

# YOLO-AF: 结合注意力机制和焦点卷积的垃圾堆检测模型

宋琦<sup>1</sup>, 苑春苗<sup>1</sup>, 曹立昆<sup>1</sup>, 黄臣臣<sup>1</sup>, 杨清永<sup>2\*</sup>

<sup>1</sup>天津工业大学软件学院, 天津

<sup>2</sup>天津中德应用技术大学软件与通信学院, 天津

收稿日期: 2024年1月26日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

## 摘要

随着城镇化进程的不断加速, 垃圾堆成为城市管理的重要问题, 智能化的垃圾堆检测算法, 可以有效提高环境管理的水平和效率。在YOLOv7检测算法的基础上, 通过引入注意力机制、改进MPConv模块以及优化损失函数对模型进行改进, 提出了YOLO-AF垃圾堆检测模型。引入注意力机制和使用焦点模块增强特征信息解决垃圾堆和背景相似导致检测精度低问题, 综合利用二元交叉熵损失函数和Kullback-Leibler散度损失函数解决垃圾分类不平衡问题。实验结果表明, 在多类型垃圾堆检测任务中, YOLO-AF垃圾堆检测模型相较于YOLOv7表现出更高的精度、召回率和mAP, 性能得到显著提升, 训练300轮次的模型mAP值为92.45%。在同等条件下, 检测效果优于当前主流的目标检测算法, 可以较好地满足垃圾堆实时工业检测的需求。

## 关键词

垃圾堆检测, YOLOv7算法, 注意力机制, 焦点模块, 特征提取

# YOLO-AF: A Garbage Heap Detection Model Combining Attention Mechanism and Focus Convolution

Qi Song<sup>1</sup>, Chunmiao Yuan<sup>1</sup>, Likun Cao<sup>1</sup>, Chenchen Huang<sup>1</sup>, Qingyong Yang<sup>2\*</sup>

<sup>1</sup>School of Software, Tiangong University, Tianjin

<sup>2</sup>School of Software and Communication, Tianjin Sino-German University of Applied Sciences, Tianjin

Received: Jan. 26<sup>th</sup>, 2024; accepted: Feb. 22<sup>nd</sup>, 2024; published: Feb. 29<sup>th</sup>, 2024

\*通讯作者。

文章引用: 宋琦, 苑春苗, 曹立昆, 黄臣臣, 杨清永. YOLO-AF: 结合注意力机制和焦点卷积的垃圾堆检测模型[J]. 计算机科学与应用, 2024, 14(2): 468-479. DOI: 10.12677/csa.2024.142047

## Abstract

With the continuous acceleration of urbanization, garbage heaps have become an important problem in urban management, and intelligent garbage heap detection algorithms can effectively improve the level and efficiency of environmental management. On the basis of the YOLOv7 detection algorithm, the YOLO-AF garbage heap detection model was proposed by introducing the attention mechanism, improving the MPConv module and optimizing the loss function. The attention mechanism is introduced and the focus module is used to enhance the feature information to solve the problem of low detection accuracy caused by the similarity of garbage and background. The binary cross entropy loss function and Kullback-Leibler divergence loss function are comprehensively used to solve the unbalanced problem of garbage classification. Experimental results show that the YOLO-AF garbage heap detection model shows higher accuracy, recall rate and mAP than YOLOv7 in multi-type garbage heap detection tasks, and the performance is significantly improved, and the mAP value of the model after 300 rounds of training is 92.45%. Under the same conditions, the detection effect is better than the current mainstream object detection algorithm, which can better meet the needs of real-time industrial detection of garbage heaps.

## Keywords

Garbage Heap Detection, YOLOv7 Algorithm, Attention Mechanism, Focus Module, Feature Extraction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着城镇化进程不断推进以及人口的集聚和生活水平的提高,城市中产生的垃圾量急剧增加,城市垃圾堆积不仅影响城市景观和市容,更给环境卫生和居民生活带来了严重问题[1]。传统的人工巡查方式效率低下,成本高昂,而且容易忽略一些隐蔽的垃圾积累点,难以应对这一快速增长的垃圾产生情况。为提高对非法倾倒垃圾监测效率,减少人工监督或者完全取代人工监督,将深度学习技术中的目标检测(Object Detection)算法应用在垃圾堆的识别与检测中,可以实现垃圾违规倾倒自动化检测[2],及时发现并处理垃圾积累点,避免其进一步扩大,有力地减轻了城市环境的压力,有助于提高城市管理的精细化水平,为城市规划和决策提供可靠的数据支持[3]。通过智能识别垃圾堆的分布情况,城市管理者可以更有针对性地制定垃圾清理和治理策略,优化资源配置,提高管理效能。

由于垃圾堆的多样性和环境的复杂性问题,经常导致检测的精度准确率不高,因此我们提出了YOLO-AF模型来实现更加准确的垃圾堆检测。该模型以YOLOv7 [4]为基础模型,通过引入注意力机制(Attention Mechanism)和使用焦点模块(Focus Module)来解决一些垃圾堆和背景相似问题,通过对损失函数优化解决垃圾分类不平衡问题,提高了对垃圾堆检测与识别的精准性和效率。

