



基于改进 YOLOv7 算法的建筑垃圾分类检测

刘进军¹, 李磊磊¹, 张昊宇¹, 荣辉², 王少坡³, 赵坚^{1,✉}

1. 天津城建大学控制与机械工程学院, 天津 300384; 2. 天津城建大学材料科学与工程学院, 天津 300384; 3. 天津城建大学环境与市政工程学院, 天津 300384

摘要 基于深度学习算法的建筑垃圾分类检测技术对建筑垃圾回收和资源再利用具有重要意义。提出了改进的 YOLOv7 算法实现对建筑垃圾的分类检测。改进算法采用内容感知特征重组 (content-aware reassembly of features, CARAFE) 上采样算子替换 YOLOv7 中最邻近插值方式的上采样算子, 从而提高了目标检测精度; 引入分布移位卷积 (distribution shifting convolution, DSConv) 模块替换 YOLOv7 的头部网络中部分传统卷积, 实现了模型的轻量化。结果表明, 改进算法的 mAP 值达到了 90.7%, 模型计算量仅为 96G。该方法具有准确率高、稳健性强等特点, 在建筑垃圾分类检测实际场景中具有较高的应用价值。

关键词 建筑垃圾检测; YOLOv7; CARAFE; DSConv

随着我国城市城镇化加速, 新旧建筑的更迭也不断加快, 导致建筑垃圾数量增长迅速。建筑垃圾对环境造成了巨大危害, 因此建筑垃圾的回收与利用对环境保护具有重要的意义。目前堆放和填埋作为主要处理建筑垃圾的方式, 资源回收率低, 环境危害大。建筑业作为碳排放最高的三大行业之一, 节能减排的任务迫在眉睫。我国对建筑垃圾的处理方式相对单一, 建筑垃圾的处理方式主要是采用不分拣直接处理的方式, 多种建筑垃圾混杂在一起, 造成利用率变低、资源浪费。目前, 已有企业开始对建筑垃圾中混凝土块进行回收, 然后粉碎用于建筑施工、路面铺设中, 有效促进了资源回收利用。然而建筑垃圾中的砖块、泡沫板、塑料等垃圾直接影响到回收材料的使用性能, 因此在回收生产过程中需要对其进行分拣。目前由于生产工况和成本的限制, 生产线中通常采用传统人工分拣的方式。这种方式不仅效率低, 而且垃圾处理生产线的环境较恶劣、粉尘较大, 对人体身心健康有害。为此, 针对建筑垃圾精准自动分拣的方法是一项值得研究的内容。

针对垃圾智能分类识别的研究, 许多学者进行了深入探索。夏景明等^[1]提出一种基于 VGG16 改进的轻量化垃圾分类网络模型 LW-GCNet, 将传统卷积替换为深度可分离卷积并添加 SE 注意力机制; 原模型中的全连接层替换成自适应最大池化层和全局平均池化层。改进模型具有轻量化、识别准确率高等优势。赵冬娥等^[2]利用光谱角填图法、主成分分析法和 Fisher 判别分析建立了垃圾分类识别模型, 实现了对高光谱图像中的垃圾进行分类识别。XIAO 等^[3]应用近红外高光谱技术对建筑垃圾进行提取和分类, 为避免高光谱数据的冗余性, 提出一种勾股小波变换 (PWT) 提取特征反射率的方法。同时, 提出一种可以适应较恶劣的工作环境的互补故障排除方法 (CT) 用于在线识别建筑垃圾。赵珊等^[4]提出一种基于改进 SSD 的建筑垃圾分类检测模型, 将隐式特征金字塔和 MobileNetV2 引入原始 SSD 模型得到改进模型, 提升了模型检测精度的同时也加快了模型检测速度。ZHANG 等^[5]提出了一种基于 ResNet18 的建筑垃圾分类模型 (CTR), 通过在残差网络中加入了自监测模块, 使得 CTR 模型可以整合所有通道映射的相关特征, 压缩空间维度的特征, 提高了特征图的表示能力, 在 TrashNet 数据集上对该模型进行了测试, 实验结果表明, 该模型的图像分类达到较高的准确率。ZHANG 等^[6]提出了一种新的两阶段垃圾识别检测算法 (W2R), 该算法分为训练识别模型 (RegM) 和通过语义检索构建识别检索模型 (RevM) 2 个阶段。实验结果表明, 由 W2R 算法训练的 RevM 的性能比一阶

收稿日期: 2023-09-19; 录用日期: 2023-12-17

基金项目: 国家自然科学基金资助项目 (52278269); 天津市重点研发计划科技支撑重点项目 (20YFZCSN00520); 天津市科技计划项目 (21YDTPIC00490)

第一作者: 刘进军 (1979—), 男, 博士, 讲师, liujj@tcu.edu.cn; ✉通信作者: 赵坚 (1966—), 女, 博士, 教授, zhaojian_tju@163.com

