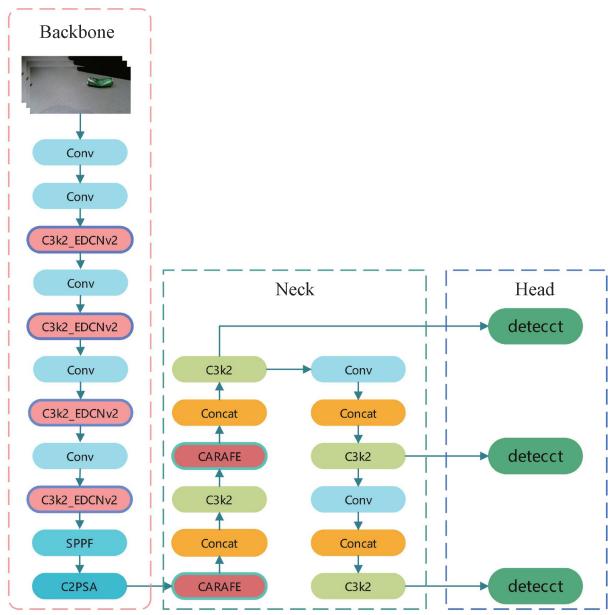


图 2 改进的 YOLO11 网络结构

Fig. 2 Network structure of the improved YOLO11

C3k2 使用固定卷积核,其感受野难以动态适应生活垃圾目标的多样形变,故而无法精准捕捉形变区域的细节特征。针对这一问题,本文设计了 EDCNv2 来弥补 C3k2 各尺度分支间缺乏动态注意力交互的缺点。重构后的 C3k2_EDCNv2 结构图如图 3 所示。

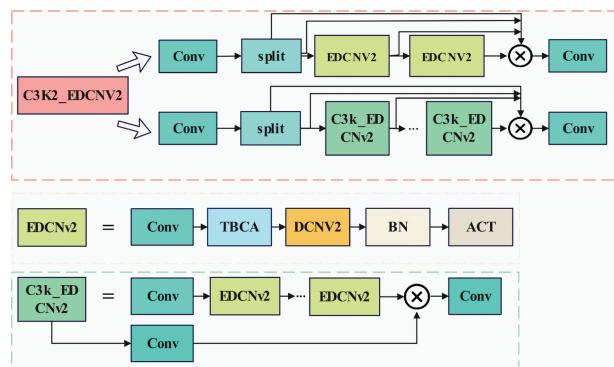


图 3 C3k2_EDCNv2 模块

Fig. 3 C3k2_EDCNv2 module

EDCNv2 由本文提出的三分支坐标注意力机制 TBCA 改进 DCNv2 得到,DCNv2 模块结构如图 4 所示。

DCNv2 旨在提升模型对几何形变、尺度变化及复杂空间结构的建模能力,其核心在于可变形卷积的调制机制与多尺度特征融合。可变形卷积即卷积位置可变,可以更有效地提取不规则图像的图像特征及边缘信息。可变形卷积原理可视化如图 5 所示。

在 DCNv2 中,传统卷积的固定采样网格被替换为可学习的偏移量(Δp_n)与调制因子(Δm_n)。偏移量通过附加的卷积层从输入特征中预测,使采样点能够自适应目标形

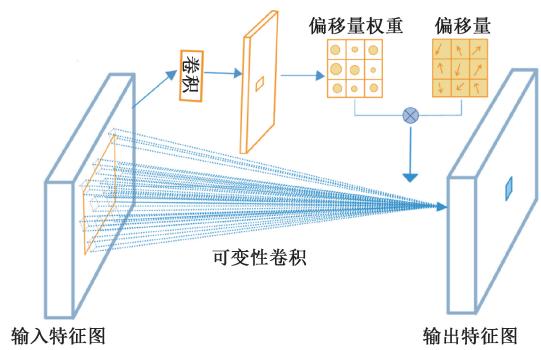


图 4 DCNv2 模块

Fig. 4 DCNv2 module

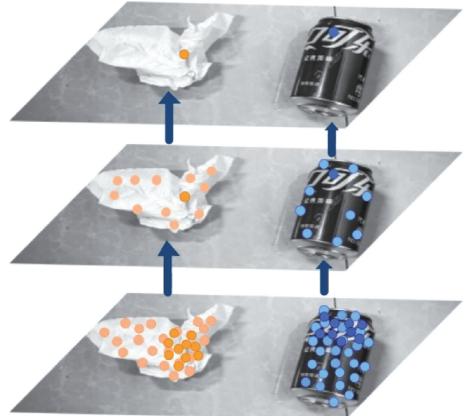


图 5 可变形卷积原理可视化

Fig. 5 Visualization of the deformable convolution principle

变,数学表达式如(1)所示。

$$\Delta p_n = W_{\text{offset}} * F_{in} \quad (1)$$

其中, W_{offset} 为偏移预测卷积核, F_{in} 为输入特征图。每个采样点 p_n 会动态偏移到新位置 $p_n + \Delta p_n$ 从而适应目标形状。同时,调制因子通过 Sigmoid 函数约束至[0,1]区间,输出特征 $F_{out}(p_0)$ 是加权采样值的总和,如式(2)所示。 p_0 是输出位置, w_n 是卷积权重, N 是采样点总数。

$$F_{out}(p_0) = \sum_{n=1}^N w_n \times F_{in}(p_0 + p_n + \Delta p_n) \times \Delta m_n \quad (2)$$

