文章编号　1004-924X(2024)10-1595-11

改进 YOLOv7 的服务机器人密集遮挡目标识别

陈仁祥[1*]，邱天然[1]，杨黎霞[2]，余腾伟[1]，贾　飞[1]，陈　才[3]

(1. 重庆交通大学 交通工程应用机器人重庆市工程实验室，重庆 400074；
2. 重庆科技学院 工商管理学院，重庆 401331；
3. 重庆智能机器人研究院，重庆 400714)

摘要：针对服务机器人视觉抓取时待识别目标存在密集遮挡导致识别效果差的问题，提出改进 YOLOv7 的服务机器人密集遮挡目标识别方法。首先，为改善密集遮挡目标特征信息丢失导致识别困难的问题，使用深度过参数化卷积构建深度过参数化高效聚合网络，利用不同卷积核对每个通道进行运算，增强网络感知能力，使网络关注目标未遮挡区域特征；其次，为抑制密集遮挡目标边界不易区分对识别造成的影响，将坐标注意力机制嵌入主干网络中，使网络获取目标位置信息并更好地关注特征图中的重要区域，增强网络特征提取能力；最后，使用 Ghost 网络进行轻量化改进，减少计算量并降低模型内存占用。在自建数据集与公共数据集分别对模型进行对比实验，实验结果表明，改进后模型 mAP 分别达到 92.9%，87.8%。本文模型在降低内存占用的同时，识别精度和识别效率提升，整体性能更优。

关　键　词：密集遮挡；改进 YOLOv7；服务机器人；目标识别

中图分类号：TP391　　**文献标识码**：A　　**doi**：10.37188/OPE.20243210.1595

A method for dense occlusion target recognition of service robots based on improved YOLOv7

CHEN Renxiang[1*], QIU Tianran[1], YANG Lixia[2], YU Tenwei[1], JIA Fei[1], CHEN Cai[3]

(1. *Chongqing Engineering Laboratory of Traffic Engineering Application Robot, Chongqing Jiaotong University, Chongqing* 400074, *China*;
2. *School of Business Administration, Chongqing University of Science and Technology, Chongqing* 401331, *China*;
3. *Chongqing Intelligent Robot Research Institute, Chongqing* 4000714, *China*)
* *Corresponding author, E-mail: manlou. yue*@126. *com*

Abstract: Aiming at the problem of poor recognition effect due to dense occlusion of the target to be recognized during visual grasping of service robots, we propose to improve the dense occlusion target recognition method for service robots with YOLOv7. First, in order to improve the problem of recognition difficulties caused by the loss of feature information of densely occluded targets, a deep over-parameterized

收稿日期：2023-11-14；**修订日期**：2024-01-04.

基金项目：国家自然科学基金(No. 51975079)；重庆市教委科学技术研究项目(No. KJZD-M202200701)；重庆市自然科学基金创新发展联合基金(No. CSTB2023NSCQ-LZX0127)；重庆市研究生联合培养基地项目(No. JDLHPYJD2021007)；重庆市专业学位研究生教学案例库(No. JDALK2022007)；重庆市研究生科研创新项目(No. CYS23509)；重庆科技大学科研启动项目(No. ckre202212030)

convolution was used to construct a deep over-parameterized high-efficiency aggregation network, and different convolution kernels were used to operate on each channel to enhance the network sensing ability, so that the network focused on the features of the target's uncovered area; second, in order to suppress the influence caused by dense occlusions and indistinguishable target boundaries on recognition, the coordinate attention mechanism was embedded into the backbone network. This enabled the network to obtain target position information and paid more attention to the important areas in the feature map, thereby enhancing the capability of the network to extract features; finally, the Ghost network was used to improve the lightweighting, reduce the number of parameters of the network model and the number of floating-point operations to realize the lightweighting, reduce the memory occupation of the model, and increase the model operation efficiency. Comparison experiments were conducted on the model in the self-constructed dataset and the public dataset respectively, and the experimental results show that the improved model achieves a mAP of 92.9% on the self-constructed dataset and 87.8% on the public dataset, which is better than the original method and the other commonly used methods. In this paper, the model reduces the memory footprint while the recognition accuracy and recognition efficiency are improved, and the overall performance is better.

Key words: dense occlusion; improved YOLOv7; service robots; target recognition

1 引　言

近年来机器人广泛应用于工业、食品、服务等行业,依靠视觉传感器采集并实现目标识别是服务机器人的重要能力[1]。抓取物常出现密集和相互遮挡为机器人准确识别和抓取带来了挑战。同时,机器人算力资源有限。故提升服务机器人对密集遮挡目标的识别精度并实现模型轻量化意义重大[2]。

目前常用的识别模型可分为双阶段模型、单阶段模型和基于transformer的方法。双阶段模型分为定位与分类两个阶段,如Faster R-CNN[3],Mask R-CNN[4]等。基于transformer的方法如DETR[5]等。双阶段模型与基于transformer的方法精度较高,但整体计算量较大,增加了训练难度且难以满足轻量化要求。单阶段模型则是根据提取出的特征直接对目标进行预测与定位,精度稍低但识别效率高,如SSD(Single Shot MultiBox Detector)[6],YOLO(You Only Look Once)[7]等。杨坚[8]等提出改进YOLOv4方法(YOLOv4-tiny-x)以提高被遮挡目标的识别准确率。滕臻[9]等使用YOLOv5完成对不同药瓶类型的识别,但未考虑遮挡程度对识别产生的影响。张伏[9]等构建了基于改进型YOLOv4-LITE轻量级神经网络模型,对被遮挡目标具有比YOLOv4原模型更好的识别效果,但未考虑目标特征信息丢失问题。

