

# 基于 YOLOv5 的废旧塑料瓶双目视觉识别定位方法

doi:10.20269/j.cnki.1674-7100.2026.1011

李欣妍 巩雪 付斌  
崔功卓 刘京宇 杨鸿阳

哈尔滨商业大学  
轻工学院  
黑龙江 哈尔滨 150028

**摘要:** 为了提高废旧塑料瓶视觉识别定位的准确度, 提出一种基于改进 YOLOv5 的轻量化双目视觉目标检测算法。首先, 在主干网络中分别使用 GhostBottleNeck 模块和 GhostNetV2 模块替换原始的 CBS 和 C3 模块; 其次, 在颈部网络引入 GSConv 卷积和基于 GSConv 的 VOVGCSP 结构。通过自制数据集训练改进模型, 并结合双目相机标定与立体匹配技术实现目标测距。结果表明: 1) 改进后的 YOLOv5 模型与原模型对比, 准确率由 87.92% 提升到 95.39%, 参数量由 7 012 888 下降到 5 933 320, 帧率由 96 帧/s 提升到 105 帧/s。改进网络有效减少了模型参数量, 且准确率提高了 7.47%。2) 通过双目测距实验得到, 检测的最大误差在 7 mm 左右, 相对误差在 1% 以内, 此误差符合要求。该方法的识别定位速度和精准度均达到实时处理要求, 能为智能回收设备开发提供技术支持。

**关键词:** 轻量化; YOLOv5 模型; 废旧塑料瓶; 双目标定

**中图分类号:** TP391.4; TB489

**文献标志码:** A

**文章编号:** 1674-7100(2026)01-0098-09

**引文格式:** 李欣妍, 巩雪, 付斌, 等. 基于 YOLOv5 的废旧塑料瓶双目视觉识别定位方法 [J]. 包装学报, 2026, 18(1): 98-106.

## 1 研究背景

随着我国经济的增长和居民消费水平的提高, 在大规模生产和大消费的时代背景下, 不免会产生大量的垃圾, 进而给生态环境造成负担。目前, 垃圾分类主要靠人工处理, 然而人工处理劳动量大, 且容易导致分类错误<sup>[1-3]</sup>。因此, 开展自动化垃圾分类方法的研究工作已刻不容缓。

近年来, 深度学习和计算机视觉技术的快速发展为垃圾分类提供了新的解决方案。研制具备高识别率与高定位精度的智能化垃圾处理设备, 如智能回收垃圾桶、智能检测垃圾无人机、回收垃圾机器

人等智能化装置, 能够在众多实际场景中得到广泛应用<sup>[4-7]</sup>。吴建等<sup>[8]</sup>利用传统图像处理方法, 通过手动提取图像特征来完成垃圾与背景的分离。Kang Z. 等<sup>[9]</sup>用图像多特征融合技术对 ResNet-34 进行了改进, 使整个网络的运行速度和检测精度得到提升。向伟等<sup>[10]</sup>通过 CaffeNet 模型来减少复杂环境对垃圾识别的影响。文灿华等<sup>[11]</sup>使用 Faster RCNN 对生活中较为常见的 6 种生活垃圾进行分类, 并通过实验对比分析 3 种不同主干网络的性能差异。戴康<sup>[12]</sup>先对数据集进行归一化处理, 以通过数据增强等策略来提升垃圾分类检测方法的鲁棒性, 再优化 YOLOv3, 在最后一组残差块中新增一组残差块,

收稿日期: 2025-10-21

基金项目: 黑龙江省属高等学校基本科研业务费项目 (2023-KYYWF-1011)

作者简介: 李欣妍, 女, 硕士生, 主要研究方向为智能化装备开发, E-mail: 1371167247@qq.com

通信作者: 巩雪, 女, 副教授, 主要从事智能化装备开发研究, E-mail: kahnannie@163.com

以实现不同尺度特征融合。改进后的算法可以生成 4 种尺度特征图, 检测不同尺度的对象。王心鹏<sup>[13]</sup>提出了基于 YOLOv3 融合注意力机制并采用 SoftNMS 的垃圾检测算法。该算法仅引入了少量参数, 使检测精度提升。目前, 垃圾分类的智能化研究仍存在一些问題, 如: 规模较小的数据集易导致模型在面对复杂环境时的鲁棒性不足, 在实际应用中存在识别精度不高、误检率高、训练成本高等问题<sup>[14]</sup>。因此, 如何进一步优化模型结构、提升检测精度、降低计算资源消耗, 是当前的研究重点。

综上, 本研究拟改进 YOLOv5 模型, 即从注意力机制、网络结构和损失函数 3 个维度对模型进行改进, 以减少参数量, 提升其在垃圾分类任务中的表现, 提高分类准确率与识别速度。此外, 在优化算法的基础上, 利用双目视觉识别定位方法对废旧塑料瓶进行测距。

