

基于注意力机制的轻量 YOLOv5 识别定位算法

宋建辉, 李亚洲, 刘砚菊, 刘晓阳

(沈阳理工大学 自动化与电气工程学院, 沈阳 110159)

摘要: 为解决医疗看护环境下桌面生活物品检测效果不佳、定位误差较大的问题, 提出一种基于 YOLOv5 的改进模型。首先, 在主干网络末端使用坐标注意力 (coordinate attention, CA) 机制, 使算法能够捕获跨通道、跨方向和位置的信息, 提高算法的识别精度; 然后, 引入 GhostConv 卷积减少模型参数量, 使模型更加轻量化, 提高检测速度; 最后, 使用 SIoU 替换原算法的定位损失函数, 使定位损失计算考虑到真实框与预测框的方向差异, 有助于提升模型的稳定性。在 COCO 数据集部分物品种类上进行多次对比实验, 结果表明, 与原算法相比较, 改进算法的精确率和召回率分别提高了 4.1% 和 1.3%, 在交并比为 0% ~ 50% 和 50% ~ 90% 时的均值平均精度分别提高了 2.7% 和 3.9%, 参数量减少了 16.9%, 每秒传输帧数提高了 0.47 帧, 平均定位误差在 X 轴方向上减小了 0.29 mm、在 Y 轴方向上减小了 0.14 mm。

关键词: YOLOv5; GhostConv 卷积; 注意力机制; 损失函数; 目标识别

中图分类号: TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1003-1251.2024.03.002

Lightweight YOLOv5 Identification and Localization Algorithm Based on Attention Mechanism

SONG Jianhui, LI Yazhou, LIU Yanju, LIU Xiaoyang

(Shenyang Ligong University, Shenyang 110159, China)

Abstract: In order to solve the problem of poor detection effect and large positioning error of desktop living objects in medical nursing environment, an improved model based on YOLOv5 is proposed. Firstly, the coordinate attention (CA) mechanism is used at the end of the backbone network to enable the algorithm to capture information across channels, directions and locations, and improve the identification accuracy of the algorithm. Then, GhostConv convolution is introduced to reduce the number of model parameters, make the model more lightweight, and improve the detection speed. Finally, the positioning loss function of the original algorithm is replaced with SIoU, so that the positioning loss calculation takes into account the direction difference between the real box and the predicted box, which is helpful to improve the stability of the model. The results show that compared with the original algorithm, the improved algorithm increases the accuracy and recall by 4.1% and 1.3%, respectively, the mean accuracy of 0% ~ 50% and 50% ~ 90% in the intersection and union ratio is improved by 2.7% and 3.9% respectively, the number of parameters by 16.9%, the number of frames per second is increased by 0.47 frames, and the average positioning error by 0.29 mm in the X-axis direction, which is reduced by 0.14 mm in the Y-axis direction.

Key words: YOLOv5; GhostConv convolution; attention mechanism; loss function; target identification

示非线性激活。

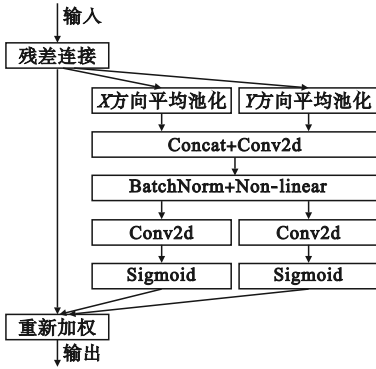


图2 CA注意力机制结构

Fig.2 CA attention mechanism structure

首先在空间层面将特征图按照 X 、 Y 两个方向对通道信息进行编排,保留特征的位置信息和空间层面之间的联系;然后对 X 、 Y 方向的特征信息进行拼接和一次卷积,经过非线性操作后再次拆分为两个方向的特征信息,分别经过激活函数得到不同空间方向的特征权重信息;最后将输入的特征信息和得到的不同空间特征信息相融合,提高算法对图像特征的代表能力。

