DOI: 10.13382/j.jemi.B2307049

融合改进 YOLO 和语义分割的遮挡目标抓取方法*

林 哲 潘慧琳 陈 丹
(福州大学电气工程与自动化学院 福州 350108)

摘 要:针对遮挡目标的机器人抓取存在的遮挡干扰问题,提出了改进的 YOLO-CA-SD 和语义分割的遮挡目标检测模型及抓取方法,完成多目标及非目标物互相遮挡干扰情况下的抓取。首先,该模型在 YOLOv5l 中添加坐标注意力,在损失函数基础上考虑检测框匹配方向的问题,增加框之间的角度信息,并对原模型检测部分进行解耦,减少耦合造成的信息丢失。其次,提出了改进的 DeeplabV3+目标分割模型,用 MobileNetV2 替换 DeeplabV3+原主干网络,减小模型复杂度,在空洞空间金字塔池化结构中添加 CA 模块融合像素坐标信息提高分割精度,解决了遮挡干扰问题。最后,利用点云配准得到目标姿态相对于模板姿态的末端旋转角及最优抓取点。在 2 750 张自主构建的常用工具遮挡数据集上进行性能测试,结果表明:改进后的模型在 mAP@0.5,mAP@0.5:0.95、60%目标物体遮挡率数据集及 60%非目标物体遮挡率数据集上的检测精度分别提高了 0.052%、0.968%、6.000%、7.400%。此基础上改进的语义分割模型分割速度和 MIOU 分别提升了 33.45%和 0.625%,并且通过 ABB IRB1200 机械臂实现遮挡目标的抓取实验,验证了该方法的可行性与实用性。

关键词:遮挡目标检测;目标抓取;YOLO;DeeplabV3+模型;点云配准
中图分类号:TP242.2;TN919.5 **文献标识码**:A **国家标准学科分类代码**:460.5030

Grasp method for occlusion method by fusing improved YOLO with semantic segmentation

Lin Zhe Pan Huilin Chen Dan
(School of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: For the problem of occlusion interference in robot grasping of occluded targets, an improved YOLO-CA-SD and semantic segmentation occluded target detection model and grasping method are proposed to complete the grasping when multiple targets and non-target objects occlude and interfere with each other. Firstly, the model adds a coordinate attention to YOLOv5l, considers the problem of detection frame matching direction based on the loss function, adds angle information between frames, and detects the original model decoupling is partially performed to reduce information loss caused by coupling. Secondly, an improved DeeplabV3+ target segmentation model was proposed. The original DeeplabV3+ backbone network was replaced by MobileNetV2 to reduce the model complexity. A CA module was added to the Atrous Spatial Pyramid Pooling structure to fuse pixel coordinate information to improve segmentation accuracy and solve the occlusion interference problem. Finally, the end rotation angle of the target poses relative to the template pose and the optimal grasping point are obtained by point cloud registration. The performance test is carried out on the self-built 2 750 commonly used tool occlusion data set. The experimental results show that the improved model improves the detection accuracy by 0.052%, 0.968%, 6.000%, and 7.400% on mAP@0.5, mAP@0.5:0.95, 60% target object occlusion rate and 60% non-target object occlusion rate datasets. The improved semantic segmentation model on this basis improves the segmentation speed and MIOU by 33.45% and 0.625%, and the ABB IRB1200 robotic arm is used to realize the experiments on the grasping of obscured targets, which verified the feasibility and practicality of the method.

Keywords: obscured target detection; target grasping; yolo; deeplabV3+ mod; point cloud registration

收稿日期:2023-11-15 Received Date:2023-11-15
*基金项目:国家自然科学基金(61973085)、福建省自然科学基金面上项目(2022J01114)资助

0　引　言

在复杂的环境中,经常要求机器人能够自动识别遮挡目标并完成抓取任务。在这个过程中,需要机器人协同视觉系统完成对遮挡目标的检测、位姿识别、目标抓取等工作。最常用于物体检测的模型主要包括单阶段检测算法和双阶段检测算法[1],其中单阶段检测算法包括:YOLO[2-3]、SSD[4]等。双阶段检测算法包括 Faster R-CNN[5]、R-FCN[6]等。在检测框及损失函数方面,Zheng 等[7]对损失函数进行改进,利用了边界回归框中心点的位置信息,从而可以在物体的非核心部分被遮挡的情况下实现检测框生成。Bodla 等[8]以交并比(intersection over union,IOU)来更新置信度,从而减少框被误删概率。Ning 等[9]对符合阈值的预测框坐标进行加权融合,生成一个新的预测框,该方法可以得到很高的精确率和召回率。在特征提取方面,Deng 等[10]设计了多尺度混合注意力模块,增强目标的局部信息,对不相关的遮挡信息进行抑制,从而减小干扰。吴文涛等[11]利用二维目标检测器结合二维检测框和几何投影关系获取物体的三维点云,由欧氏聚类方法获得聚类点云,实现了三维目标检测。Li 等[12]设计了 YOLO-ACN 网络,在每个残差块的通道和空间维度中都引入注意力机制,该方法可以加强网络对遮挡目标和小目标的关注。尽管使用上述方法取得了显著成就,但对于遮挡目标检测仍存在以下困难:1)在遮挡环境下,由于检测框之间可能存在严重重叠的情况,所以可能导致训练过程收敛速度很慢,且检测框可能出现误删、误检等情况。2)由于遮挡的存在,网络无法很好区分目标独有的特征,当目标之间存在非常相似的特征时,检测器很难精确地对严重遮挡的目标进行分类。