服务机器人密集遮挡目标识别的关键在于:目标边界可能因密集产生边界不易区分问题,目标受到遮挡导致部分特征信息丢失,影响识别精度[2,11];文献[12]方法对存在遮挡的目标识别精度有提升,但对未遮目标,识别精度反而下降。

因此,提出了改进YOLOv7的服务机器人密集遮挡目标识别方法。即:(1)针对密集遮挡目标特征信息丢失导致识别困难的问题,构建深度过参数化高效聚合网络,利用不同卷积核对每个通道进行运算,增强网络感知能力,使网络关注目标未遮挡区域特征信息。(2)针对密集遮挡目标边界不易区分的问题,在网络关键位置引入坐标注意力机制,为网络提供目标位置信息并使网络关注特征图中的重要区域,提升网络特征提取能力。(3)为实现模型轻量化,使用Ghost卷积对主干网络进行优化,利用特征之间的相关性和冗余性,降低模型计算量。最后,通过实验证明了所提方法的可行性和有效性。

2 改进YOLOv7网络模型原理

2.1 改进YOLOv7网络模型设计

改进YOLOv7(Improved-YOLOv7,Im-YOLOv7)网络模型的结构如图1所示。主要包括以

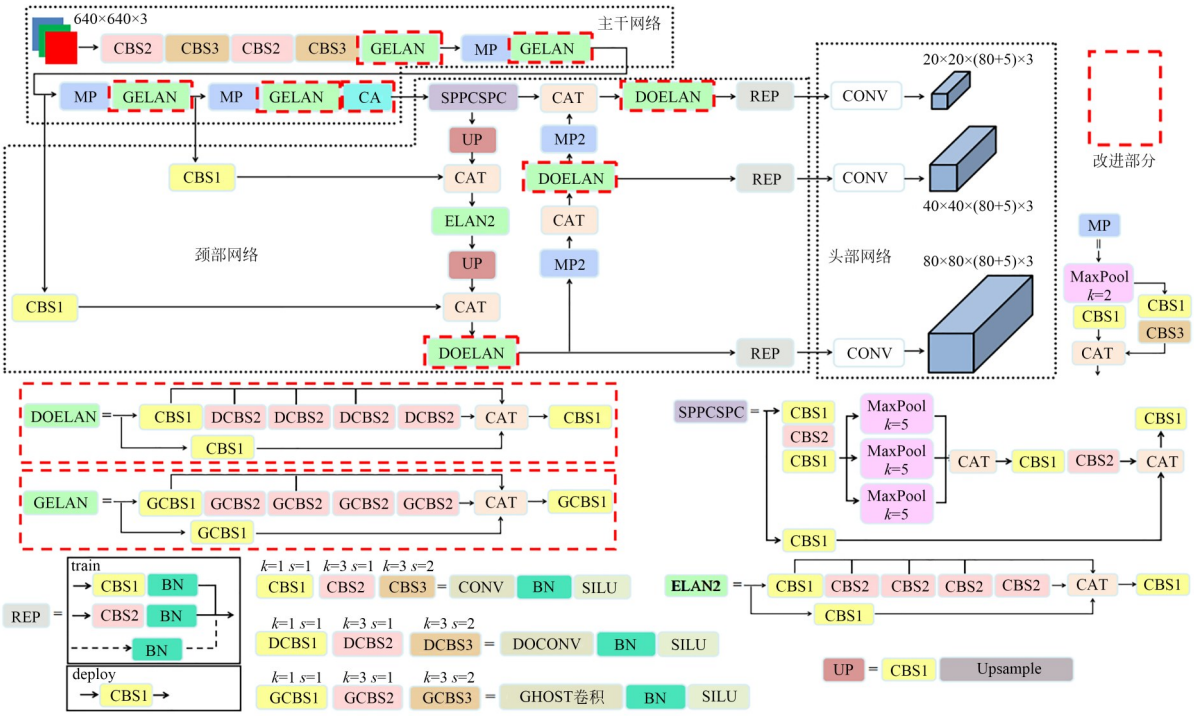

图 1　改进 YOLOv7 结构

Fig. 1　Improve YOLOv7 structure

下部分：(1)主干网络：主干网络由数个 CBS，ELAN 和 MP 结构组成，负责提取输入图像的特征。设计轻量化高效聚合模块对主干网络进行轻量化改进，并在特征提取关键位置加入坐标注意力模块，为网络提供目标位置信息并使网络更加关注重要区域，优化网络特征提取能力。(2)颈部网络：用于进一步整合和提炼特征，从主干网络的最后一层输出中提取特征图，然后经过 SPPCSP 模块进行通道数的调整，再通过自顶向下和自底向上的方式将特征图融合，本文在颈部网络构建深度过参数化高效聚合网络，在 Rep 模块前利用不同卷积核对每个通道进行运算，增强网络感知能力。(3)头部网络：负责输出目标类别与边界框。