1.2 引入轻量化 GhostConv 卷积模块

相比改进前算法,添加 CA 注意力机制后的 YOLOv5 算法计算复杂度增大,故引入 GhostConv 替换 Conv,在不影响检测精度的情况下减少模型的参数量,提高检测速度。

由常规卷积 Conv 获取的特征层存在某些特征图的特征信息较为相似的情况,而 GhostConv 卷积则通过消除冗余特征信息减少获取特征信息的参数量。GhostConv 结构如图 3 所示。GhostConv 先通过常规卷积提取特征信息,获取到一般特征图,然后采用廉价的线性变换 $\phi_1, \phi_2, \dots, \phi_i$ 得到 Ghost 特征图,拼接基础特征图和 Ghost 特征图作为输出。

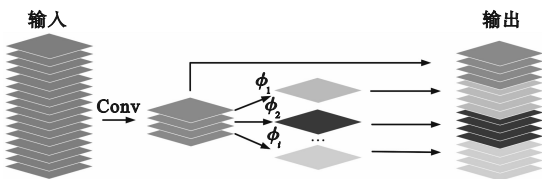


图3 GhostConv 结构图

Fig.3 GhostConv structure diagram

1.3 改进定位损失函数

YOLOv5 算法的损失函数包括类别损失、置信度损失和定位损失,其中定位损失函数为 CIoU,该损失函数没有考虑到最终输出目标框和标

注框之间的方向差异,导致定位损失收敛速度较慢,而 SIoU^[13] 引入了两个框的角度差异,可以加快定位损失收敛速度,提高精确度。SIoU 定位损失函数由四部分组成:角度损失 α 、距离损失 β 、形状损失 γ 、交并比 (IoU) 损失。各损失计算式如式 (1) ~ (4) 所示。

$$\alpha = 1 - 2\sin^2(\omega - \pi/4) \quad (1)$$

$$\beta = \sum_{t=x,y} (1 - e^{-(2-\alpha)\rho_t^2}) \quad (2)$$

$$\gamma = \sum_{s=w,h} (1 - e^{-k_s})^\theta \quad (3)$$

$$\text{IoU} = (B_F \cap B_L) / (B_F \cup B_L) \quad (4)$$

式中: ω 表示预测框中心像素点和标注框中心像素点的连线与水平方向之间的夹角; x 和 y 分别表示 X 轴和 Y 轴方向; ρ_t 表示预测框和标注框中心像素坐标在 X 轴(或 Y 轴)方向上的差与两框外接长方形水平边(或垂直边)长的比值; θ 表示对形状损失的关注度; w 和 h 分别表示宽度和高度方向; k_s 表示预测框宽(或高)和标注框宽(或高)的差值绝对值与最大宽矩(或高矩)的比值; B_F 表示标注框; B_L 表示预测框。

定位损失 L_b 计算式为

$$L_b = 1 - \text{IoU} + \frac{\gamma + \beta}{2} \quad (5)$$

1.4 目标定位

目标定位需将像素坐标转换为世界坐标,像素坐标和世界坐标的关系为

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{A} \times \mathbf{B} \times \mathbf{Z}_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (6)$$

式中: $\mathbf{A} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0} & 1 \end{bmatrix}$,为相机外参矩阵,其中 \mathbf{R} 为旋

转矩阵, \mathbf{T} 为偏移矩阵; $\mathbf{B} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$,为相

机内参矩阵, f_x 和 f_y 分别表示相机水平和垂直方向的固有焦距参数, u_0 和 v_0 分别表示相机主点在图像水平和垂直方向上的偏移量; $[u \ v \ 1]^T$ 是目标检测框的中心像素坐标; $[X_w \ Y_w \ Z_w \ 1]^T$ 是目标的世界坐标; Z_c 为成像距离,本文使用的相机相对于桌面的高度为 520 mm,故成像距离为固定值,即 $Z_c = 520 \text{ mm}$ 。