此外,为了实现机器人的精准抓取,还需要确定目标的位姿,定位精度、抓取效率是影响工业机器人作业的主要因素[13]。机器人获取目标位姿的方法主要有分析法和数据驱动法。其中分析法主要依靠人工建立物理学模型获取物体特征,或依据物体的 3D 模型,计算得到其最优抓取点,该方法准确率取决于人工设计的方案,泛化能力差,无法获得未知目标物体的抓取位姿。基于数据驱动方法是目前主流的抓取方法,主要分为抽样评估法和端到端方法。抽样评估法通过神经网络生成一个物体的多个可抓取位置,从中选取最优抓取框。端到端的方法对目标图像提取特征信息,通过神经网络直接输出抓取位姿[14]。Zeng 等[15]和 Morrison 等[16]提出了一种端到端的神经网络,输出图像中所有像素的抓取质量分布。Cheng 等[17]提出了一种随机裁剪集成神经网络,解决了相似物体遮挡的检测。Wang 等[18]提出了一个基于点云的 PointNetRGPE 模型来解决姿势估计问题。上述方法应用于机器人抓取仍存在一些问题:1)抽样评估法在选择抓取位置时,需要对所有抓取框完成评估,这将导致抓取检测效率低。2)端到端方法得到抓取角度的方式过于繁琐,导致整体抓取时间过长,且泛化能力不强,易受杂乱场景中物体遮挡的影响,很难达到精确抓取。

针对上述遮挡环境下目标检测和抓取存在的问题,提出一种改进的 YOLO 遮挡目标检测模型 YOLO-CA-SD,并提出了改进的 DeeplabV3+目标分割与点云配准抓取策略。主要的贡献和创新点如下:1)采用了结合坐标注意力(coordinate attention,CA)的 YOLOv5 模型,并对定位损失函数进行改进,增加检测框之间的角度信息,提升检测器对遮挡目标的定位能力;2)对检测部分进行解耦使得分类、回归任务分开,最终使得检测精度得到提升;3)利用 DeeplabV3+解决遮挡干扰,将其主干网络替换为 MobileNetV2,提高抓取检测效率,在空洞空间金字塔池化(atrous spatial pyramid pooling,ASPP)结构中添加 CA 模块融合像素坐标信息,提高分割精度。采用基于 FPFH 的粗配准及点到面的精配准进行点云配准得到目标相对于模板的位姿,相较于数据驱动法,该方法无需对抓取框进行评估且利用语义分割解决遮挡干扰后,不易受杂乱场景遮挡影响,可以达到较高精度抓取。

1　改进的 YOLO 算法

针对 YOLOv5 对遮挡目标不能很好提取特征信息,且物体间重叠易造成特征混合,从而降低检测器性能的问题,在 YOLOv5 基础上,对主干网络、损失函数及检测部分做了改进,得到如图 1 所示的改进后模型。

主要在 Backbone 经过多个 CBS 和 C3 模块完成对图像的特征提取后,添加注意力机制 CA,以提高对遮挡目标的检测能力。最后通过 SPPF 输出,相较原 SPP 模块,在输出结果尺度不变下提高运算速度。在 Prediction 中对 Neck 输出的 3 个特征张量进行不同尺度特征图检测,相较原模型,不再耦合处理,利用解耦头(decoupled head,DH)将分类、回归任务解耦输出。

1.1　注意力机制 CA

位置信息对于模型捕捉关键信息是非常重要的,特别是对于遮挡目标而言,获取目标的坐标信息往往能更好的检测到遮挡目标。

CA[19]模块对目标横纵坐标信息进行融合并提取,其结构如图 2 所示,在后续通过语义分割网络对遮挡目标进行分割时,CA 模块的位置信息提取能力也能提升网络分割精度。