DCNv2 通过并行分支预测多组偏移量与调制因子,分别捕捉不同尺度下的形变模式。通过特征图通道维度的分组机制,多尺度形变特征被高效融合,从而提升对复杂场景的表征能力。

尽管 DCNv2 通过可变形卷积的调制机制显著提升了模型对几何形变的建模能力,但 DCNv2 的偏移量预测依赖于局部卷积操作,缺乏对全局空间分布的显式建模。调制因子的生成仅通过单一路径的卷积与 Sigmoid 函数实现,未充分融合通道注意力与空间注意力机制。这导致对长距离依赖关系的捕捉不足,尤其在复杂场景中易受局部噪声干扰。此外,并行分支间缺乏动态注意力引导的交互机制,导致特征互补性未充分挖掘。

本文提出的 TBCA 通过全局-局部特征联合校准设计,有效弥补 DCNv2 的上述缺陷。该模块整合了纵向(H)与横向(W)方向的双路径平均池化特征,并在通道(channel,C)层面引入全局通道分支以增强语义表征,最终构建了空间-通道联合响应的特征建模机制,其结构图如图 6 所示。

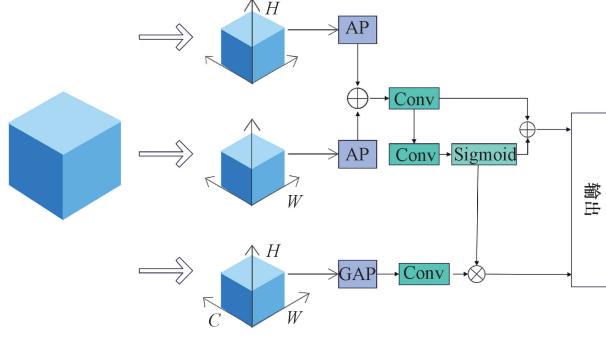


图 6 TBCA 模块

Fig. 6 TBCA module

图 6 中,纵向池化分支通过平均池化来压缩空间信息,保留纵向的空间分布特性;横向池化分支与纵向池化分支对称,通过池化保留横向结构,并通过转置操作使其与垂直方向特征排列一致来进行拼接操作;全局通道分支采用标准的全局平均池化,获得整张特征图的全局通道语义,随后通过卷积进行通道特征重构,使得该分支具备对通道维度的表达与调控能力。在纵向特征图与横向特征图沿纵向维度拼接后,通过卷积提取联合空间依赖以获得更具全局建模能力的空间表示,再将该特征图通过一层卷积,生成两个方向上的注意力权重。这两个注意力权重作用于对应的纵向池化分支特征,实现方向感知的特征增强。通道分支特征图通过与空间注意力权重相乘,进一步融合空间与通道的协同信息,最后将 3 组注意力分别作用于输入特征图以形成输出特征。

TBCA 通过分离 H 和 W 维度的局部池化,保留了空间结构细节,避免了传统单一路径的维度混淆问题。利用卷积与 Sigmoid 生成注意力权重,同时嵌入空间敏感性与通道重要性,提升了特征选择粒度。全局特征为局部校准提供了语义先验,局部特征为全局上下文注入空间细节,形成了闭环优化。

2.2 颈部网络中上采样的改进

传统 YOLO 的上采样方法仅利用亚像素邻域信息,无法捕捉丰富的语义信息。本文将 CARAFE 集成至 YOLO 的上采样层。CARAFE 是一种基于内容感知的特征上采样算子,其核心是通过动态生成适应性重组核与上下文信息聚合,实现高效且语义敏感的特征重建。

CARAFE 结构如图 7 所示,一方面将通道数为 C 的特征图通过 1×1 卷积进行压缩降维,得到通道数为 C_m 的特征图,并采用 3×3 卷积进行内容编码生成初始的重组核。为有效降低计算复杂度,利用核归一化器在每个空间位置

应用 Softmax 函数实现权重归一化。另一方面,CARAFE 通过窗口展开操作提取低分辨率特征图的局部邻域信息,然后利用归一化后的重组核对这些信息进行加权融合,最终输出细节更丰富的高分辨率特征图。