摄像头镜头中心相对于机械臂底座中心的坐标为 $(0, 125, 520)$,故偏移矩阵 $\mathbf{T} = [0 \ 125 \ 520]$ 。

摄像机镜头的坐标方向和机械臂底座的坐标方向相反,则旋转矩阵为

$$\mathbf{R} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (7)$$

将 \mathbf{T} 矩阵和 \mathbf{R} 矩阵带入式(6),得到相机的外参矩阵为

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 125 \\ 0 & 0 & -1 & 520 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

相机标定是确定相机内参的过程,本文实验采用张正有标定法^[14]对相机进行标定。标定板样式设定为 10×7 方格,相机标定需要多张不同视角的标定板图,使用 Python 程序进行标定。先提取标定板上的每个角点,通过每个标定板角点坐标之间的差值比较,计算得到相机的内参矩阵为

$$\mathbf{B} = \begin{bmatrix} 517.043 & 807 & 0 & 321.322 & 014 & 0 \\ 0 & 520.712 & 097 & 239.061 & 626 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

本文目标检测结束后得到目标框中心的像素坐标,将像素坐标、外参矩阵 \mathbf{A} 、内参矩阵 \mathbf{B} 和 Z_c 代入式(6),即可得到对应的世界坐标。

2 模型训练

2.1 数据集介绍

COCO 数据集是一个可用于目标定位、语义分割和图像标题生成的大规模数据集,共有 33 多万张图像,其中 22 万张是有标注的图像,包含 150 万个目标、80 个目标类别。本文对桌面物体进行近景俯拍、检测和定位,在 COCO 数据集中挑选 9 500 张包含苹果、橘子、勺子和手机的图片,将最终的数据集按照 6:2:2 的比例分配到训练集、验证集和测试集。

2.2 训练环境介绍

本文训练过程使用的硬件配置为 10 GB NVIDIA GeForce RTX3080 显卡,深度学习框架为 PyTorch1.10.0,Python 版本为 3.8.16,CUDA 版本为 12.0,操作系统为 Ubuntu 系统。

2.3 算法训练损失参数和网络训练参数设置

算法的总损失 L_g 计算式为

$$L_g = L_j + L_b + L_s \quad (10)$$

式中: L_s 表示置信度损失; L_b 表示预测的定位损失; L_j 表示类别损失。

为保证对比实验的数据结果有效,本文使用相同的数据集和训练参数分别对原 YOLOv5 算法和改进 YOLOv5 算法进行训练。训练参数设置:训练轮数统一采用 300,每次训练抓取的数据样本数量设置为 64,超参数采用默认数值。

2.4 定位实验各坐标系参数设置

目标定位是将像素坐标转换为对应的世界坐标,图 4 为相机视角俯视图,像平面宽为 640 像素,高为 480 像素,图中 A 点为像素平面中心点(320,240),且与相机光轴重合。图 5 为整体结构侧视图,由图 5 可见,相机坐标系与机械臂底座坐标系的 X 轴和 Z 轴方向相反且处于同一平面,两坐标系原点在 X 轴和 Z 轴上分别有 125 mm 和 520 mm 的距离,相机 Z 轴和机械臂 X 轴的延长线交于世界坐标系的原点(B 点)。

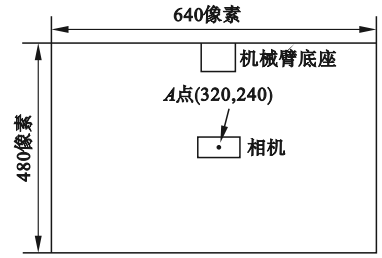


图 4 相机视角俯视图

Fig. 4 Top view of camera perspective

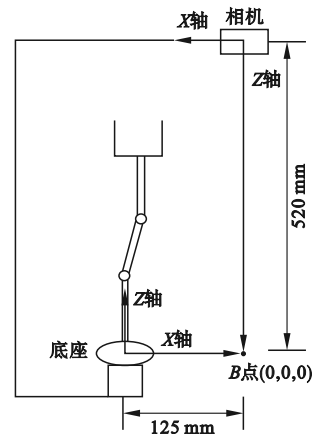


图 5 整体结构侧视图

Fig. 5 Side view of the overall structure

2.5 评价指标

本文旨在提高算法精确度和减小定位误差,为验证模型的性能,采用精确率(P)、召回率(R)和均值平均精度(mAP)评估模型的检测性能,采用平均定位误差(PE)评价模型的定位性能。