段分类模型 (ClfM) 和手动分拣实验 (MS) 更好且鲁棒性更高。高明等^[7]提出了一种迁移学习网络架构 GANet, 并将 GANet 应用于垃圾分类识别中, 将 EfficientNet-B5 作为主干网络; 引入像素级空间注意力机制 PSATT; 应用 RAdam 优化方法、标签平滑正则化与 Focal loss 相结合的损失函数、改进的阶梯型 OneCycle 学习率控制法、SWA 权重平滑等训练策略, GANet 在云端部署测试中达到了较好的推理速度和较高的分类准确率。康庄等^[8]提出了一种基于特征提取网络 Inception v3 和迁移学习相结合的方法, 并将模型部署在树莓派 3B+上, 实现了对垃圾的自动分类回收。马雯等^[9]将 Faster R-CNN 中特征提取网络 VGG16 替换为 ResNet50, 同时使用 Soft-NMS 算法代替原模型中非极大值抑制算法, 得到一种改进的 Faster R-CNN 模型用于垃圾检测, 改进模型对垃圾具有较高的检测精度。LIN 等^[10]提出了一种 CVGGNet 模型, 即 VGG 结构 (VGGNet-11、VGGNet-13、VGGNet-16 和 VGGNet-19) 基于迁移学习, 结合数据扩充和循环学习率对建筑垃圾进行分类。实验结果表明, 循环学习率和迁移学习可以提高 VGGNet 的性能。NOWAKOWSKI^[11]等利用卷积神经网络 (CNN) 对垃圾的类型进行分类, 并利用区域卷积神经网络 (R-CNN) 检测垃圾的类别和大小, 所采用的分类和检测算法均显示出较高的识别效率。此外, 张睿萍等^[12]将 Mask R-CNN 中特征提取主干网络 ResNet 替换为 ResNeXt101 得到改进模型, 解决了垃圾多分类任务中分类检测效果不佳等问题。LU 等^[13]应用语义分割模型 DeepLabv3+开发了建筑垃圾分割模型。邢洁洁等^[14]将 YOLOv5s 模型进行了轻量化改进, 并将改进模型部署到边缘计算机设备上应用于农田垃圾检测中。上述方法在不同程度上提升了垃圾分类检测的性能。建筑垃圾作为垃圾的一种, 在分类检测方法上具有相通性, 因此以上垃圾分类检测方法的提出为建筑垃圾分类检测的可行性提供了理论铺垫。

WANG 等^[15]提出了 YOLOv7 目标检测模型, 并通过与以往目标检测模型检测性能进行对比, 验证了 YOLOv7 具有更高的检测精度和更快的检测速度。由于建筑垃圾分拣任务中需要建筑垃圾分类检测算法具有较高的实时性和准确性, 因此本研究选用 YOLOv7 作为基础模型对建筑垃圾进行分类检测。但由于建筑垃圾形状不规则且颜色差异大, 故存在误检漏检的情况, 因此本研究提出一种基于 YOLOv7 的改进模型 DC-YOLOv7。本研究的创新点包括: 1) 改进 YOLOv7 上采样模块, 用 CARAFE 上采样算子替代最邻近插值上采样算子, 提高了模型的检测精度, 以实现了对建筑垃圾的准确分类识别; 2) 对 YOLOv7 模型进行轻量化设计, 选用 DSConv 模块替换 YOLOv7 的 Head 中部分传统卷积, 降低了模型计算量的大小, 保持较高的分类检测速度; 3) 将 DC-YOLOv7 模型应用在建筑垃圾分拣任务中, 并通过实验验证了其有效性。

1 材料与方法

1.1 YOLOv7 算法概述

目前 YOLOv7 有 7 种不同的网络模型, 针对边缘 GPU 设计了轻量级的 YOLOv7-tiny 模型; 针对普通 GPU 设计了 YOLOv7 模型, 然后通过对 YOLOv7 的深度和宽度进行缩放, 又得到了 YOLOv7-X 模型; 针对高性能云 GPU 设计了 YOLOv7-W6 模型, 通过对 YOLOv7-W6 缩放得到了 YOLOv7-E6 和 YOLOv7-D6 两种模型, 然后将 E-ELAN 用于 YOLOv7-E6 模型中, 得到了 YOLOv7-E6E 模型^[15]。在建筑垃圾分拣生产线上, 权衡模型检测性能和成本等问题, 建筑垃圾分类检测模型主要部署在普通 GPU 服务器中, 因此本研究不再讨论适用于云 GPU 的 YOLOv7-W6、E6、D6 和 E6E 四种模型。虽然 YOLOv7-tiny 模型是面向边缘 GPU 部署而设计的, 但也可部署在普通 GPU 上。因此对于普通 GPU 服务器, YOLOv7-tiny、YOLOv7 与 YOLOv7-X 三种模型均可适用。为对比三者的检测性能, 选择在 Microsoft COCO 通用数据集上训练三种模型, 检测性能和参数对比结果见表 1。从表 1 可知, YOLOv7-tiny 比 YOLOv7 的参数量和 FLOPs 分别低 29.3M、90.8G, 检测速度每秒快 123 帧, 但是 mAP 值却比 YOLOv7 低 14.6 个百分点。YOLOv7-X 比 YOLOv7 的 mAP 值虽然高 1.3 个百分点, 但是模型的参数量和 FLOPs 分别比 YOLOv7 高 32.8M、85.2G, 检测速度也每秒慢 35 帧。综合考虑网络检测精度、速度和轻量化要求后, 结合模型应用场景, 本研究选

表 1 不同模型的检测性能和参数对比结果

Table 1 The detection performance and parameter comparison results of different models

模型	评价指标			
	mAP/%	参数量/M	FLOPs/G	FPS/(帧·s ⁻¹)
Yolov7-tiny	54.5	5.9	13.7	212
Yolov7	69.1	35.2	104.5	89
Yolov7-X	70.4	68.0	189.7	54

