

基于 YOLOv7 的垃圾检测方法研究

陈君, 赵小会, 王博士, 季虹, 李维乾

(西安工程大学 计算机科学学院 陕西省服装设计智能化重点实验室, 西安 710048)

摘要: 随着社会经济的发展, 人们的生活水平持续提高, 生活垃圾量急剧攀升; 为了有效应对垃圾分拣效率低、准确率差等问题, 提出一种以 YOLOv7 网络为基础模型的垃圾检测算法; 该算法对 YOLOv7 网络进行了一系列改造, 首先, 在 Head 模块添加了注意力机制 SimAM, 增强了模型的感知能力和自适应能力, 从而提高检测精度; 其次, 在主干网络中改进了非极大值抑制算法 (soft-NMS) 去除冗余的检测框, 再次改进了损失函数为边框回归损失函数 SIOU, 提高了检测的精度和速度; 最后, 采用 C3 模块替换 YOLOv7 有的 ELAN-W 模块, 提升网络对较小目标的检测能力; 通过数据集对改进的网络进行测试, 平均准确度为 98.93%, 高于原模型的 96.31%, 实验结果也表明改进算法的检测效果有较为明显的提升。

关键词: 深度学习; 目标检测; 注意力机制; 非极大值抑制; 垃圾分类

Research on Garbage Detection Method Based on YOLOv7

CHEN Jun, ZHAO Xiaohui, WANG Boshi, JI Hong, LI Weiqian

(The Shaanxi Key Laboratory of Clothing Intelligence, School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: With the development of social economy and the continuous improvement of people's living standard, the amount of domestic garbage has rapidly increased. In order to effectively deal with the low efficiency and poor accuracy of garbage sorting, a garbage detection algorithm based on YOLOv7 network as a base model is proposed. The algorithm made a series of modifications to the YOLOv7 network, firstly, the attention mechanism SimAM was added to the head module, which enhanced the model's perceptual ability and adaptive ability so as to improve the detection accuracy; Secondly, non-maximum suppression (soft-NMS) was replaced in the backbone network to remove redundant detection frames; Then, the loss function was improved to be the edge regression loss function SIOU, improving the accuracy and speed of detection; Finally, the C3 module was used to replace the ELAN-W module in the YOLOv7, promoting the network's detection ability for smaller targets. Through experiment on the data-set, the average accuracy of the improved network is 98.93%, which is better than the 96.31% of the original model. Experimental results show that the improved algorithm has a more obvious enhancement in detection.

Keywords: deep learning; target detection; attention mechanism; non maximum suppression; garbage classification

0 引言

随着国民经济的快速发展, 人们的生活水平得到了很大的提高, 生活方式也发生了巨大的变化, 人们的物质需求日益增加, 加上我国人口基数大, 使得我国每年产生的生活垃圾也急剧增长。这些垃圾包括一些可回收利用的塑料、玻璃、金属、纸张等, 具有一定的经济价值。但是这些生活垃圾大多数被堆放在城市周围, 严重影响了人们的日常生活环境。近年来我国国民素质不断提高, 绿色发展环保理念逐渐深入人心, 如何合理有效地处理日渐增多的生活垃圾已经成为我国亟需解决的问题。如果不能及时对生活垃圾进行合理的处理, 不仅会造成环境污染, 还会引发各种疾病的传播, 同时对一些可回收垃圾的不当处理也

会造成严重的资源浪费。我国在处理垃圾方面, 与过去相比虽然取得一些进步, 但也仅对三分之二左右的垃圾进行了较为有效的处理, 且采用的处理方式大多是传统的垃圾处理方法, 比如对垃圾进行焚烧、填埋等, 利用这些方法处理垃圾, 会对空气、水资源、土壤等产生很多弊端, 对垃圾进行分类检测和合理地回收处理等可以有效地解决这一问题。目前主要是用人工进行垃圾的分拣, 这种垃圾分拣的方式存在着不可忽略的缺点: 人工分拣不仅成本比较高、分拣的效率比较低, 而且恶劣的工作环境会对分拣工人的身体健康情况产生严重影响^[1]。通过垃圾智能化检测技术, 能够有效降低成本、提高垃圾分拣的效率, 同时能够极大减少在人工分拣过程存在的误拣、漏拣等问题。

收稿日期: 2023-10-10; 修回日期: 2023-12-06。

基金项目: 国家自然科学基金(62106189); 陕西省高速公路施工机械重点实验室开放基金(300102250510); 西安工程大学科研基金(BS201847)。

作者简介: 陈君(1983-), 男, 博士, 讲师。

季虹(1983-), 女, 博士, 副教授。

李维乾(1980-), 男, 博士, 副教授。

引用格式: 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.

