



DOI:10.12404/j.issn.1671-1815.2402281

引用格式:李利,梁晶,陈旭东,等.基于TLF-YOLOv8的堆叠垃圾实例分割算法[J].科学技术与工程,2025,25(5):2009-2018.

Li Li, Liang Jing, Chen Xudong, et al. Stacked garbage instance segmentation algorithm based on TLF-YOLOv8[J]. Science Technology and Engineering, 2025, 25(5): 2009-2018.

## 基于 TLF-YOLOv8 的堆叠垃圾实例分割算法

李利<sup>1,2,3</sup>, 梁晶<sup>1,2</sup>, 陈旭东<sup>1,2</sup>, 潘红光<sup>1,2</sup>, 寇发荣<sup>1,3</sup>

(1. 西安科技大学电气与控制工程学院, 西安 710054; 2. 西安市电气设备状态监测与供电安全重点实验室, 西安 710054; 3. 西安科技大学机械工程学院, 西安 710054)

**摘要** 相较于一般场景下的图像实例分割,复杂堆叠场景下的实例分割受到严重遮挡、同类别待测物体堆叠等复杂情况的影响,使得其实例分割具有更大的难度。针对具有复杂堆叠场景下的垃圾实例分割问题,提出了一种融合 YOLOv8 与双层特征网络策略的实例分割算法。首先,在数据预处理部分进行特征数据分层,并通过双层图卷积网络(graph convolutions network, GCN)实现双分支特征融合,减弱堆叠情况对被遮挡物体特征的影响,从而解决复杂堆叠遮挡下的实例分割问题。同时,为了解决同类待测物体易混淆的问题,融入了软阈值化非极大值抑制算法和新的交并比算法。最后,根据应用场景和数据集的复杂性,优化了主干网络部分的特征提取模块,并在主干网络部分引入了多尺度注意力机制,有效提高了模型的检测性能。实验使用遮挡垃圾分类实例分割数据集,实验结果表明该方法的平均准确率、交并比阈值为 0.5 时的平均准确率( $AP_{50}$ )、交并比为 0.5~0.95 时的平均准确率( $AP_{50-95}$ )等指标较之前的其他方法更优。相较于原 YOLOv8 算法,检测  $AP_{50}$  提高了 7.9%,分割  $AP_{50}$  提高了 5.4%,具有更好的检测和分割效果。

**关键词** 垃圾堆叠; 双层特征解耦融合; YOLOv8 算法; 软阈值化非极大值抑制; 动态非单调聚焦机制; 期望最大化注意力  
**中图分类号** TP391.4; **文献标志码** A

### Stacked Garbage Instance Segmentation Algorithm Based on TLF-YOLOv8

LI Li<sup>1,2,3</sup>, LIANG Jing<sup>1,2</sup>, CHEN Xu-dong<sup>1,2</sup>, PAN Hong-Guang<sup>1,2</sup>, KOU Fa-rong<sup>1,3</sup>

(1. School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;  
2. Xi'an Key Laboratory of Electrical Equipment Status Monitoring and Power Supply Safety, Xi'an 710054, China;  
3. School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

**[Abstract]** Compared to image instance segmentation in general scenes, instance segmentation in complex stacked scenes is affected by complex situations such as severe occlusion and stacking of similar objects, making instance segmentation more difficult. To solve the problem of garbage instance segmentation in complex stacking scenarios, an instance segmentation algorithm combining YOLOv8 and two-layer feature network strategy was proposed. Firstly, the feature data was layered in the data preprocessing part, and the two-branch feature fusion was realized through the graph convolutions network (GCN), which reduces the influence of stacking on the features of the occluded objects, thus solving the instance segmentation problem under complex stack occlusion. At the same time, in order to solve the problem that similar objects are easily confused, a soft threshold non-maximum suppression algorithm and a new intersection ratio algorithm were integrated. Finally, according to the complexity of application scenarios and data sets, the feature extraction module of the backbone network was optimized, and the multi-scale attention mechanism was introduced in the backbone network, which effectively improves the detection performance of the model. In the experiment, examples of occlusive garbage classification were used to segment the dataset. The experimental results show that this method outperforms other methods in terms of average accuracy, average accuracy when the intersection to union ratio threshold is 0.5 ( $AP_{50}$ ), and average accuracy when the intersection to union ratio is 0.5~0.95 ( $AP_{50-95}$ ). Compared with the original YOLOv8 algorithm, the detection  $AP_{50}$  is increased by 7.9% and the segmentation  $AP_{50}$  is increased by 5.4%, which has better detection and segmentation effects.

**[Keywords]** garbage stacking; double layer feature decoupling fusion; YOLOv8 algorithm; soft non-maximum suppression; wise-in-intersection over union; expectation-maximization attention

收稿日期: 2024-03-31 修订日期: 2024-11-18

基金项目: 国家自然科学基金(61603295); 陕西省自然科学基金基础研究计划(2024JC-YBQN-0726); 陕西省教育厅科研计划(23JK0550); 西安市科技计划(23DCYJSGG0025-2022)

第一作者: 李利(1991—),男,汉族,山东枣庄人,博士,讲师,硕士研究生导师。研究方向: 网络化系统控制、模式识别及应用、故障诊断等。  
E-mail: lilixiansen@163.com。

随着社会的急速发展和人们消费水平的日益增长,人类产生的废物垃圾越来越多,这为垃圾处理带来巨大压力,处理不当可能会给生态系统带来不可逆转的灾难,包括气候变暖、土壤污染、海洋污染等<sup>[1]</sup>。垃圾分类是一项重要的环保行动,通过对垃圾进行分类投放、收集、运输和处理,可以实现垃圾减量化、资源化和无害化,对保护环境、节约资源和促进可持续发展具有重要意义<sup>[2-3]</sup>。垃圾分类可看作是一种复杂堆叠场景,分类前提是要进行垃圾的识别,只通过自然图像难以实现堆叠场景下的垃圾识别<sup>[4]</sup>。由于X射线技术具有高透射性和保留边缘轮廓信息的特点<sup>[5]</sup>,故X射线技术被用于垃圾焚烧发电厂和城市固体废弃物处理中心等场所的垃圾分类处理<sup>[6]</sup>。使用X射线技术,可以对垃圾进行分类和筛选,将可回收物、有害垃圾和其他垃圾分开,实现垃圾的减量化和资源化。因此,研究堆叠场景下基于X射线的垃圾实例分割算法具有重要的意义。