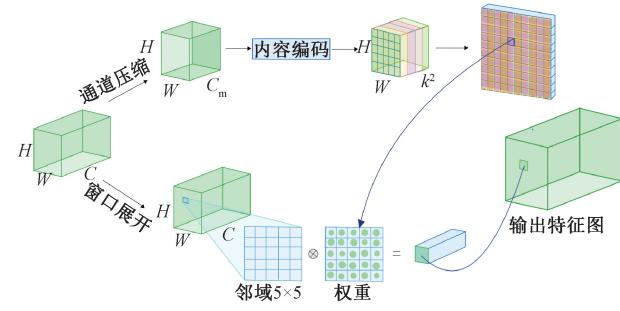


图 7 CARAFE 原理图

Fig. 7 CARAFE schematic diagram

2.3 损失函数部分的改进

CIoU(complete intersection over union)是 YOLO11 系列目标检测算法中用于边界框回归的损失函数,它的核心在于引入中心点距离和长宽比一致性作为惩罚项,来实现更全面的回归优化。如式(3)所示。

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3)$$

其中, IoU 为交并比,计算预测框 b 和真实框 b^{gt} 的重叠区域。 $\rho^2(b, b^{gt})$ 为预测框与真实框中心点的欧氏距离。 c 是最小包围框的对角线长度, v 为长宽比一致性项, α 为自适应权重。CIoU 通过联合优化重叠区域、中心点距离和长宽比提升了回归精度,但其静态权重分配机制无法适应训练过程中样本质量的变化。在实际垃圾检测任务中 CIoU 仍有一定局限性。

为解决这一问题,本文引入 EMASlideLoss 损失函数。EMASlideLoss 是一种基于动态权重调节的目标检测边界框回归损失函数,其核心由 4 部分组成:IoU 损失项 (L_{IoU}),它衡量预测框与真实框的重叠程度;距离损失项 (L_{dist})为约束预测框与真实框中心点的欧氏距离;宽高比损失项 (L_{aspect})为惩罚预测框与真实框的长宽比差异;EMA 平滑项 (L_{EMA})是历史损失指数移动平均值,用于稳定训练过程。EMASlideLoss 函数如式(4)所示。

$$L = \alpha(t)L_{IoU} + \beta(t)L_{dist} + \gamma(t)L_{aspect} + \lambda(t)L_{EMA} \quad (4)$$

其中, $\alpha(t)$ 、 $\beta(t)$ 、 $\gamma(t)$ 、 $\lambda(t)$ 为时变权重系数,通过 EMA-IoU 机制动态调整过程如下:

首先,基于历史 IoU 值的 EMA 计算当前样本的阈值 Q_t ,其中 η 为平滑因子, IoU_t 表示当前样本在当前时间步的交并比。阈值的计算如式(5)所示。

$$Q_t = \eta Q_{t-1} + (1 - \eta)IoU_t \quad (5)$$

其次,根据样本当前 IoU 与阈值 Q_t 的关系划分区域。设过渡区宽度为 Δ ,对低 IoU 区域降低权重,减少低质量

样本的梯度贡献;对于滑动过渡区域,通过平滑连接低/IoU 区域避免权重突变;对高 IoU 样本采取指数衰减权重的方式防止模型过度优化高 IoU 样本,增强模型对精确检测的学习。

上述划分区域规则如式(6)所示,其中 α_{base} 为低质量区域的基准权重, γ 为低质量衰减指数, α_{mid} 为过渡区域起始权重, λ 为高质量衰减系数。

$$\left\{ \begin{array}{l} \alpha(t) = \alpha_{base} \left(\frac{IoU_t}{Q_t - \Delta} \right)^{\gamma}, \quad IoU < Q_t - \Delta \\ \alpha(t) = \alpha_{mid} + (\alpha_{high} - \alpha_{mid}) \times \sin \left(\frac{\pi}{2} \times \frac{IoU_t - (Q_t - \Delta)}{2\Delta} \right), \quad Q_t + \Delta \leqslant IoU \leqslant Q_t + \Delta \\ \alpha(t) = \alpha_{high} \times e^{-\lambda(IoU_t - (Q_t + \Delta))}, \quad IoU > Q_t + \Delta \end{array} \right. \quad (6)$$

EMASlideLoss 的权重调节机制如图 8 所示。阈值左侧为低 IoU 区域(即低质量区)可能严重偏离物体的定位,对该区域采用滑动过渡策略平滑处理潜在的难样本;阈值右侧为样本 IoU 显著高于阈值的区域,对该区域采用权重衰减策略,以防止模型过度优化简单样本,从而提升泛化能力。