基于深度学习的目标检测算法因其良好的跨场景能力和泛化能力等特点得到了广泛的应用^[2], 许多的国内外学者聚焦于深度学习研究, 不断提出基于深度学习的目标检测方法, 用于垃圾的检测与分类研究。文献 [3] 利用区域卷积神经网络对目标检测方法进行研究, 对图片、视频等网络电子垃圾进行检测, 取得了良好的检测精度; 文献 [4] 在深度学习的基础上构建了一种智能系统, 用于垃圾的检测; 文献 [5] 提出了 FSSD 算法, 它可以看作是一种增强的 SSD 算法, 但改进后的网络检测精度稍有下降; 文献 [6] 提出了基于 SSD 的改进的网络结构, 改进了传统的卷积方法, 对可回收垃圾进行检测研究, 检测效果较好; 文献 [7] 在 SSD 算法的基础上加深网络层数, 并使用通道叠加方式来加强信息传递, 提升了垃圾检测的速率和准确率; 文献 [8] 提出了 YOLO TrashNet 网络检测模型, 实现了对视频垃圾进行分类检测。文献 [9] 提出一种基于 YOLOv2 改进的检测网络, 识别的准确率较高, 但其没有对标注的垃圾数据集进行细分, 因而无法区分垃圾的种类; 文献 [10] 提出一种基于 YOLOv3 的轻量级垃圾目标检测算法, 有效地降低了网络的参数量; 文献 [11] 提出在 YOLO 模型中添加注意力机制和 PANeT 增强特征提取能力, 能够快速准确地识别垃圾; 文献 [12] 提出一种基于 Faster RCNN 的检测模型, 使用 ResNet 作为特征提取网络, 该模型在作者自制数据集检测精度达到了 89%, 但数据集没有对垃圾种类进行细分, 实用性较差; 文献 [13] 提出了一种基于 VGG16 和 Faster RCNN 网络相结合的模型, 对垃圾进行检测, 提升了小目标检测的精度; 文献 [14], 提出了一种改进的 Mask R-CNN 算法模型用于生活垃圾的检测, 提高了垃圾检测的准确度; 文献 [15] 提出了物联网下基于机器学习的垃圾分类检测管理装置。文献 [16] 提出了基于物联网技术的智能垃圾分类系统设计; 文献 [17] 提出一种命名为 CompostNet 的网络检测模型, 实现了厨余垃圾进行分类检测; 文献 [18] 在已有网络的基础上提出了胶囊神经网络 (CapsuleNet), 对塑料垃圾数据集进行检测, 取得了较高的分类准确率。上述的研究工作对于垃圾检测都提出了自己的想法和见解, 对现有的算法进行不同的改进。但他们的研究工作还存在一些缺点, 首先垃圾检测相对于普通目标检测有更多的目标密集场景, 许多垃圾排列紧凑, 很容易发生目标漏检现象; 其次是垃圾检测中存在大量的负样本, 当数据集中各个类别样本数量不均衡时, 样本数目较少的类别的负样本数目相比于样本数目多的类别的负样本数目更多, 那么少数目的类别检测效果会更差。

基于深度学习的垃圾检测算法虽然已经取得了一定的效果, 但由于检测场景的多样性、复杂性和不确定性, 垃圾检测仍然处于优化和进一步提升阶段。本研究在结合前人经验的基础上, 以 YOLOv7 深度学习模型为基础, 通过改进, 用于垃圾的分类及检测, 相比以往的检测方法提高了检测的速度和检测的精度, 为环境治理提供决策依据。

1 基于深度学习的算法理论

1.1 基于深度学习的目标检测方法

随着深度学习技术的不断发展, 目标检测逐渐引起研究人员的注意。目标检测就是识别出图片中物体的类别和位置、用矩形框圈出来物体并给出相应的置信度。目标检测技术的发展经历了从传统的目标检测到基于深度学习的目标检测的过程。

传统的目标检测的特点是用手工提取特征进行目标的检测, 然后把提取到的特征输入到相关算法中, 进而判定物体的位置和类别。如 Viola Jones 检测器^[19]、HOG 检测器^[20]等。由于手工提取的特征有限, 并不能提取到图片中的所有特征, 故对图片的检测能力不高, 加之图片中有很多大小形状不一的目标物体, 需要很多的候选框, 这样就带来了计算量问题, 并且传统的目标检测需要大量的人工操作, 因此适应的场景会存在很多限制, 所以研究人员开始考虑使用深度学习来进行目标检测。基于深度学习的目标检测算法一般分为两段 (Two-stage) 目标检测算法, 二者最主要的区别是有没有单独的候选框筛选阶段。Two-stage 算法是将整个检测过程分成两个步骤进行, 第一步将输入的图片分成数量较多的待检测区域, 利用特征提取网络对输入的图像进行特征提取, 并且进行位置的回归; 第二步利用训练好的分类器对提取到的图像特征进行分类, 得到检测图像的目标物体所对应的类别, 这种方法虽然检测精度比较高, 但是在第一步进行特征提取时会产生大量的冗余计算, 无法达到实时性的要求。两阶段的代表算法有 R-CNN 系列^[21]、Mask-RCNN^[22]等。One-stage 算法是在一个卷积神经网络中融合了分类、位置回归两个过程, 实现了真正意义上端到端的目标检测。检测过程先利用骨干特征提取网络对输入图像进行特征提取, 在特征图上生成一定数量的候选区域, 然后对这些候选区域计算置信度得分, 剔除冗余的候选框, 最后进行目标分类和位置回归。这类算法在运算速度上占有优势。One-stage 代表算法有 YOLO 系列^[23]、SSD^[24]等。

1.2 YOLOv7 网络模型

YOLOv7 是在 YOLOv4 和 YOLOv5 的基础上做了改进, 但在网络的整体架构上、文件配置的涉及推理和验证等过程上大部分都继承自 YOLOv5。但和以往的 YOLO 系列不同的是, YOLOv7 拥有更多的卷积层, 网络层数更深、采用动态标签分配策略, 进行标签的分配、在 ELAN 的基础上进行改进, 提出了 ELAN-W 这一新的网络架构。YOLOv7 网络包括输入端 (Input)、主干网络 (Backbone)、颈部 (Neck) 和检测头 (Head) 4 部分^[25]。Input 部分的作用是负责数据预处理, 包括图片的缩放、数据增强技术和自适应锚框计算。Backbone 为骨干网络, 主要由 CBS、ELAN、MP-1 组成, 它的作用是负责特征提取。其中 ELAN 模块由最长和最长的梯度路径构成, 通过最短路径堆叠更多 block, 学习更多特征, 以高效为主。Neck 部分负责多尺度特征图的信息融合^[26], 主要由 FPN^[27]和 PAN^[28]