择 YOLOv7 作为基础模型，并对其进行优化，以实现建筑垃圾的分类检测。

YOLOv7 模型主要由输入 (Input)、主干网络 (Backbone) 和头部网络 (Head) 三部分组成^[16]。其主干网络是由 CBS、ELAN 和 MP-1 等模块组成，头部网络是由 CBS、Conv、Upsample、SPPCSPC、MP-2、ELAN'和 RepConv 等模块组成。YOLOv7 模型结构如图 1 所示。输入图像经过一系列数据预处理后通过输入模块把尺寸缩放至固定尺寸，随后进入主干网络 (Backbone) 提取图像特征，将图像特征送入头部网络 (Head) 中进行特征融合，得到大、中、小 3 种尺寸的特征，最后将融合后的特征送入到检测头，输出检测结果。

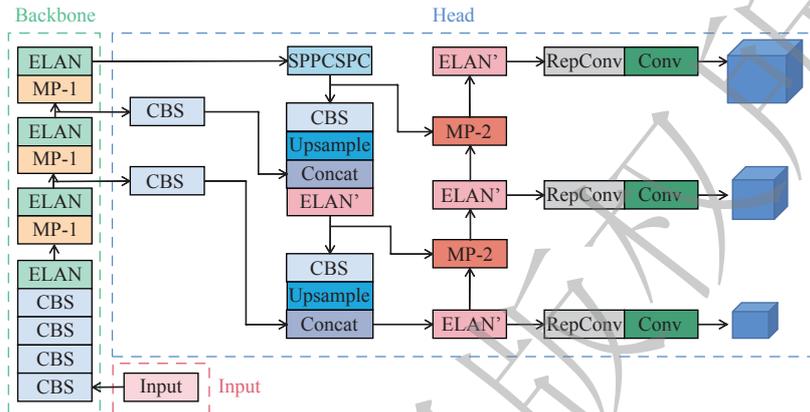


图 1 YOLOv7 模型结构

Fig. 1 Structure diagram of YOLOv7

在 YOLOv7 模型中，CBS 模块由卷积、批量归一化和 SiLU 激活函数组成^[17]。ELAN 和 ELAN'模块通过控制梯度路径，使模型学习到更多的特征。SPPCSPC 模块通过增大感受野，使算法适应不同分辨率的图像。MP-1 和 MP-2 模块主要作用是执行下采样。Upsample 模块是上采样的模块，它使用的上采样方式是最近邻插值法。RepConv 模块作用是对不同尺度特征进行通道数调整^[16]。YOLOv7 部分基础组件结构如图 2 所示。

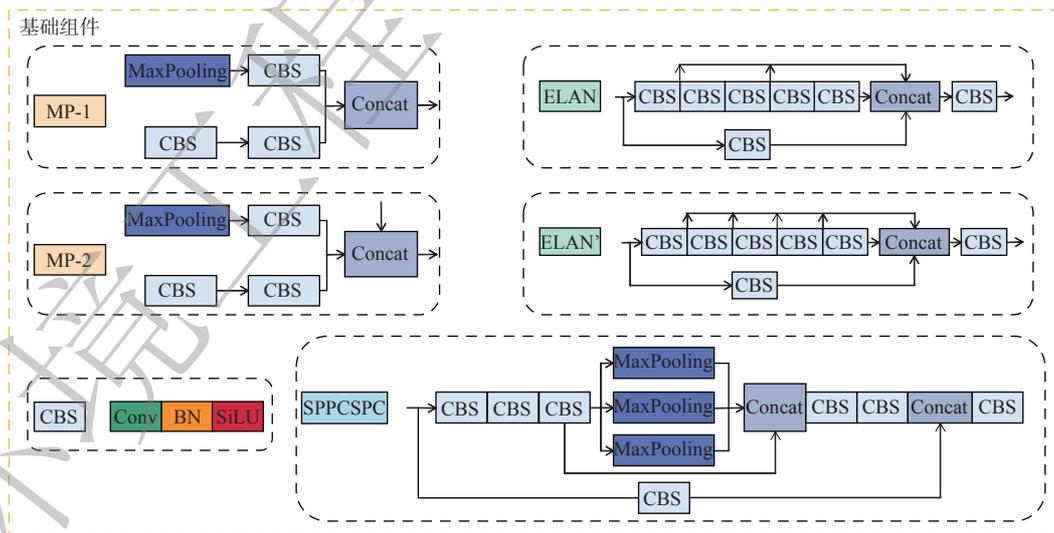


图 2 YOLOv7 基础组件结构

Fig. 2 Basic component structure diagram of YOLOv7

1.2 YOLOv7 模型改进

YOLOv7 中的上采样模块使用的是最近邻插值算法，该算法不需要计算，只需将距待求像素点最近的像素灰度值赋值给它。最近邻插值法具有计算量小、计算简单等优点，但由于未考虑周围像素点的影响，采样

后损失了图像的质量，图像失真严重。

考虑到周围像素点影响，减少图像质量的损失，使用可在更大感受野内聚合上下文信息的 CARAFE 模块替换 YOLOv7 中最邻近插值算法的上采样模块；为提高模型精度的同时保持模型的轻量化，同时，引入 DSConv 模块对 YOLOv7 中 Head 部分进行轻量化改进，由于 DSConv 中采用整数运算，因而实现了内存使用量的减少并提高运算速度。最终改进的模型名称取自 DSConv 和 CARAFE 的首字母并结合 YOLOv7，称为 DC-YOLOv7。DC-YOLOv7 的模型结构如图 3 所示，其改进组件结构图如图 4 所示。