针对堆叠场景下的垃圾识别,要解决的关键问题是如何对该场景中遮挡部分的有效特征进行补偿、如何改善由于场景的密集性所导致的候选框混淆问题,以及通过何种方式提高算法的检测精度<sup>[7-8]</sup>。目前,许多研究人员针对不同场景中的实例分割问题进行了研究,取得了许多成果。针对一般场景的实例分割算法通常可分为双阶段和单阶段两大类,其中双阶段算法主要由目标检测和实例分割两阶段组成。He等<sup>[9]</sup>基于先检测后分割的思路提出了双阶段算法中最为经典的掩码区域卷积神经网络(mask region-based convolutional neural network, Mask-RCNN)算法。Kirillov等<sup>[10]</sup>进一步细化了Mask-RCNN产生的粗糙掩码边缘分割算法。Rendle等<sup>[11]</sup>通过后处理细化模块提高Mask-RCNN的边界质量。Zhang等<sup>[12]</sup>利用边缘信息和语义分割信息提出了细化Mask-RCNN生成的粗糙掩码边缘的RefineMask算法。不同于双阶段算法,单阶段实例分割算法主要是通过目标检测和像素分割并行的方式实现实例分割,这种设计思路使得该类算法具有较少的网络参数,实时性远高于双阶段算法。Tian等<sup>[13]</sup>提出了一种无锚框的候选框生成算法。Wang等<sup>[14]</sup>基于单阶段算法思路提出了一种划分区域后直接分割的基于位置的目标分割(segmenting objects by locations, Solo)算法。Bolya等<sup>[15]</sup>通过设计解耦分支进一步减小了分割网络的整体参数量,进而大大提升了网络的实时性。另外,针对类似堆叠场景中的实例分割算法,目前学术界在这方面研究较少,基于多尺度特征融合的方式提高

分割效果是目前此类场景提高分割精度的主要研究方法。Ling等<sup>[16]</sup>基于全卷积单阶段目标检测(fully convolutional one-stage object detection, FCOS)算法,提出了一种遮挡感知下的双图层实例分割网络(bilayer convolutional network, BCNet),将图像中感兴趣区域(region of interest, RoI)建模为两个重叠图层,并在mask和boundary预测的同时考虑遮挡关系的相互影响。Ke等<sup>[17]</sup>基于proposal-free的思路,结合重叠场景下的X射线检测需求,提出了一种由“易”到“难”策略(easy-to-hard instance segmentation network, ETHSeg),通过置信度高低划分图层,再进行特征融合。然而,受其网络结构局限,算法性能较低且未考虑到该类场景中的其他因素对实例分割效果的影响。总的来说,这种方法实质上只是通过增加梯度流和模型参数的方式提高检测性能,对该类场景的针对性不强,效果也并不理想。

综上所述,现有算法没有考虑到堆叠场景特性对实例分割效果的影响,其分割精度和分割效果仍有提升的空间<sup>[18-20]</sup>。现针对堆叠场景下的垃圾分类问题,基于X射线透射图数据特性和YOLOv8、BCNet等基本方法,提出一种双层特征融合YOLOv8(two layer feature fusion YOLOv8, TLF-YOLOv8)实例分割算法,解决堆叠场景下垃圾分类检测和检测精度不高的难题。首先,为了减弱堆叠场景对图像实例分割的影响,对输入训练集图像进行特征分层,再使用特征提取网络提取图像特征。然后,针对堆叠场景候选框易混淆的问题,在主干网络中引入软阈值化非极大值抑制(soft non-maximum suppression, soft-NMS)算法和动态非单调聚焦机制(wise-intersection over union, WIoU),保留同类堆叠待测物体的候选框信息。其次,针对模型检测精度问题,引入期望最大化注意力机制(expectation-maximization attention, EMA),同时优化主干网络部分的特征提取(CSPLayer-2Conv, C2f),用于提高后续的检测和分割精度。最后,使用一组遮挡垃圾分类实例数据集进行实验验证。

## 1 YOLOv8 介绍

YOLOv8骨干网络和Neck部分参考了YOLOv7 ELAN设计思想,将YOLOv5的C3结构换成了梯度流更丰富的C2f结构,并对不同尺度模型调整了不同通道数,大幅提升了模型性能。Head部分相比YOLOv5改动较大,换成了目前主流的解耦头结构,将分类和检测头分离,同时也从Anchor-Based换成了Anchor-Free,如图1所示。



卷积网络 GCN 层能够学习场景中的上下文依赖关系。此外,在检测头之前进行特征融合,避免增加太多的计算成本和模型参数量。同时,只对检测头的一个输入通道进行特征融合,可以最大限度地减少对原始网络稳定性的影响。本文算法考虑到场景中候选框密集和测试实例容易变形的特点,还对部分结构中的算法进行了优化。

### 2.1 双层解耦特征融合

在堆叠场景进行实例分割时,由于物体互相堆叠,导致主干特征提取网络难以提取到被遮挡物体的部分有效特征。为了减弱上层物体对下层待测物体的这种影响,引入了一种双层特征解耦融合的思想,通过对上下层特征解耦,再通过图卷积网络算法对两层特征进行融合,使得特征提取网络能够学习到被遮挡区域的部分特征,从而在一定程度上减弱了堆叠场景对下层待测物体实例分割的影响,整体流程如图 3 所示。