## 2. 相关工作

### 2.1. 基于深度学习的垃圾检测

在垃圾检测领域,Mittal G等人[5]介绍了一款名为SpotGarbage的Android应用,用于检测和定位用

户点击的地理标记图像中的垃圾，他们提出了一种名为 GarbNet 的 CNN 架构，该架构在 GINI 数据集上进行训练，用于实时垃圾检测。Camero 等人[6]探索了一种深度神经演化技术，构建并训练了一个循环神经网络(RNN)。RNN 预测垃圾容器的填充水平，以跟踪实际城镇产生的垃圾量。Ruiz 等人[7]通过在 Trashnet 数据集上训练 VGG、Inception 和 ResNet 等几种标准 CNN 架构进行比较，以寻找对多种垃圾物品进行分类的最佳模型。结果表明 Inception-ResNet 的组合架构取得了 88.6%的准确性，取得了最佳结果。Wang 等人[1]基于深度学习的架构，提出基于 Faster-RCNN 以及 ResNet 的垃圾检测算法，用于实时检测城市中的垃圾，还采用了数据融合和增强技术，提高了精度和准确性。Anjum 等人[8]研究了一种名为 GarbConvNet (GCN)的深度 CNN 架构，该架构在 Garbage Image Dataset (GIDset)上进行训练。GarbConvNet 模型以令人期待的准确性检测、定位和分类垃圾和非垃圾图像。VJ A 等人[9]介绍了一种名为 Automatic Waste Segregator (AWS)的系统，旨在将垃圾分为三大类别：金属、有机和塑料，以实现固体废物的高效管理。

## 2.2. 基于物联网 IoT 技术的垃圾检测

超声波传感器和 GSM 模块也被用于跟踪整个垃圾收集过程。Singh A 等人[10]使用一种基于物联网 (IoT)的技术，通过红外传感器从垃圾桶获取实时数据。他们使用 RaspberryPi2 开发板，当垃圾桶满时，向垃圾管理员发送通知，并建议垃圾收集的最佳路径。Mahankali S 等人[11]提出通过使用 Haar 级联分类器、轮廓和 Blob 创建以检测移动车辆的垃圾处理。基于车辆的车牌，识别车主的身份，使用 OpenALPR 库进行垃圾识别。

## 3. 方法

### 3.1. YOLOv7

YOLO [12]具有检测速度快、精度高的特点。将目标检测任务作为目标区域预测和类别预测的回归问题。相对于其他目标检测和识别方法，如 Faster R-CNN [13]，YOLO 的目标区域预测和目标类别预测被集成到单一的神经网络模型中，更适合在实际应用环境中实现高准确度的快速目标检测和识别。YOLOv7 采用了扩展高效远程注意力网络(E-ELAN)、基于级联模型的模型缩放和卷积重参数化等策略，在检测效率和准确性之间取得了很好的平衡。

YOLOv7 由四个模块组成，分别是输入、骨干、头部和预测。输入模块将输入图像缩放到统一的像素大小，以满足骨干网络的输入大小要求。骨干模块由若干个 BConv 卷积层、E-ELAN 卷积层和 MPConv 卷积层组成，其中 BConv 由卷积层、批量归一化(BN)层和 LeakyReLU 激活函数组成，用于提取不同尺度的图像特征。E-ELAN 卷积层保持了原有的 ELAN 设计架构，通过引导不同特征组的计算块来学习更多样化的特征，在不破坏原有梯度路径的情况下提高网络的学习能力。MPConv 卷积层在 BConv 层上增加了一个 Maxpool 层，形成两个分支，以提高网络的特征提取能力。头部模块采用传统的 PAFPN 结构，引入自下而上的路径，便于将底层信息向高层传递，从而实现不同层次特征的高效融合。预测模块通过 REPCov 结构调整 PAFPN 输出的不同尺度的 P3、P4 和 P5 特征的图像通道数，最后通过卷积进行预测。

垃圾堆检测模型需要同时满足实时性和准确性的要求。鉴于检测精度和速度之间的良好平衡，我们选择 YOLOv7 作为基线模型，针对目前垃圾检测存在的问题进行改进，以实现实时准确的垃圾检测性能。

### 3.2. 注意力机制

在复杂的现场环境中检测垃圾需要考虑许多因素，可变照明条件、遮挡、垃圾和背景颜色相似等问题，都将导致垃圾堆检测不准确。注意力机制是指通过对网络输入部分赋予不同的权重，使模型忽略不

相关的信息，专注于重要信息，可以有效提高模型在复杂背景下的特征提取能力。因此，我们将注意力机制模块 SE 嵌入到 YOLOv7 的网络结构中来学习通道间的重要性权重，以增强对特定特征的关注度，SE 模块如图 1 所示。

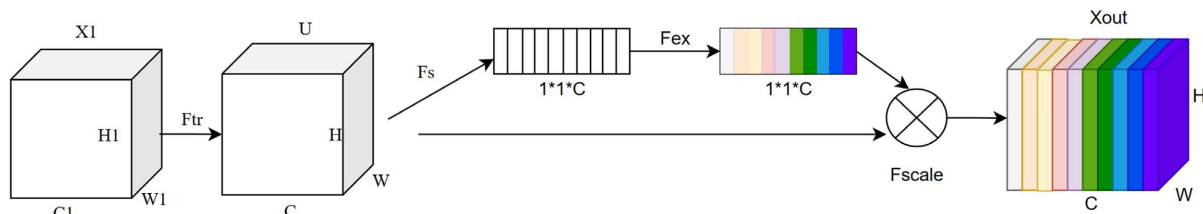


Figure 1. SE module diagram

图 1. SE 模块图

SE 模块主要由 Squeeze 和 Excitation 两个模块组成。Squeeze 步骤通过全局平均池化将每个通道的特征进行降维，得到一个全局垃圾堆特征向量，表示了每个通道的整体贡献。Excitation 模块通过学习一个激励向量，可以调整每个通道的权重，从而强化对不同通道对垃圾堆重要特征的关注度。这个激励向量的学习是通过两个全连接层来实现的，这些层可以自适应地学习权重，以适应不同任务的需要。最后，SE 模块将每个通道的激励与原始垃圾特征相乘，以生成具有注意力机制的垃圾堆特征图，从而突出垃圾堆的关键特征，增强提取垃圾堆关键特征的能力，减少了复杂背景对垃圾堆检测的干扰。