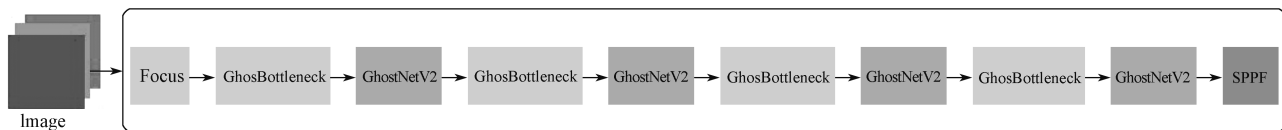


图 1 YOLOv5 主干网络结构图

Fig. 1 YOLOv5 backbone network architecture

### 1) GhostBottleneck 模块

Ghost 通过简单的线性变换生成与标准卷积相似的特征图, 从而显著降低模型计算量。基于 Ghost 的 GhostBottleneck 模块 (结构见图 2), 借鉴了 ResNet 的残差连接机制, 通过跨层恒等映射有效缓解梯度消失问题, 同时抑制过拟合现象。

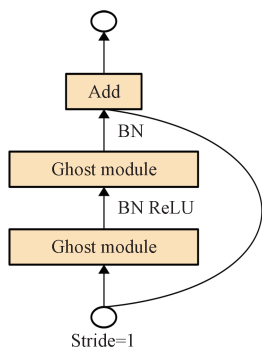


图 2 GhostBottleneck 结构图

Fig. 2 GhostBottleneck structure diagram

GhostBottleneck 模块主要由两个 Ghost 模块构成。第 1 个 Ghost 模块作为扩展层, 旨在增加特征图的通道维度。第 2 个 Ghost 模块则作为压缩层, 将通

## 2 目标检测

YOLOv5<sup>[15]</sup> 的模块化结构 (如 CBS、C3) 易于轻量化改造, 可灵活替换为 GhostBottleneck、GSConv 等轻量化组件, 同时其在参数量与计算量方面具备显著优势, 符合移动端或嵌入式设备的实时性要求。因此, 本研究用 GhostNetV2 和解耦全连接注意力机制 (decoupled fully connected, DFC) 改进 YOLOv5, 以实现模型轻量化, 减少参数的同时提升检测精度。

### 2.1 主干网络优化

从两方面改进 YOLOv5 的主干网络, 其结构见图 1。首先, 用基于 Ghost<sup>[16-17]</sup> 的 GhostBottleneck 模块替代原始 CBS 模块, 以降低计算量; 其次, 引入结合 Ghost 模块与 DFC<sup>[18]</sup> 注意力机制的 GhostNetV2 模块替换 C3 模块, 以增强特征表达能力。

道维度恢复至与原始输入相匹配, 从而保持跨层连接通道的一致性。此外, 借鉴 MobileNet 的设计思想, 在第 2 个 Ghost 模块后取消 ReLU 激活函数, 但其他层均包含批量归一化 (BN) 和 ReLU 激活函数。出于计算效率考量, Ghost 模块的初始卷积为点卷积。

### 2) GhostNetV2 模块

GhostNet 也使用了两个 Ghost 模块, 使特征维度先上升再下降, 但是通过廉价操作得到特征, 一定程度上损坏了模型的表达力和容量。因此, 对 GhostNet 进行优化设计, 得到 GhostNetV2 模块, 结构如图 3 所示。将 DFC 注意力机制与第 1 个 Ghost 模块并行设计, 可实现对长距离空间位置依赖关系的捕捉, 提升模型的特征表达能力。此外, DFC 注意力机制的计算复杂度较高, 若直接引入会增加计算成本。故对 DFC 注意力机制进行优化。首先对特征图在水平与垂直方向进行  $2 \times$  下采样 (长宽各缩减为原尺寸的  $1/2$ ), 使 DFC 在低分辨率特征空间运行, 此时计算量降至原来的  $1/4$ ; 随后通过上采样恢复特征图分辨率, 确保其与 Ghost 分支的输出尺寸匹配。这种设计在保证 DFC 功能完整性的同时, 有效控制了



$$O = Sigmoid(A) \odot V(X). \quad (1)$$


Diagram illustrating the proposed parallel attention mechanism:

- input**: A 7x7 grid with a single dark brown patch (focused patch) at the center.
- Ghost branch**: Processes the input to produce a 5x5 grid with a 3x3 dark brown patch (focused patch) at the center.
- Attention branch**: Processes the input to produce a 7x7 grid with a 3x3 dark brown patch (focused patch) and a 5x5 light orange patch (participated patches) surrounding it.
- multiply**: The outputs of the Ghost branch and Attention branch are multiplied to produce the final **output**.
- output**: A 7x7 grid with a 3x3 dark brown patch (focused patch) and a 5x5 light orange patch (participated patches) surrounding it.

Legend:

- Dark brown square: focused patch;
- Light orange square: participated patches in ghost branch;
- Light orange square: participated patches in attention branch.