1.2　损失函数的改进

传统的定位损失函数仅考虑预测框和真实框之间的

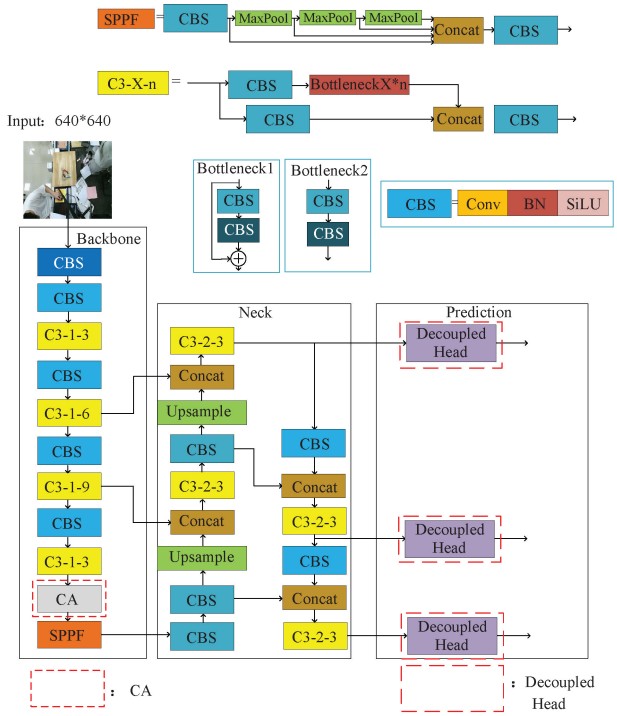


图 1　改进的 YOLOv5 网络结构

Fig. 1　Improved YOLOv5 network structure

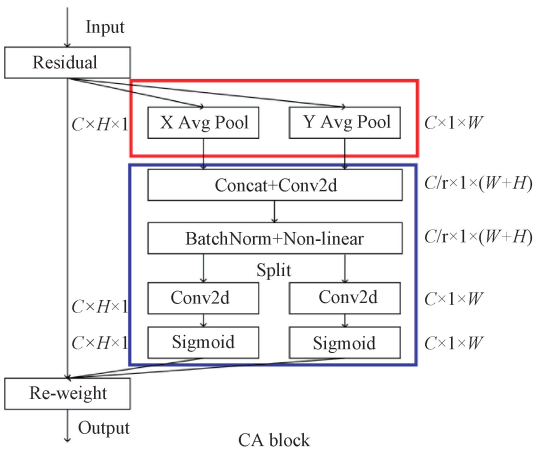


图 2　CA 模块

Fig. 2　CA module

中心点距离、最小包围框和长宽比等条件。完全交并比(complet-iou, CIOU)在距离交并比(distance-iou, DIOU)基础上加入了纵横比的惩罚项,平衡因子仅考虑纵横比的相对值,而 SCYLLA 交并比(scylla-iou,SIOU)将宽与高单独进行考虑。为了考虑框匹配方向的问题,有必要在原损失函数基础上增加框之间的角度信息[20]。

SIOU 损失函数,记为 L_{SIOU},一共由 4 个惩罚项组成,包括距离、角度、形状、交并比,如式(1)所示。

$$L_{SIOU} = 1 - IOU + \frac{\Delta + \Omega}{2} \tag{1}$$

其中,Δ 为距离惩罚项,Ω 为形状惩罚项,IOU 为交并比。

1)距离惩罚项 Δ

距离惩罚项 Δ 中包含了角度惩罚项 Λ,其定义如式(2)、(3)所示。

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma\rho_t}) \tag{2}$$

$$\begin{cases} \rho_x = \left(\dfrac{\boldsymbol{b}_{c_x}^{gt} - \boldsymbol{b}_{c_x}}{c_w}\right)^2 = \left(\dfrac{C_w}{c_w}\right)^2, t = x \\ \rho_y = \left(\dfrac{\boldsymbol{b}_{c_y}^{gt} - \boldsymbol{b}_{c_y}}{c_h}\right)^2 = \left(\dfrac{C_h}{c_h}\right)^2, t = y \\ \gamma = 2 - \Lambda \end{cases} \tag{3}$$

其中,$\boldsymbol{b}_{c_x}^{gt}$,$\boldsymbol{b}_{c_y}^{gt}$ 为真实框的中心点横纵坐标,$\boldsymbol{b}_{c_x}$,$\boldsymbol{b}_{c_y}$ 为预测框的中心点横纵坐标。如图 3 所示,C_w 代表检测框与真实框中心点横坐标距离,C_h 代表检测框与真实框中心点纵坐标距离,c_w 代表最小包围矩阵的宽,c_h 代表最小包围矩阵的高,Λ 代表角度惩罚项。

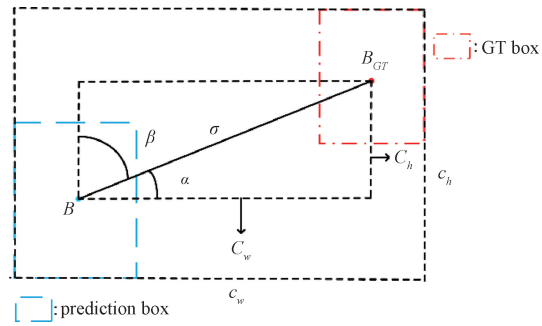


图 3　角度惩罚项示意图

Fig. 3　Angle penalty term schematic diagram

2)角度惩罚项 Λ

其意义在于将预测框快速收敛至真实框的某一个轴,使得两框中心点之间夹角 $\partial = 0$。如图 3 所示,将预测框和真实框中心点作为矩阵的对角线顶点,由于对角线与矩阵两边夹角满足 $\alpha + \beta = \dfrac{\pi}{2}$,可计算得到如式(4)、(5)所示。

$$\Lambda = 1 - 2 \times \sin^2\left(\partial - \frac{\pi}{4}\right) \tag{4}$$

$$\begin{cases} \sin(\partial) = \dfrac{C_h}{\sigma} \begin{cases} if\alpha \leqslant \dfrac{\pi}{4}, \partial = \alpha \\ if\beta \leqslant \dfrac{\pi}{4}, \partial = \beta \end{cases} \\ \sigma = \sqrt{(\boldsymbol{b}_{c_x}^{gt} - \boldsymbol{b}_{c_x})^2 + (\boldsymbol{b}_{c_y}^{gt} - \boldsymbol{b}_{c_y})^2} \\ C_h = \max(\boldsymbol{b}_{c_y}^{gt}, \boldsymbol{b}_{c_y}) - \min(\boldsymbol{b}_{c_y}^{gt}, \boldsymbol{b}_{c_y}) \end{cases} \tag{5}$$