在 Squeeze 模块[14]中，通过全局平均池化来压缩每个通道的垃圾堆特征图，将其降维成一个固定大小的特征向量，表示每个通道的全局重要性，即该通道在整个垃圾堆特征图中的贡献程度。Squeeze 步骤的如式(1)：

$$z_i = \frac{1}{H \times W} \sum_{j=1}^H \sum_{k=1}^W x_{ijk} \quad (1)$$

其中， $z_i$  为  $i$  通道压缩后的特征值， $H$  和  $W$  分别为垃圾堆特征图的高度和宽度， $x_{ijk}$  为  $i$  通道在垃圾堆特征图中  $(j, k)$  位置的特征值。

在 Excitation 模块[15]中，SE 模块学习一个权重向量，用于调整每个通道的特征图。这个权重向量被视为每个通道的“激励”，它决定了每个通道在最终垃圾堆特征图中的重要性。这个激励向量通过一个全连接层来学习，然后通过 Sigmoid 函数进行归一化，以确保权重在 0 到 1 之间。Excitation 步骤的如式(2)所示：

$$s_i = \sigma(W_2 \delta(W_1 z_i)) \quad (2)$$

式中， $s_i$  为第  $i$  个通道的激励值， $W_1$  和  $W_2$  为两个全连接层的权重矩阵， $\sigma$  为 Sigmoid 函数和  $\delta$  为 ReLU 激活函数。

最后，SE 模块将每个通道的激励与原始垃圾堆特征映射相乘，生成具有注意机制的特征映射。如等式(3)所示：

$$y_i = s_i \cdot x_i \quad (3)$$

其中， $y_i$  是最终输出垃圾堆特征图中的第  $i$  通道， $s_i$  是该通道的激活值， $x_i$  是原始垃圾堆特征图中的第  $i$  通道。

我们将 SE 模块添加到 YOLOv7 的骨干网络(如 ResNet50)的末端，以增强图像特征的代表能力。SE 模块可以自动识别并增强图像中垃圾堆的重要特征，同时抑制噪声和不相关信息，从而提高垃圾堆检测

准确性。

### 3.3. MPCConv 模块改进

在进行实际垃圾检测中，由于会出现垃圾堆的特征不明显，干扰检测效果。为了让模型更全面地捕捉目标的特征，提高目标检测的性能，更好地适应垃圾堆检测任务，我们通过在 YOLOv7 的 MPCConv 模块中集成多个卷积核，从而扩展感受野的范围，使其能够有效地捕捉不同尺度下的特征，多尺度信息融合后有助于提高算法对垃圾堆在不同细粒度上的表示能力，增强了检测的全局感知能力。YOLOv7 中 MPCConv 模块的两个分支的下分支使用  $3 \times 3$  卷积内核进行卷积操作[16]。如图 2 所示，当步长为 2 时，一些特征信息可能会丢失，网络中可能会出现低效的特征学习。

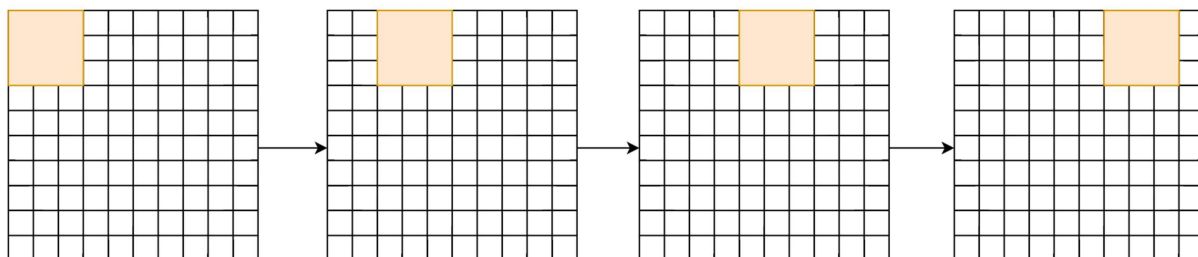


Figure 2. MPCConv convolution process

图 2. MPCConv 卷积过程

受 YOLOv5 中焦点模块的启发，我们用 Focus 模块[17]来代替 MPCConv 下分支中的  $3 \times 3$  卷积内核，如图 3 所示。Focus 模块将对特征图进行切片操作，在一张特征图片中每隔一个像素取一个值，类似于邻近下采样，这样就得到了四张特征图片，四张图片互补，并且没有信息丢失，使得特征输出通道扩充了 4 倍，之后将得到的新图片再经过卷积操作，最终得到了没有信息丢失的二倍下采样特征图。通过 Focus 模块使得垃圾堆的特征图增加，增强了复杂背景下的垃圾检测性能，改进后的 MPCConv 模块我们称之为 F-MPCConv。