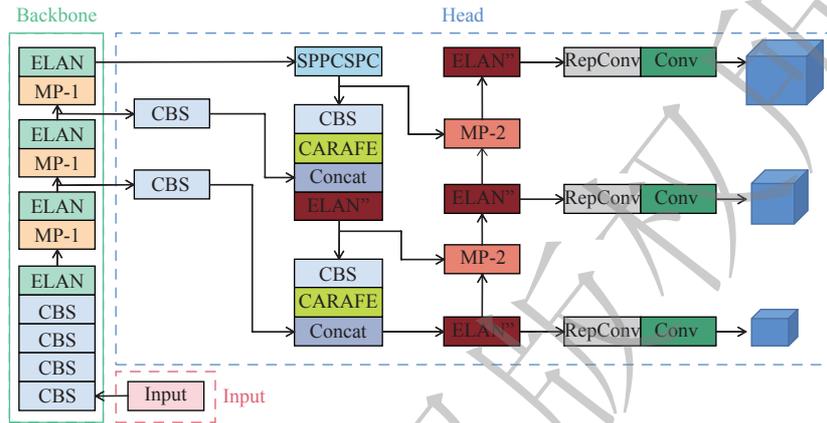


图 3 DC-YOLOv7 模型结构

Fig. 3 Structure diagram of DC-YOLOv7

1) CARAFE 上采样算子

CARAFE 是轻量级上采样算子，在重组的时候可以较大的感受野，能更好地利用周围的信息，会根据输入特征来指导重组过程，且不会引入过多的参数和计算量。在不同的任务中，使用 CARAFE 上采样算子的模型的检测效果都取得了明显提升，同时其带来少量额外参数和计算量。

CARAFE 算子主要由上采样核预测模块和特征重组模块组成，其结构图见文献 [18]。输入特征图 x ，其大小为 $C \times H \times W$ ，首先上采样核预测模块会预测每个位置上的上采样核，然后特征重组模块会完成上采样操作，得到输出特征图 x' ，其大小为 $C \times \sigma H \times \sigma W$ ， x' 中任意目标位置 $l' = (i', j')$ 与 x 中的源位置 $l = (i, j)$ 是一一对应的，其中 σ 为上采样率， $i = i' / \sigma$ ， $j = j' / \sigma$ 。

上采样核预测模块主要完成特征图通道压缩、内容编码及上采样核预测、上采样归一化三个步骤。为了减少后续的计算量，对于形状为 $C \times H \times W$ 的输入特征图，通过 1×1 卷积将它的通道数压缩到 C_m ，假设上采样核尺寸为 $K_{up} \times K_{up}$ ，对于压缩后的特征图使用 $K_{encoder} \times K_{encoder}$ 的卷积预测上采样核，将通道维在空间维展开得到上采样核，其形状为 $K_{up}^2 \times \sigma H \times \sigma W$ ，随后通过 softmax 对上采样核进行归一化操作^[19]。其每个目标位置 l' 的上采样核 $w_{l'}$ 由上采样核预测模块 ψ 基于 x_l 的邻域来预测，如式 (1) 所示。

$$w_{l'} = \psi(N(x_l, k_{encoder})) \tag{1}$$

特征重组模块会将 x' 中所有位置都映射到 x 中，以目标位置为中心，取大小为 $K_{up} \times K_{up}$ 的区域，将 $K_{up} \times K_{up}$ 与该点预测的上采样核做点积运算得到输出，重组过程如式 (2) 所示，其中 ϕ 是将 x_l 的邻域与上采样核 $w_{l'}$ 进行重组的内容感知模块。

$$x'_{l'} = \phi(N(x_l, k_{up}), w_{l'}) \tag{2}$$

式中： $N(x_l, k)$ 表示以 l 为中心 x 的 $k \times k$ 区域。

2) DSConv 分布偏移卷积

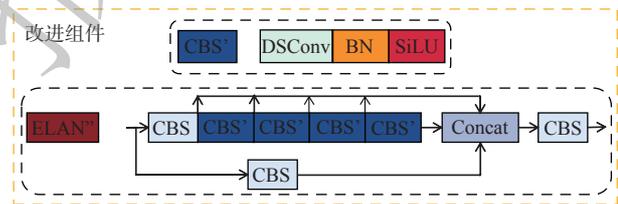


图 4 DC-YOLOv7 改进组件结构图

Fig. 4 Improved component structure diagram of DC-

DSConv 称为分布偏移卷积, 其总体目标是通过使用量化和分布偏移来模拟卷积层的行为, DSConv 相较于传统卷积, 其内存使用量更低、计算速度更快。DSConv 将传统的卷积内核分解为可变量化内核 (VQK) 和分布偏移, VQK 仅保存可变位长整数值, 它具有与原始卷积张量相同的大小 (C_o, C_i, K_h, K_w) , 其中 C_o 是下一层的通道数, C_i 是当前层中的通道, K_h 和 K_w 分别是内核的宽度和高度。分布偏移由两个分布移位器张量组成, 包括内核中的分布偏移 (KDS) 和通道中的分布偏移 (CDS), 通过应用 KDS 和 CDS 移动 VQK, 使 DSConv 的输出和原始权重张量的值相匹配。