结构组成。Head 部分的作用是进行损失函数计算, 主要由 CBS、SPPCSPC、E-ELAN、MP-2、RepConv 组成, 实现了不同特征尺度信息的融合, 并且减少了计算量。YOLOv7 是目前 YOLO 系列较先进的算法, 在准确率和速度上都超越了以往的 YOLO 系列。YOLOv7 模型的网络结构如图 1 所示。

1.3 注意力机制

注意力机制也被称为“神经网络的注意力”, 或者更简单的: “注意力”。人脑在工作和学习时, 其实是有一定的注意力的, 比如我们在浏览器上搜索时, 大部分的注意力都集中在搜索结果的左上角, 这说明大脑在处理信号的时候是有一定权重划分的, 而注意力机制的提出正是模仿了大脑的这种特性。深度学习中的注意力机制源自对人类视觉注意力机制的原理研究, 它和人类选择重要信息过程相似, 网络在处理数据时, 将注意力集中在重要部分。

注意力机制有两种, 一种是软注意力 (Soft Attention), 另一种则是强注意力 (Hard Attention)。软注意力更关注区域或者通道, 而且软注意力是确定性的注意力, 学习完成后直接可以通过网络生成, 最关键的地方是软注意力是可微的, 这是一个非常重要的地方。可以微分的注意力就可以通过神经网络算出梯度并且前向传播和后向反馈来学习得到注意力的权重。强注意力与软注意力不同点在于, 首先强注意力是更加关注点, 也就是图像中的每个点都有可能延伸出注意力, 同时强注意力是一个随机的预测过程, 更强调动态变化。当然, 最关键的是强注意力是一个不可微的注意力, 训练过程往往是通过增强学习 (Reinforcement Learning) 来完成的。

将注意力机制应用于神经网络中, 目的是进行计算资源的分配, 根据任务的重要程度分配计算资源。

在神经网络中引入注意力机制的原因有两个: 1) 计算

能力的限制: 目前计算能力依然是限制神经网络发展的瓶颈, 当输入的信息过多时, 模型也会变得更复杂, 通过引入注意力, 可以减少处理的信息量, 从而减少需要的计算资源; 2) 算法优化的限制: 虽然目前可以通过对一些算法进行优化, 来简化神经网络, 但是对于神经网络来说, 长距离的“记忆”能力, 一直是摆在人们面前的一个大难题。

通过在神经网络中加入注意力机制, 可以使网络把更多的注意力定位到最感兴趣的信息上, 同时过滤掉无用信息、提取出更多的有关重要信息的细节, 且可以在弱化无效目标干扰的同时, 节省计算资源。

2 垃圾分类检测算法框架

2.1 改进 YOLOv7 的算法结构

YOLOv7 是目前 YOLO 系列较先进的算法, 也是一个很好的目标检测模型, 在准确率和速度上都超越了以往的 YOLO 系列。因此本研究选择 YOLOv7 为基础模型, 并在 YOLOv7 的基础上进行了 4 个部分的改进: 1) 在 Head 模块添加 SimAM 注意力机制模块, 它不需要额外的参数进行学习, 而是通过计算相似度得分来确定每个键向量的重要性, 从而实现了轻量化和高效的注意力机制, 可以有效地提升模型性能; 2) 采用更加通用的非极大值抑制算法 soft-NMS 替换原 YOLOv7 中的非极大值抑制算法 NMS, 有效地提升密集遮挡数据的检测精度; 3) 采用 Siou 损失函数替换原 YOLOv7 中的损失函数 CIou; 4) 采用 C3^[29] 模块替换原 YOLOv7 有的 ELAN-W 模块, 提升网络对较小目标的检测能力。以上这些改进措施能提升模型整体的检测精度。改进后模型如图 2 所示。

2.2 融合 SimAM 注意力机制

SimAM 模块是由文献 [30] 提出的一种无参数的注意力模块, 它和现有的通道注意力机制和空间注意力机制是不同的, 如 CBAM、SE 等。通道和模块注意力机制大多都