**Fig. 4** Information flow processes

The diagram illustrates the proposed channel-wise fusion module. It takes an input of  $c_1$  channels and splits it into two parallel paths. The top path uses a convolutional layer (Conv) to reduce the number of channels to  $c_1/2$ . The bottom path uses a depthwise convolutional layer (DWConv) to also reduce the number of channels to  $c_1/2$ . The outputs from both paths are then concatenated (concat) and finally shuffled (shuffle) to produce the output of  $c_2$  channels.

**Fig. 5 GSConv Schematic**

The diagram illustrates the GSCNet architecture. It starts with an input of  $c_1$  channels with spatial dimensions  $h$  and  $w$ . This input passes through a Conv layer. The output of this Conv layer is then processed by two GSCConv layers in sequence. A residual connection branches off from the output of the first Conv layer and is added to the output of the second GSCConv layer at a Concat layer. Finally, the output of the Concat layer passes through a second Conv layer to produce the output with  $c_2$  channels.

**Fig. 6** VOVGCSP structure diagram

$$\frac{T - (x_l - x_r)}{T} = \frac{Z - f}{Z}, \quad (2)$$

– 100 –

$$Z = \frac{fT}{x_l - x_r} = \frac{fT}{d}, \quad (3)$$

式中,  $d$  为左右相机的像素点视差值。

由视差值即可解得  $P$  的深度信息  $Z$ 。

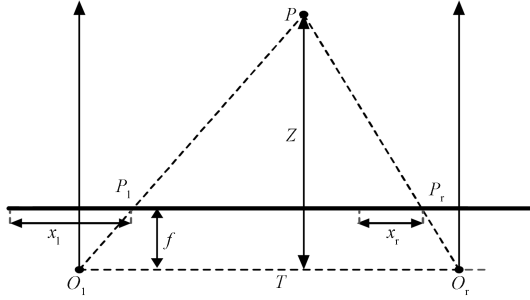


图 7 双目测距原理图

Fig. 7 Binocular ranging principle diagram

### 3.2 双目相机参数标定

在 MATLAB 中, 采用张正友标定法来标定双目相机, 结果如图 8 所示。通过棋盘格的角点分布和标定板的空間位置来验证标定参数的准确性。若标定后的角点坐标与理论值高度吻合, 则表明相机内参和外参的标定结果可靠, 进而为后续的双目测距实验提供基础。如图 9 所示, 标定后的相机误差在 0.4 个像素左右, 小于 1 个像素, 表明标定结果准确可靠。

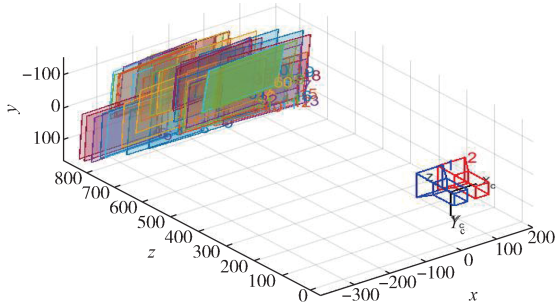


图 8 双目相机的标定结果

Fig. 8 Binocular camera calibration results

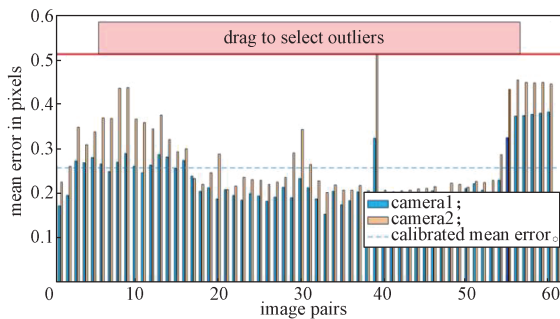


图 9 误差结果

Fig. 9 Error result

双目相机参数如下:

左右摄像机的内参矩阵  $M_r$  和  $M_l$  分别为

$$M_r = \begin{bmatrix} 600.3692 & -2.4631 & 657.2625 \\ 0 & 602.1253 & 301.4275 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

$$M_l = \begin{bmatrix} 592.9991 & -3.4200 & 637.3050 \\ 0 & 594.0315 & 300.5259 \\ 0 & 0 & 1 \end{bmatrix}; \quad (5)$$

以左相机坐标系作为参考坐标系, 旋转矩阵 ( $R$ ) 和平移矩阵 ( $T$ ) 表示右相机相对于左相机的空间位姿关系, 即

$$R = \begin{bmatrix} 0.9998 & 0.0002 & -0.0178 \\ -0.0003 & 0.9999 & 0.0015 \\ 0.0178 & -0.0016 & 0.9999 \end{bmatrix}, \quad (6)$$

$$T = [-60.9817 \ 0.2011 \ -2.5191]. \quad (7)$$

### 3.3 双目立体矫正

在双目视觉系统中, 由于相机安装时难以保证绝对水平, 两相机的成像平面往往存在一定角度偏差。这种非共面性会导致立体匹配困难, 进而影响深度计算的准确性。为此, 采用立体校正方法, 通过数学变换将左右相机的成像平面重投影至同一平面上。立体校正示意图如图 10 所示。