在 DSConv 的量化过程中, VQK 和每块尾数张量 (Mantissa Tensor) 进行点积运算, KDS 和原始卷积的指数张量 (Exponent Tensor) 数值进行相加运算, 将点积运算结果与相加运算结果进行点积运算, 得到最终输出结果。DSConv 的量化过程结构图见文献 [20]。

在量化过程中, 首先需要对每个卷积层权重进行缩放, 将所有的权重量化到最接近的整数 w_q 并放到 VQK 中, w_q 的范围如式 (3) 所示。

$$w_q \in \mathbb{Z}, b \in \mathbb{N} | -2^{b-1} \leq w_q \leq 2^{b-1} - 1 \quad (3)$$

式中: w_q 是张量中的参数值; b 是网络输入位数。

随后, KDS 需要将 VQK 匹配原始的卷积张量, KDS 用两个张量在两个域中移动来做到这一点, 在这一过程中, 对两个张量进行初始化, 采用最小化相对熵的方法, 求出原始权重分布和 KDS 之间最小的信息损失。通过位移后的 VQK 和原始分布的 softmax 值来计算 ξ (ξ 为该块的 KDS 值), 如式 (4) 所示。

$$\xi = \min_{\xi} \sum_j T_j \log\left(\frac{T_j}{I_j}\right) \quad (4)$$

上述公式中位移后的 VQK 和原始分布的 softmax 值分别如式 (5)、式 (6) 所示。

$$T_j = \frac{e^{w_j}}{\sum_i e^{w_i}} \quad (5)$$

$$I_j = \frac{e^{\hat{\xi} - w_{qj}}}{\sum_i e^{\hat{\xi} - w_{qi}}} \quad (6)$$

式中: j 为当前层的第 j 块卷积核; i 为第 i 层卷积; $\hat{\xi}$ 为偏差, 默认为 0; T_j 为当前第 j 块的 VQK 的初始值; I_j 为当前第 j 块的 softmax 值。

通过上述方法得到了 KDS 和 VQK, 从而实现了原始卷积的量化。

2 结果与讨论

2.1 实验环境与参数设置

本实验基于 Windows11 操作系统, 采用 Python3.7.6 编程语言, Pytorch1.11.0 深度学习框架。实验相关硬件配置及模型参数如表 2 所示。

2.2 实验数据集

本实验采用自制的建筑垃圾数据集, 实验数据集图片示例如图 5 所示。数据集采集于不同光照条件和不同角度下, 共采集了砖、泡沫板、塑

表 2 相关硬件配置及模型参数

Table 2 Related hardware configuration and model parameters

名称	配置	名称	数值
GPU	RTX3060	图像大小/像素	640×640
CPU	R7 5800H	学习率	0.01
CUDA(版本)	11.3	迭代次数	300
内存	16G	批量大小	4

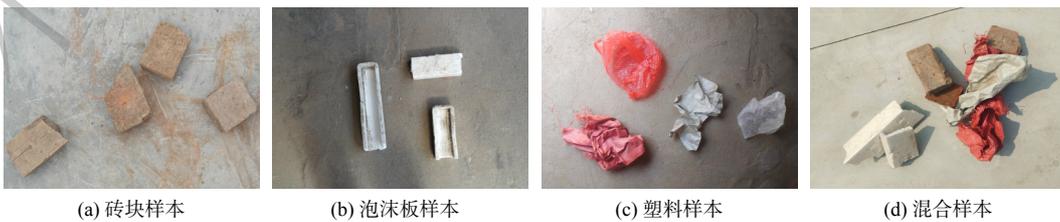


图 5 实验数据集图片示例

Fig. 5 Example images of experimental dataset

料 3 类建筑垃圾图片，其中包含不同建筑垃圾相互堆叠的图片 (见图 5(d))，每类采集了约 200 张 RGB 图片，共约 800 张，分辨率均为 640×640 像素。本实验以 7:1:2 的比例划分的训练集、验证集和测试集。在该数据集上进行对比实验，训练并验证改进模型的有效性。

2.3 评价指标

本实验使用精确率 P (Precision)、召回率 R (Recall)、平均精度均值 mAP (mean Average Precision)、每秒检测帧数 FPS (Frame Per Second)、参数量 $Params$ 和反应模型复杂度的计算量 $FLOPs$ 作为评价指标，来衡量建筑垃圾的检测效果^[21]。

1) 精确率 P 表示预测的正样本有多少是预测正确的，如式 (7) 所示。

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

2) 召回率 R 表示样本中的正样本有多少是预测正确的^[22]，如式 (8) 所示。

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

上述两式中：针对每类建筑垃圾， FP 表示误检的建筑垃圾个数，个； TP 表示检测正确的建筑垃圾的个数，个； FN 表示漏检的建筑垃圾个数，个。

3) mAP 指所有类别建筑垃圾的平均精度 AP (Average Precision) 的均值。在坐标系中， AP 可由计算 $P-R$ 曲线围成的面积得到， AP 衡量的是训练模型在每个类别上的检测效果，如式 (9) 所示。

$$AP = \int_0^1 P(R) dR \quad (9)$$

在多类别识别的目标检测任务中，常使用 mAP 指标。本研究使用的 mAP 是 IoU 阈值为 0.5 时得到的，它也被称为 $mAP@.5$ ，该评价指标常用于评价目标检测模型的整体检测性能， mAP 越高表示目标检测网络模型整体效果越好，如式 (10) 所示。