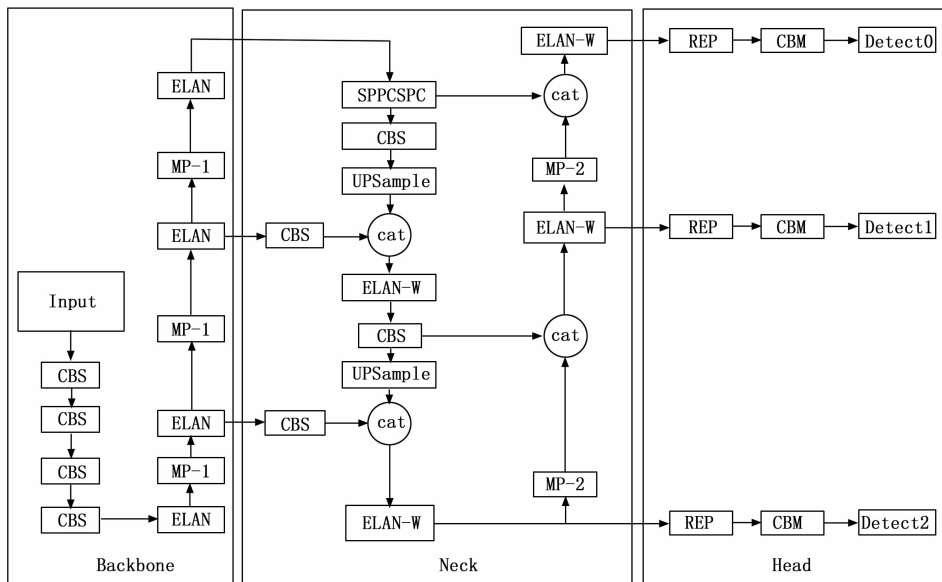


图 1 YOLOv7 网络结构

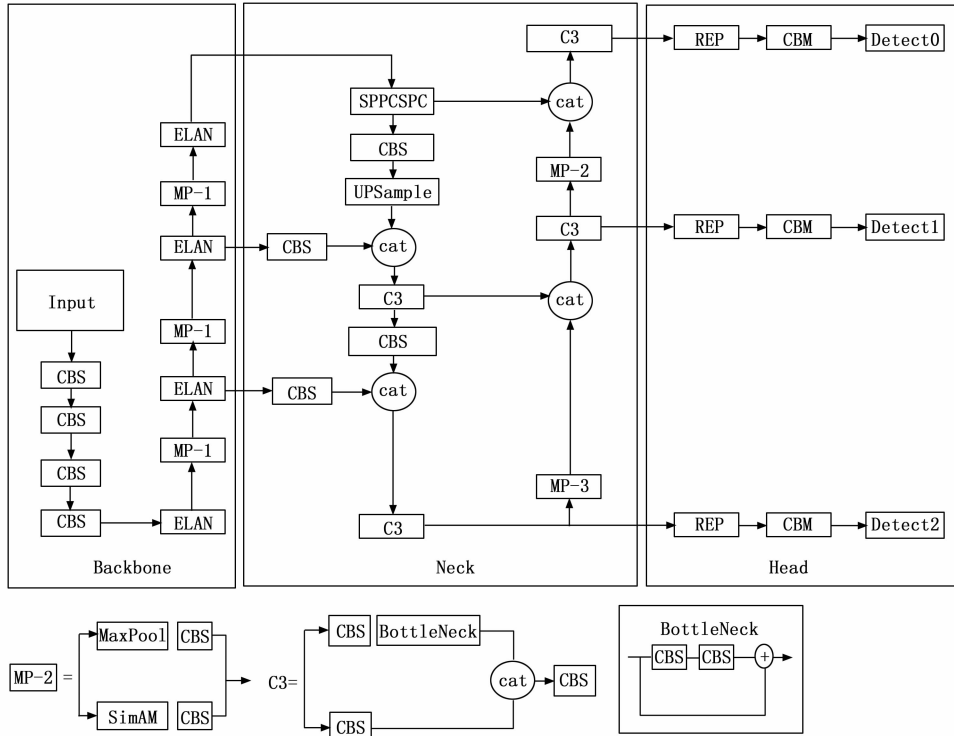


图 2 改进后的 YOLOv7 网络结构

是利用权值生成的方法和额外的子网络生成注意力的权值，它们只能在通道或者空间层面上细化特征信息，但是这种方式限制了它们学习和辨别更多特征的能力，因此会使它们在空间变化和跨通道的情况下灵活性变差^[31]通道注意力机制和空间注意力机制分别如图 3 中 (a) (b) 所示。

SimAM 注意力机制同时关注到空间维度和通道维度，定义了统一空间维度和通道维度的三维注意力权重，如图 3 (c) 所示。它能在不需要额外参数的情况下推导出 3D 注意力权重值，SimAM 的操作相较于其他的注意力机制，更加简单高效，避免了过多的网络结构调整造成的模型参数增加的问题。该注意力机制的作用原理是在神经科学理论的基础上优化能力函数，通过能量函数计算注意力机制的权值。能量函数的定义为：

$$e_t(\mathbf{w}_t, b_t, y, x_t) = (y_t - \hat{t}) + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2$$

$$\hat{t} = \mathbf{w}_t t + b_t$$

$$\hat{x}_i = \mathbf{w}_i x_i + b_i$$

\mathbf{w}_t 为权重， b_t 为偏置， y, y_t, y_0 均表示标量数量， t 和 x_i 分别为输入特征的目标神经元和其他神经元， M 为神经元个数。最小化公式 (1) 等价于训练统一通道内神经元 t 与其他神经元之间的线性可分性。采用二值标签，并在式 (1) 中添加正则项 λ ，最终能量函数定义为：

$$e_t(\mathbf{w}_t, b_t, y, x_t) =$$

$$\frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (\mathbf{w}_i x_i + b_i)]^2 + [1 - (\mathbf{w}_t t + b_t)]^2 + \lambda \mathbf{w}_t^2$$

(2)