改进部分在图中使用虚线框标出。

2.2　深度过参数化高效聚合网络

针对密集遮挡时目标特征信息存在丢失问题，利用深度过参数化模块(如图 2)，构建深度过参数化高效聚合网络。

首先对输入特征进行 depthwise 操作，从而得到中间变量：

$$p' = D \circ p. \tag{1}$$

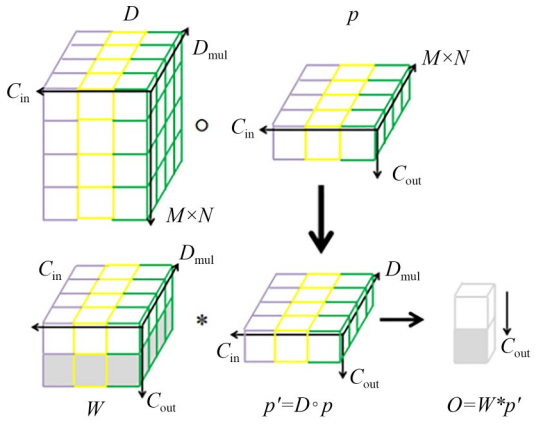

图 2　深度过参数化模块

Fig. 2　Depthwise Over-parameterized module

其次对中间变量进行卷积，得到结果：

$$O = W*p'. \tag{2}$$

最后，深度过参数化卷积的输出结果可以表示为：

$$O = W*(D \circ p), \tag{3}$$

其中：p 表示输入特征，D 表示深度卷积内核，$\circ$ 表示深度过参数化操作，M 和 N 表示输入向量的空间维度。C_{in}，C_{Out} 表示输入和输出向量通道数，$*$ 表示常规卷积运算符[13]。

对网络中的CBS模块实施过参数化操作，每个通道采用不同卷积核进行运算，构成DCBS模块。利用DCBS在三个Rep操作前构建深度过参数化高效聚合网络(DOELAN)，DOELAN结构如图1中所示，目的是在颈部网络输出不同尺寸特征图前，通过额外的深度卷积来增强卷积层，使每个通道用不同内核进行卷积，增强模型特征感知能力，使模型关注目标未遮挡区域特征信息。

与原结构相比，DOELAN结构将训练后所得到的线性运算通过单个卷积层表示，在推理阶段仅使用单层计算，从而将计算量减少为与常规层相近。

2.3　坐标注意力机制特征融合网络

坐标注意力机制(Coordinate Attention，CA)分为信息嵌入与CA生成两个步骤[14]，如图3。XPOOL与YPOOL分别沿着x方向与y方向通过平均池化提取特征信息，接着通过Concat聚合特征信息层，将其拼接为$C\times 1\times(H+W)$的特征图，接着使用1×1卷积获得远程依赖关系，将通道维数压缩，使用ReLU激活函数，可以得到水平与垂直注意张量，之后沿着两个方向进行split操作，通过卷积将通道数恢复，并进行Sigmoid非线性激活。最后进行重新加权操作，完成坐标注意力的施加，增强模型对重要区域的关注。

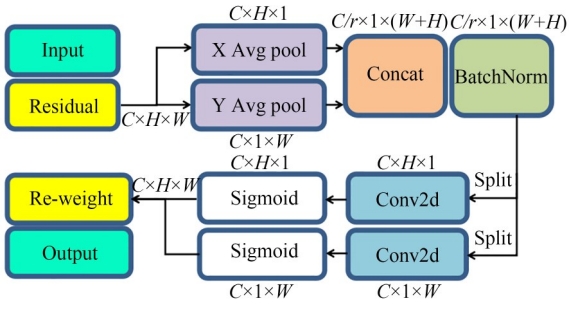

图3　坐标注意力模块

Fig. 3　Coordinate attention module

服务机器人依靠视觉抓取时，待识别目标可能因密集产生边界不易区分问题。在特征提取部分的关键位置加入坐标注意力模块，坐标注意力机制使模型关注到图像特征间的关联性并获取精确的位置信息，令模型能更好地关注特征图中的重要区域，投入更多注意力资源，增强模型对待识别物品的敏感度并增强模型特征提取能力；同时，模型利用获取到的目标位置信息，判断目标在图像中的位置，抑制密集遮挡目标边界不易区分对识别造成的影响。

2.4　Ghost轻量化网络

Ghost卷积首先使用1×1卷积对输入特征图进行特征提取，生成基础特征层。然后对特征层的每一个通道进行卷积，生成Ghost特征图。

当网络输入为：$h\times w\times c$，输出为：$h'\times w'\times c'$，卷积核大小为k，Ghost特征图为n个。

则普通卷积的计算量为：

$$h'\times w'\times c'\times k\times k\times c. \tag{4}$$

而Ghost卷积的计算量为：

$$h'\times w'\times c'\times k\times k\times \frac{c+n-1}{n}. \tag{5}$$

通过式(4)和式(5)可以看出，Ghost卷积具有更少的计算量。移动终端等边缘设备存储能力和计算能力有限，Ghost卷积使模型计算量减小，训练生成的模型占据的物理存储空间较小，为模型的部署提供便利[15]。

使用Ghost卷积对主干网络进行轻量化改进，使用Ghost卷积构建轻量化高效聚合网络(GELAN)，结构如图1中所示。利用GELAN对主干网络结构进行优化。优化后特征层大小不发生变化，保证了网络的一致性。GELAN使用线性变换取代部分常规卷积，减少推理计算量并保持网络性能，完成模型的轻量化。