其中,α 为对角线与矩阵的长之间的夹角,β 为对角

线与矩阵的宽之间的夹角。若 $\alpha \leqslant \frac{\pi}{4}$，对 α 进行优化，否则对 β 进行优化。σ 为预测框和真实框中心点之间的欧式距离。

3）形状惩罚项 Ω

形状惩罚项 Ω 主要是对比预测框和真实框之间长宽的差值，如式（6）、（7）所示。

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^{\theta} \tag{6}$$

$$\begin{cases} w_w = \dfrac{|w - w^{gt}|}{\max(w, w^{gt})} \\ w_h = \dfrac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \tag{7}$$

其中，$\theta \in (2,6)$，其值对于每个数据集而言是一个固定值，可以控制形状惩罚项对整个损失函数的贡献程度。w_t 代表检测框之间宽和高的比值关系，w，h 代表预测框的宽和高，w^{gt}，h^{gt} 代表真实框的宽和高。

4）交并比惩罚项 IOU

交并比惩罚项 IOU 描述预测框和真实框之间的重合度，如式（8）所示。

$$IOU = \frac{B \cap B_{GT}}{B \cup B_{GT}} \tag{8}$$

SIOU 在考虑中心点距离、检测框之间长宽比等基础上，结合了角度信息，加快了模型收敛速度。用形状惩罚项代替原损失函数中的纵横比，使得检测框之间纵横比可解释性更强。

1.3　检测部分的改进

YOLOv5 的检测部分将分类、回归二者不分开，两者理论上是需要独立进行，若耦合会造成信息丢失[21]。因此有必要对原模型耦合头（coupled head，CH）进行解耦得到 DH。

YOLOv5 从主干网络中提取到的特征张量共有 3 个，每个特征张量经过一个 1 × 1 卷积得到类别（classification，Cls）预测结果、检测框回归（regression，Reg）信息以及置信度（objectness，Obj）的信息。其中 n_a 代表 anchor 的数量，C 为类别数。假设输入张量尺寸为 $H \times W \times 256$，进行 DH 操作，如图 4 所示。

DH 对输入利用一个 1×1 卷积调整其通道数。得到两个平行分支，上分支用于分类检测，下分支用于定位和置信度检测，每个分支通过两个 3 × 3 卷积和一个 1 × 1 卷积后，得到检测结果。

2　基于 DeeplabV3+的遮挡目标分割改进算法

目前，一般的抓取策略包括抽样评估法和端到端方法，这些方法需要对多个抓取位置进行评估，计算量过大，或者无法很好解决遮挡干扰，精度不高。对语义分割模型 DeeplabV3+进行改进，实现对遮挡目标的分割，解决遮挡干扰。并通过与模板点云配准得到目标的抓取点坐标与末端旋转角。

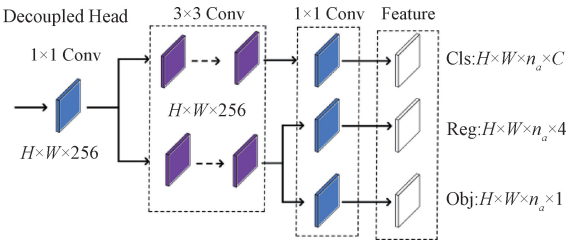


图 4　DH 结构

Fig. 4　DH structure

2.1　改进的 DeeplabV3+遮挡目标分割模型

DeeplabV3+分割速度较慢，原因在于其主干网络模型参数量大，用轻量级 MobileNetV2 取代原有的主干网络可以提高分割效率，并添加 CA 模块，改进后的 DeeplabV3+网络结构如图 5 所示，主要包括编码器和解码器两部分。

当图像输入编码器中，由 MobileNetV2 生成两个不同深度的特征层，分别为浅层特征层和深层特征层，其中将 MobileNetV2 的第 5 个下采样层（网络第 7 层）的步距设为 1，因为下采样层过多会失去更多像素特征，最终会导致网络特征信息获取过少，所以为了减少像素特征丢失，可以采用 4 个下采样层。在深层特征层，ASPP 使用膨胀率为 6、12、18 的 3×3 膨胀卷积进行信息提取，使网络可以融合多种不同尺度语义信息，同时利用 CA 模块进一步增强特征信息提取能力和定位能力。最后将特征层进行堆叠，再经过 1×1 卷积降维，将输出的高语义特征信息输入解码器中进行上采样。

将浅层特征层输入解码器中经过 1×1 卷积后与编码器的输出进行特征融合，再经过 3×3 的卷积完成特征提取，再利用上采样调整输出图像的尺寸，完成分割。

2.2　确定目标抓取位姿

在有遮挡情况下，通过改进的 DeeplabV3+模型对目标进行分割，遍历得到目标检测框范围内对应像素值的像素点坐标，将分割后的目标转换成对应的点云，与模板点云（机械臂抓取时末端关节角为 0°时的目标点云）进行配准，得到变换矩阵，计算目标在 Z 轴方向上相对于模板位姿的旋转角以及最优抓取点坐标。