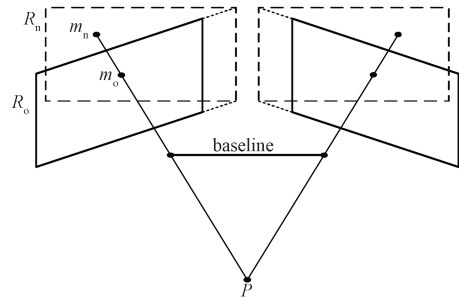


图 10 立体校正示意图

Fig. 10 Stereoscopic correction diagram

通过标定中得到的参数来实现相机矫正, 即通过旋转矩阵  $R$  调整左右相机位置, 使其处于同一水平面上。为得到左右旋转矩阵, 引入新矩阵  $R_{\text{rect}}$ ,

$$R_{\text{rect}} = \begin{bmatrix} \vec{e}_1^T \\ \vec{e}_2^T \\ \vec{e}_3^T \end{bmatrix}, \quad (8)$$

$$\vec{e}_1 = \frac{\vec{T}}{\|\vec{T}\|}, \quad (9)$$

$$\bar{e}_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} (-T_y, T_x, 0)^T, \quad (10)$$

$$\bar{e}_3 = \bar{e}_1 \times \bar{e}_2, \quad (11)$$

式中,  $T_x$ 、 $T_y$  为横纵轴上的分量。

由此求得左右旋转矩阵分别为:

$$R_l = R_{\text{rect}} r_l, \quad (12)$$

$$R_r = R_{\text{rect}} r_r. \quad (13)$$

式中:  $R_l$  和  $R_r$  为左右相机在立体校正后的旋转矩阵;  $r_l$  和  $r_r$  为左右相机在标定阶段得到的原始外参旋转矩阵。

相机矫正主要是消除图像畸变。立体校正双目图像结果如图 11 所示。可见, 矫正后的图像畸变得到较好改善, 图像几乎位于同一平面上。

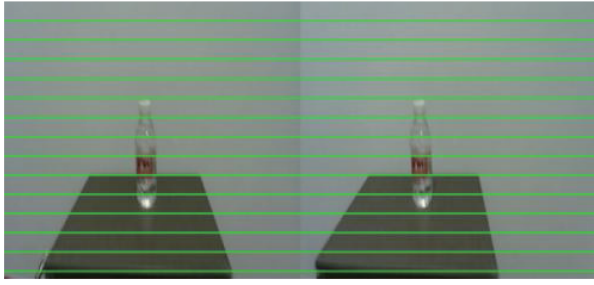


图 11 立体校正双目图像

Fig. 11 Stereoscopic correction binocular images

## 4 实验与结果分析

### 4.1 评价指标

为了验证模型改进效果, 本研究自制了废旧塑料瓶数据集, 并基于此做消融实验。使用平均准确度 (average precision, AP) 和平均精度均值 (mean average precision, mAP)<sup>[20]</sup> 作为评价模型性能的指标, 使用帧率 (frame per second, FPS) 即每秒内可以处理的图像数量作为评估模型检测速度的指标。平均准确度确定了目标的准确率 ( $P$ ) 和召回率 ( $R$ )。

$$P = \frac{TP}{FP + TP}, \quad (14)$$

$$R = \frac{TP}{TP + FN}, \quad (15)$$

$$AP = \int_0^1 P(R) dR, \quad (16)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i. \quad (17)$$

式中:  $TP$  表示正确样本中被正确识别的样本数;  $FP$  表示错误样本被错误的识别为正确样本的数量, 即为

误检;  $FN$  表示正确样本被识别为错误样本的数量, 即假的负样本;  $P(R)$  表示准确率关于召回率的函数。

### 4.2 实验环境及参数配置

实验环境如表 1 所示。采用 64 位 Windows11 操作系统, 硬件配置为 i7-12700H+RTX 3060+6G。

表 1 实验环境

Table 1 Experimental environment

名称	版本
操作系统	Windows 11
CPU	Intel Core i7-12700H
GPU	NVIDIA RTX 3060
RAM	6G
CUDA	11.0
深度学习框架	Pytorch 1.8.1

YOLOv5 系列模型的默认图像输入尺寸为  $640 \times 640$ , 这一尺寸在计算效率和特征提取能力之间取得了较好的平衡。在 GPU 显存容量限制下, Batch Size 设为 3, 以平衡训练稳定性和梯度更新准确性。YOLOv5 在 COCO 等大型数据集上的默认训练 Epoch 为 300, 这一设置能够保证模型充分收敛。本研究的自制数据集规模较小, 但通过数据增强扩充了样本多样性, 因此仍采用 300 Epoch, 以确保模型学习到鲁棒特征。