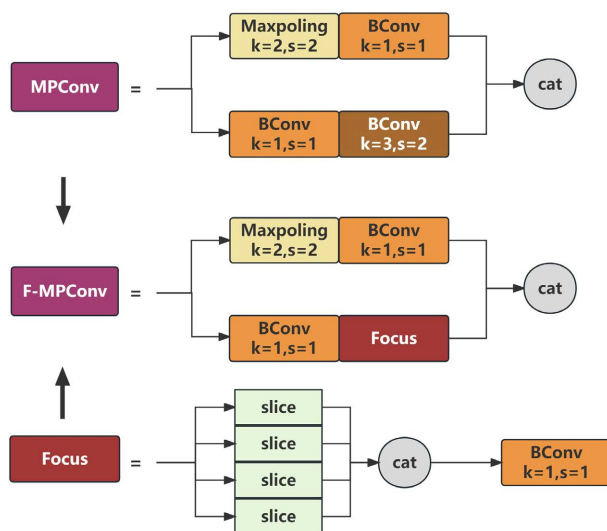


Figure 3. F-MPCConv module

图 3. F-MPCConv 模块

综上所述,改进后的YOLOv7模型如图4所示,我们称之为YOLO-AF。

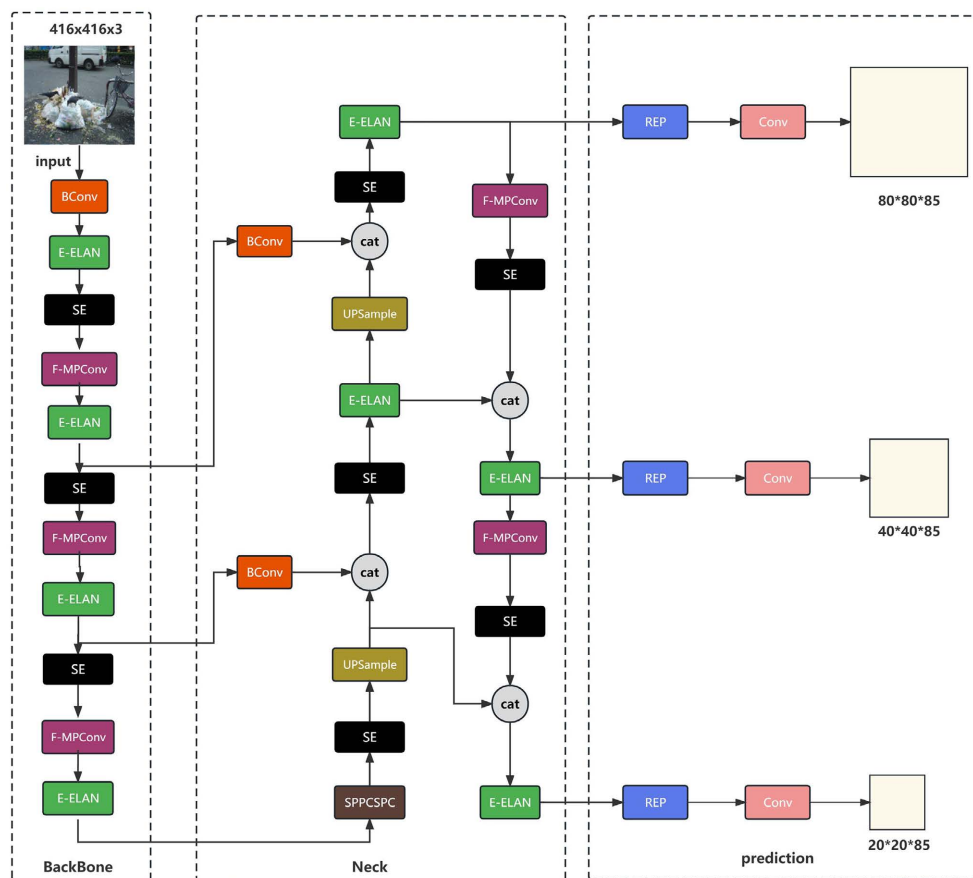


Figure 4. Schematic diagram of the YOLO-AF model

图4. YOLO-AF模型示意图

### 3.4. 损失函数优化

对于深度学习模型,通过损失反向传播[18]来训练模型中的神经网络权值,度量了模型的预测与实际标签之间的差异,损失函数的选择直接影响了检测结果。传统的损失函数,如二元交叉熵损失函数[19],常用于二分类任务的损失函数,它衡量了模型的输出与真实标签之间的差异。对于每个像素,二元交叉熵损失函数如式(4)所示:

$$L_{BCE}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)) \quad (4)$$

其中  $N$  为像素总数,  $y_i$  为真实标签,  $\hat{y}_i$  为模型的预测。虽然二值交叉熵损失函数在大多数检测任务中表现良好,但在垃圾分类不平衡方面可能存在一些缺点。

Kullback-Leibler 散度损失函数[20]是用于测量两个概率分布之间差异的一种损失函数。在目标检测中可以将模型的输出看作是一个概率分布,表示每个像素属于不同的类别。而真实标签也可以看作是一个概率分布,其中只有一个类别的概率为1,其余都为0。Kullback-Leibler 散度损失函数如式(5)所示:

$$L_{KLdiv}(p, q) = \sum_i p_i \cdot \log\left(\frac{p_i}{q_i}\right) \quad (5)$$