采用快速点特征直方图（fast point feature histograms，FPFH）作为粗配准方法，对某点及其领域内的点进行统计分析，给出该点的几何特征 φ，ϕ，θ 描述如式（9）所示。

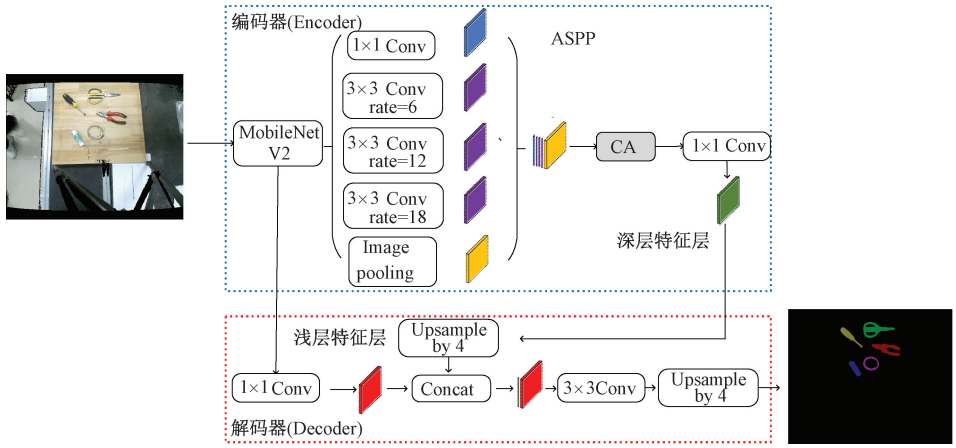


图 5　改进后的 DeeplabV3+结构图

Fig. 5　Improved DeeplabV3+ structure diagram

$$\begin{cases}\varphi = [(p_1 - p_0) \times \boldsymbol{n}_0] \cdot \boldsymbol{n}_1 \\ \phi = \boldsymbol{n}_0 \cdot \dfrac{p_1 - p_0}{d} \\ \theta = \arctan([\boldsymbol{n}_0 \times (p_1 - p_0) \times \boldsymbol{n}_0] \cdot \boldsymbol{n}_1, \boldsymbol{n}_0 \cdot \boldsymbol{n}_1)\end{cases} \tag{9}$$

其中，p_0、p_1 为如图 6 所示的源点云 P 上的点，$\boldsymbol{n}_0$ 和 $\boldsymbol{n}_1$ 分别为 p_0 和 p_1 垂直于点表面的法向量，d 为 p_0 和 p_1 的欧氏距离。

将所有近邻点对于 p_0 的特征描述 φ,ϕ,θ 表示成简化点特征直方图(simplified point feature histograms, SPFH)。计算每个近邻点 $p_1,p_2,\cdots,p_k$ 的 SPFH，并以相对于 p_0 的距离 $w_k = |p_0 - p_k|$ 作为权重，将所有特征描述加权融合作为 p_0 的特征。具体如式(10)所示。

$$FPFH(p_0) = SPFH(p_0) + \frac{1}{k}\sum_{i=1}^{k}\frac{1}{w_k} \cdot SPFH(p_k) \tag{10}$$

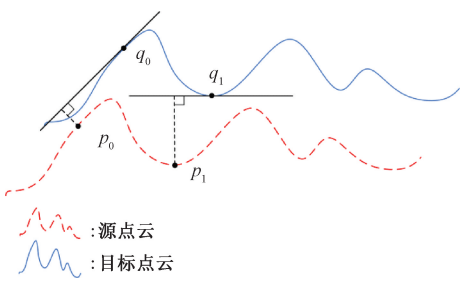


图 6　点到面的优化函数示意图

Fig. 6　Point to surface optimization function diagram

将模板点云与目标点云体素化后基于 FPFH 进行粗配准，将其结果作为精配准的初始位姿。由于基于点到点的精配准仅考虑点间的欧氏距离，不同的对应点之间会产生干扰，而点到面的精配准方法以点到对应平面距离作为约束，不易受点间干扰，且收敛速度相比点到点更快。点到面的方式如图 6 所示，以源点云 P 中的点 p_i 到目标点云 Q 中对应点 q_i 的切平面的距离平方和作为优化函数，如式(11)所示，并采用非线性最小二乘方法求解这个优化函数。

$$f(M) = argmin_M \sum_{i=1}^{n}((\boldsymbol{M} \cdot \boldsymbol{p}_i - \boldsymbol{q}_i) \cdot \boldsymbol{n}_i)^2 \tag{11}$$

其中，$\boldsymbol{n}_i$ 为目标点云中点 q_i 的法向量，$\boldsymbol{p}_i = (p_{ix}, p_{iy}, p_{iz}, 1)^{\mathrm{T}}$，$\boldsymbol{q}_i = (q_{ix}, q_{iy}, q_{iz}, 1)^{\mathrm{T}}$。