2.5　密集遮挡目标识别流程

所提出改进YOLOv7的服务机器人密集遮挡目标识别流程如图4所示。具体步骤如下：

(1)采集图像、标注构建数据集并划分数据集。该步骤中应注意目标遮挡形式多样性，且保证不同种类目标数据样本足够。

(2)构建Im-YOLOv7服务机器人密集遮挡目标识别网络并初始化参数。

(3)将训练集与验证集数据输入到Im-YOLOv7服务机器人密集遮挡目标识别网络中，通过计算损失、反向传播不断训练，更新网络参数并保存最终模型。

(4)调用训练完成的模型，将测试集数据输入，得到识别目标区域候选框以及目标类别。

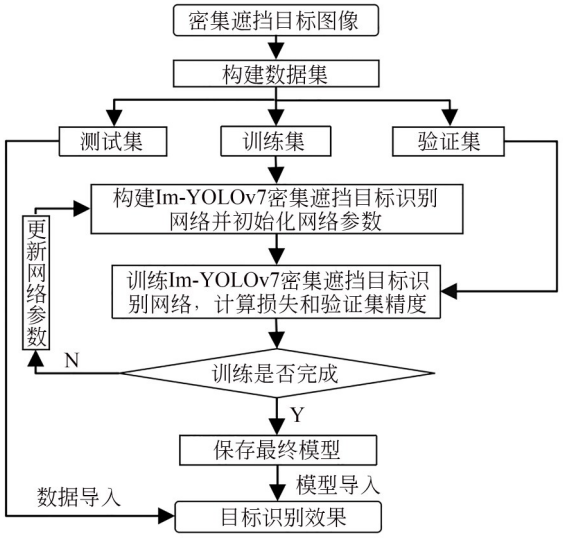

图 4　识别流程

Fig. 4　Overall flowchart of Recognition

3　实验与分析

3.1　实验环境

实验环境为 WIN10 操作系统；GPU 为 NVIDIA3070；Python 版本为 3.8；CUDA 版本为 11.6；Pytorch 版本为 1.11.0。批样本（batch size）数量为 4，动量因子（Momentum）为 0.937，权值初始学习率（learning rate）为 0.01，迭代次数设置为 600 次。

3.2　评价指标

为客观衡量模型的目标识别效果，使用准确率（Precision，P）、召回率（Recall，R）、平均精度（Average Precision，AP）、平均精度均值（Mean Average Precision，mAP）、模型内存占用来评价训练后的模型[16]。

$$Precision=\frac{TP}{TP+FN}\times 100\%, \tag{6}$$

$$Recall=\frac{TP}{TP+FP}\times 100\%, \tag{7}$$

$$AP=\int_0^1 (P(R)\mathrm{d}R), \tag{8}$$

$$mAP=\frac{\sum(AP)}{n}, \tag{9}$$

其中：TP 为真实的正样本数量，FN 为虚假的负样本数量，FP 为虚假的正样本数量。n 为目标类别数。准确率与召回率曲线与坐标轴围成的面积大小等于 AP 值大小，$P(R)$ 表示准确率与召回率曲线。mAP@0.5 表示 IOU 阈值为 0.5 时对所有类别的 AP 取平均值，在实验中用 mAP 表示[17]。

3.3　自建数据集

为测试服务机器人工作场景中改进 YOLOv7 模型的性能。将生活中常见物品模拟服务机器人抓取对象[18]。对 Biscuit，Sausage，Luncheon meat，Chocolate，Vitamin 五类生活中常见物品的图像进行采集，为保证数据多样性，其中 201 张图像数据来自 Intel RealSense D435i 深度相机拍摄、504 张图像来自互联网，在 LabelImg 工具上进行手工标注，目标间存在接触定义为密集样本；目标被遮挡面积在 5%～60% 时，定义为遮挡样本；目标被遮挡面积大于 60% 时，将不进行标注。数据集遮挡情况及其数量如表 1 所示。

表 1　密集遮挡情况

Tab. 1　Dense occlusion of dataset

	无密集遮挡样本个数	密集遮挡样本个数
训练集	1 188	2 264
验证集	108	230
测试集	100	166

接着使用大小变换、剪切、旋转，等方法进行数据增强，最终得到 1 540 张图像，按照 8∶1∶1 的比例将数据集划分为训练集、验证集以及测试集。训练集用于模型训练，验证集用于修正训练过程，测试集用于对模型性能进行测试[19]。

3.3.1　消融实验

为分析不同改进对模型产生的影响，证明改进过程的有效性，设计消融实验，实验使用相同数据集，相同参数及相同环境。实验结果如表 2 所示。改进 1 表示利用 DOELAN 结构对网络改进，改进 2 表示将 CA 注意力机制嵌入网络关键位置，改进 3 表示使用 Ghost 轻量化网络对主干网络进行改进。

改进 1 提升模型的 mAP 至 91.1%。改进后的 DOELAN 结构增加了额外的深度卷积用于增强卷积层，提升模型感知能力，使网络关注目标未遮挡区域特征。改进 2 在信息嵌入阶段获取物品的关键特征信息，增强模型的特征提取能力，抑制密集遮挡对目标识别造成的影响，在保证准确率与召回率的情况下，mAP 较 YOLOv7 模型

提升 2.6%。改进 1,2 在提升模型性能的同时，增加了计算量，使网络识别效率出现轻微程度的下降，而改进 3 通过减少网络模型参数数量及运算次数来实现轻量化，使模型内存占用较原模型减小，FPS 提升明显，提高了模型的实时性与识别效率，且 mAP 值达到了 89.2%。在改进 1,3 或改进 2,3 同时施加在模型时，模型 mAP 值与单独施加改进 1 或改进 2 结果相近，进一步证明了改进 3 在轻量化模型的同时，未对模型精度造成负面影响。