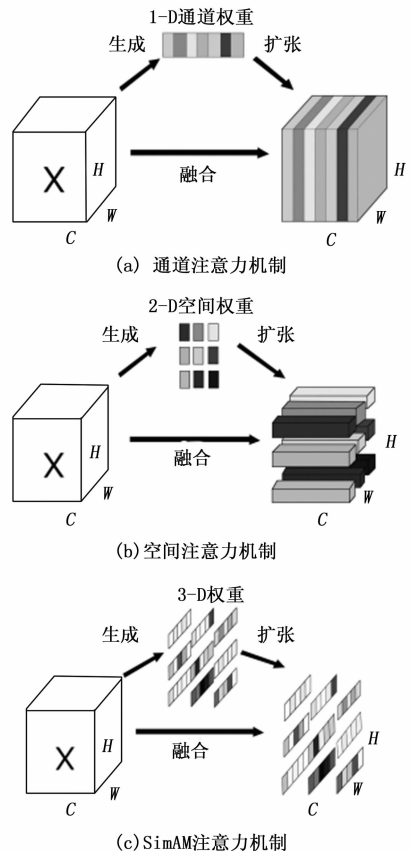


图 3 注意力机制图

公式的解析解为:

$$\begin{aligned} w_i &= \frac{2(t - u_i)}{(t - u_i)^2 + 2\sigma_i^2 + 2\lambda} \\ b_i &= -\frac{1}{2}(t + u_i)w_i \end{aligned} \quad (3)$$

其中:

$$\begin{aligned} u_i &= -\frac{1}{M-1} \sum_{i=1}^{M-1} X_i \\ \sigma_i^2 &= \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - u_i)^2 \end{aligned} \quad (4)$$

通过代入化简可得最小能量值公式:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \mu)^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (5)$$

通过最小能量公式得出, 神经元 t 的能量越低, 则它与周围神经元的区别越大, 重要性也越高。因此神经元的重要性可以通过 $1/e_i^*$ 得到。最终得到 SimAM 模块公式如下:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \otimes X \quad (6)$$

X 为输入的特征, E 为 e_i^* 在所有通道和空间维度的总和; 式 (6) 中, 添加 sigmoid 函数的作用是限制 E 的过大值。

2.3 基于 soft-NMS 的非极大值抑制

YOLOv7 算法中, 原始的非极大值抑制 (NMS, non-maximum suppression) 原理是对高于设定阈值的预测框进行筛选, 在每一轮筛选中, 如果置信度最高的预选框和其他预测框的并交比 (IoU) 大于了所设定的阈值, 那些预选框的置信度就会被赋为 0, 从而将置信度最高的预测框附近重合面积较高的其他预选框删掉, 经过多次重复的筛选过程, 最终保留下来的预选框作为检测结果。NMS 的表达式为:

$$s_i = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ 0, & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (7)$$

其中: s_i 为当前预选框的置信度, N_i 为设定的 IoU 阈值, M 为置信度最高的检测框, b_i 为其他预选框。

NMS 算法会过滤掉与基准框的交并比超过设定阈值的检测框, 但是这一操作会对重叠目标造成漏检, 使得网络的检测精度降低。因此本文采用 Soft-NMS 算法代替 NMS 算法, 在 Soft-NMS 算法中, 不是直接将大于阈值的预测框置 0 删除, 而是将它们置信度进行衰减, 这种方式能有效地降低漏检率, 提高模型的检测效果。Soft-NMS 的表达式为:

$$s_i = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ 0, & [1 - \text{IoU}(M, b_i)], \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (8)$$

2.4 基于 SIoU 的损失函数

YOLOv7 网络中, 采用 CIoU^[32] 作为位置损失函数, 但是 CIoU 存在一个不可回避的缺陷, 因为它的纵横比描述的不是预测框与置信度之间的真实差异, 而是一个相对值。相较于传统的损失函数, CIoU 会导致网络的收敛速度下降, 也会导致网络的收敛效率变低。考虑到 CIoU 存在的缺陷, 本研究采用更有优势的损失函数 SIoU^[33] 替换 YOLO7 网络模型

中的损失函数 CIoU。SIoU 的参数如图 4 所示, 它包括角度损失 (Angle Cost)、距离损失 (Distance Cost)、形状损失 (Shape Cost)、交并比损失 (Intersection Over Union) 4 部分, 其具体定义如下。SIoU 的参数如图 4 所示。

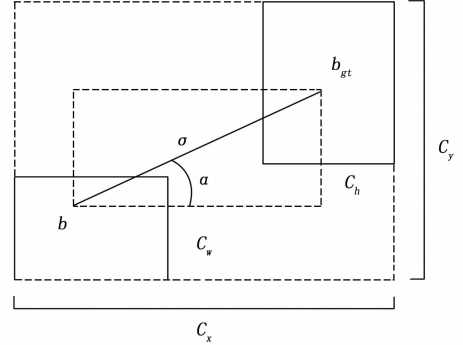


图 4 损失函数 SIOU 的计算

角度损失公式如下:

$$\Delta = \cos\left\{2\left[\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right]\right\} \quad (9)$$

其中: c_h 为真实框和预测框中心点的高度差, σ 为真实框和预测框中心点的距离。

距离损失公式如下:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma t}) = 2 - e^{-\gamma p_x} - e^{-\gamma p_y} \quad (10)$$

$$\rho_x = \left(\frac{b_x^{gt} - b_x}{c_x}\right)^2, \rho_y = \left(\frac{b_y^{gt} - b_y}{c_h}\right)^2, \gamma = 2 - \Delta \quad (11)$$

b_x^{gt} 和 b_y^{gt} 为真实框中心坐标, b_x 和 b_y 为预测框中心坐标, c_x 和 c_y 为真实框和预测框最小外接矩形的宽和高。

形状损失公式如下:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta = (1 - e^{-w_x})^\theta + (1 - e^{-w_y})^\theta \quad (12)$$

$$W_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, W_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (13)$$