算法精确率、召回率和均值平均精度的计算式分别为

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$mAP = \frac{1}{d} \sum_{i=1}^d AP_i \quad (13)$$

式中: TP 表示被正确检测出的物体个数; FP 表示被错误检测出的物体个数; FN 表示未被检测出的物体个数; d 表示种类数; AP_i 表示种类 i 的平均检测精度。

算法平均定位误差 (PE) 是改进算法与原算法平均定位误差的差值, 其计算式为

$$PE = \frac{1}{n} \sum_{j=1}^n (PE_L - PE_F) \quad (14)$$

式中: n 表示实验定位的次数; PE_L 表示改进算法定位坐标与实际物体定位坐标之间的误差; PE_F 表示原算法定位坐标与实际物体定位坐标之间的误差。

3 实验结果与分析

3.1 改进算法实验结果

四种桌面物品在改进框架下的训练结果如表 1 所示, 其中 mAP_{50} 和 mAP_{50-95} 分别表示交并比在 0% ~ 50% 和 50% ~ 95% 时对应的 mAP 。

表 1 改进框架训练结果

Table 1 Improved frame training results

类别	标签数	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$
苹果	1 025	81.3	64.4	71.9	51.3
橘子	1 086	80.1	65.2	71.4	53.7
勺子	1 071	79.5	52.8	61.4	38.8
手机	1 209	85.7	64.4	71.6	49.0

由表 1 可知, 验证集中四个种类含有的标签数较为接近, 且各种类的 P 、 R 、 mAP_{50} 和 mAP_{50-95} 值分别高于 79%、52%、61% 和 38%, 表明改进算法对验证集有较好的识别效果。

图 6 是改进 YOLOv5 算法在训练 300 轮后的损失收敛图。由图 6 可知, 定位损失、类别损失和置信度损失分别在 50 轮、150 轮和 30 轮之前快速收敛, 且分别在 150 轮、200 轮和 100 轮之后保持稳定速率收敛。

图 7 为改进算法的一些可视化检测结果, 可见, 检测置信度最高达到 97%。

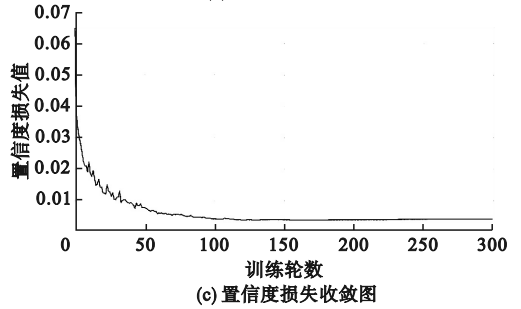
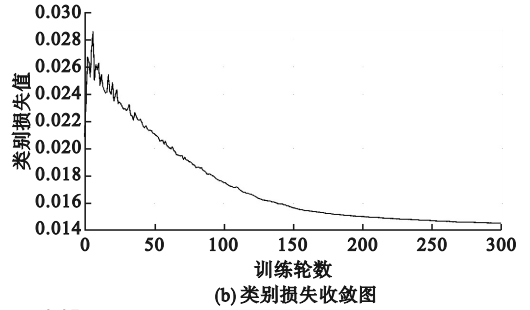
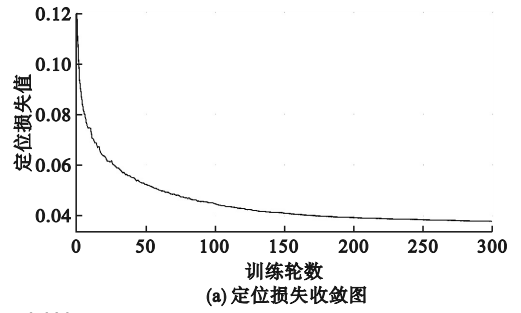