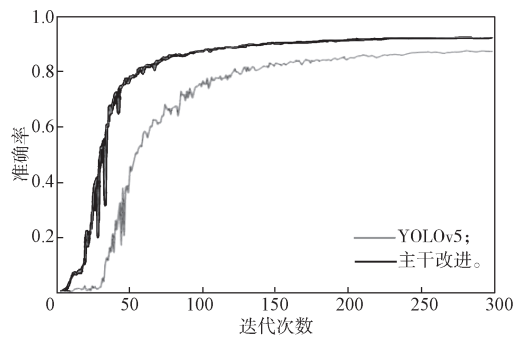
### 4.3 消融实验

#### 4.3.1 改进算法与 YOLOv5 对比分析

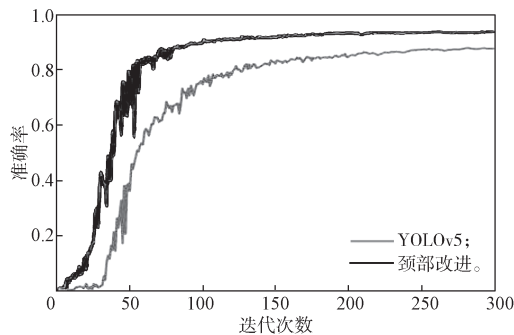
为验证改进 YOLOv5 模型的有效性, 将改进的 YOLOv5 模型与原始的 YOLOv5 模型在自制的废旧塑料瓶数据集上进行对比消融实验。原始数据集包含不同拍摄角度、光照条件及复杂场景的废旧塑料瓶图像, 其中单目标图像约占 60%, 多目标图像约占 40%。为扩大数据量, 通过图像的几何变换等方法进行数据增强, 如翻转变换、旋转变换、缩放变换、对比度变换等, 并进行了像素值归一化和标准化处理。此外, 使用 LabelImg 工具标注边界框。按 7:3 比例将自制的废旧塑料瓶数据集划分为训练集和测试集, 并通过人工校验确保标注质量, 剔除模糊或畸变的无效图像, 最终得到训练集 3150 张、测试集 1350 张。

#### 1) 主干网络改进

分析主干网络两种改进方式前后的模型性能。首先将原模型中的 CBS 模块替换为 GhostBottleneck; 在此基础上, 再引入 GhostNetV2。图 12 展示了模型改进前后的性能, 具体数据如表 2 所示。



a) GhostBottleneck



b) GhostBottleneck+GhostNetV2

图 12 主干网络改进前后的性能对比图

Fig. 12 Performance comparison before and after backbone network improvement

表 2 主干网络改进前后的数据对比

Table 2 Data comparison before and after backbone network improvement

模型	mAP/%	参数量 / 个	FPS/( 帧·s <sup>-1</sup> )
YOLOv5	87.92	7 012 888	96
GhostBottleneck	91.15	6 714 768	98
GhostBottleneck+GhostNetV2	92.66	6 414 768	99

由图 12 和表 2 可知, GhostBottleneck 模块的引入提升了模型性能, mAP 由 87.92% 提升到 91.15%, 参数量由 7 012 888 下降到 6 714 768, FPS 由 96 帧/s 提升到 98 帧/s; GhostBottleneck 与 GhostNetV2 的同时引入, 模型整体性能进一步提升, mAP 提升到 92.66%, 参数量下降到 6 414 768, FPS 提升到 99 帧/s。这证明了 GhostBottleneck+GhostNetV2 改进主干网络的有效性。

## 2) 颈部网络改进

对于颈部网络, 将 GSConv 与基于 GSConv 的 VOVGCSP 分别替换原模型中的 CBS、C3 模块。图 13 展示了颈部网络改进前后的性能, 具体数据如表 3 所示。从图 13 和表 3 可以看出, 颈部网络改进后, 模型的 mAP 由 87.92% 提升到 93.80%, 参数量

由 7 012 888 下降到 6 313 298, FPS 由 96 帧/s 提升到 101 帧/s。可见, 相对于主干网络的改进, 颈部网络改进后模型性能更好。

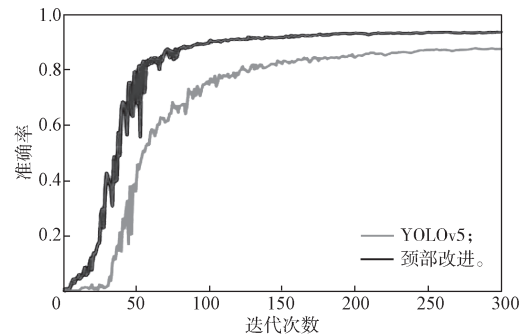


图 13 颈部网络改进前后的性能对比图

Fig. 13 Performance comparison before and after neck network improvement

表 3 颈部网络改进前后的数据对比图

Table 3 Data comparison before and after neck network improvement

模型	mAP/%	参数量	FPS/( 帧·s <sup>-1</sup> )
YOLOv5	87.92	7 012 888	96
GSConv+VOVGCSP	93.80	6 313 298	101