交并比损失公式如下:

$$\text{IoU} = \frac{B \cap B^{gt}}{B \cup B^{gt}} \quad (14)$$

B 和 B^{gt} 分别为预测框和真实框的面积。

综上, 可得 SIOU 的损失函数计算公式如下:

$$L_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \quad (15)$$

2.5 C3 结构

YOLOv7 网络中使用了大量的高效层聚合网络 (ELAN) 模块, 在网络层面, 它属于梯度路径设计网络的范畴。设计 ELAN 的主要目的是解决在执行模型缩放时深度模型的收敛性会逐渐恶化的问题。ELAN 主要由 vovNet 和 CSPNet 组成, 并利用计算块中的堆栈结构优化了整个网络的梯度长度。由于 ELAN 模块需要在每个卷积层中添加自适应卷积用来增强模型性能, 因此网络中会有大量 ELAN 模块堆叠, 这些堆叠的模块不仅会导致网络的参数量和复杂度增大, 还会使得网络的稳定状态和性能变差。

为了解决上述提到的问题，本文采用 C3 模块代替 YOLOv7 网络模型中原来特征融合网络中的 ELAN 模块。C3 是 YOLOv5 网络模型中的关键模块，它主要由 CBS 模块和 Bottleneck Block 模块组成，其中 CBS 结构是由卷积层、批量归一化层和 SiLU 激活函数组成。C3 结构采用了多尺度特征融合和跨通道信息传递机制来提高模型的表达能力，同时通过残差连接和特征图拼接来保持信息流的连续性。相比于传统的残差块结构，C3 结构可以更有效地学习多尺度特征，并能够在不影响网络性能的情况下提高模型的学习能力和稳定性。此外，C3 结构的参数量相对较少，可以有效地减少网络的计算量，从而提高目标检测的速度和精度。

3 实验方案

3.1 数据集

此次垃圾分类与检测实验采用的是极市平台上公开的数据集，该数据集主要分为 3 个大类：纸类 (Paper)、塑料 (Plastic)、其他垃圾 (Other-waste)，共 15 000 张图片，按照 7 : 2 : 1 的将数据集划分为训练集、测试集和验证集 3 部分。数据集中的部分图片如图 5 所示。

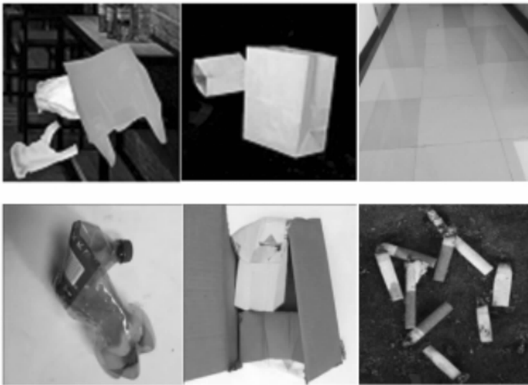


图 5 数据集样本

3.2 实验环境配置

改进算法的实验环境为：在 GPU: NVIDIA GeForce RTX 3060 进行网络训练和测试，CUDA 版本为 11.6，Python 版本为 3.8.0，实验环境的具体信息如表 1 所示。

表 1 实验环境

配置	Windows 10 64-bit
CUDA	11.6
GPU	NVIDIA GeForce RTX 3060
CPU	Intel Core i5-11400
Python	3.8.0
Pytorch	1.10.2

3.3 实验相关评价指标

为了评估模型的检测性能，本文选用准确率 (P, precision)、召回率 (R, recall)、平均精度 (AP)、平均精度均值 (mAP)、训练时间 (Time) 等评价指标进行对比。

准确率是评价算法的重要指标之一。准确率越高，说

明算法预测的结果与实际的结果越接近，表明算法的正确率越高。其定义如下：

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

其中：TP 表示实际为真且预测为真，FP 表示实际为假但预测为真。

召回率表示一个模型所分出的正样本占真实情况下正样本的比例，召回率越高，说明算法能够正确检测出更多的实际存在的物体。其定义如下：

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

其中：FN 表示实际为真，预测为假。

平均精度可体现为 P 曲线与 R 曲线和坐标轴围成的面积，PR 曲线围成的面积越大，表示模型的性能越好，它反映模型对某一类别的识别效果，其定义如下：

$$AP = \int_0^1 P(R) dR \quad (18)$$

将所有的精度值相加除以所检测到的类别数，即求平均数，得到的值就是目标检测网络模型的均值平均精度，可表示神经网络模型在用于实验数据集上的训练结果。mAP 是 AP 的平均值，是模型综合性能的体现，定义如下：

$$mAP = \frac{1}{n} \sum_{k=1}^{K=n} AP_k \quad (19)$$

其中：n 为总的种类数，AP_k 为第 k 个类别的精确度。

4 实验结果与分析

4.1 消融实验

为研究改进后模型的检测效果，以及具体的改进方法起到的作用，本次实验设置了 4 个消融实验。第一个实验 (v7-0) 为原始的 YOLOv7 模型；第二个实验 (v7-0-c3) 在 YOLOv7 网络基础上用 C3 模块替换了原网络的 ELAN 模块；第三个实验 (v7-1) 在 YOLOv7 网络基础上添加注意力机制 SimAM；第四个实验 (v7-1-c3) 用 C3 模块替换了 YOLOv7 网络的 ELAN 模块，并添加注意力机制 SimAM；第五个实验 (v7-2-c3) 用 C3 模块替换了 YOLOv7 网络的 ELAN 模块、添加注意力机制 SimAM 和更换损失函数为 SioU；第六个实验 (Ours) 为用 C3 模块替换了 YOLOv7 网络的 ELAN 模块、添加注意力机制 SimAM、更换损失函数为 SioU、更换非极大值抑制函数为 soft-NMS，即文中所提的改进模型。具体结果如表 2 所示。