图 6 改进 YOLOv5 算法的训练损失收敛图

Fig. 6 Convergence diagram of training loss of the improved YOLOv5 algorithm

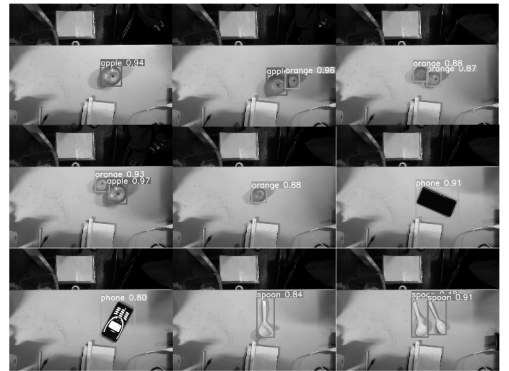


图 7 改进 YOLOv5 算法的可视化检测结果

Fig. 7 Visual detection results of the improved YOLOv5 algorithm

3.2 检测算法对比实验

原 YOLOv5 算法与改进 YOLOv5 算法的对比实验结果如表 2 所示, 表中 FPS 表示每秒传输帧数, 用于评估目标检测速度。由表 2 可知, 改进 YOLOv5 模型的 P 值和 R 值较原算法分别提升了 4.1%、1.3%, mAP_{50} 值和 mAP_{50-90} 值分别提升了

2.7% 和 3.9%, 参数量减少了 16.9%, FPS 提高了 0.47 s⁻¹。

表 2 COCO 部分数据集训练对比结果

Table 2 Training comparison results of some COCO data sets

模型名称	P/%	R/%	mAP ₅₀ /%	mAP ₅₀₋₉₀ /%	参数量/B	FPS/s ⁻¹
YOLOv5	78.5	60.9	67.2	46.4	7 030 417	46.17
改进 YOLOv5	81.7	61.7	69.0	48.2	5 843 745	46.64

YOLOv5 原算法和改进算法的精确率 (P) 收敛曲线对比如图 8 所示。可以看出,改进算法的 P 值在 100 轮之后始终高于原算法,在 130~230 轮之间差距较大。

YOLOv5 原算法和改进算法的召回率 (R) 收敛曲线对比如图 9 所示。可以看出,改进算法的 R 值在 240 轮之后始终高于原算法,在 240 轮之前差距不大。