## 3) 主干网络与颈部网络改进

将主干网络 (GhostBottleneck+GhostNetV2) 和颈部网络 (GSConv+VOVGCSP) 的优化策略同时应用, 构建了轻量化改进模型 YOLOv5-LW。YOLOv5-LW 的性能如图 14 所示, 具体数据见表 4。

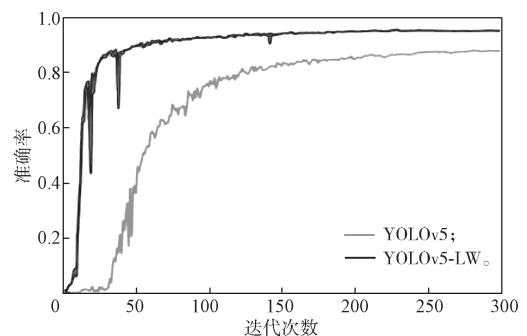


图 14 轻量化改进前后的性能对比图

Fig. 14 Performance comparison before and after lightweight improvement

表 4 轻量化改进前后的数据对比图

Table 4 Data comparison before and after lightweight improvement

模型	mAP/%	参数量	FPS/( 帧·s <sup>-1</sup> )
YOLOv5	87.92	7 012 888	96
YOLOv5-LW	95.39	5 933 320	105



由图 14 和表 4 可知, YOLOv5-LW 的 mAP 达到 95.39%, 较原始 YOLOv5 模型的提升了 7.47%。这一结果表明, 主干网络与颈部网络的协同优化显著增强了模型的特征提取与融合能力。YOLOv5-LW 在精度、参数量与速度 3 项指标上均优于原模型, 证实了本文改进策略的有效性。本模型可为资源受限场景(如嵌入式设备)下的实时目标检测提供技术参考。

#### 4.3.2 检测对比

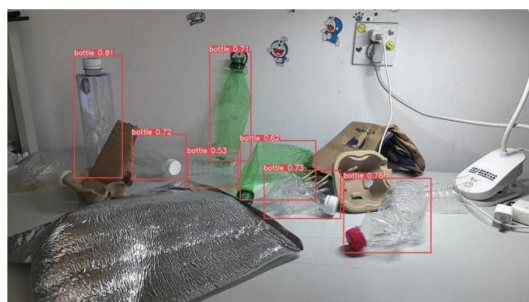
用 YOLOv5 与 YOLOv5-LW 对自制的废旧塑料瓶数据集中测试集 1000 张图像进行检测。其中, YOLOv5 的识别率仅为 51%, 而 YOLOv5-LW 的识别率达到 83%。此外, 用 YOLOv5 与 YOLOv5-LW 还进行了多目标检测。通过网络搜集以及手机拍摄获得 1000 张多目标图像用于测试。某图像的识别结果

如图 15 所示。由此可见, 在诸如遮挡、严重变形等特殊情况下, YOLOv5-LW 对废旧塑料瓶的检测能力均优于 YOLOv5。

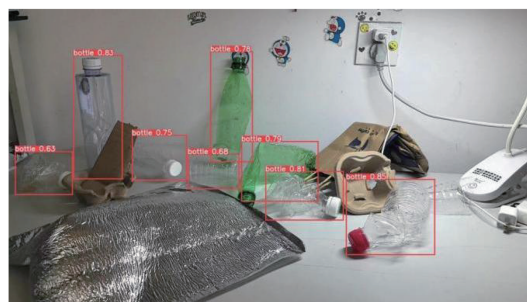
#### 4.4 双目测距实验

先将标定得到的参数应用到双目测距实验中, 再使用 YOLOv5-LW 对自制的废旧塑料瓶数据集中不同距离(450, 550, 650, 750, 850, 950, 1050 mm)的目标进行检测, 通过立体匹配 BM(bidirectional matching)算法得到被测目标深度。双目测距实验结果如图 16~17 所示, 具体数据如表 5 所示。

在深度图 16 中, 可以看到较为完整的废旧塑料瓶轮廓。由表 5 可知, 实验最大误差为 7 mm。可见, 改进的深度学习算法 YOLOv5-LW 可以在不同距离下正确地识别目标, 整体方案可行。