Kullback-Leibler 散度损失函数的优点是能够处理类别不平衡问题，因为对于每个像素都考虑了真实标签和模型输出之间的差异。为了解决垃圾分类不平衡问题，我们综合利用了二元交叉熵损失函数和 Kullback-Leibler 散度损失函数的优点，将它们权重相加，并为每个损失函数分配一个权重，选择将权重均分为 0.5，以平等地考虑它们的贡献。我们提出的垃圾检测模型的损失函数如式(6)所示，我们将其称之为 *rdrop loss*。

$$L(y, \hat{y}, p, q) = 0.5 \cdot L_{BCE}(y, \hat{y}) + 0.5 \cdot L_{KLdiv}(p, q) \quad (6)$$

通过将二值交叉熵损失函数的像素级预测与 Kullback-Leibler 散度损失函数的类不平衡处理能力相结合，我们的模型可以解决类不平衡问题，得到更准确的垃圾堆检测结果。

## 4. 实验

### 4.1. 数据集

论为了涵盖垃圾堆在不同环境和场景中的多样性，我们在城市、郊区和乡村等不同地域进行了广泛的数据采集，并标注了相应的数据集。通过高分辨率的数字相机和无人机等设备，获取了包括公共场所、居民区等不同场景下的垃圾堆图像。这些图像覆盖了垃圾的各种类型、规模和形状，确保了数据集的多样性。我们选取不同环境和不同时间的垃圾堆图片作为研究对象，图像分辨率为  $640 \times 640$  像素。为了得到多样性的图像数据，采集的图像包括五个场景：生活垃圾、建筑垃圾、白天、夜晚和遮挡，垃圾堆数据图像如图 5 所示。



**Figure 5.** Examples of garbage heap images: (a) Domestic rubbish at night; (b) Domestic rubbish during the day; (c) Obscured domestic rubbish; (d) Building rubbish during the day

**图 5.** 垃圾堆图片示例：(a) 夜晚的生活垃圾；(b) 白天的生活垃圾；(c) 遮挡的生活垃圾；(d) 白天的建筑垃圾

为了增加数据集的多样性和训练算法的鲁棒性，采用随机旋转、镜像翻转、亮度调整等方式，生成更多样化的训练样本，提高算法对于光照、角度和噪声等变化的适应能力，使得垃圾堆检测在实际场景中更为可靠。整个数据集共 3370 张图片，划分为训练集、验证集和测试集，三者的比例为 8:1:1，数据集结构如表 1 所示。

**Table 1.** Garbage heap dataset information  
**表 1.** 垃圾堆数据集信息

数据集	图像分辨率(像素)	图像数量
训练集	640	2696
验证集	640	337
测试集	640	337

## 4.2. 实验环境及评价指标

实验配置如下：操作系统为 Windows 10 的 64 位操作系统的专业版；深度学习框架为 Pytorch1.7；CUDA 版本为 11.1；编程语言为 Python 3.9。训练时输入图片的大小为  $640 \times 640$ 。为了减少训练过程时间，获得更好的训练效果，将学习率设置为 0.001，权重衰减系数设置为 0.0005，使用 Adam 优化器，训练批次设置为 16，迭代周期设置为 300 个 epoch。

为了全面、客观地评估改进 YOLOv7 算法的性能，采用精度(P)、召回率(R)、F1 得分(F1)、平均精度(AP)和 mAP 对检测性能进行评价，涵盖了模型在不同方面的表现[21]。具体公式如下(7)~(11)。

### 1) 精度(Precision)

精度是指模型在所有被预测为正类别的样本中，实际为正类别的比例更高的精度意味着算法更少地将负样本错误地分类为正样本，具有更好的准确性。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

### 2) 召回率(Recall)

召回率衡量了模型能够正确检测出所有正样本的能力。高召回率表示算法对于真实存在的垃圾堆有更好的捕捉能力，避免了漏检的情况。

$$R = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

### 3) F1 得分(F1 Score)

F1 得分是精度和召回率的调和平均，综合考虑改进 YOLOv7 模型的准确性和全面性。追求更平衡的 F1 得分，以确保算法在准确性和召回率之间取得良好的平衡。

$$F1 = 2 \frac{PR}{P + R} \times 100\% \quad (9)$$

### 4) 平均精度(Average Precision, AP)

平均精度综合模型在不同类别上的精度表现，考量每个类别的精度，从而更全面地评估了算法在各类别上的性能。

$$AP = \int_0^1 P(R) dR \times 100\% \quad (10)$$

### 5) 平均精度均值(mean Average Precision, mAP)

mAP 是 AP 的平均值，提供对整个数据集的综合评估。mAP 综合考虑了模型在不同类别上的性能，为算法的整体效果提供了直观的度量。

$$mAP = \frac{1}{n} \sum AP_i \times 100\% \quad (11)$$



其中, TP 是被检测为垃圾堆的垃圾堆, FP 是被检测为垃圾堆的背景, FN 是未被检测的垃圾堆。P 和 R 可以从 TP、FP 和 FN 计算, F1 可以从 P 和 R 计算。