假设短时间内，点云间旋转角度很小，存在近似条件：绕 x,y,z 轴的旋转角 $\alpha,\beta,\gamma \to 0$，并且根据极限条件：$\sin\theta \approx \theta$，$\cos\theta \approx 1$，忽略高阶无穷小 θ^2。变换矩阵 $\boldsymbol{M}$ 可写为式(12)所示。

$$\boldsymbol{M} = \begin{bmatrix} 1 & \alpha\beta - \gamma & \alpha\gamma + \beta & t_x \\ \gamma & \alpha\beta\gamma + 1 & \beta\gamma - \alpha & t_y \\ -\beta & \alpha & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

$\boldsymbol{M}$ 根据极限条件可以表示为近似变换矩阵 $\boldsymbol{M}'$ 如式(13)所示。

$$\boldsymbol{M}' = \begin{bmatrix} 1 & -\gamma & \beta & t_x \\ \gamma & 1 & -\alpha & t_y \\ -\beta & \alpha & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

其中，t_x,t_y,t_z 为平移量。

3　实验与分析

遮挡目标检测与抓取的整体流程如图 7 所示。首先完成对遮挡目标的检测，同时完成遮挡目标的语义分割。最后对分割后的目标点云与模板点云进行配准，得到抓取目标的姿态角和抓取点坐标实现机器人精确抓取。

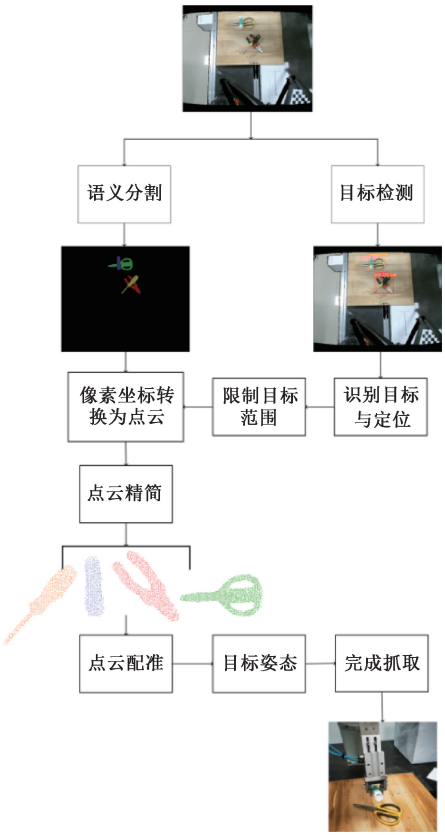


图 7　机器人抓取实验流程图

Fig. 7　Schematic diagram of the process of robot grasping experiment

3.1　实验数据集

目标检测训练数据包含 2 750 张图像(包括 250 张不包含任何目标物体的背景图),其中各目标物体出现次数分别为:钳子 1 192 次、剪刀 1 174 次、螺丝刀 1 193 次、固体胶 1 174 次、胶带 1 127 次。验证集为真实采样的 600 张 20%~40%遮挡率的图像。

语义分割训练集采用彩色与深度图像配准后的配准图像,如图 8 所示,训练集共 800 张图像,其中验证集共 200 张图像,每张图片均包含 5 种目标物体。

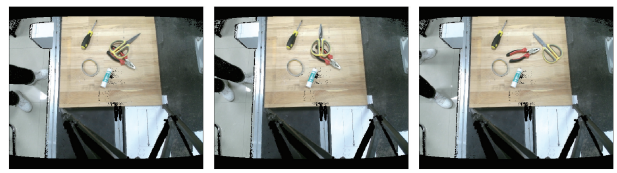


图 8　语义分割数据集图像示例

Fig. 8　Semantic segmentation dataset image example

目标检测测试集一共 6 个:$T_{20\%}$ 为 600 张 20%遮挡率的图像,$T_{40\%}$ 为 600 张 40%遮挡率的图像,$T_{60\%}$ 为 500 张 60%左右遮挡率的图像,$T^B_{20\%}$ 为 500 张非目标物 20%遮挡率的图像,$T^B_{40\%}$ 为 500 张非目标物 40%遮挡率的图像,$T^B_{60\%}$ 为 500 张非目标物 60%~80%遮挡率的图像,如图 9(a)~(f)所示。

(a) 20%遮挡数据集$T_{20\%}$　(b) 40%遮挡数据集$T_{40\%}$　(c) 60%遮挡数据集$T_{60\%}$

(a) 20% occlusion dataset $T_{20\%}$　(b) 40% occlusion dataset $T_{40\%}$　(c) 60% occlusion dataset $T_{60\%}$

(d) 非目标20%遮挡数据集$T^B_{20\%}$　(e) 非目标40%遮挡数据集$T^B_{20\%}$　(f) 非目标60%遮挡数据集$T^B_{20\%}$

(d) Non-target 20% occlusion dataset $T^B_{20\%}$　(e) Non-target 40% occlusion dataset $T^B_{20\%}$　(f) Non-target 20% occlusion dataset $T^B_{20\%}$

图 9　不同遮挡率下测试图像

Fig. 9　Test images under different occlusion ratios

3.2　评价指标

采用精确率、召回率、平均精确度均值和准确度来衡量检测效果,其中:

1)精确率(precision,P)

$$P = \frac{TP}{TP + FP} \tag{14}$$

2)召回率(Recall,R)

$$R = \frac{TP}{TP + FN} \tag{15}$$

其中，*TP* 代表预测结果和真实结果均为真的样本数，*FP* 代表预测结果为真但真实结果为假的样本数，*FN* 代表预测结果为假但真实结果为真的样本数。

3）平均精确度均值（mean average precision，mAP）

AP_i 等于第 i 个类别目标 PR 曲线，即以 R 作为横坐标，P 作为纵坐标的曲线，所围成的面积。*mAP* 代表所有 AP_i 平均值：

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{16}$$

4）准确率（accuracy，A）

A 从整体角度描述检测器的优劣，其含义为检测结果与真实标签一致的数量占总样本数的比例，如式（17）所示。

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{17}$$

其中，*TN* 是预测结果与真实结果均为假的样本数。

5）均交并比（mean intersection over union，MIOU）

语义分割网络常将 MIOU 作为评价指标，如式（18）所示。

$$MIOU = \frac{1}{k+1}\sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{18}$$

其中，p_{ij} 是将 i 类预测为 j 类的数目。$k+1$ 代表共 k 类目标和 1 个背景类。MIOU 取值范围在［0，1］之间，值越大代表分割精度越高。

3.3　遮挡目标检测实验结果与分析

为了提高检测准确率，实验中采用的原模型及改进模型都是以 YOLOv5l 作为基础模型，训练代数均为 120 代，Batch size 为 16。在对测试集进行测试时，为了与实际机器人工作时条件一致，所有测试集实验中置信度阈值均为 0.2，交并比阈值均为 0.1。主要对比不同模型在 60%重遮挡率测试集下的检测能力，以此检验模型的泛化能力及对严重遮挡目标检测精度。

遮挡目标检测实验结果如表 1 所示，由表 1 可知，YOLO-CA-SD 在 60%遮挡测试集上准确率 *A* 相比 YOLO-CA-S 模型提高 3.80%，相比原模型提高 6.00%，说明对检测部分进行解耦后对于困难样本的检测能力得到了很大的提升，因对简单样本存在过拟合的原因，导致 20%遮挡测试检测精度有所下降。YOLO-CA-SD 相比原模型在 mAP@0.5:0.95 提高 0.968%（mAP@0.5:0.95 指 IOU 阈值从 0.5 到 0.95，以 0.05 为一个步长求所有 mAP 的均值作为结果。）虽然在 60%的重遮挡率测试集下的检验能力与 YOLOv7 相差不大，但是在 40%遮挡率上，YOLO-CA-SD 取得了更好的准确率，可以看出其在不同遮挡率情况下，其均能保持较高准确率。

表 1　不同模型下的检测实验结果对比

Table 1　Comparison of test results under different models

模型	*mAP* @0.5	*mAP* @0.5:0.95	$T_{20\%}$ *A*/%	$T_{40\%}$ *A*/%	$T_{60\%}$ *A*/%
YOLOv5l	0.992 02	0.803 44	99.33	97.83	80.60
YOLO-CA	0.992 15	0.808 05	99.50	**98.17**	81.40
YOLO-CA-S	0.992 31	0.808 11	**99.83**	**98.17**	82.80
YOLO-CA-SD	0.992 54	**0.813 12**	99.67	**98.17**	86.60
YOLOv7	**0.995 9**	0.786 6	99.33	96.33	**87.60**

四者在实际测试样本上的对比如图 10（a）~（l）所示，由图 10（b）~（f）和（j）~（l）可知，当背景物和检测目标相互严重遮挡时，原始模型、YOLO-CA-S 以及 YOLOv7 模型都会出现漏检以及误检。同时在分辨率更低的配准图的检测上，原模型出现了漏检、YOLOv7 模型出现了误检，进一步说明检测头耦合易造成漏检和误检。而 YOLO-CA-SD 在严重遮挡时仍能识别目标，从图 10（g）~（i）的检测结果可得，YOLO-CA-SD 对检测框定位相比原模型更精确，对于分辨率不高且存在噪声干扰的配准图仍能精确地检测出目标。

为了验证模型在非目标遮挡情况下的检测效果，利用原模型及改进模型分别对不同遮挡率的非目标遮挡测试集进行验证，以此检验模型的泛化能力及对非目标物体遮挡的检测精度，实验结果如表 2 所示。

表 2　不同模型下的非目标物体遮挡检测实验结果对比

Table 2　Comparison of experimental results of non-target object occlusion detection under different models

模型	$T^B_{20\%}$ *A*/%	$T^B_{40\%}$ *A*/%	$T^B_{60\%}$ *A*/%
YOLOv5l	94.60	92.40	86.60
YOLO-CA	98.40	94.80	92.00
YOLO-CA-S	**98.60**	95.20	91.60
YOLO-CA-SD	95.60	**95.60**	**94.00**
YOLOv7	94.40	90.60	86.80

由表 2 可知，YOLO-CA-SD 在非目标物体 60%遮挡率测试集上准确率 *A* 比 YOLO-CA-S 提高 2.40%，比原模型提高 7.40%。对比原模型及 YOLOv7，YOLO-CA-SD 在非目标物体遮挡检测精度和多目标物体遮挡检测精度有所提升，进一步证明 YOLO-CA-SD 相较基线网络对困难样本检测能力有较大提升。