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i \quad (10)$$

式中： c 为数据集中的总类别数； i 为检测次数。

4) FPS 表示目标检测网络每秒检测图片的数量，常用来评估目标检测的速度，如式 (11) 所示。

$$FPS = \frac{FrameCount}{ElapsedTime} \quad (11)$$

式中： $FrameCount$ 为检测图片的总数，帧； $ElapsedTime$ 为检测图片的总时长，s。

5) 参数量 $Params$ 表示模型中包含的所有参数总量，用来衡量模型的大小(空间复杂度)，单位为 M。

6) 计算量 $FLOPs$ 用来衡量算法的复杂度(计算时间复杂度)，单位为 G。

2.4 YOLOv7 与改进模型实验对比

为了研究 CARAFE 上采样算子中 $K_{encoder}$ 和 K_{up} 值的改变对改进 YOLOv7 模型性能的影响，在建筑垃圾数据集上进行对比实验，实验结果见表 3。从实验结果可以看出增加 K_{up} 的同时也需要更大的 $K_{encoder}$ ，因为内容编码器需要更大的感受野来预测更大的重组内核。本实验主要对比 mAP 值，总体来说，同时增加 $K_{encoder}$ 和 K_{up} 可以提高性能，而仅增加其中之一则不会提高性能。通过实验结果可以总结出公式 $K_{encoder} = K_{up} - 2$ ，与文献^[18]中得出的结论一致，按照这个公式设置 $K_{encoder}$ 和 K_{up} 值得到的模型效果更佳。为了权衡模型检测精度和检测速度之间的关系，在后续实验中，

表 3 不同 $K_{encoder}$ 和 K_{up} 值的模型检测效果

Table 3 Detection performance of the improved mode with different $K_{encoder}$ and K_{up} values

$K_{encoder}$	K_{up}	$P/\%$	$R/\%$	$mAP/\%$	$FPS/(帧 \cdot s^{-1})$
1	3	90.8	82.6	90.7	61
1	5	95.9	74.6	90.6	54
3	3	94.1	84.9	90.4	61
3	5	93.9	81.4	91.0	55
3	7	96.5	76.1	91.0	43
5	5	96.2	80.4	90.7	53
5	7	97.0	84.2	91.7	43

CARAFE 上采样算子中的 K_{encoder} 和 K_{up} 的值折中设置为 $K_{\text{encoder}} = 3$ 和 $K_{\text{up}} = 5$ 。

为了验证本研究提出的网络模型 (DC-YOLOv7) 改进的效果, 在建筑垃圾数据集上进行消融实验, 将 DC-YOLOv7 与 YOLOv7 和改进模型 I 进行对比。改进模型 I 是在 YOLOv7 的基础上引入了 CARAFE 算子。为保证实验的准确性, 模型训练的过程中均采用相同的参数, 实验结果见表 4。

通过表 4 可以看出, 在 YOLOv7 中引入 CARAFE 上采样算子, 对比 YOLOv7 网络模型, 虽然模型的整体参数量略有增加, 改进模型 I 比 YOLOv7 的参数量增加了 0.3%, 计算量增加了 0.2%, 但是 mAP 值得到大幅度提升, 提升了 2.6%, 验证了 CARAFE 对模型精度提升的优越性。同时, 在 YOLOv7 中 Head 部分引入 DSConv 模块后, 与 YOLOv7 对比, DC-YOLOv7 的 mAP 值提升了 2.3%, 计算量降低了 8.7%, 这表明 DC-YOLOv7 既提高了检测精度又保持着模型的轻量化。此外, 由于引进的 DSConv 模块不如原始卷积模块提取的特征丰富, 降低了特征提取的精准度, 因而 DC-YOLOv7 略低于改进模型 I 的 mAP 值。相较于 YOLOv7, 虽然 mAP 值提升的少一些, 但是 DC-YOLOv7 更加轻量化, 便于部署。总体而言, 该实验验证了将 CARAFE 上采样算子和 DSConv 模块组合使用在 YOLOv7 网络模型中的有效性。

在测试集中选取了 5 张建筑垃圾的图片, 分别使用 YOLOv7 模型、改进模型 I 和 DC-YOLOv7 模型 3 种网络模型对建筑垃圾图片进行目标检测, 不同模型的检测结果如图 6 所示, 其中图 6 第⑤行为不同建筑垃圾堆叠图片的检测结果。从检测结果图中可以看出, 与 YOLOv7 相比, 改进模型 I 和 DC-YOLOv7 的漏检率和错检率均有降低, 如 YOLOv7 漏检了图 6 第①行中的泡沫板、图 6⑤行中的塑料, 但两张图中漏检的目标均被改进模型 I 和 DC-YOLOv7 成功检测出来。在图 6 第②行中 YOLOv7 将光线误检成了泡沫板, 在图 6 第③行和图 6 第④行中 YOLOv7 也发生了误检的情况, 这些误检的目标均被改进模型 I 和 DC-YOLOv7 正确检出。由于图 6 第①行到图 6 第⑤行中受光照、相互遮挡、物体形状不规则且颜色有差异等问