表 2　消融实验结果

Tab. 2　Results of ablation experiment

模型	P/%	R/%	mAP/%	模型体积/MB	FPS/(frame•s^{-1})
YOLOv7	87.9	82.8	88.8	72	31.3
改进 1	86.5	82.8	91.1	74	29.5
改进 2	86.5	88.3	91.4	72	30.2
改进 3	82.1	85.3	89.2	48	38.0
改进 1+2	93.0	86.0	92.5	74	28.6
改进 1+3	93.9	86.4	91.6	50	36.4
改进 2+3	90.0	84.0	91.5	48	37.0
Im-YOLOv7	92.8	88.7	92.9	50	35.8

将改进 1 与改进 2 同时施加在模型时，模型获得了较高的 mAP 值，达到了 92.5%。说明两种改进可以同时对模型性能做出提高。将改进 1,2,3 同时施加在模型时，本文的 Im-YOLOv7 识别方法具有更优的性能，mAP 达到了 92.9%，相较于原模型提升 4.1%，模型内存占用仅为 50 MB，比原模型减小 30.5%，FPS 达到了 35.8，较原模型提升，综上所述，Im-YOLOv7 识别方法在性能提升的同时实现了模型轻量化，并提高了识别实时性。

3.3.2　模型识别可视化实验

为分析改进对模型产生的作用，对图像识别结果可视化，如图 5(彩图见期刊电子版)。以 JET 格式色彩表示得分大小，颜色越接近红色表示得分越高，网络判断高分区域更可能包含待识别图像。以锚框显示数字表示置信度大小，置信度越高表示网络预测的锚框包含某个物体的可能性越大[20]。第一行图片为遮挡目标，YOLOv7 原网络得分较高区域较小，识别置信度最低。网络加入 DOELAN 后，使用额外的深度卷积增强卷积层，增强模型特征感知能力，可以看出，在目标未遮挡区域有较多得分高区域，识别置信度明显提高。说明相较于原网络，DOELAN 能使模型感知到更多未遮挡部分的特征信息，帮助网络判断目标类型，改善识别效果。

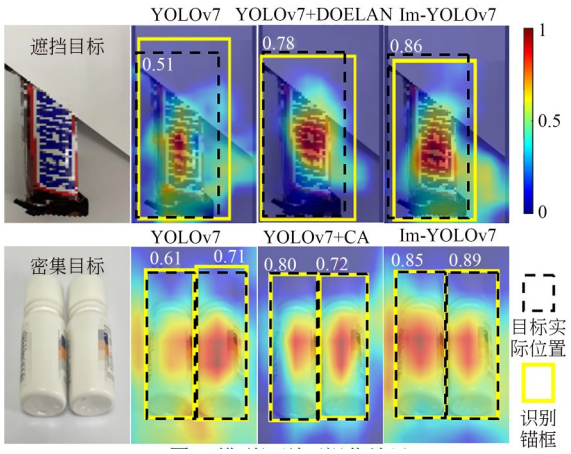

图 5　模型识别可视化结果

Fig. 5　Model recognition visualization results

第二行图片为密集目标，YOLOv7 原网络未能很好区分目标边界，得分较高区域朝相邻物体延伸，导致识别置信度低。加入坐标注意力机制，为网络提供目标位置信息，帮助网络判断目标在图像中的位置，并对重要区域进行关注，抑制密集目标边界不易区分对识别造成的影响。由图可以看出，目标实际所在位置得分较高且边界位置得分较低，说明坐标注意力机制能使网络准确判断目标位置，并区分目标边界。在遮挡与密集目标识别中，Im-YOLOv7 在待识别区域有最多得分高区域，并取得最高识别置信度，锚框

位置更精确，说明 Im-YOLOv7 可同时有效发挥深度过参数化高效聚合网络、坐标注意力机制的作用，使网络对密集遮挡目标的识别精度提升。

3.3.3　无密集遮挡实验

服务机器人工作场景中，密集、遮挡与未密集遮挡目标可能同时存在，为进一步验证 Im-YOLOv7 模型的有效性。设计实验，测试不存在密集遮挡的目标识别效果。实验使用训练完成的 Im-YOLOv7 与 YOLOv7 模型，分别对完全不存在密集遮挡的物品进行识别，测试包括 140 张无遮挡场景图片，包括 245 个样本个数。实验结果如表 3 所示。

深度过参数化高效聚合网络在识别过程中提升网络感知能力，坐标注意力机制提供位置信息，弥补常规卷积结构由于平移不变性导致的位置信息缺失问题，使网络更加关注目标所在区域。

表 3　无密集遮挡测试数据

Tab. 3　No dense occlusion test data

测试数据	方法	P/%	R/%	mAP/%
无遮挡场景	YOLOv7	95.7	92.4	95.6
无遮挡场景	Im-YOLOv7	92.2	95.0	97.5

实验表明，Im-YOLOv7 在不存在遮挡的场景中，mAP 相较于 YOLOv7 提升了 1.9%。证明深度过参数化高效聚合网络与坐标注意力机制使网络性能提升，改进在密集遮挡场景和一般场景均生效，且在密集遮挡场景中，识别精度提升更为明显。进行无密集遮挡识别可视化实验，结果如图 6。可以看出，识别无密集遮挡物体时，Im-YOLOv7 较原 YOLOv7 网络在目标实际所在位置高分区域更多，锚框位置均较为精确，而 Im-YOLOv7 网络取得了更高的识别置信度，证明所提识别方法对密集遮挡目标与无密集遮挡目标均有效。