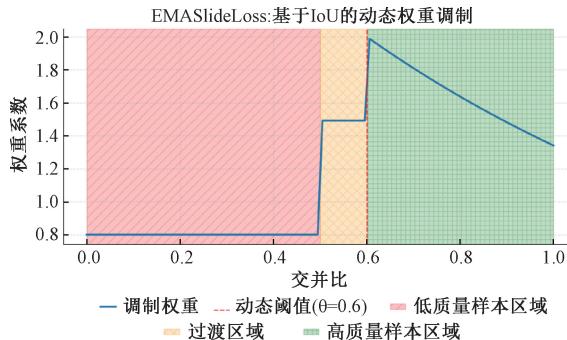


图 8 EMASlideLoss 的权重调制机制

Fig. 8 Weight modulation mechanism of EMASlideLoss

3 实验结果与分析

3.1 数据集

目前的公开生活垃圾检测数据集包括单一白色背景垃圾数据集 TranshNet^[12]、户外垃圾数据集 Taco^[13]、无人机小目标垃圾数据集 UAVVAste^[14]等。本研究考虑生活垃圾的复杂性及多样性,选择来自 2019 年华为公司垃圾分类大赛的公开数据集作为主要的图像数据。该数据集包含一次性快餐盒、书籍纸张、充电宝、剩饭剩菜等 44 类生活垃圾,具有单目标和多目标共 14 964 张垃圾图像。通过统计各类垃圾的数目,发现本研究采用的 44 类生活垃圾数据集存在显著的类别分布不平衡问题,例如陶瓷器皿样本量超过 2 000,而金属厨具仅约 100。这种长尾分布特性为模型训练与性能提升带来多重挑战,大量样本类别主导损失函数优化方向,导致模型过度拟合多数类特征,而对样本稀缺

类别的特征学习不足。

为缓解原始数据集的类别分布不平衡问题,提升模型泛化能力,本研究通过多源数据融合与数据增强技术对初始数据集进行扩充与优化。整合数据集 Taco、HGI-30^[15]与 GSA2D^[16]中的部分图像,并爬取网络垃圾图像,经人工清洗与标签标准化后,形成跨场景、多样本的初始数据集。针对样本稀缺类别,采用定向增强策略,通过旋转以及光度扰动,模拟垃圾在真实环境中的摆放状态与光线分布。经上述流程,原始数据集图像数量扩充至 17 885 张,标注框总数达 27 702 个,数据分布均衡性显著改善。

优化后的数据集包含单一背景与复杂背景、单个与多个目标、户外与室内环境,以及不同尺寸、形状的垃圾样本,多样化的数据集可提升模型的泛化性,更适应实际应用需求。数据集示例如图 9 所示。

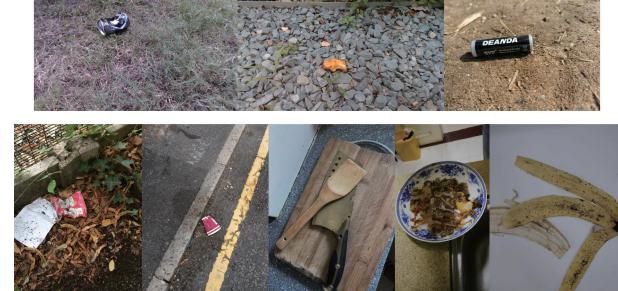


图 9 优化后的华为云数据集示例

Fig. 9 Examples from the optimized Huawei Cloud dataset

优化后数据集分布如图 10 所示,左上角为各类别实例数柱状图,可以看出各类别数据分布情况;右上角为边界框可视化图,展示数据集中目标检测框的分布情况;左下角为

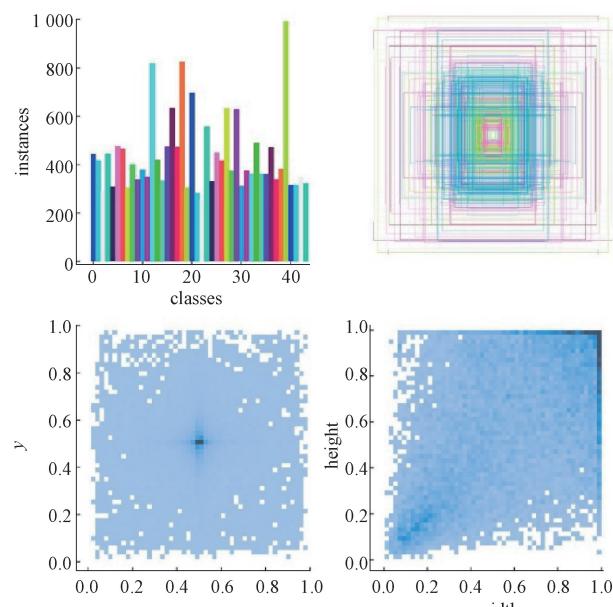


图 10 数据集分布图

Fig. 10 Distribution of the dataset

目标在图像平面上的位置分布热图,颜色深浅表示目标在对应位置出现的密度;右下角为目标边界框宽高分布热图,颜色深浅表示具有相应宽高的边界框出现的密度。本文将优化后的数据集以 7:1:2 的比例分为训练集、验证集和测试集。

3.2 实验配置

本实验平台采用 Ubuntu20.04 操作系统,搭载 Intel(R) Xeon(R) Platinum 8474C,主频为 800 MHz。系统内存为 32 GB,显卡为 Nvidia GeForce RTX 4090,显存容量为 24 GB,Pytorch 框架版本为 2.0.0+cu118,Python 版本为 3.8.10。Batch size 设置为 32,图像大小设置为 640×640,epoch 设置为 300。网络未使用预训练模型的权重且所有实验使用一致的超参数训练验证。