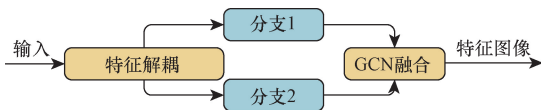


图 3 双分支特征融合

Fig. 3 Dual branch feature fusion

#### 2.1.1 特征解耦分层

如图 4 所示,在数据集的数据处理部分对标注数据进行解耦分层。具体来说,先遍历一个矩形框的列表,对于列表中的每个矩形框,计算它与其他矩形框之间的交集并生成统一掩码,记录交集的像素总数;然后,计算交集像素总数与矩形框内像素总数之间的比例,从而判断该矩形框是否与其他物体相交。

对位于交集的矩形框,通过计算该掩码中的非零像素数量,使其生成背景“对象”分割掩码。同时寻找该掩码中的所有连通区域,并对每个连通区域进行简化处理,生成包含所有连通区域的遮挡层分割掩码列表。然后,通过图卷积网络,实现对遮挡层和被遮挡层的特征关系表达,从而达到遮挡层辅助被遮挡层分割的效果。

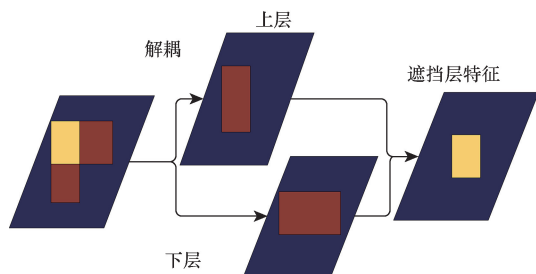


图 4 双层网络策略原理

Fig. 4 Principle of double layer network strategy

#### 2.1.2 特征融合

对数据集进行预处理,首先,解耦出两层特征图像数据后,通过特定标签区分,将两层特征图像数据分别作为两个分支的输入;然后,将原始输入的特征图与输出的特征图进行元素相加操作得到最后输出带权重的特征图。

如图 5 所示,采用双层 GCN 层网络思想,对于得到的特征,第一个 GCN 通过同时检测遮挡轮廓和掩模,显式地建模遮挡区域,提取必要的信息,指导第二个 GCN 的掩模预测。具体来说,双层 GCN 网络策略实现了两个 GCN 层的特征权重融合,算法流程如下。

**步骤 1** 首先,第一个 GCN 网络通过  $1 \times 1$  卷积对输入特征向量  $x$  进行通道数压缩;然后,将卷积后输出的、两组特征向量进行矩阵乘法得到注意力权重矩阵,再进行 softmax 操作;最后,将得到的权重矩阵  $g, \varphi, \theta$  与分支输出的特征向量进行点乘,得到带注意力的矩阵特征图,经过  $1 \times 1$  卷积得到与输入大小一致的输出特征图  $z_0$ 。

**步骤 2** 将第一个 GCN 网络得到的输出特征图  $z$  与第二个 GCN 网络的输入特征  $x$  累加,从而获得融合的遮挡感知特征,并将其作为第二个 GCN 层的输入。第二个 GCN 层得到输入后,重复步骤 1,得到输出特征  $z$ 。

**步骤 3** 将输出特征图  $z$  用于掩模分割预测。

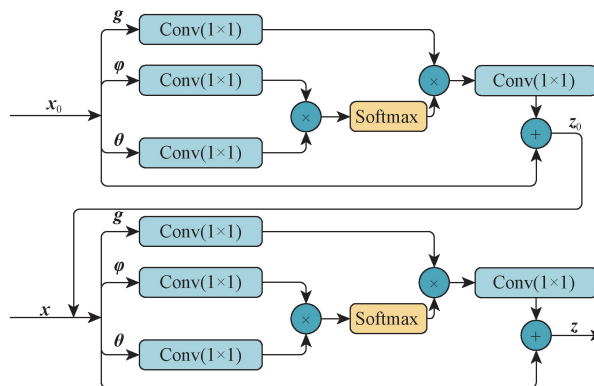


图 5 双层 GCN 层基本结构

Fig. 5 Basic structure of double layer GCN layer

### 2.2 Soft-NMS 算法优化

在实际研究中,高度重叠场景下部分待测物较为密集,难以进行同类物体的独立分割。针对此问题,考虑去掉传统的 NMS 算法。因为传统的 NMS 方法是基于分类分数的,只有最高分数的预测框能留下来,但是大多数情况下 IoU 和分类分数不是强相关,很多分类标签置信度高的框位置都不准。同时,还会有其他物体的框因和当前物体的框重合部分过多而被删掉的情况,如图 6 所示。因此,改用 Soft-NMS 算法。

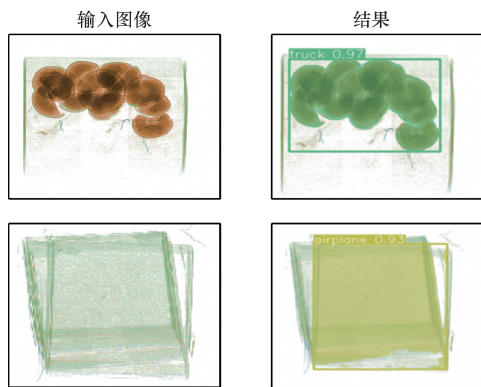


图 6 同类候选框易混淆问题

Fig. 6 Confusion problem of similar candidate boxes

Soft-NMS 通过 IoU 算法计算后,对于最高分框 overlap 大于设定阈值的候选框,不把它直接去掉,而是将它的置信度降低,如式(1)所示。同一个物体周围的框有很多,每次选择分数最高的框,抑制其周围的框,分数最高的框的 IoU 越大,抑制程度越大。一般来说,表示同一个物体的框的 IoU 是会比另一个物体的框的 IoU 大,故这样就会将其他物体的框保留下来,而同一个物体的框被去掉。通过这种方法,可以使多一些候选框被保留下来,一定程度上避免 overlap 的情况出现,提高了重叠场景下的检测效果。

$$S_i = \begin{cases} S_i, & \text{IoU}(M, b_i) < N_i \\ S_i [1 - \text{IoU}(M, b_i)], & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (1)$$