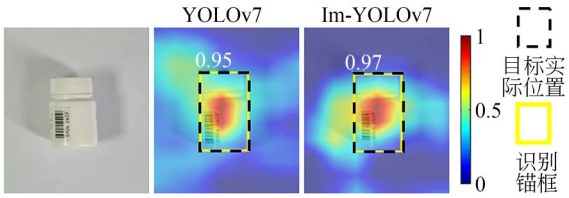

图 6　无密集遮挡识别可视化结果

Fig. 6　No dense occlusion to recognize visualization results

3.3.4　对比实验

在相同测试集下，分别与 DETR，YOLOv7，YOLOv8-l 等模型进行对比实验，统计所有模型的评价指标如表 4 所示。

可以看出，对比其他常用模型，本文提出的改进模型具有更高的识别精度，mAP 对比 YOLOv7 模型提高 4.1%，与 YOLOv8-l，YOLOv5-l，YOLOv5-s 相比，mAP 分别提高 3.3%，4.4%，6.6%。综合评价指标，DETR 模型内存占用大，识别精度较低。YOLOv5-s 具有最小的模型内存占用，但是准确率较低，mAP 值为 86.3%，难以很好满足识别精度要求。YOLOv5 改进模型针对识别对象存在遮挡的场景改进，模型内存占用比较小，准确率对比 YOLOv5-s 模型有所提升，但是 mAP 值仍然较低。YOLOv4-tiny-x 模型针对被遮挡目标进行优化检测头等多种改进，检测效果优于 YOLOv5-l。YOLOv7 模型的 mAP

表 4　自建数据集对比实验结果

Tab. 4　Self built dataset dataset comparison experiment results

模型	P/%	R/%	mAP/%	模型体积/MB	FPS/(frame•s^{-1})
DETR[5]	80.1	75.4	85.6	473	8.9
YOLOv5-s	85.3	78.0	86.3	**14**	32.2
YOLOv5 改进[21]	90.2	76.0	86.7	36	28.6
YOLOv5-l	90.4	81.2	88.5	90	30.2
YOLOv4-tiny-x[8]	89.1	82.9	88.6	82	34.4
YOLOv7	87.9	82.8	88.8	72	31.3
YOLOv8-l	92.3	78.4	89.6	84	34.1
Im-YOLOv7	**92.8**	**88.7**	**92.9**	50	**35.8**

值相对较高，但仍难以满足识别精度要求。YOLOv8-l模型准确率较高，mAP值达到了89.6%，但模型内存占用大。相较本实验中其他识别模型，提出的Im-YOLOv7识别精度较高，FPS值最高，模型处理图像速度快，实时性好。模型内存占用方面较为轻量化，取得更优的整体性能。

为测试本文模型实际应用效果，使用实验室服务机器人装有的Intel RealSense深度相机对图像进行采集，服务机器人实验平台如图7所示。

图7　服务机器人实验平台

Fig. 7　Service robot experimental platform

并将参与对比实验的七种模型分别对实际图像进行识别，图像中待识别目标存在密集与遮挡现象，被遮挡面积为5%～20%的待识别物体时定义为遮挡情况，被遮挡面积为20%以上待识别物体时定义为严重遮挡情况。效果较优的三种模型识别结果如图8所示。漏检与错检目标在图中用红圈标出。因密集遮挡导致目标边界不易区分，特征信息丢失，YOLOv7与YOLOv8-l模型难以对所有密集遮挡目标进行识别，均存在未成功识别的目标，相比于其他模型，Im-YOLOv7漏、误识别概率最低，总体置信度更优，识别效果最好。

3.4　MessyTable数据集

为研究Im-YOLOv7模型在其他数据集的适应能力，在Messytable[22]数据集上进行对比实验，Messytable数据集是研究复杂场景目标识别的数据集，由商汤与新加坡南洋理工大学联合制作，数据集中每个场景都较为复杂，包含多个对象，这些对象可能被其他物体遮挡，部分场景中对象存在密集遮挡现象，符合服务机器人工作场景。数据集包含五万余张图片，120类物体。为研究密集遮挡目标的识别效果，选出包含密集遮挡对象的共1 016张图像进行实验对比，包括：Apple，Banana，Cup，Luncheon meat，Biscuit，Instant noodles六类物品。并将整个数据集按8∶1∶1的比例划分成训练集、测试集和验证集。图9为MessyTable数据集中部分存在密集遮挡的图像。