### 4.3. 与其他算法比较

为了证明本文所提模型的优越性, 利用控制变量原理进行对比实验, 将所提模型与当前主流的目标检测器在相同实验环境和相同数据集上的检测性能进行了对比, 在我们的数据集训练 200 个 epoch 后, 训练结果显示性能趋于平稳。因此, 300 个 epoch 足以用于数据集的训练过程。表 2 显示了不同模型在 P、R、F1、mAP 和平均检测时间的训练结果, 最佳结果都以粗体突出显示。

**Table 2.** Comparison of the detection effect of YOLO-AF with other models

**表 2.** YOLO-AF 与其他模型的检测效果比较

Model	P (%)	R (%)	F1 (%)	mAP (%)	Average Detection Time (ms)
YOLOv3	85.34	83.22	84.26	86.02	23.82
YOLOv4-tiny	84.13	82.64	83.37	85.60	22.87
YOLOv4	84.87	82.75	83.80	85.44	23.04
YOLOv5s	86.45	83.39	84.89	87.91	18.91
YOLOv6	86.10	84.55	85.32	86.73	20.48
YOLOv7-tiny	88.24	83.57	85.84	88.32	<b>18.54</b>
YOLOv7	87.02	85.64	86.32	89.57	19.76
Ours	<b>91.27</b>	<b>88.47</b>	<b>89.85</b>	<b>92.45</b>	21.95

对比结果表明, 我们的模型不仅保证了较高的检测精度, 而且满足了实时检测速度要求。YOLO-AF 在精度(P)、召回率(R)、F1 得分(F1)、以及平均精度(mAP)上分别为 91.27%、88.47%、89.85%和 92.45%, 都明显超过原有的算法。特别是在 mAP 方面, 达到了 92.45%, 有着显著提升。YOLO-AF 在垃圾堆检测方面优于表 2 中的其他模型, 在检测速度和检测精度之间取得了良好的平衡。8 个模型的检测时间均小于 25 ms, YOLO-AF 虽然平均检测时间略高于 YOLOv7-tiny (21.95 ms 对比 18.54 ms), 但仍满足实时性要求, 考虑到性能的全面提升, 这种增加是可以接受的。

### 4.4. 消融实验

为了验证我们的 YOLO-AF 模型对垃圾堆检测的优化效果, 进行了多个消融实验。每个实验都使用相同的训练参数和不同的模型结构进行, 分别为: 实验 1: 嵌入 SE 注意力机制模块; 实验 2: 替换损失函数为 rdrop loss; 实验 3: 添加 F-MPCConv 模块; 实验 4: 嵌入 SE 注意力机制模块及添加 F-MPCConv 模块; 实验 5: 嵌入 SE 注意力机制模块及添加 F-MPCConv 模块和替换损失函数 rdrop loss。消融试验的结果如表 3 所示。

**Table 3.** Garbage heap dataset information

**表 3.** 消融实验结果

实验序号	方法	mAP (%)	R (%)	Average Detection Time (ms)
0	Yolo v7	89.57	85.64	19.76
1	+SE	91.87	84.65	23.37
2	+rdrop loss	91.51	83.79	21.34

续表

3	+F-MPConv	92.04	87.56	22.45
4	+SE+ F-MPConv	92.26	86.46	22.63
5	+SE+rdrop loss+ F-MPConv	92.45	88.47	21.95

由表 3 所示, 通过引入 SE 注意力机制, 整体性能有所提升, mAP 提高至 91.87%, 通过对通道进行加权, 使网络更加关注重要的特征, 有助于提升对垃圾堆的细致表达。但召回率略微下降至 84.65%, 而平均检测时间增加至 23.37 ms。

通过优化损失函数后, mAP 进一步提高至 91.51%, 但召回率略微降低至 83.79%, 而平均检测时间略有增加至 21.34 ms。这是由于优化后的损失函数更加关注难以训练的样本, 导致对于一些正样本的捕捉不足。增加的平均检测时间表明, 引入更复杂的损失函数会带来一些计算负担。

通过使用 F-MPConv 模块, 模型性能得到显著提升, mAP 和召回率分别达到 92.04% 和 87.56%, 平均检测时间为 22.45 ms。通过融合多尺度信息和增强非线性特性, 使得模型更好地捕捉垃圾堆的复杂特征, 使得性能明显得到提升。

通过将 SE 和 F-MPConv 模块结合使用后, mAP 持续提升至 92.26%, Recall 稍有下降至 86.46%, 平均检测时间为 22.63 ms。SE 注意力机制对于提高精度有一定帮助, 但在 Recall 和检测速度上有轻微的影响。

同时引入 SE、优化的损失函数和 F-MPConv 模块后, mAP 保持在 92.45%, 而 Recall 显著提高至 88.47%, 平均检测时间为 21.95 ms。可以看出, 结合 SE、优化的损失函数和 F-MPConv 模块, 整体性能相较于基准模型得到了提升, 综合考虑了精度、召回率和检测速度的权衡。