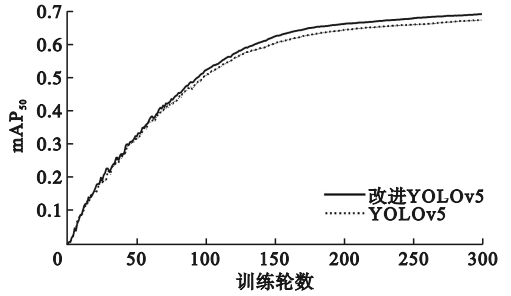


图 10 mAP₅₀ 收敛曲线

Fig. 10 mAP₅₀ convergence curve

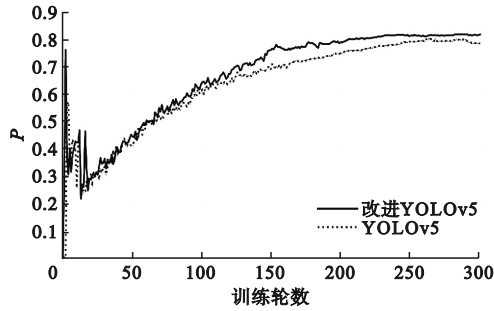


图 8 精确率收敛曲线

Fig. 8 Accuracy convergence curve

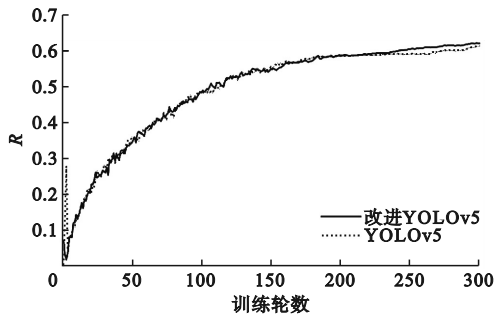


图 9 召回率收敛曲线

Fig. 9 Recall convergence curve

YOLOv5 原算法和改进算法的 mAP₅₀ 收敛曲线对比如图 10 所示。可以看出,改进算法的 mAP₅₀ 值始终高于原算法,并在 80 轮之后差距明显增大。

YOLOv5 原算法和改进算法的 P-R 曲线对比如图 11 所示。可以看出,改进算法的 P-R 曲线始终位于原算法 P-R 曲线的上方,证明改进算法较原算法有更好的检测效果。

将 YOLOv5 算法和一些常用的注意力机制结合起来,进行对比实验,实验结果如表 3 所示。

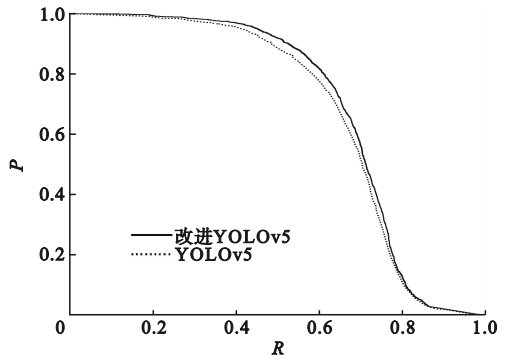


图 11 算法 P-R 曲线

Fig. 11 Algorithm P-R curve

表 3 算法常用注意力机制对比

Table 3 Comparison of common attention mechanisms of the algorithm

算法	P/%	R/%	mAP ₅₀ /%	mAP ₅₀₋₉₀ /%
YOLOv5	78.5	60.9	67.2	46.4
YOLOv5 + SE ^[15]	81.6	59.8	67.7	46.6
YOLOv5 + ECA ^[16]	79.5	60.7	68.0	47.0
YOLOv5 + CBAM ^[17-18]	79.9	61.1	67.6	46.3
YOLOv5 + CA	82.1	59.4	67.5	46.6

由表 3 可知,相较于原 YOLOv5 算法,各注意力机制对目标检测均有提升效果,但每种注意力机制提升的效果不同。加入 SE 注意力机制后,算法的 P 值和 mAP 值均有所提升,但提升程度不算最高;对 R 值提升最大的是 CBAM 注意力机制;对 mAP 提升最大的是 ECA 注意力机制;对 P 值

提升最大的是 CA 注意力机制,较原算法提升了 4.6%,该注意力机制对算法的 mAP_{50} 和 mAP_{50-90} 也有一定的提升效果,仅 R 值有所下降。本文注重提升算法的检测精确率,故 CA 注意力机制效果较佳。

3.3 消融实验

本文对原 YOLOv5 算法进行了三个方面的改

进,改进 YOLOv5 算法消融实验结果如表 4 所示。由表 4 可知:相比于其他几组实验,第 8 组实验得到的 R 值、 mAP_{50} 和 mAP_{50-90} 值均最高, P 值、参数量和 FPS 虽不是最佳,但较最佳值相差不大;第 2~8 组实验的结果均优于第 1 组实验。