表 4 基于 YOLOv7 模型的消融实验

Table 4 Ablation experiment based on YOLOv7 model

网络模型	评价指标		
	Params/M	FLOPs/G	mAP/%
YOLOv7	35.5	105.2	88.4
改进模型 I	35.6	105.4	91.0
DC-YOLOv7	35.6	96.0	90.7





图 6 不同模型检测效果图

Fig. 6 Detection effects of different models

题的影响，会造成 YOLOv7 模型误检漏检的情况，但改进模型 I 和 DC-YOLOv7 模型则没有受这些问题的影响。可以看出改进模型 I 和 DC-YOLOv7 模型检测效果比较接近，相对于 YOLOv7 检测效果均有提升。因此，采用改进模型 I 和 DC-YOLOv7 模型可以检测出更准确的建筑垃圾类别信息，降低了漏检率和错检率的同时，置信度也更高。

2.5 改进模型与其他模型检测速度对比

为了验证改进模型 DC-YOLOv7 的检测速度，在建筑垃圾数据集上进行对比实验，对比 FasterRCNN^[23]、SSD^[24]、RFBNet^[25]、RetinaNet^[26]、YOLOv7^[15] 等目标检测网络模型，实验结果如表 5 所示。

通过表 5 可以看出，DC-YOLOv7 的检测速度能达到每秒 50 帧，这是由于其采用了 DSCony 模块来简化传统的卷积模块，使网络模型更加轻量化。DC-YOLOv7 比 SSD、RFBNet、RetinaNet 模型的检测速度分别快了 117.4%、100%、85.2%，这表明了 DC-YOLOv7 在实时检测速度上的优势。虽然 DC-YOLOv7 比 YOLOv7 网络模型的检测速度慢了 16.7%，但也可以满足建筑垃圾实时检测任务的需求。

此外，在轻量化方面，DC-YOLOv7 的参数量仅有 35.6M，模型的计算复杂度为 96G。与 SSD、RFBNet 和 YOLOv7 相比，虽然参数量分别高了 11.6 M、0.1 M、0.1M，但模型的复杂度分别低了 178.5 G、221.3 G、9.2G。相比于 Faster RCNN、RetinaNet，DC-YOLOv7 的参数量和计算复杂度均低于这两个模型的这两项指标，参数量分别低了 101.2M 和 0.8M，计算复杂度分别低了 305.8G 和 50.3G。综上所述，DC-YOLOv7 足够轻量化，可以被部署在大部分硬件设备上。

2.6 实际工况下的模型检测效果

本研究建筑垃圾分类检测算法主要应用在建筑垃圾回收与再利用企业的建筑垃圾分拣生产线中。在建筑垃圾传送过程中，基于本研究提出垃圾分类检测模型实现对非混凝土块的垃圾进行识别，将识别信息传递给自动分拣设备，完成建筑垃圾的分拣操作。为实现建筑垃圾中的钢筋高效分拣，生产线中通常采用磁铁吸附的方法。最后对分拣后的混凝土块进行粉碎，完成回收再利用。因此本研究只需要识别除了钢筋、混凝土块之外的其他建筑垃圾即可。为检测出更多类型的建筑垃圾，本研究增加对木块样本的训练。为验证对实际工况下建筑垃圾识别检测效果，本节选用实际工况下分别含砖块、泡沫板、塑料、木块等建筑垃圾的图像，并对四张图片中的建筑垃圾进行识别，算法识别检测结果如图 7 所示。由图 7 可知，本研究提出的建筑垃圾识别模型对于实际工况下的建筑垃圾具有较好的识别效果。

表 5 不同网络模型性能对比结果

Table 5 Performance comparison results of different network models

网络模型	评价指标		
	Params/M	FLOPs/G	FPS/(帧·s ⁻¹)
Faster-RCNN	136.8	401.8	10
SSD	24.0	274.5	23
RFBNet	35.5	317.3	25
RetinaNet	36.4	146.3	27
YOLOv7	35.5	105.2	60
DC-YOLOv7	35.6	96.0	50



图7 实际工况下的模型检测效果

Fig. 7 Detection effects of the proposed model under actual conditions

3 结论

提出了一种基于改进 YOLOv7 的建筑垃圾检测模型 DC-YOLOv7。相比于其他目标检测模型,改进模型更好地权衡了模型检测精度和模型性能之间的关系,在具有较高检测精度和检测速度的同时保持了模型的轻量化。引入 CARAFE 上采样算子替换了 YOLOv7 中最邻近插值方式的上采样算子,虽然带来了少量额外参数和计算量,但提升了模型的检测精度;在 YOLOv7 的 Head 网络部分引入 DSCConv 模块替换部分传统卷积,由于 DSCConv 模块中采用整数运算,因此,相比于传统卷积中的浮点运算,DSCConv 模块内存使用量更少、运算速度更快。在建筑垃圾测试集上测试,本研究提出的建筑垃圾检测模型 DC-YOLOv7 的 mAP 值达到了 90.7%,检测速度可达到每秒 50 帧,计算量降至 96G,综合性能高于常见的目标检测网络。该模型能够快速准确地检测出常见种类的建筑垃圾,克服了建筑垃圾形状不规则、颜色差异大等特征导致的精度低、检测速度慢等问题,可为提升建筑垃圾的回收和资源化利用提供技术支撑。