### 2.3 WIoU 算法优化

堆叠场景下的实例分割存在候选框选取难区分的问题,而 YOLOv8 原有纵横比-交并比算法 (complete-intersection over union, CIoU) 只考虑了预测框和真实框的重叠部分,没有考虑两者之间的区域,导致在评估结果时可能存在偏差。在 Head 部分引入了 WIoU 算法,如图 7 所示。

WIoU 通过考虑预测框和真实框之间的区域对 IoU 进行加权,解决了传统 IoU 在评估结果时可能存在的偏差问题。具体而言,WIoU 计算方法如下。

**步骤 1** 计算预测框和真实框的 IoU 得分。

**步骤 2** 计算两个框之间的区域:用预测框和真实框的边框中心点计算它们之间的距离,并将这个距离作为两个框之间的最大距离,进而计算两个

框之间的区域。

**步骤 3** 根据两个框之间的区域,计算权重系数,该系数衡量了两个框之间的关系,可以用于加权 IoU 得分。

**步骤 4** 通过引入框之间的区域和权重系数,WIoU 可以更准确地评估目标检测结果,避免了传统 IoU 的偏差问题。

在 WIoU 中,每个物体框的权重取决于其与真实标注框的重叠程度。重叠程度越大的物体框权重越高,重叠程度越小的物体框权重越低。通过这种方式,WIoU 能够更好地评估检测结果,并且在存在大小物体不平衡的情况下也能给出更准确的评价,同时也更贴合实际研究场景。

### 2.4 C2f\_DCN 算法优化

考虑到堆叠场景及对应数据集中部分实例的不规则性,且样本更加复杂多变,为了使网络尽可能保留更多有效信息,在网络中引入了可变形卷积 (deformable convolutional networks, DCN)。其原理主要用额外的偏移量来增加模块中的空间采样位置,并学习来自目标任务的偏移量,无须额外监督。传统的卷积结构可以定义为

$$y(P_0) = \sum_{P_n \in \mathbf{R}} w(P_n)x(P_0 + P_n) \quad (2)$$

式(2)中: $y$  为输出特征图的每个点,与卷积核中心点对应; $P_n$  为  $P_0$  在卷积核范围内的每个偏移量。

可变形卷积则在式(2)的基础上为每个点引入了一个偏移量  $\Delta P_n$ ,偏移量是由输入特征图与另一个卷积生成,通常是小数,如式(3)所示。

$$y(P_0) = \sum_{P_n \in \mathbf{R}} w(P_n)x(P_0 + P_n + \Delta P_n) \quad (3)$$

此外,为了保证算法的检测精度,在主干特征提取网络 C2f 模块中的残差连接模块 Bottleneck 中引入 DCN,再将 Bottleneck\_DCN 引入 C2f 模块中,组成 C2f\_DCN,如图 8 所示。

通过这种方式,引入 DCN 替代了原模型中的卷积算法。相比于传统卷积,DCN 能更好地提高主干网络对复杂形变物体特征轮廓的表达能力。同时,为了不影响整体网络的收敛性,仅在主干特征提取网络的 C2f 中加入这种卷积方式,而未更改后续网络部分 C2f 的卷积方式,保证了算法的稳定性。

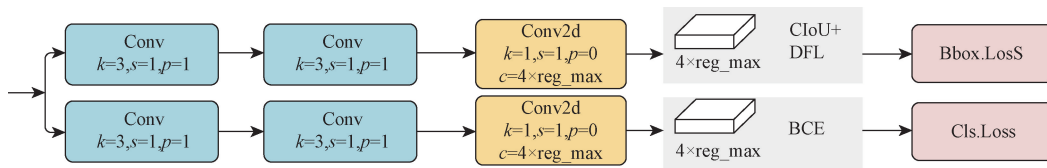


图 7 优化 IoU 算法后的 Decoupled-Head 模块

Fig. 7 Decoupled head module optimized by IoU algorithm

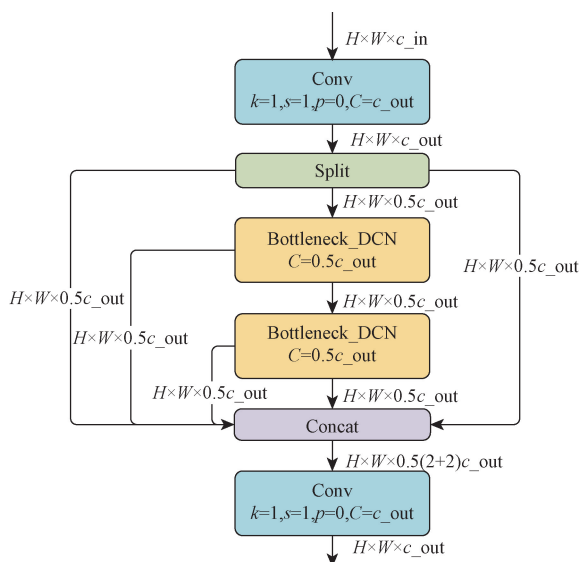


图8 基于 Bottleneck\_DCN 改进的 C2f\_DCN

Fig. 8 C2f\_DCN improved based on Bottleneck\_DCN

## 2.5 EMA 注意力机制

针对本文算法的实例分割精度问题,引入了EMA注意力机制,实现了在原有全局特征图的基础上通过期望最大化重建特征图,从而提高算法的图像分割性能。如图9所示,期望最大化注意力机制EMA,摒弃了在全图上计算注意力图的流程,转而通过期望最大化(EM)算法迭代出一组紧凑的基,在这组基上运行注意力机制,从而大大降低了复杂度。其中,E步更新注意力图,M步更新该组基。E、M交替执行,收敛之后用来重建特征图。