3.3 模型评价指标

在目标检测任务中,评估模型的性能需依赖多维度指标以全面反映其检测能力与效率。

精确率(precision, P)衡量模型预测为正样本的准确性。召回率(recall, R)则评估模型对真实目标的覆盖能力。

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

平均精度(average precision, AP)通过计算不同召回率阈值下的精确率均值,评估模型在特定交并比 IoU 阈值下的综合性能。其中,AP@0.5 以 IoU=0.5 为判定标准,广泛用于衡量宽松匹配场景下的检测能力;而 AP@0.5:0.95 则在 IoU 阈值 0.5~0.95 范围内取均值,反映模型对严格定位要求的适应性。平均精度均值(mean average precision, mAP)通过平均所有类别的 AP 值,提供全局性能评估,其中 N 为类别总数。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (9)$$

实际应用中,还需结合推理速度指标如 FPS、计算复杂度指标 GFLOPs 和参数量 Para,以评估模型在实时性与资源占用上的适用性。

3.4 对比试验

1) Backbone 改进效果对比

为评估 EDCNv2 所采用的 TBCA 对 DCNv2 的优化效果,本文对比了简单注意力模块 SimAM^[17]、混合局部通道注意力 MLCA^[18]、自适应细粒度通道注意力 AFGCA^[19]、上下文锚注意力 CAA^[20]在优化后的华为云垃圾数据集上的性能表现。由表 1 可知,与其他注意力机制相比,基于 TBCA 的 EDCNv2 性能更优。

表 1 不同注意力机制的对比

Table 1 Comparison of different attention mechanisms

注意力机制(+CARAFE)	mAP@0.5/%	mAP@0.5:0.95/%	P	R	Para	GFLOPs
DCNv2_SimAM	0.751	0.634	0.798	0.682	2 783 966	6.6
DCNv2_MLCA	0.750	0.654	0.831	0.654	2 784 002	6.6
DCNv2_AFGCA	0.752	0.639	0.788	0.687	2 788 526	6.6
DCNv2_CAA	0.755	0.635	0.814	0.671	2 796 926	6.8
EDCNv2	0.759	0.643	0.816	0.667	2 806 322	6.6

表 2 展示了不同的改进 C3k2 方法在优化后数据集上的性能对比结果。使用不同文献中提出的卷积模块重构 C3k2,多样化分支块 DBB^[21]通过引入不同结构和尺寸的卷积分支来增加卷积层的复杂性;细节增强卷积 DEConv^[22]通过多尺度特征融合与残差学习来强化高频细节恢复;全维度动态卷积 ODConv^[23]是通过 SE 注意

力在通道等维度实现动态性来提升模型精度。实验表明,相较于 C3k2 及其他改进方法,C3k2_EDCNv2 在关键指标上均取得显著提升:mAP@0.5 提升 0.5%,mAP@0.5:0.95 提高 0.6%。此外,C3k2_EDCNv2 的 Recall 提升至 0.691,验证了其在复杂场景下的性能优势。

表 2 C3k2 改进模块对比

Table 2 Comparison of C3k2 improvement variants

卷积块	mAP@0.5/%	mAP@0.5:0.95/%	P	R	Para	GFLOPs
Baseline	0.747	0.629	0.814	0.661	2 590 732	6.4
C3k2_DBB	0.740	0.616	0.804	0.657	2 590 732	6.4
C3k2_DEConv	0.747	0.623	0.790	0.678	2 591 116	6.4
C3k2_ODConv	0.728	0.597	0.798	0.636	2 641 098	5.4
C3k2_EDCNv2	0.752	0.635	0.784	0.691	2 666 218	6.4

2) 损失函数效果对比

基于前两个改进,给出损失函数在生活垃圾检测上的

对比实验,实验结果如表 3 所示。将定位损失替换为平均极坐标距离交并比 MPDIoU^[24]与形状感知交并比

ShapeIoU^[25]都没有得到很好的检测效果,在 SlideLoss^[26]上加入指数移动平均滑动动态调整样本权重,使难样本的

损失贡献更平滑地变化。在参数量与 GFLOPs 不增加的同时,mAP50 上升 0.6%,较其他损失函数表现更优。

表 3 不同损失函数对比

Table 3 Comparison of different loss functions

损失函数	mAP@0.5/%	mAP@0.5:0.95/%	P	R	Para	GFLOPs
CIOU	0.759	0.643	0.816	0.667	2 806 322	6.6
MPDIOU	0.753	0.639	0.811	0.665	2 788 526	6.6
ShapeIoU	0.759	0.643	0.791	0.699	2 806 322	6.6
SlideLoss	0.759	0.644	0.811	0.682	2 796 926	6.8
Ours	0.765	0.646	0.798	0.697	2 806 322	6.6