#### 4.5. 检测效果分析

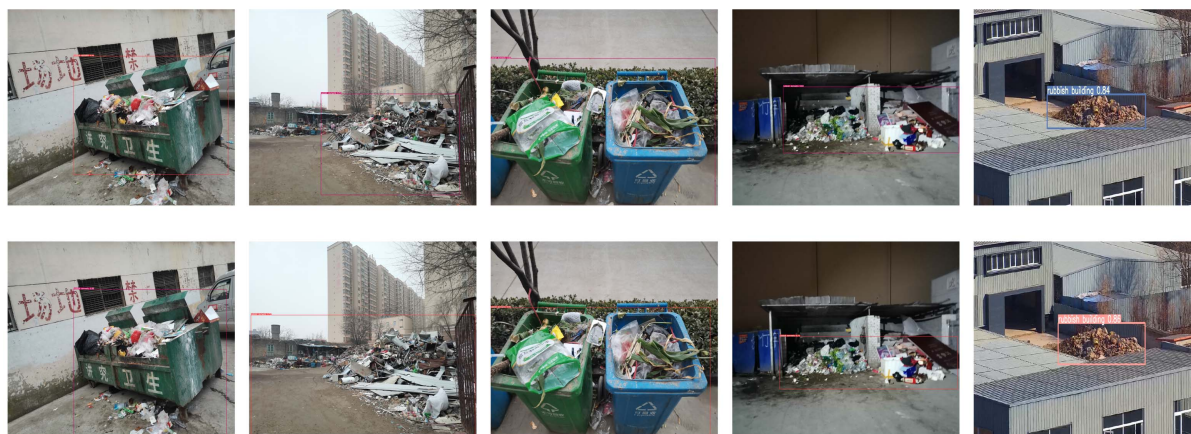
为了探究 YOLO-AF 模型与原 YOLOv7 模型在不同场景下的垃圾堆检测性能的差异, 从测试集中选取 50 张不同场景的垃圾堆图片进行测试。测试结果如表 4 所示。从表 4 可以看出, 待测试的垃圾堆数量为 61 个生活垃圾, 56 个建筑垃圾, 55 个白天的垃圾堆, 58 个夜晚的垃圾堆以及 60 个被遮挡的垃圾堆。将 YOLO-AF 模型输出的检测结果与图片的标注数据进行对比, 可以看出, YOLO-AF 模型对五种场景的垃圾堆的检测结果优于原始模型。

**Table 4.** Comparison results of the YOLO-AF model and YOLOv7 in the detection of garbage heaps under different types  
**表 4.** YOLO-AF 模型与 YOLOv7 在不同类型下检测垃圾堆的对比结果

场景	模型	待测试的垃圾堆数量	识别正确		识别错误	
			数量	正确率(%)	数量	错误率(%)
生活垃圾	YOLOv7	61	53	86.89	3	4.92
	Ours	61	57	93.44	2	3.28
建筑垃圾	YOLOv7	56	49	87.50	2	3.57
	Ours	56	51	91.07	1	1.79
白天	YOLOv7	58	51	87.93	3	5.17
	Ours	58	53	91.38	1	1.72
夜晚	YOLOv7	56	48	85.71	4	7.14
	Ours	56	52	92.86	1	1.79
遮挡	YOLOv7	60	50	83.33	5	8.33
	Ours	60	54	90.00	2	3.33

对于不同垃圾类别，YOLO-AF 模型可以准确区分生活垃圾和建筑垃圾，识别的正确率分别达到了 93.44% 和 91.07%，相较于 YOLOv7 的 86.89% 和 87.50% 有明显提升。针对不同光线下的场景，在光线充足的白天，YOLO-AF 模型识别的正确率为 91.38%，而 YOLOv7 为 87.93%；而在光线不足的夜晚，YOLO-AF 模型识别的正确率为 92.86%，而 YOLOv7 仅为 85.71%。在遮挡情况下的检测任务中，YOLO-AF 模型同样表现出卓越的性能，识别正确率达到了 90.00%，相对于 YOLOv7 的 83.33% 有显著提升。

这些结果不仅显示改进后的 YOLOv7 模型在不同垃圾类别下的更高准确性，而且在识别错误率方面也相对较低。总体而言，YOLO-AF 模型相较于 YOLOv7 在实际垃圾堆检测中展现了更为可靠和准确的性能，为其在实际应用场景中的广泛推广提供了有力的支持，因此，YOLO-AF 更适合复杂环境中的垃圾堆检测。然而，在某些情况下，我们的算法无法成功检测，这些检测失败的主要原因是无人机拍摄的角度和距离与现实监控摄像存在一定差距，而且在现实环境中，存在可以模仿垃圾堆特征的不规律摆放的物体。这种相似性可能会导致错误检测。不同场景下垃圾堆的检测结果如图 6 所示。



**Figure 6.** Comparison of detection results: The first line is an example of YOLOv7 detection results. The second line is an example of YOLO-AF detection results

**图 6.** 检测结果比较：第一行为 YOLOv7 检测结果示例，第二行为 YOLO-AF 检测结果示例

## 5. 结论

本文提出的 YOLO-AF 主要是实现复杂环境下对垃圾堆的高效检测。该模型使用在各种复杂环境中收集的不同类别垃圾堆作为数据集。通过增加注意力机制，对原有网络模块进行改进，优化损失函数，提高了模型的检测速度和精度。实验结果表明，本文改进的策略可以显著提高复杂环境中不同类型垃圾堆的检测精度，而不会显著影响检出率。

本文提出的模型在计算性能较强的个人计算机上表现良好，但其模型复杂度较高，在低配计算平台上无法达到实时检测的要求，在弱光条件且高度遮挡的场景中需要提高检测性能。未来，我们计划探索模型压缩和轻量级设计，目的是在移动和嵌入式设备上部署模型，同时平衡准确性和速度。我们将研究使用立体相机计算的点云图像，这可能会更准确地识别和定位垃圾堆。