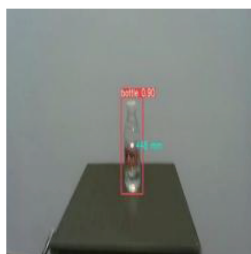
a) YOLOv5



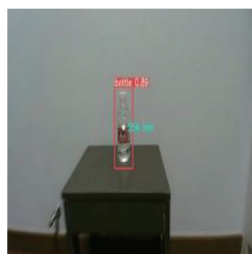
b) YOLOv5-LW

图 15 两种模型的多目标检测结果

Fig. 15 Multi-object detection results for two models



a) 450 mm



b) 550 mm



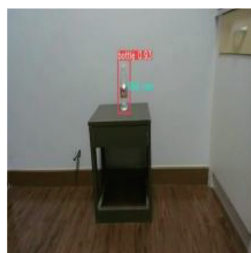
c) 650 mm



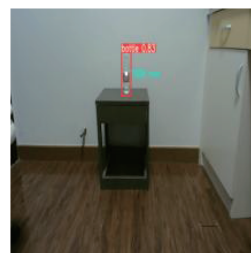
d) 750 mm



e) 850 mm



f) 950 mm



g) 1050 mm

图 16 不同距离的目标检测结果

Fig. 16 Object detection results at different distances

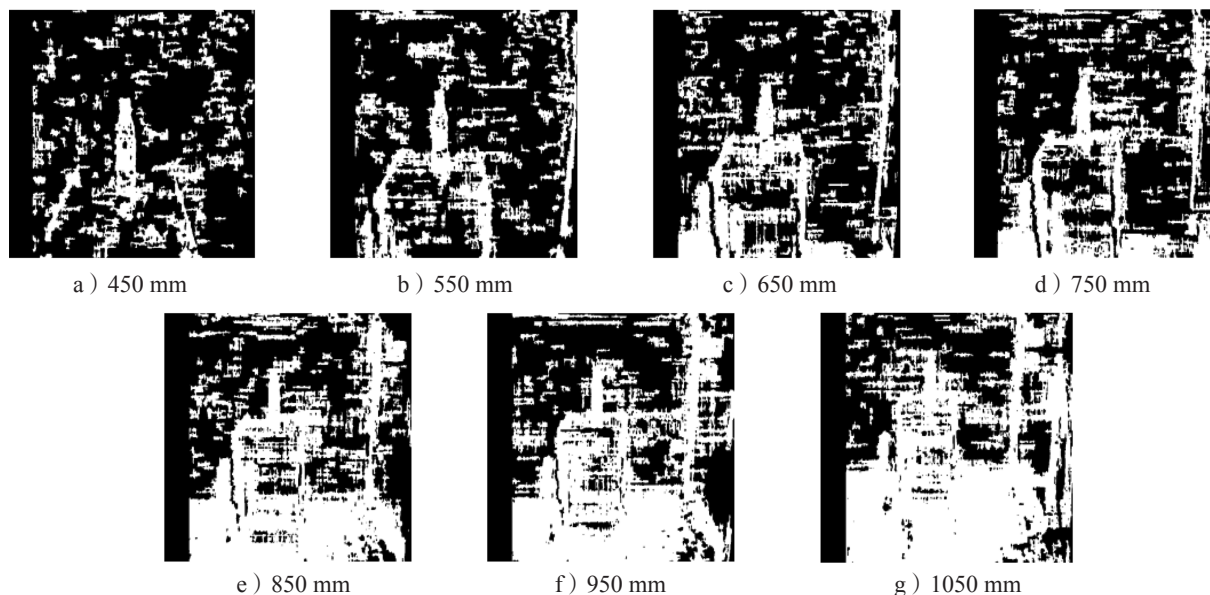


图 17 不同距离的目标检测深度图

Fig. 17 Target detection depth map at different distances

表 5 测距结果

Table 5 Ranging result

实际距离 /mm	测量距离 /mm	绝对误差 /mm
450	448	-2
550	554	4
650	657	7
750	751	1
850	853	3
950	952	2
1050	1056	6

## 5 结论

针对 YOLOv5 的模型参数多、检测速度慢, 以及无法在移动设备中快速检测目标的缺点, 本研究对 YOLOv5 进行了轻量化改进。

1) 在主干网络, 将 GhostBottleneck 替换 CBS 模块, 并引入 GhostNetV2; 在颈部网络, 将 GSConv 与基于 GSConv 的 VOVGCSP 分别替换 CBS、C3 模块; 最终构建了轻量化改进模型 YOLOv5-LW。在自制的废旧塑料瓶数据集上进行消融实验。实验结果表明, 与 YOLOv5 相比, YOLOv5-LW 的 mAP 由 87.92% 提升到 95.39%, 参数量由 7 012 888 下降到 5 933 320, FPS 由 96 帧/s 提升到 105 帧/s。可见, YOLOv5-LW 在检测准确度、检测速度上均有一定程度的提升。在复杂情景下, YOLOv5-LW 对多目标的检测性能也优

于 YOLOv5。

2) 将标定得到的参数校正相机, 对距离在 450~1050 mm 之间采集到的图像进行立体匹配。双目测距实验结果表明, 检测的最大误差为 7 mm, 相对误差约为 1%, 此误差符合要求。

可见, 所提出的双目视觉识别与定位方法能够在满足精度要求的前提下, 实现对废旧塑料瓶的稳定检测与距离测量, 为后续废旧塑料瓶的自动分拣与回收提供了可靠的技术支持。