3)与其他算法对比

图 11 和表 4 对比了主流目标检测算法在优化后的华为生活垃圾数据集上的表现,包括 YOLO 系列^[27-29]、SSD^[30]及多项基于 YOLO 的目标检测改进方法。SDS-YOLO^[31]是基于 YOLO11 部署在移动端的目标检测研究,FE-YOLO^[32]是轻量化模型研究,荀雨薇等^[33]是在生活垃圾检测场景下着力于提升检测准确率。通过对比不同模型,显示了本文改进算法在生活垃圾检测场景下的适

用性。图 11 是各算法指标对比图,其中实线为本文算法。从图 11 中可以看出,在 200 轮迭代之前,各算法表现各有优势;但在 200 轮后,本文算法始终高于其他算法。由表 4 可知,基线模型参数量较小,在此模型上的改进很好地平衡了精度与轻量化,mAP@0.5 达到 76.5%(较基线模型提升 1.8%),mAP@0.5:0.95 提升至 64.6%(较基线模型提升 1.7%)。实验结果表明,改进算法在精度提升与计算资源占用之间实现了有效平衡。

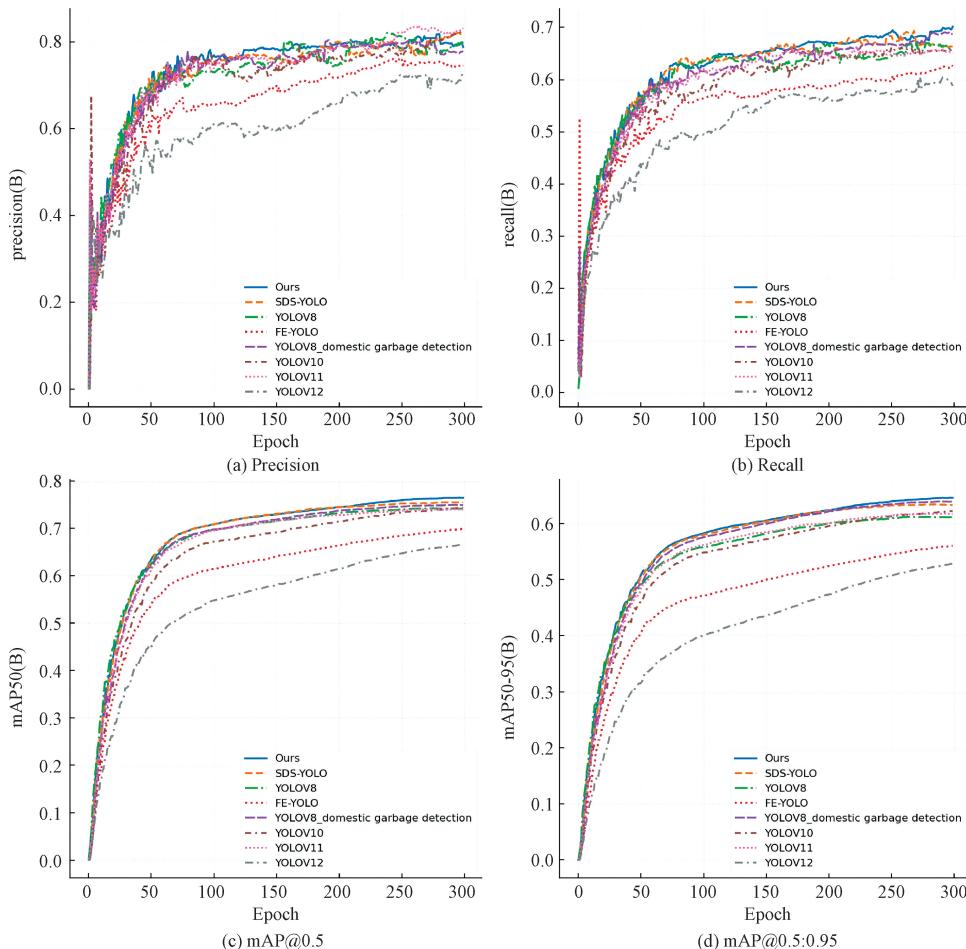


图 11 各算法指标对比图

Fig. 11 Performance comparison of different detection algorithms

表 4 本文算法与其他算法的性能对比

Table 4 Comparison of different loss functions

模型	mAP@0.5/%	mAP@0.5:0.95/%	P	R	Para	GFLOPs
SSD	0.705	—	0.723	0.64	29 359 712	65.037
YOLOv8	0.742	0.612	0.779	0.668	3 014 228	8.2
YOLOv10	0.743	0.622	0.821	0.657	2 711 576	6.6
YOLO11n	0.747	0.629	0.814	0.661	2 590 732	6.4
YOLOv12	0.666	0.529	0.733	0.587	2 565 308	6.4
SDS-YOLO	0.755	0.635	0.814	0.666	2 603 084	6.4
FE-YOLO	0.698	0.56	0.747	0.624	2 303 812	6.1
文献[33]	0.749	0.64	0.778	0.686	7 325 040	12.8
Ours	0.765	0.646	0.798	0.697	2 806 322	6.6