表 4 改进算法消融实验结果

Table 4 Ablation experimental data of the improved algorithm

实验组数	YOLOv5	注意力机制	GhostConv 模块	修改损失函数	$P/\%$	$R/\%$	$mAP_{50}/\%$	$mAP_{50-90}/\%$	参数量/B	FPS/ s^{-1}
1	√				78.5	60.9	67.2	46.4	7 030 417	46.17
2	√	√			82.1	59.4	67.5	46.6	7 056 065	45.06
3	√		√		79.9	61.4	68.1	47.2	5 818 097	47.37
4	√			√	79.5	60.4	67.7	46.3	7 030 417	46.93
5	√	√	√		83.0	60.1	68.0	47.6	5 843 745	46.49
6	√	√		√	81.8	60.7	68.4	47.0	7 056 065	45.82
7	√		√	√	81.7	61.6	68.6	47.7	5 818 097	47.47
8	√	√	√	√	81.7	61.7	69.0	48.2	5 843 745	46.64

3.4 定位算法对比实验

目标定位的世界坐标由目标检测得到的中心像素坐标经转化公式得到,其与实际世界坐标偏差为定位误差。每轮训练的定位误差结果如表 5 所示。根据表 5 数据经计算可知:由原算法定位物体中心像素坐标求出的世界坐标与实际世界坐

标的 X 轴平均误差为 0.72 mm、 Y 轴平均误差为 0.71 mm;由改进算法定位物体中心像素坐标求出的世界坐标与实际世界坐标的 X 轴平均误差为 0.43 mm、 Y 轴平均误差为 0.57 mm;改进算法较原算法的目标定位平均误差在 X 轴上减小了 0.29 mm、在 Y 轴上减小了 0.14 mm。

表 5 目标定位世界坐标对比误差表

Table 5 Target positioning world coordinate comparison error table

实验序号	原算法检测的目标图像中心像素坐标	改进算法检测的目标图像中心像素坐标	目标实际世界坐标/mm	原算法定位的世界坐标/mm	改进算法定位的世界坐标/mm	原算法与实际世界坐标的误差/mm	改进算法与实际世界坐标的误差/mm
1	(338.5,272.0)	(337.5,274.0)	(16.77,33.89)	(17.28,32.89)	(16.27,34.89)	(0.50,-1.00)	(-0.50,1.00)
2	(341.5,274.5)	(341.5,275.0)	(19.79,35.89)	(20.29,35.39)	(20.29,35.89)	(0.50,-0.50)	(0.50,0.00)
3	(399.0,282.0)	(398.5,282.0)	(77.12,43.88)	(78.12,42.88)	(77.62,42.88)	(1.01,-1.00)	(0.50,-1.00)
4	(409.0,315.0)	(409.5,315.5)	(89.19,75.83)	(88.18,75.83)	(88.68,76.33)	(-1.01,0.00)	(-0.50,0.50)
5	(470.5,300.0)	(470.0,301.5)	(149.53,61.85)	(150.03,60.86)	(149.53,62.35)	(0.50,-1.00)	(0.50,0.50)
6	(342.0,290.5)	(343.5,290.0)	(21.80,50.87)	(20.80,51.37)	(22.30,50.87)	(-1.01,0.50)	(0.50,0.00)
7	(418.5,280.0)	(418.5,280.0)	(97.23,39.88)	(97.73,40.88)	(97.73,40.88)	(0.50,1.00)	(0.50,1.00)

4 结论与展望

本文提出了一种改进的 YOLOv5 目标检测算法:将主干网络最后的 Conv 替换为 CA 注意力机制,使算法能够捕获跨通道、跨方向和位置的信息,提高了算法的识别精度;引入 GhostConv 卷

积,减少了模型参数量,提高了检测速度;对算法的定位损失函数进行更改,将 CIoU 损失函数改为 SIoU 损失函数,考虑了目标框和真实框之间的角度差异,加快了损失收敛速度。实验结果表明,改进算法比原算法在识别精确率、召回率和均值平均精度 (mAP_{50} 和 mAP_{50-90}) 方面分别提高了 4.1%、1.3% 和 2.7%、3.9%,平均定位误差较原