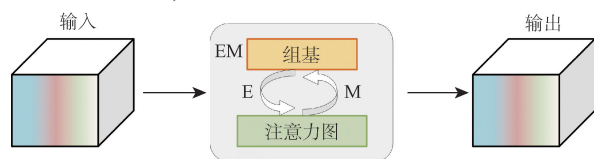


图9 EMA 注意力机制原理

Fig. 9 The principle of EMA attention mechanism

## 3 实验仿真

### 3.1 数据集与评价指标

实验使用了一个由香港中文大学提出的5038张X射线图像(共30881个垃圾检测对象)组成的真实数据集作为基准数据集。在此基础上,在训练集与测试集中筛选并补充了部分搜集到的X射线垃圾图像数据。

实验采用mAP、AP<sub>50</sub>、AP<sub>50-95</sub>作为算法性能的评价指标。其中,mAP表示所有类别的平均准确率(mean average precision);AP<sub>50</sub>表示交并比为0.5时的mAP,AP<sub>50-95</sub>表示交并比为0.5~0.95时的mAP。mAP计算公式如式(4)所示。

$$mAP = \frac{1}{C} \sum_{i=1}^C \int_0^1 P(R) dR \quad (4)$$

式(4)中:C为数据集的类别数量;P(precision)为每一类的准确率;R(recall)为召回率,即算法正确识别的目标与所有目标的比值。mAP实质上是P-R曲线下的面积。当R=0.5时,所求得的mAP即为AP<sub>50</sub>,同理可得AP<sub>50-95</sub>。

### 3.2 训练硬件环境

实验所使用电脑的操作系统为Linux,GPU为英伟达2080Ti,显存12G。同时,将BATCH SIZE设置为6,较高的学习率可能导致训练不稳定,而较低的学习率可能导致训练速度过慢或模型陷入局部最优,因此将初始学习率设置为常用的0.01。Mosaic数据增强算法默认最后10轮关闭,使得模型更好地适应原始数据分布,减少过拟合的风险。

### 3.3 对比实验分析

#### 3.3.1 模型参数及改进效果

从模型复杂度角度分析,YOLOv8n网络模型共有225个layers,共计3157200个参数。沿用基准模型YOLOv8n基本网络结构,提出了新的TLF-YOLOv8网络。改进后的网络模型共有278个layers,共计3893736个参数。模型参数量有了一定增加,网络计算复杂度由原模型的12.6GFLOPs增长为13.2GFLOPs,但检测效果提升显著,其中GFLOPs(giga floating-point operations per second),即每秒10亿(10<sup>9</sup>)次的浮点运算数。

检测效果如图10所示,可以明显看出,改进后的TLF-YOLOv8算法对处于堆叠场景下的待测物体的检测效果有了较明显的提高,同时也能较好地对待堆叠物体的边缘部分进行分割。

模型训练损失函数的收敛曲线如图11所示,损失函数在50个epochs里未有明显波动,模型损失函数数值趋于稳定,即损失函数收敛。其中,训练时损失函数在最后10个epochs时,由于关闭Mosaic数据增强算法,使得损失函数有了小幅下降,属正常现象。

#### 3.3.2 损失函数超参对比分析

改进后的交并比算法损失函数到受超参数 $\alpha$ 和 $\&$ 影响,且最优的一组超参数往往要受到数据集和网络结构的影响。为选取合适的超参数,增加超参数对比试验,得出参数组为 $\alpha = 1.8, \& = 4$ 时,模型性能达到最佳,实验结果如表1所示。

#### 3.3.3 $F_1$ 曲线评估

为了更加全面地评估模型的稳健性,实验采用 $F_1$ 作为模型性能评价指标之一。 $F_1$ 综合考虑了模型的精确度和召回率,在这两个指标的基础上进行计算,计算公式如式(5)所示。