## 参考文献

- [1] Wang, Y. and Zhang, X. (2018) Autonomous Garbage Detection for Intelligent Urban Management. *MATEC Web of Conferences*, **232**, Article 01056. <https://doi.org/10.1051/mateconf/201823201056>
- [2] Zeng, D., Zhang, S., Chen, F., et al. (2019) Multi-Scale CNN Based Garbage Detection of Airborne Hyperspectral Data. *IEEE Access*, **7**, 104514-104527. <https://doi.org/10.1109/ACCESS.2019.2932117>

- [3] Patel, D., Patel, F., Patel, S., *et al.* (2021) Garbage Detection Using Advanced Object Detection Techniques. 2021 *International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, 25-27 March 2021, 526-531. <https://doi.org/10.1109/ICAIS50930.2021.9395916>
- [4] Zhao, H., Zhang, H. and Zhao, Y. (2023) Yolov7-Sea: Object Detection of Maritime UAV Images Based on Improved Yolov7. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, Waikoloa, HI, 3-7 January 2023, 233-238. <https://doi.org/10.1109/WACVW58289.2023.00029>
- [5] Mittal, G., Yagnik, K.B., Garg, M., *et al.* (2016) Spotgarbage: Smartphone App to Detect Garbage Using Deep Learning. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, 12-16 September 2016, 940-945. <https://doi.org/10.1145/2971648.2971731>
- [6] Camero, A., Toutouh, J., Ferrer, J., *et al.* (2019) Waste Generation Prediction in Smart Cities through Deep Neuroevolution. *Ibero-American Congress of Smart Cities, ICSC-CITIES 2018*, Soria, 26-27 September 2018, 192-204. [https://doi.org/10.1007/978-3-030-12804-3\\_15](https://doi.org/10.1007/978-3-030-12804-3_15)
- [7] Ruiz, V., Sánchez, Á., Vélez, J.F., *et al.* (2019) Automatic Image-Based Waste Classification. *International Work-Conference on the Interplay between Natural and Artificial Computation*, Almería, 3-7 June 2019, 422-431. [https://doi.org/10.1007/978-3-030-19651-6\\_41](https://doi.org/10.1007/978-3-030-19651-6_41)
- [8] Anjum, M. and Umar, M.S. (2018) Garbage Localization Based on Weakly Supervised Learning in Deep Convolutional Neural Network. 2018 *International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, 12-13 October 2018, 1108-1113. <https://doi.org/10.1109/ICACCCN.2018.8748568>
- [9] Aleena, V.J., Balakrishnan, K., Rosmi, T.B., *et al.* (2016) Automatic Waste Segregator and Monitoring System. *Journal of Microcontroller Engineering and Applications*, **3**, 1-7.
- [10] Singh, A., Aggarwal, P. and Arora, R. (2016) IoT Based Waste Collection System Using Infrared Sensors. 2016 *5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, 7-9 September 2016, 505-509. <https://doi.org/10.1109/ICRITO.2016.7785008>
- [11] Mahankali, S., Kabbin, S.V., Nidagundi, S., *et al.* (2018) Identification of Illegal Garbage Dumping with Video Analytics. 2018 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, 19-22 September 2018, 2403-2407. <https://doi.org/10.1109/ICACCI.2018.8554678>
- [12] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [13] Ren, S.Q., He, K.M., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [14] Chao, X., Hu, X., Feng, J., *et al.* (2021) Construction of Apple Leaf Diseases Identification Networks Based on Xception Fused by SE Module. *Applied Sciences*, **11**, Article 4614. <https://doi.org/10.3390/app11104614>
- [15] Jiang, Y., Chen, L., Zhang, H., *et al.* (2019) Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks with Small SE-ResNet Module. *PLOS ONE*, **14**, e0214587. <https://doi.org/10.1371/journal.pone.0214587>
- [16] Zheng, Y., Cui, Y. and Gao, X. (2023) An Infrared Dim-Small Target Detection Method Based on Improved YOLOv7. *Proceedings of the 2023 Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, Phuket, 28-30 April 2023, 1-5. <https://doi.org/10.1145/3596286.3596289>
- [17] Weng, Z., Hu, R. and Zheng, Z. (2023) Study on Individual Identification Method of Cow Based on CD-YOLOv7. *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing*, Sanya, 10-12 February 2023, 169-175. <https://doi.org/10.1145/3592686.3592717>
- [18] Da Cruz, L.B., Júnior, D.A.D., Diniz, J.O.B., *et al.* (2022) Kidney Tumor Segmentation from Computed Tomography Images Using DeepLabv3+ 2.5D Model. *Expert Systems with Applications*, **192**, Article 116270. <https://doi.org/10.1016/j.eswa.2021.116270>
- [19] Ruby, U. and Yendapalli, V. (2020) Binary Cross Entropy with Deep Learning Technique for Image Classification. *International Journal of Advanced Trends in Computer Science and Engineering*, **9**, 5393-5397. <https://doi.org/10.30534/ijatcse/2020/175942020>
- [20] Van Erven, T. and Harremoës, P. (2014) Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory*, **60**, 3797-3820. <https://doi.org/10.1109/TIT.2014.2320500>
- [21] Fang, J., *et al.* (2022) Single-Target Detection of *Oncomelania hupensis* Based on Improved YOLOv7s. *Frontiers in Bioengineering and Biotechnology*, **10**, Article 861079. <https://doi.org/10.3389/fbioe.2022.861079>