算法在X轴方向上减小了0.29 mm、在Y轴方向上减小了0.14 mm。

为使病房看护系统更好地为病人服务,还需进行以下探索:

1)使算法与机械臂结合起来,实现机械臂的自动抓取和放置;

2)增加机械臂与人的互动功能,可以是语音互动或按键互动;

3)为算法打造一款可视化界面,使算法与外界的交流更加方便。

参考文献(References):

- [1] LEE H W. The study of mechanical arm and intelligent robot [J]. *IEEE Access*, 2020, 8: 119624 – 119634.
- [2] TAO Y T, ZHOU J. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking [J]. *Computers and Electronics in Agriculture*, 2017, 142: 388 – 396.
- [3] NAKANO A, NAGAMUNE K. A development of robotic scrub nurse system: detection for surgical instruments using faster region-based convolutional neural network [J]. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2022, 26(1): 74 – 82.
- [4] LAWAL M O. Tomato detection based on modified YOLOv3 framework [J]. *Scientific Reports*, 2021, 11: 1447.
- [5] HAO Q A, GUO X, YANG F. Fast recognition method for multiple apple targets in complex occlusion environment based on improved YOLOv5 [J]. *Journal of Sensors*, 2023, 2023: 1 – 13.
- [6] SONG C Y, WANG K, WANG C, et al. TDPPL-net: a lightweight real-time tomato detection and picking point localization model for harvesting robots [J]. *IEEE Access*, 2023, 11: 37650 – 37664.
- [7] CHEN Z Y, LI X L, WANG L, et al. An object detection and localization method based on improved YOLOv5 for the teleoperated robot [J]. *Applied Sciences*, 2022, 12(22): 11441.
- [8] 司永胜, 乔军, 刘刚, 等. 苹果采摘机器人果实识别与定位方法 [J]. *农业机械学报*, 2010, 41(9): 148 – 153.
- SI Y S, QIAO J, LIU G, et al. Recognition and location of fruits for apple harvesting robot [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(9): 148 – 153. (in Chinese)

- [9] 项荣, 应义斌, 蒋焕煜, 等. 基于双目立体视觉的番茄定位 [J]. *农业工程学报*, 2012, 28(5): 161 – 167.
- XIANG R, YING Y B, JIANG H Y, et al. Localization of tomatoes based on binocular stereo vision [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28(5): 161 – 167. (in Chinese)
- [10] 解则晓, 陈文柱, 迟书凯, 等. 基于结构光视觉引导的工业机器人定位系统 [J]. *光学学报*, 2016, 36(10): 400 – 407.
- XIE Z X, CHEN W Z, CHI S K, et al. Industrial robot positioning system based on the guidance of the structured-light vision [J]. *Acta Optica Sinica*, 2016, 36(10): 400 – 407. (in Chinese)
- [11] JOCHER G. YOLOv5 [EB/OL]. (2020 – 06 – 17) [2023 – 04 – 23]. <https://github.com/ultralytics/YOLOv5>.
- [12] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 13708 – 13717.
- [13] GEVORGYAN Z. Siou loss: more powerful learning for bounding box regression [EB/OL]. arXiv:2205.12740 (2022 – 05 – 25) [2023 – 04 – 23]. <https://arxiv.org/abs/2205.12740>.
- [14] ZHANG Z. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330 – 1334.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132 – 7141.
- [16] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11531 – 11539.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV). Cham, Switzerland: Springer, 2018: 3 – 19.
- [18] 宋建辉, 顾天宇, 刘砚菊, 等. 基于注意力机制的全卷积抓取姿态预测方法 [J]. *沈阳理工大学学报*, 2023, 42(2): 1 – 7.
- SONG J H, GU T Y, LIU Y J, et al. A fully convolutional grasp posture prediction method based on attention mechanism [J]. *Journal of Shenyang Ligong University*, 2023, 42(2): 1 – 7. (in Chinese)

(责任编辑:宋颖韬)