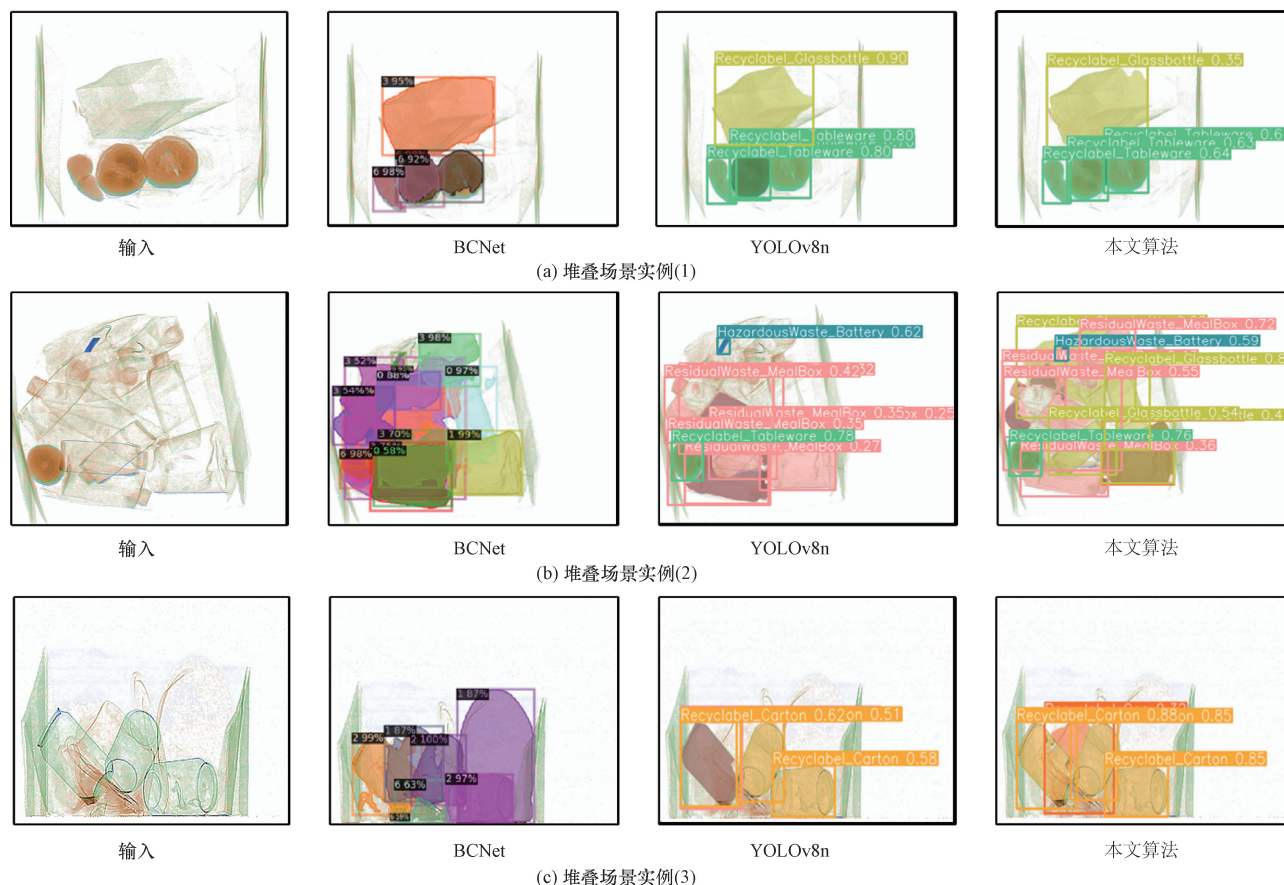


图 10 模型堆叠实例的分割效果对比

Fig. 10 Comparison of segmentation effects for stacked instances of models

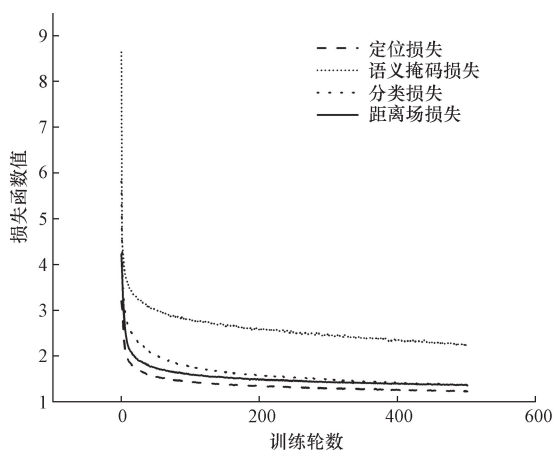


图 11 损失函数收敛曲线

Fig. 11 Convergence curve of loss function

表 1 超参数对比实验

Table 1 Comparison experiment of hyperparameters

算法	AP <sub>50</sub> (检测)	AP <sub>50</sub> (分割)
CIoU	75.7	65.4
WIoU(α=2.4, &=2)	76.5	65.5
WIoU(α=2.0, &=3)	76.8	65.6
WIoU(α=1.8, &=4)	77.1	65.6
WIoU(α=1.5, &=5)	76.9	65.3

$$F_1 = 2 \frac{PR}{P + R} \tag{5}$$

$F_1$  曲线如图 12 所示,图(a)为检测  $F_1$ ,图(b)为分割  $F_1$ 。其中每个图中的 12 条细实线为数据集对应的 12 个类别的分数,粗实线代表模型总体的  $F_1$ ,两图的模型  $F_1$  分数都达到 0.5 以上,表明模型有较好的稳健性。

### 3.3.4 主流算法对比分析

为进一步验证本文算法的有效性,使用一些相同领域类的主流算法进行对比,结果如表 2 所示。

相比于原始 YOLOv8 等主流网络算法,TLF-YOLOv8 融合网络算法的检测平均准确率、检测 AP<sub>50</sub>、分割平均准确率、分割 AP<sub>50</sub> 等指标均有了较大提高。从检测角度分析,检测平均准确率提高了 10.5%,检测 AP<sub>50</sub> 提高了 7.9%,检测 AP<sub>50-95</sub> 提高了 2.8%。从分割角度分析,分割平均准确率提高了 6.1%,分割 AP<sub>50</sub> 提高了 5.4%,分割 AP<sub>50-95</sub> 提高了 3.3%。因此,相较于原算法,改进后的算法针对堆叠实例的实例分割性能有了明显改善。

### 3.4 消融实验分析

为验证各个改进策略的效果,在 X 射线垃圾识别数据集上进行了多组消融实验,实验结果如表 3 所示。

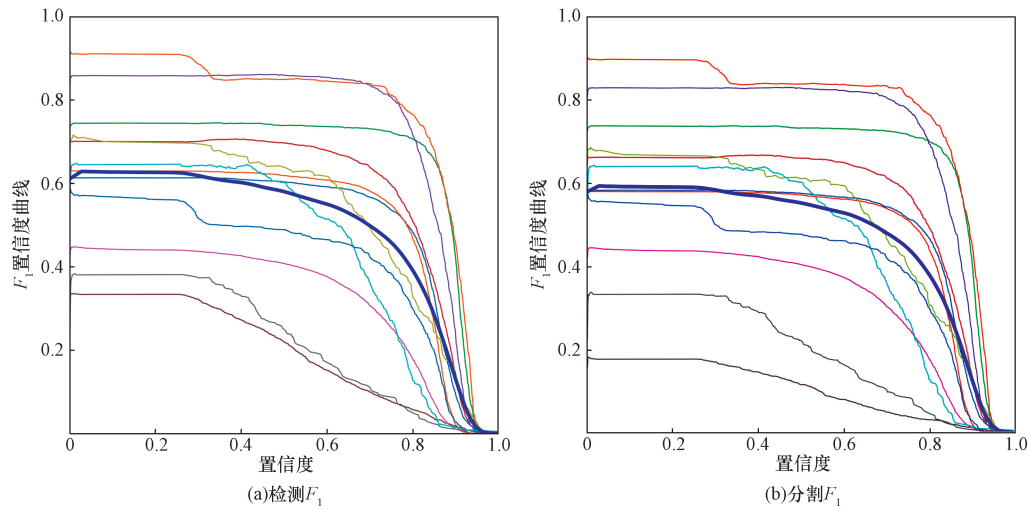
图 12  $F_1$  曲线图Fig. 12  $F_1$  curve chart