表 2 消融实验对比表

组别	评价指标					
	P	R	mAP@0.5 / %	mAP@0.5 : 0.95 / %	参数量 / M	计算量 / B
V7-0	0.971 1	0.948 1	96.31	91.72	141.93	105.1
V7-0-c3	0.986 0	0.951 1	98.47	91.77	128.87	40.3
V7-1	0.974 1	0.952 4	98.44	91.87	141.87	104.9
V7-1-c3	0.976 0	0.953 1	98.47	91.95	128.77	40.3
V7-2-c3	0.979 2	0.959 6	98.62	92.89	128.83	40.3
Ours	0.983 1	0.960 8	98.93	92.64	126.11	39.7

表 3 与其他模型的对比实验

组别	评价指标						
	P	R	$mAP_{0.5}/\%$	$mAP@0.5:0.95/\%$	参数量/M	计算量/B	FPS/(帧·s ⁻¹)
YOLOv4	0.833 1	0.875 2	87.56	85.32	47.83	35.7	62.3
YOLOv5	0.951 6	0.936 0	0.964 9	0.907 4	26.80	16.0	89.3
YOLOv7	0.971 1	0.949 8	0.984 1	0.917 2	141.93	105.1	35.4
SSD	67.20	43.16	0.908 8	0.340 7	23.75	18.5	46
Retinanet50	71.13	52.34	0.853 7	0.764 2	157.50	128.23	34
YOLOv7-tiny	0.956 3	0.909 0	0.823 6	0.823 6	22.96	13.2	59.1
本文模型	0.983 0	0.960 1	0.989 3	0.936 4	126.11	39.7	78.1

消融实验的实验数据如表 2 所示, 由表 2 可知: 在第一个实验 (v7-0) 和第二个实验 (v7-0-c3) 可知, 用 C3 结构替换 YOLOv7 网络模型中的 ELAN 结构, 会使网络层数变少, 导致网络整体的参数量变小和计算量变小; 由图 6 可知, C3 模块的加入提高了模型对小目标的检测能力; 第二个、第三个、第四个、第五个实验的 P 、 R 、 $mAP_{0.5}$ 、 $mAP_{0.5:0.95}$ 指标值均高于原始模型, 但参数量都比原始模型小, 有效提升了模型的精度。由此可知, 针对原 YOLOv7 网络模型的改进, 在垃圾检测数据集上都起到了良好的效果。



(a) V7-0检测图

(b) V7-0-c3检测图

图 6 模型检测样本

4.2 对比实验

为了选择最合适的基准网络模型以及进行更为全面的对比, 选择了一些常见其他架构的检测模型进行实验比较, 为了更好地对比实验, 实验均采用相同的数据集、相同的设备, 设置相同的参数, 结果如表 3 所示。

综合这些模型的检测结果来看: 相比 YOLOv4、YOLOv5、SSD、Retinanet50、YOLOv7-tiny 模型, YOLOv7 模型速度更快且精度更高, 因此选择以 YOLOv7 为基本网络。由表 3 可得所提改进模型, 检测的精确度最高且速度最快。综上, 所提改进模型考虑了精度与参数量的权衡, 在垃圾检测方面具有优势。

4.3 实验分析

通过设置对比实验和消融实验, 验证了所选 YOLOv7 为基准模型的必要性, 以及改进算法的有效性。SimAM 的操作增强了模型的特征提取能力且更加简单高效, 避免了过多的网络结构调整造成的模型参数增加的问题; soft-

NMS 算法能有效地提升密集遮挡数据的检测精度; 损失函数的改进提升了模型整体的检测精度; 最后 C3 模块的加入提升了网络对小目标的检测能力。从表 2 和表 3 的实验结果验证了以上这些改进措施的效果。综上, 所提改进方法提升了模型整体的检测速度和检测精度。

5 结束语

本文利用深度学习技术, 在 YOLOv7 网络模型的基础上提出了一种改进的算法用于垃圾的分类与检测。该算法通过添加注意力机制 SimAM, 增强模型提取有效特征的能力; 其次改进了非极大值抑制函数为 soft-NMS, 加强对重叠目标的识别能力, 提高了检测的精度, 同时模型仍保持较高的检测速度; 然后用 SioU 损失函数替换了原模型中的损失函数有效避免了预测框的游荡现象, 提升了算法的收敛速度; 最后采用 C3 模块替换原 YOLOv7 有的 ELAN-W 模块, 提升网络对较小目标的检测能力。

由对比实验结果和消融实验结果可得, 所提的改进算法在垃圾检测任务中检测精度方面表现出更好的效果, 其 mAP 值达到 98.93%, 且相比原模型训练时间更短, 具有更高的准确性和鲁棒性, 也表明了所提改进方法的有效性。

文中所提出的改进方法检测性能相比于原 YOLOv7 算法有了一定的提升, 也在所选数据集上证明了其检测性能的优越性, 但是还存在一些缺陷, 如: 数据集种类不够多等。未来的研究可以考虑引入更多的垃圾分类检测数据集, 提高本文方法的通用性, 增强本文算法的泛化能力。

参考文献:

- [1] LUO Q, SUN Y, LI P, et al. Generalized completed local binary patterns for time-efficient steel surface defect classification [J]. IEEE Transactions on Instrumentation and Measurement, 2018, 68 (3): 667-679.
- [2] 吕凡, 郝丽萍, 章骅, 等. 病毒在环境卫生作业环境中的存活潜力及感染风险防控探讨 [J]. 环境卫生工程, 2020, 28 (1): 1-9.
- [3] NOWAKOWSKI P, PAMULA T. Application of deep learning object classifier to improve waste collection planning [J]. Waste Management, 2020, 109: 1-9.