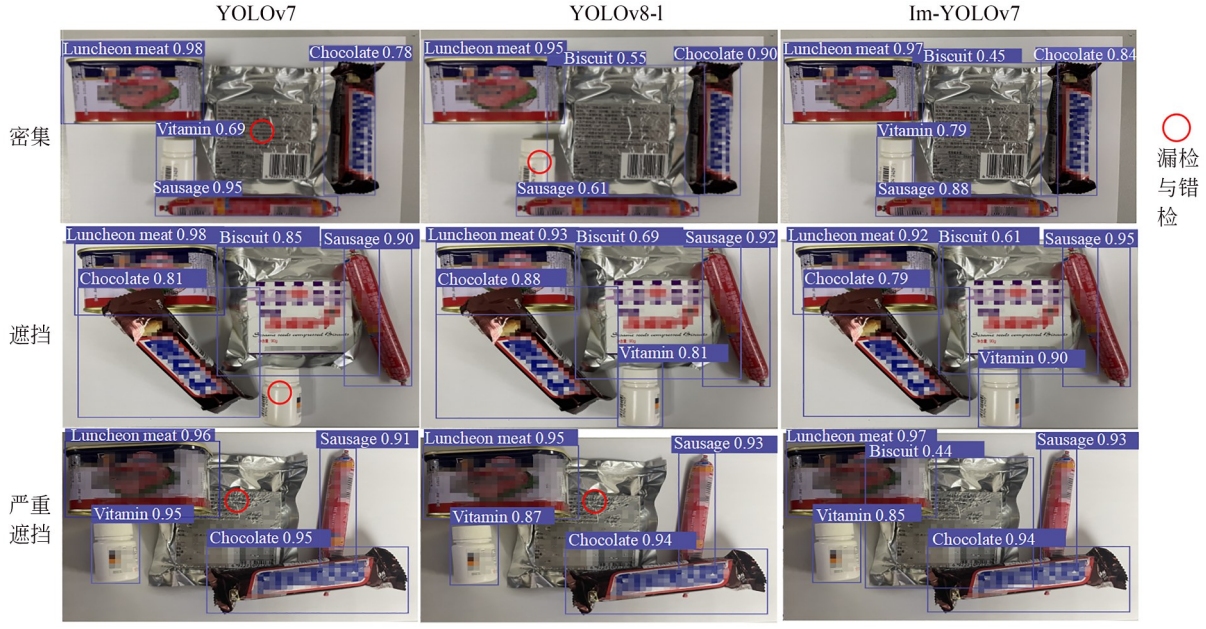

图8　实际识别效果对比

Fig. 8　Comparison of actual recognition effects

图 9　MessyTable 数据集密集遮挡图像

Fig. 9　Dense occlusion images from MessyTable dataset

为验证所提方法识别效果。在相同测试集中，与几种常见模型进行对比，对比结果如表 5 所示。

可以看出 DETR 模型 mAP 值较低，内存占用最大。YOLOv5-s 模型内存占用最小，但识别精度较低。YOLOv5 改进模型内存占用比较小，准确率较高，但 mAP 值仍然较低。YOLOv5-l 模型准确率最高，达到了 90.0%，但召回率和 mAP 值仍然较低并且模型内存占用较大。YOLOv4-tiny-x 模型召回率低，mAP 值优于 YOLOv5-l，但模型内存占用同样较大。

表 5　MessyTable 数据集对比实验结果

Tab. 5　MessyTable dataset comparison experiment results

模型	P/%	R/%	mAP/%	模型体积/MB
DETR[5]	83.4	74.9	80.6	473
YOLOv5-s	85.7	72.4	81.2	**14**
YOLOv5 改进[21]	86.2	76.6	82.2	36
YOLOv5-l	**90.0**	74.8	83.0	90
YOLOv4-tiny-x[8]	85.5	74.4	84.4	82
YOLOv7	82.8	81.1	84.6	72
YOLOv8-l	89.6	71.9	84.1	84
Im－YOLOv7	84.7	**83.3**	**87.8**	50

YOLOv7 模型通过增加网络宽度使自身能学习更多特征信息，其召回率较高，mAP 值达到了 84.6%。相比 YOLOv5-l 模型有所提高。YOLOv8-l 模型，准确率达到了 89.6%。但召回率和 mAP 值仍然不够理想。所提出的 Im-YOLOv7 模型召回率为 83.3%，mAP 值达到了 87.8%，对比 YOLOv7 基础模型，mAP 提升了 3.2%，并且有着更小的模型内存占用，模型轻量化的同时提升识别精度，总体性能更优，证明了 Im-YOLOv7 模型在不同数据集上的有效性。

4　结　论

服务机器人依靠视觉实现物品抓取时，由于待识别目标容易存在密集遮挡情况，导致识别精度下降，针对该问题，提出 Im-YOLOv7 识别方法，所提方法的优势在于：

(1)构建深度过参数化高效聚合网络，令每个通道使用不同的内核进行卷积，增强模型特征感知能力，改善密集遮挡目标特征信息丢失导致识别困难的问题；在主干网络中加入坐标注意力模块，提升网络特征提取能力，提供目标位置信息使模型更好地关注特征图中的重要区域，抑制密集遮挡目标边界不易区分对识别造成的影响；使用 Ghost 卷积改进主干网络，在不损失识别精度的基础上对网络模型进行轻量化改进。

(2)相较于其他几种常用模型，Im-YOLOv7 模型在降低内存占用的同时，做到了识别精度提升，使用不同数据集进行实验，在自建数据集与 MessyTable 数据集 mAP 分别达到 92.9% 和 87.8%。证明该模型具有良好的适应能力。

参考文献：

[1]　崔永成，田国会，周昭旭，等. 智能空间下面向动作序列生成的服务机器人指令解析方法[J]. 机器人，2024，46(1)：1-15.

CUI Y C，TIAN G H，ZHOU Z X，*et al*. A ser-

vice robot instruction parsing method for action sequence generation in intelligent space[J]. *Robot*, 2024, 46(1): 1-15. (in Chinese)

[2] 陶永, 刘海涛, 王田苗, 等. 我国服务机器人技术研究进展与产业化发展趋势[J]. 机械工程学报, 2022, 58(18): 56-74.
TAO Y, LIU H T, WANG T M, *et al*. Research progress and industrialization development trend of Chinese service robot[J]. *Journal of Mechanical Engineering*, 2022, 58(18): 56-74. (in Chinese)

[3] REN Y, ZHU C R, XIAO S P. Object detection based on fast/faster RCNN employing fully convolutional architectures[J]. *Mathematical Problems in Engineering*, 2018, 2018: 3598316.