## 参考文献:

- [1] 王策仁. 基于深度迁移学习垃圾分类算法研究应用 [D]. 贵阳: 贵州大学, 2022.
- [2] 徐 丽. 基于深度学习的垃圾分类系统的设计与开发 [D]. 杭州: 浙江大学, 2021.
- [3] 蔡 勇. 面向绿色环境的垃圾检测算法研究 [D]. 南京: 南京邮电大学, 2022.
- [4] 董利君, 杨 婧, 付雯雯, 等. 基于垃圾分拣人工及自动化的研究 [J]. 电脑知识与技术, 2022, 18(2): 118-119, 129.
- [5] 宋 科, 叶 礼, 崔旭升. 基于视觉定位的机器人天地盒自动成型机设计 [J]. 包装与食品机械, 2023, 41(5): 91-96.
- [6] 高思岩. 基于无人机遥感的建筑垃圾机器辨识与变化检测方法研究 [D]. 北京: 北京建筑大学, 2022.
- [7] 陈佳欣, 仲子鑫. 基于机器视觉的家庭智能垃圾桶设

- 计研究[J]. 网络安全技术与应用, 2021(7): 152-153.
- [8] 吴 健, 陈 豪, 方 武. 基于计算机视觉的废物垃圾分析与识别研究[J]. 信息技术与信息化, 2016(10): 81-83.
- [9] KANG Z, YANG J, LI G L, et al. An Automatic Garbage Classification System Based on Deep Learning[J]. IEEE Access, 2020, 8: 140019-140029.
- [10] 向 伟, 史晋芳, 刘桂华, 等. 改进 CaffeNet 模型在水面垃圾识别中的应用[J]. 传感器与微系统, 2019, 38(8): 150-152, 156.
- [11] 文灿华, 李 佳, 董 雪. 基于 Faster RCNN 的生活垃圾智能识别[J]. 激光与光电子学进展, 2020, 57(20): 139-145.
- [12] 戴 康. 复杂场景下的垃圾分类检测方法研究[D]. 南京: 南京理工大学, 2021.
- [13] 王心鹏. 基于 YOLOv3 的垃圾检测算法研究[D]. 重庆: 重庆邮电大学, 2022.
- [14] 孙潇鹏, 林 建, 郭海龙, 等. 基于“深度”神经网络的贮藏期柚果品质检测模型研究[J]. 包装与食品机械, 2023, 41(5): 9-16, 22.
- [15] 杨 萌, 张爱军, 潘文松. 基于改进 YOLOv5 算法的纸袋缺陷检测[J]. 组合机床与自动化加工技术, 2024(3): 105-108, 113.
- [16] 陈 实. 轻量化改进的 YOLOv5 与道路病害实时检测[D]. 武汉: 武汉大学, 2022.
- [17] 史徐睿. 基于深度学习的火灾实时检测研究[D]. 杭州: 浙江大学, 2022.
- [18] 陈 锦, 蒋锦华, 庄丽萍, 等. 融入部分卷积和解耦全连接注意力的轻量级人体姿态估计[J]. 现代信息科技, 2023, 7(23): 93-98, 105.
- [19] 张佳璐. 基于深度学习的道路车辆检测算法研究[D]. 临沂: 临沂大学, 2023.
- [20] 王晓红, 张 微. MNTH-YOLOv8: 一种用于食品包装中蚊虫高效检测的深度学习算法[J]. 包装学报, 2024, 16(3): 91-98.

(责任编辑: 邓 彬)

## Binocular Visual Identification and Positioning Method for Used Plastic Bottles Based on YOLOv5

LI Xinyan, GONG Xue, FU Bin, CUI Gongzhuo, LIU Jingyu, YANG Hongyang

(School of Light Industry, Harbin University of Commerce, Harbin 150028, China)

**Abstract:** In order to improve the accuracy of visual recognition and localization of used plastic labeled bottles, a lightweight binocular visual target detection algorithm based on the improved YOLOv5 is proposed. Firstly, in the backbone network, the original CBS and C3 modules are respectively replaced by the GhostBottleNeck module and the GhostNetV2 module. Secondly, GSConv and GSConv-based VOVGCSP modules are introduced into the neck network. The dataset is trained by field real measurements to be used for training the improved YOLOv5 model. Based on the optimized algorithm, a binocular camera is used to investigate the used plastic bottle ranging system. Results show that 1) the improved YOLOv5 model increases the accuracy from 87.92% to 95.39%, the number of parameters decreases from 7 012 888 to 5 933 320, and the FPS improves from 96 frame/s to 105 frame/s when comparing with the original model. The replacement network effectively reduces the number of parameters of the model, and the accuracy of the improved model improves by 7.47%. 2) The maximum error of detection is about 7 mm through binocular ranging experiment, and the relative error is within 1%, which meets the requirements. The identifying and locating accuracy both meet the requirements of real-time processing, which can provide technical support for the development of intelligent recycling equipment.

**Keywords:** lightweight; YOLOv5 model; used plastic bottle; binocular calibration