参考文献

- [1] 夏景明,徐子峰,谈玲.轻量化网络 LW-GCNet 在垃圾分类中的应用[J].环境工程,2023,41(2):173-180.
- [2] 赵冬娥,吴瑞,赵宝国,等.高光谱成像的垃圾分类识别研究[J].光谱学与光谱分析,2019,39(3):921-926.
- [3] XIAO W, YANG J H, FANG H Y, et al. A robust classification algorithm for separation of construction waste using NIR hyperspectral system[J]. *Waste Management*, 2019, 90: 1-9.
- [4] 赵珊,刘子路,郑爱玲,等.基于 MobileNetV2 和 IFPN 改进的 SSD 垃圾实时分类检测方法[J].计算机应用,2022,42(S1):106-111.
- [5] ZHANG Q, ZHANG X, MU X, et al. Recyclable waste image recognition based on deep learning[J]. *Resources, Conservation and Recycling*, 2021, 171(99): 105636.
- [6] ZHANG S, CHEN Y M, YANG Z L, et al. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification[J]. *Resources, Conservation and Recycling*, 2021, 169: 105543.
- [7] 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.
- [8] 康庄,杨杰,郭濠奇.基于机器视觉的垃圾自动分类系统设计[J].浙江大学学报(工学版),2020,54(7):1272-1280+1307.
- [9] 马雯,于炯,王潇,等.基于改进 Faster R-CNN 的垃圾检测与分类方法[J].计算机工程,2021,47(8):294-300.
- [10] LIN K S, ZHOU T, GAO X F, et al. Deep convolutional neural networks for construction and demolition waste classification: VGGNet structures, cyclical learning rate, and knowledge transfer[J]. *Journal of Environmental Management*, 2022, 318: 115501.
- [11] NOWAKOWSKI P, PAMUA T. Application of deep learning object classifier to improve e-waste collection planning[J]. *Waste Management*, 2020, 109: 1-9.
- [12] 张睿萍,宁芊,雷印杰,等.基于改进 Mask R-CNN 的生活垃圾检测[J].计算机工程与科学,2022,44(11):2003-2009.
- [13] LU W S, CHEN J J, XUE F. Using computer vision to recognize composition of construction waste mixtures: A semantic segmentation approach[J]. *Resources, Conservation and Recycling*, 2022, 178: 106022.
- [14] 邢洁洁,谢定进,杨然兵,等.基于 YOLOv5s 的农田垃圾轻量化检测方法[J].农业工程学报,2022,38(19):153-161.
- [15] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.
- [16] 赵元龙,单玉刚,袁杰.改进 YOLOv7 与 DeepSORT 的佩戴口罩行人跟踪[J].计算机工程与应用,2023,59(6):10.
- [17] WANG Y, FU B, Fu L W, et al. In situ sea cucumber detection across multiple underwater scenes based on convolutional neural networks and image enhancements[J]. *Sensors*, 2023, 23(4): 2037.
- [18] WANG J Q, CHEN K, LIU Z W, et al. CARAFE: content-aware reassembly of features[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3007-3016.
- [19] 陈范凯,李士心.改进 yolov5 的无人机目标检测算法[J].计算机工程与应用,2023,59(18):218-225.
- [20] GENNARI M, FAWCETT R, PRISACARIU V A. DSCConv: efficient convolution operator[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5148-5157.

- [21] 王浩, 吕晓琪, 谷宇. 基于语义融合与多尺度注意力的红外行人检测[J/OL]. 激光杂志. <http://kns.cnki.net/kcms/detail/50.1085.tn.20230213.1802.006.html>.
- [22] 肖振久, 林瀚翰, 曲海成. 改进YOLOv7的SAR舰船检测算法[J]. 计算机工程与应用, 2023, 59(15): 243-252.
- [23] REN S Q, HE K W, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [24] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [25] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection[C]//European Conference on Computer Vision, 2018: 385-400.
- [26] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE international conference on computer vision, 2017: 2980-2988.

(责任编辑: 金曙光)

Construction waste classification and detection algorithm based on improved YOLOv7

LIU Jinjun¹, LI Leilei¹, ZHANG Haoyu¹, RONG Hui², WANG Shaopo³, ZHAO Jian^{1*}

1. School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China; 2. School of Materials Science and Engineering, Tianjin Chengjian University, Tianjin 300384, China; 3. School of Environmental and Municipal Engineering, Tianjin Chengjian University, Tianjin 300384, China

*Corresponding author, E-mail: zhaojian_tju@163.com

Abstract The construction waste classification and detection technology based on deep learning algorithm is of great significance to the construction waste recycling and resource reuse. In this paper, the improved YOLOv7 algorithm was proposed to realize the classification and detection of construction waste. Firstly, the content aware reassembly of features (CARAFE) upsampling operator was used to replace the nearest interpolation upsampling operator in YOLOv7, which reduced the loss of image quality in the upsampling process and aggregates contextual information within a large receptive field, thereby improving the detection accuracy of construction waste. Secondly, the distribution shifting convolution (DSCConv) module was introduced to replace the traditional convolution in the head network of YOLOv7, achieving lightweight of the model. The experimental results showed that the *mAP* value of the improved model reached 90.7%, and the computational complexity was only 96G. The improved model had higher accuracy and stronger robustness performance. It has high application value in the field of construction waste classification and detection.

Keywords construction waste detection; YOLOv7; CARAFE; DSCConv