[4] HE K M, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN[C]. 2017 *IEEE International Conference on Computer Vision* (*ICCV*). *Venice*, *Italy*. IEEE, 2017: 2980-2988.

[5] CARION N, MASSA F, SYNNAEVE G, *et al*. *End-to-End Object Detection with Transformers*[M]. Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 213-229.

[6] YANG J, HE W Y, ZHANG T L, *et al*. Research on subway pedestrian detection algorithms based on SSD model[J]. *IET Intelligent Transport Systems*, 2020, 14(11): 1491-1496.

[7] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Las Vegas*, *NV*, *USA*. IEEE, 2016: 779-788.

[8] 杨坚, 钱振, 张燕军, 等. 采用改进YOLOv4-tiny的复杂环境下番茄实时识别[J]. 农业工程学报, 2022, 38(9): 215-221.
YANG J, QIAN Z, ZHANG Y J, *et al*. Real-time recognition of tomatoes in complex environments based on improved YOLOv4-tiny[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(9): 215-221. (in Chinese)

[9] 滕臻, 崔国华, 高鹏, 等. 基于机器视觉及深度学习的静脉药物调配机器人药瓶识别[J]. 机床与液压, 2022, 50(5): 33-37.
TENG Z, CUI G H, GAO P, *et al*. Drug bottle recognition of intravenous drug dispensing robot based on machine vision and deep learning[J]. *Machine Tool & Hydraulics*, 2022, 50(5): 33-37. (in Chinese)

[10] 张伏, 陈自均, 鲍若飞, 等. 基于改进型YOLOv4-LITE轻量级神经网络的密集圣女果识别[J]. 农业工程学报, 2021, 37(16): 270-278.
ZHANG F, CHEN Z J, BAO R F, *et al*. Recognition of dense cherry tomatoes based on improved YOLOv4-LITE lightweight neural network[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37(16): 270-278. (in Chinese)

[11] 高毅, 何森. 密集遮挡条件下的步态识别[J]. 光学 精密工程, 2023, 31(2): 263-276.
GAO Y, HE M. Gait recognition algorithm in dense occlusion scene[J]. *Opt. Precision Eng.*, 2023, 31(2): 263-276. (in Chinese)

[12] LI Y D, GUO K, LU Y G, *et al*. Cropping and attention based approach for masked face recognition[J]. *Applied Intelligence*, 2021, 51(5): 3012-3025.

[13] CAO J M, LI Y Y, SUN M C, *et al*. DO-conv: depthwise over-parameterized convolutional layer[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2022, 31: 3726-3736.

[14] 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.

[15] HAN K, WANG Y H, TIAN Q, *et al*. GhostNet: more features from cheap operations[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Seattle*, *WA*, *USA*. IEEE, 2020: 1577-1586.

[16] 李大湘, 苏仲恒, 刘颖. 基于改进YOLOv4的道路交通标志识别[J]. 光学 精密工程, 2023, 31(9): 1366-1378.
LI D X, SU Z H, LIU Y. Road traffic sign recognition algorithm based on improved YOLOv4[J]. *Opt. Precision Eng.*, 2023, 31(9): 1366-1378. (in Chinese)

[17] 司永胜, 孔德浩, 王克俭, 等. 基于CRV-YOLO的苹果中心花和边花识别方法[J]. 农业机械学报, 2024, 55(2): 278-286.
SI Y S, KONG D H, WANG K J, *et al*. Recognition of apple king flower and side flower based on CRV-YOLO[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(2):

278-286. (in Chinese)

[18] 石杰，周亚丽，张奇志. 基于改进 Mask RCNN 和 Kinect 的服务机器人物品识别系统[J]. 仪器仪表学报，2019，40(4)：216-228.
SHI J，ZHOU Y L，ZHANG Q Z. Service robot item recognition system based on improved Mask RCNN and Kinect[J]. *Chinese Journal of Scientific Instrument*，2019，40(4)：216-228. (in Chinese)

[19] 刘颖，姜威，李冠典，等. 基于标签嵌入方法的纺织品瑕疵识别网络[J]. 光学 精密工程，2023，31(10)：1563-1579.
LIU Y，JIANG W，LI G D，*et al*. Textile defect recognition network based on label embedding[J]. *Opt. Precision Eng.*，2023，31(10)：1563-1579. (in Chinese)

[20] SELVARAJU R R，COGSWELL M，DAS A，*et al*. Grad-CAM：visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*，2020，128(2)：336-359.

[21] YU Z，HUANG H，CHEN W，*et al*. YOLO-FaceV2：a scale and occlusion aware face detector[J]. *arXiv preprint arXiv*：*2208*. 02019，2022：1-18.

[22] CAI Z A，ZHANG J Z，REN D X，*et al*. *MessyTable*：*Instance Association in Multiple Camera Views*[M]. Computer Vision-ECCV 2020. Cham：Springer International Publishing，2020：1-16.

作者简介：

陈仁祥(1983－)，男，四川盐源人，教授，博士生导师，2007 年和 2012 年于重庆大学分别获得学士学位和博士学位，现为重庆交通大学机电与车辆工程学院副院长，主要从事机器视觉、智能测试技术与信号处理方面的研究。E-mail：manlou.yue@126.com

通讯作者：

邱天然(1996－)，男，贵州贵阳人，硕士研究生，2018 年于杭州电子科技大学获得学士学位，主要从事计算机视觉与机器人控制方法研究。E-mail：979895217@qq.com