表 2 主流算法性能对比

Table 2 Performance comparison of mainstream algorithms

算法	主干网络	mAP(检测)	AP <sub>50</sub> (检测)	AP <sub>50-95</sub> (检测)	mAP(分割)	AP <sub>50</sub> (分割)	AP <sub>50-95</sub> (分割)
Mask-R-CNN	ResNet-101-FPN	45.3	63.9	50.0	42.9	59.7	47.1
Cascade R-CNN	ResNet-101-FPN	46.3	63.8	52.5	—	—	—
BCNet	ResNet-101-FPN	48.5	65.6	52.1	45.1	61.3	49.2
ETHSeg	ResNet-101-FPN	48.7	66.7	53.3	46.9	63.2	51.0
双层 DCT-Mask	ResNet-101-FPN	52.6	68.7	—	49.2	65.9	—
YOLOACT	ResNet-101-FPN	46.9	64.6	47.2	44.7	61.7	59.1
YOLOv8n(seg)	DarkNet-53	76.6	75.7	60.6	71.6	65.4	56.4
Solov2	ResNet50vd-DCN	46.7	64.3	46.9	47.5	63.2	59.5
本文算法	DarkNet-53	87.1	83.6	63.4	77.7	70.8	59.7

表 3 消融实验结果

Table 3 Results of ablation experiment

算法	mAP(检测)	AP <sub>50</sub> (检测)	mAP(分割)	AP <sub>50</sub> (分割)
YOLOv8n	76.6	75.7	71.6	65.4
+ TLF	80.7	77.5	75.8	68.8
+ EMA	82.4	78.3	75.9	70.1
+ C2f-DCN	85.1	79.7	76.7	70.5
+ Soft-NMS	85.7	82.2	76.9	70.6
+ WIoU	87.1	83.6	77.7	70.8

(1) 加入双层特征解耦策略的 TLF-YOLOv8 算法的性能明显优于原始 YOLOv8 网络,检测平均准确率提高了 4.1%,检测 AP<sub>50</sub>提高了 1.8%;分割平均准确率提高了 4.2%,分割 AP<sub>50</sub>提高了 3.4%。

(2) 使用加入 EMA 注意力机制的改进网络和未加之前的网络模型进行对比,检测平均准确率提高了 1.7%,检测 AP<sub>50</sub>提高了 0.8%;分割平均准确率未有明显变化。

(3) 引入优化后的 Bottleneck\_DCN 组成 C2f\_DCN 模块后,检测的平均准确率提高了 2.7%,检测 AP<sub>50</sub>提高了 1.4%,改进网络的检测性能明显优于

原始网络。

(4) 将 YOLOv8 中原始 NMS 优化为 Soft-NMS,优化后的网络的检测性能明显优于原始网络,其中检测的平均准确率提高了 0.6%,检测 AP<sub>50</sub>提高了 2.5%。由于 Soft-NMS 只是对候选框选取方法做了优化,保留了部分低置信度的候选框,所以改进后模型的平均准确率提升效果不明显,但其在一定程度上减弱了同类待测物体在堆叠场景下候选框易混淆的问题。同时,保留的部分候选框在后续的分类回归中得到了一定分数,从而使 AP<sub>50</sub>有了较明显的提升。

(5) 引入 WIoU 替换 YOLOv8 中的 CIoU 后,优化后的网络的检测性能明显优于原始网络,其中检测的平均准确率提高了 1.4%,检测 AP<sub>50</sub>提高了 1.4%,分割性能指标变化不明显。

## 4 结论

针对复杂堆叠垃圾场景的实例分割问题,提出了一种 TLF-YOLOv8 实例分割算法。在具有严重堆

叠的垃圾识别场景下,能够有效减弱堆叠现象对图像实例分割的影响,从而准确地对图像待测目标进行实例分割。首先,在 YOLOv8 算法中引入双层特征融合的思路,通过对数据集输入数据解耦,划分上下层,再通过建立双分支进行特征融合,从而尽可能保持参数成本,并减弱堆叠场景对图像实例分割的严重影响。其次,引入了 Soft-NMS,有效保留了同类别堆叠物体的独立候选框信息。为了提高复杂场景下的实例分割效果,在基准模型骨干中采用 Bottleneck\_DCN 结构对主干特征提取网络部分的 C2f 模块进行优化;同时引入了 EMA 注意力机制,提高了算法的图像分割性能。最后,根据场景及数据集特点,将 WIoU 损耗作为边界框回归损失,间接改善了候选框混淆问题。以 X 射线垃圾处理数据集为实验验证对象,设计了对比实验、消融实验等,从多个角度证明了所提出改进方法的可行性和有效性。同时,在实验过程中,发现该网络的改进结构存在冗余问题,即在实际模型推理中,通常只需要一个分支图像输入,则第二分支部分结构难以形成有效权重,且算法复杂度有了一定增加。

研究成果主要贡献如下。

(1) 在 YOLOv8 中融合双层特征解耦思想,对数据集部分标注特征进行分层。通过该方法,对下层待测物体由于遮挡所引起的特征缺失现象进行了一定补偿,减弱了上层遮挡物体对下层被遮挡物体实例分割的影响