3.5 消融实验

为验证各改进模块的有效性,本文在相同实验条件下进行了系统的消融实验。实验设置如下:1)采用原始 YOLO11n 作为基线模型;2)在基线模型中加入 EDCNv2 模块;3)在实验 2)基础上引入 CARAFE 上采样算子;4)在实验 3)基础上进一步整合 EMASlideLoss 损失函数。

消融实验的结果如表 5 所示,基线模型在保持轻量化的同时,取得了 mAP50 为 0.747 和 mAP50-95 为 0.629 的

基准性能。引入 EDCNv2 模块后,检测精度有所提升,mAP50 和 mAP50-95 分别提高了 0.5% 和 0.6%。继续加入 CARAFE 上采样算子后,虽然模型参数量略有增加,但检测性能进一步提升,mAP50 和 mAP50-95 分别增长了 1.2% 和 1.4%。当采用 EMASlideLoss 损失函数时,模型性能获得进一步优化,mAP50 和 mAP50-95 分别提升 1.8% 和 1.7%。这一系列实验充分验证了各改进模块的有效性。

表 5 消融实验

Table 5 Ablation study

基线模型	EDCNv2	CARAFE	EMASlideLoss	Param/M	GFLOPs	mAP@0.5/%	mAP@0.5:0.95/%
✓				2.59	6.4	0.747	0.629
✓	✓			2.63	6.3	0.752	0.635
✓	✓	✓		2.80	6.6	0.759	0.643
✓	✓	✓	✓	2.80	6.6	0.765	0.646

3.6 检测可视化

为进一步验证本文改进后的模型对于生活垃圾的检测性能,采用测试集中的垃圾图像对基线模型和改进模型进行对比实验,实验检测结果如图 12 所示。图中第 1 行为基线模型在测试集上的检测结果,第 2 行为改进模型在测试集上的检测结果。

对比图 12 第 1 行的两张图片可知,在对重叠图像的检测中,改进模型对重叠物体分别做出了识别,但基线模型只检测出了部分物体;对比图中 2 行两张图片可知,改进模型提高了检测物体的精确度,减少了误检的发生;对比图中第 3 行可知,改进模型降低了对小目标的漏检;对比图中第 4 行可知,改进模型提高物体检测的置信度。综上,改进算法通过增强的上下文感知模块与动态特征融合策略,有效提升了重叠物体的区分能力以及检测的精度,这些改进验证了算法在几何形变适应性与场景泛化性方面的优势,为复杂环境下的鲁棒检测提供了可靠解决方案。



图 12 YOLO11 与改进后的模型检测效果对比

Fig. 12 Detection results comparison between YOLO11 and the improved model

4 结 论

本文针对复杂背景下的生活垃圾识别困难、检测精度

低、人工检测耗时耗力等问题,提出了基于 YOLO11 的生活垃圾检测改进算法。基于华为云 44 类生活垃圾,融合 TACO、HGI-30、GSAD 等来源的数据,构建了一个具有复杂背景的单目标与多目标结合的生活垃圾数据集,使数据类别得到了均衡。在 YOLO11 中引入 DCNv2 模块,并设计 TBCA 注意力机制改进 DCNv2 得到 EDCNv2;引入 CARAFE 上采样模块与 EMASlideLoss 损失函数,实现了模型在生活垃圾数据集上的良好检测性能。与基线模型相比,mAP-50、mAP50-95 分别提升了 1.8%、1.7%。综上,本文提出的基于 YOLO11 的生活垃圾改进模型在优化后的华为云生活垃圾数据集上有良好的检测性能,可以用于基于轻量化的边缘设备的智能垃圾检测分类系统中。