- [4] VERMA V, GUPTA D, GUPTA S, et al. A deep learning-based intelligent garbage detection system using an unmanned aerial vehicle [J]. *Symmetry*, 2022, 14 (5): 960.
- [5] LI Z X, ZHOU F Q. FSSD: Feature fusion single shot multibox detector [J/OL]. *ArXiv*, vol. abs/1712.00960, 2017.
- [6] 白 阳, 刘 斌, 李彦彤. 基于 SSD 的可回收垃圾检测研究 [J]. *陕西科技大学学报*, 2020, 38 (6): 152-158.
- [7] 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.
- [8] DE CAROLIS B, LADOGANA F, MACCHIARULO N. Yolo trashnet: Garbage detection in video streams [C] //2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). *IEEE*, 2020: 1-7.
- [9] LIU Y, GE Z, LÜ G, et al. Research on automatic garbage detection system based on deep learning and narrowband internet of things [J]. *Journal of Physics: Conference Series*, 2018: 1069-1076.
- [10] 许 伟, 熊卫华. 一种改进的轻量级垃圾目标检测算法 [J]. *计算机技术与发展*, 2022, 32 (2): 63-68.
- [11] 张 伟, 刘 娜, 江 洋, 等. 基于 YOLO 神经网络的垃圾检测与分类 [J]. *电子科技*, 2022, 35 (10): 45-50.
- [12] WANG Y, XU Z. Autonomous garbage detection for intelligent urban management [C] //MATEC Web of Conferences, EDP Sciences, Shanghai, 2018: 1506-1060.
- [13] 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.
- [14] 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.
- [15] SHAMIN N, FATHIMAL P, PRAKASH K, et al. Smart garbage segregation & management system using internet of things (IoT) & machine learning (ML) [C] //International Conference on Innovations in Information and Communication Technology, Chennai, India, 2019: 1-6.
- [16] TAN Y, GUO W, YANG K, et al. Design of intelligent garbage classification system based on internet of things technology [J]. *Journal of Physics: Conference Series*. IOP Publishing, 2022, 2187 (1): 012020.
- [17] FROST S, TOR B, AGRAWAL R, et al. Compostnet: An image classifier for meal waste [C] //2019 IEEE Global Humanitarian Technology Conference (GHTC). *IEEE*, 2019: 1-4.
- [18] SREELAKSHMI K, AKARSH S, VINAYAKUMAR R, et al. Capsule neural networks and visualization for segregation of plastic and non-plastic wastes [C] //Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019: 631-636.
- [19] 贾海鹏, 张云泉, 袁 良, 等. 基于 OpenCL 的 Viola-Jones 人脸检测算法性能优化研究 [J]. *计算机学报*, 2016, 39 (9): 1775-1789.
- [20] 李澎林, 邹嘉程, 李 伟. 基于 HOG 和特征描述子的人脸检测与跟踪 [J]. *浙江工业大学学报*, 2020, 48 (2): 133-140.
- [21] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Pro. of IEEE Conference on Computer Vision and Pattern Recognition Piscataway, NJ: IEEE Press, 2014: 580-587.
- [22] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [EB/OL]. (2018-01-24) [2023-07-06]. <https://arxiv.org/abs/1703.06870>.
- [23] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2016: 779-788.
- [24] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //Proc. of European Conference on Computer Vision, Berlin: Springer, 2016: 21-37.
- [25] WANG C, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag of freebies sets new state of the art for real-time object detectors [EB/OL]. (2020-07-06) [2022-05-10]. <https://arxiv.org/abs/2207.02696>.
- [26] 吕 辉, 董 帆. 基于 YOLOv4 的复杂交通状况下多目标检测算法 [J]. *国外电子测量技术*, 2022, 41 (12): 41-47.
- [27] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, *IEEE*, 2017: 936-944.
- [28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, *IEEE*, 2018: 8759-8768.
- [29] 孙月莹, 陈俊霖, 张胜茂, 等. 基于改进 YOLOv7 的毛虾捕捞渔船作业目标检测与计数方法 [J]. *农业工程学报*, 2023, 39 (10): 151-162.
- [30] YANG L, ZHANG R Y, LI L, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks [C] //International Conference on Machine Learning. PMLR, 2021: 11863-11874.
- [31] CHEN Y X, WANG X Y, CHEN Z B, et al. Classification of camellia oleifera diseases in complex environments by attention and multi-dimensional feature fusion neural network [J]. *Plants*, 2023, 12 (14): 2701.
- [32] GAO J F, CHEN Y, WEI Y M, et al. Detection of specific building in remote sensing images using a novel YOLO-S-CIOU model. Case: Gas station identification [J]. *Sensors*, 2021, 21 (4): 1375.
- [33] 高 倩, 潘 杨, 朱 磊, 等. 基于 SIOU 函数的改进 YOLOv5 遥感目标检测方法 [J]. *长江信息通信*, 2022, 35 (11): 5-8.