(2) 针对堆叠场景下多个处于部分重叠状态的同类物体候选框易混淆的问题,引入了软阈值化 NMS 和 WIoU 算法,即在原有纵横比-交并比算法 CIoU 的基础上,引入了基于候选框距离的惩罚因子,改善了候选框易混淆的问题。

(3) 基于 YOLOv8n 基准模型,引入了注意力机制,优化了主干网络部分的 C2f 模块,有效提高了模型的检测性能。

#### 参 考 文 献

- [1] 胡滨, 许向勇, 邵建强, 等. 我国生活垃圾分类现状及改进对策研究[J]. 中国资源综合利用, 2023, 41(10): 120-122.  
Hu Bin, Xu Xiangyong, Shao Jianqiang, et al. Research on the current situation and improvement strategies of domestic waste classification in China [J]. China Comprehensive Utilization of Resources, 2023, 41 (10): 120-122.
- [2] 李宇聪, 潘越, 王耀, 等. 中国厨余垃圾处理现状及展望[J]. 中国资源综合利用, 2023, 41(6): 119-121.  
Li Yucong, Pan Yue, Wang Yao, et al. The current situation and prospects of kitchen waste treatment in China [J]. China Comprehensive Utilization of Resources, 2023, 41(6): 119-121.
- [3] 成浪, 敬超, 陈文鹏. 基于神经网络架构搜索的 X 射线图像违禁品检测算法[J]. 科学技术与工程, 2024, 24(2): 665-675.  
Cheng Lang, Jing Chao, Chen Wenpeng, et al. Detection algorithm of contraband in X-ray image based on neural network architecture search [J]. Science Technology and Engineering, 2024, 24 (2): 665-675.
- [4] 邓健志, 支佩佩, 张峰铭, 等. 结合拆分注意力特征融合的病理科图像分割网络[J]. 科学技术与工程, 2023, 23(7): 2922-2931.  
Deng Jianzhi, Zhi Peipei, Zhang Fengming, et al. Pathological image segmentation network combined with split attention feature fusion [J]. Science Technology and Engineering, 2023, 23 (7): 2922-2931.
- [5] 包锐. X 射线透射方法在矿产品检测中的应用[J]. 现代工业经济和信息化, 2023, 13(1): 132-133, 136.  
Bao Rui. The application of X-ray transmission method in mineral product detection [J]. Modern Industrial Economy and Information Technology, 2023, 13 (1): 132-133, 136.
- [6] 李启明, 阙祖航. 基于改进 YOLOv5 的 X 射线图像危险品检测[J]. 科学技术与工程, 2023, 23(4): 1598-1606.  
Li Qiming, Que Zuhang. X-ray image dangerous goods detection based on improved YOLOv5 [J]. Science Technology and Engineering, 2023, 23(4): 1598-1606.
- [7] 高新阳, 魏晟, 温志庆, 等. 改进 YOLOv5 轻量级网络的柑橘检测方法[J]. 计算机工程与应用, 2023, 59(11): 212-221.  
Gao Xinyang, Wei Sheng, Wen Zhiqing, et al. Improved citrus detection method for YOLOv5 lightweight network [J]. Computer Engineering and Applications, 2023, 59(11): 212-221.
- [8] 吕岳, 周浙泉, 吕淑静. 基于双层解耦策略和注意力机制的遮挡目标分割[J]. 电子与信息学报, 2023, 45(1): 335-343.  
Lü Yue, Zhou Zhequan, Lü Shujing. Occlusive target segmentation based on double-layer decoupling strategy and attention mechanism [J]. Journal of Electronics and Information Science, 2023, 45 (1): 335-343.
- [9] He K M, Gkioxari G, Piotr D, et al. Mask-RCNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2961-2969.
- [10] Kirillov A, Wu Y, He K, et al. Pointrend: image segmentation as rendering [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9799-9808.
- [11] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback [J]. ArXiv Preprint ArXiv, 2012: 1205.2618.
- [12] Zhang G, Lu X, Tan J, et al. Refinemask: towards high-quality instance segmentation with fine-grained features [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 6861-6869.
- [13] Tian Z, Shen C, Chen H, et al. FCOS: a simple and strong anchor-free object detector [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.
- [14] Wang X, Zhang R, Kong T, et al. Solov2: dynamic and fast instance segmentation [J]. Advances in Neural Information Processing Systems, 2020, 33: 17721-17732.
- [15] Bolya D, Zhou C, Xiao F, et al. Yolact: real-time instance segmentation [C]//Proceedings of the IEEE/CVF International Con-

- ference on Computer Vision. Seoul: IEEE, 2019: 9157-9166.
- [16] Ling T Q, Zhang Y X, Wang K L. ETHSeg: an amodel instance segmentation network and a realworld dataset for X-ray waste inspection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 2283-2292.
- [17] Ke L, Tai Y W, Tang C K, et al. Deep occlusion-aware instance segmentation with overlapping bilayers [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 4019-4028.
- [18] 倪汉杰, 初秀民, 张彬鹏, 等. 基于改进 Cascade-RCNN 的内河航标检测算法[J]. 中国航海, 2022, 45(3): 99-106.  
Ni Hanjie, Chu Xiumin, Zhang Binpeng, etc. An improved Cascade RCNN based inland navigation mark detection algorithm [J]. China Navigation, 2022, 45(3): 99-106.
- [19] 赵旭, 王宏. 基于 MaskRCNN 改进的全自动脑肿瘤分割[J]. 首都师范大学学报: 自然科学版, 2021, 42(6): 1-7.  
Zhao Xu, Wang Hong. Improved fully automatic brain tumor segmentation based on MaskRCNN [J]. Journal of Capital Normal University: Natural Science Edition, 2021, 42(6): 1-7.
- [20] 游江川. 基于改进 Mask-RCNN 的路面裂缝检测[J]. 电视技术, 2022, 46(6): 7-9, 19.  
You Jiangchuan. Road crack detection based on improved Mask RCNN [J]. Television Technology, 2022, 46(6): 7-9, 19.