参考文献

- [1] CHEN Y Y, SUN J SH, BI SH J, et al. Multi-objective solid waste classification and identification model based on transfer learning method[J]. Journal of Material Cycles and Waste Management, 2021, 23(6): 2179-2191.
- [2] 李博威,侯明,李擎,等.改进的 SSD 生活垃圾检测算法[J].机械设计与制造,2023(9): 157-162.
LI B W, HOU M, LI Q, et al. Improved SSD algorithm for domestic waste detection[J]. Machinery Design & Manufacture, 2023(9): 157-162.
- [3] 范金豪,崔立志.基于 YOLO_ES 的垃圾分类目标检测模型[J].电子测量技术,2023, 46(1): 160-166.
FAN J H, CUI L ZH. Waste classification object detection model based on YOLO_ES[J]. Electronic Measurement Technology, 2023, 46(1): 160-166.
- [4] 涂成凤,易安林,姚涛,等.轻量化 YOLOv5n 的高精度垃圾检测算法[J].计算机工程与应用,2023, 59(10):187-195.
TU CH F, YI AN L, YAO T, et al. High-precision garbage detection algorithm based on lightweight YOLOv5n [J]. Computer Engineering and Applications, 2023, 59(10): 187-195.
- [5] 陈君,赵小会,王博士,等.基于 YOLOv7 的垃圾检测方法研究[J].计算机测量与控制,2024, 32(12):1-8.
CHEN J, ZHAO X H, WANG B SH, et al. Research on garbage detection method based on YOLOv7[J]. Computer Measurement & Control, 2024, 32(12): 1-8.
- [6] 陈思羽,舒航,陈宇阳,等.基于 GAM-YOLOv8 算法的生活垃圾检测[J].人工智能与机器人研究,2024, 13(2):194-202.
CHEN S Y, SHU H, CHEN Y Y, et al. Domestic waste detection based on GAM-YOLOv8 algorithm[J]. Artificial Intelligence and Robotics Research, 2024, 13 (2): 194-202.
- [7] XIA ZH Y, ZHOU H K, YU H M, et al. YOLO-MTG: A lightweight YOLO model for multi-target garbage detection [J]. Signal, Image and Video Processing, 2024, 18(6): 5121-5136.
- [8] 林哲,潘慧琳,陈丹.融合改进 YOLO 和语义分割的遮挡目标抓取方法[J].电子测量与仪器学报,2024, 38(12):190-201.
LIN ZH, PAN H L, CHEN D. Occluded object grasping method integrating improved YOLO and semantic segmentation [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (12): 190-201.
- [9] ZHU X ZH, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 9300-9308.
- [10] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-aware reassembly of features[C]. IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [11] JIANG T, ZHOU J, XIE B B, et al. Improved YOLOv8 model for lightweight pigeon egg detection[J]. Animals, 2024, 14(8): 1226.
- [12] YANG M, THUNG G. Classification of trash for recyclability status[J]. CS229 Project Report, 2016, 2016(1): 3.
- [13] PROENÇA P F, SIMÕES P. Taco: Trash annotations in context for litter detection[J]. ArXiv preprint arXiv:2003.06975, 2020.
- [14] KRAFT M, PIECHOCKI M, PTAK B, et al. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle[J]. Remote Sensing, 2021, 13(5): 965.
- [15] WU ZH Z, LI H Y, WANG X F, et al. New benchmark for household garbage image recognition[J]. Tsinghua Science and Technology, 2022, 27(5): 793-803.
- [16] XU N, LIAO J F, MENG Q W, et al. Garbage segmentation and attribute analysis by robotic dogs[J]. Arxiv preprint arXiv: 2404.18112, 2024.
- [17] QIN X Y, LI N, WENG C, et al. Simple attention module based speaker verification with iterative noisy label detection[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2022: 6722-6726.
- [18] WAN D H, LU R SH, SHEN S Y, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023,

- 123: 106442.
- [19] SUN H, WEN Y, FENG H J, et al. Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing[J]. Neural Networks, 2024, 176: 106314.
- [20] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [21] DING X H, ZHANG X Y, HAN J G, et al. Diverse branch block: Building a convolution as an inception-like unit[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10886-10895.
- [22] CHEN Z X, HE Z W, LU ZH M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.
- [23] LI CH, ZHOU A J, YAO A B. Omni-dimensional dynamic convolution [J]. ArXiv preprint arXiv: 2209.07947, 2022.
- [24] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. Arxiv preprint arXiv: 2307.07662, 2023.
- [25] ZHANG H, ZHANG S J. Shape-IoU: More accurate metric considering bounding box shape and scale[J]. Arxiv preprint arXiv: 2312.17663, 2023.
- [26] YU Z P, HUANG H B, CHENA W J, et al. YOLO-facev2: A scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.
- [27] SHARMA A, KUMAR V, LONGCHAMPS L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species[J]. Microelectronics Journal, 2024, 9: 100648.
- [28] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[C]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [29] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[J]. ArXiv preprint arXiv: 2502.12524, 2025.
- [30] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi-box detector[C]. Computer Vision-ECCV: 14th European Conference, 2016, PartII14: 21-37.
- [31] WANG D R, TAN J S, WANG H, et al. SDS-YOLO: An improved vibratory position detection algorithm based on YOLOv11 [J]. Measurement, 2025, 244: 116518.
- [32] YANG Y ZH, LI Y X, TAO M H. FE-YOLO: A lightweight model for construction waste detection based on improved YOLOv8 model[J]. Buildings, 2024, 14(9): 2672.
- [33] 荀雨薇,吴丽丽. 基于改进 YOLOv8 模型的生活垃圾检测[J]. 软件导刊, 2025, 24(4):56-61.
XUN Y W, WU L L. Domestic waste detection based on improved YOLOv8 model[J]. Software Guide, 2025, 24(4): 56-61.

作者简介

任梦晗,硕士,主要研究方向为人工智能与垃圾分类检测。

E-mail:15624280321@163.com

赵海燕(通信作者),硕士,教授,主要研究方向为人工智能与计算机视觉。

E-mail:zhaohaiyan@imun.edu.cn

宋佳智,博士,讲师,主要研究方向为人工智能与生物信息学。

E-mail:songjz671@nenu.edu.cn