对非目标物体不同遮挡率的检测结果对比如图 11（a）~（l）所示，可看出在非目标物体遮挡情况下，当遮挡率 20%左右时，4 个模型都能很好的检测出目标物，但 YOLOv5l 和 YOLOv7 会出现误检出的情况，这也进一步说明 YOLO 模型检测部分信息耦合对目标检测易产生误差。当遮挡率在 40%时，原始模型开始出现漏检，且 YOLOv7 仍出现误检。在遮挡率在 60%左右时，YOLO-

(a) YOLOv5检测彩色图1
(a) YOLOv5 detection Figure 1

(b) YOLOv5检测彩色图2
(b) YOLOv5 detection Figure 2

(c) YOLOv5配准图
(c) YOLOv5 detection registration diagram

(d) YOLO-CA-S检测彩色图1
(d) YOLO-CA-S detection color Figure 1

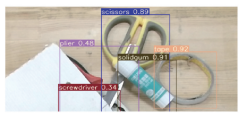
(e) YOLO-CA-S检测彩色图2
(e) YOLO-CA-S detection color Figure 2

(f) YOLO-CA-S检测配准图
(f) YOLO-CA-S detection registration diagram

(g) YOLO-CA-SD检测彩色图1
(g) YOLO-CA-SD detection color Figure 1

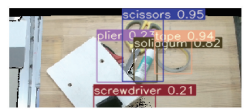
(h) YOLO-CA-SD检测彩色图2
(h) YOLO-CA-SD detection color Figure 2

(i) YOLO-CA-SD检测配准图
(i) YOLO-CA-SD detection registration diagram



(j) YOLOv7检测彩色图1
(j) YOLOv7 detection Figure 1

(k) YOLOv7检测彩色图2
(k) YOLOv7 detection Figure 2



(l) YOLOv7检测配准图
(l) YOLOv7 detection registration diagram

图 10　背景及目标间相互严重遮挡的检测结果对比

Fig. 10　Comparison of detection results of severe occlusion between background and target

(a) YOLOv5在20%遮挡率时
(a) YOLOv5 at 20% occlusion

(b) YOLOv5在40%遮挡率时
(b)YOLOv5 at 40% occlusion

(c) YOLOv5在60%遮挡率时
(c) YOLOv5 at 60% occlusion

(d) YOLO-CA-S在20%遮挡率时
(d) YOLO-CA-S at 20% occlusion

(e) YOLO-CA-S在40%遮挡率时
(e) YOLO-CA-S at 40% occlusion

(f) YOLO-CA-S在60%遮挡率时
(f) YOLO-CA-S at 60% occlusion

(g) YOLO-CA-S在20%遮挡率时
(g) YOLO-CA-S at 20% occlusion

(h) YOLO-CA-S在40%遮挡率时
(h) YOLO-CA-S at 40% occlusion

(i) YOLO-CA-S在60%遮挡率时
(i) YOLO-CA-S at 60% occlusion

(j) YOLOv7在20%遮挡率时
(j) YOLOv7 at 20% occlusion

(k) YOLOv7在40%遮挡率时
(k) YOLOv7 at 40% occlusion

(l) YOLOv7在60%遮挡率时
(l) YOLOv7 at 60% occlusion

图 11　非目标物体不同遮挡率的检测结果对比

Fig. 11　Comparison of detection results of different occlusion rates of non-target objects

CA-SD 仍能在高遮挡率下完成检测，且保持较好的定位精度，而另外 3 个模型已出现漏检。

目标物体增大遮挡面积时 YOLO-CA-SD 模型检测结果如图 12(a)～(e)所示，可分析出不同目标物受到非目标物的遮挡超过 70%后，检测器的精度会开始明显下降，当遮挡率在 80%～90%时，检测器会出现明显漏检，且即

使检测出目标，也会出现定位偏差。综上所述，YOLO-CA-SD 在非目标物体遮挡率在 70%～80%时，仍能保持一定的检测准确度。

(a) 固体胶 (a) Solid glue　(b) 双面胶 (b) Double-sided tape　(c) 钳子 (c) Plier

(d) 剪刀 (d) Scissors　(e) 螺丝刀 (e) Screwdriver

图 12　目标物体增大遮挡面积时 YOLO-CA-SD 模型检测结果

Fig. 12　YOLO-CA-SD model detection results when the target object increases the occlusion area

通过上述对比实验，可得出 YOLO-CA-SD 模型的有效性，同时也能为后续完成遮挡目标分割奠定基础。

3.4　语义分割实验结果与分析

对原模型 DeeplabV3+Xce、DeeplabV3+Mob1（5 个下采样层）、DeeplabV3+Mob（4 个下采样层）、DeeplabV3+Mob+SE 以及 DeeplabV3+Mob+CA 进行训练。以模型计算量（flops）及模型参数量（params）来衡量模型的复杂程度。

原模型的训练因显存大小限制，Batch Size（BS）只能设置为 2，为了保证模型训练充分，进行两组不同 Batch size 大小的实验，具体实验结果如表 3 所示。

表 3　不同模型语义分割对比实验结果

Table 3　Comparison of experimental results with different semantic segmentation models

模型	MIOU/%		FLOPs	Params
	BS=2	BS=8		
DeeplabV3	88.392	–	46 574 089 216	54 709 702
DeeplabV3+Mob1	88.921	88.449	**7 652 286 720**	**5 814 294**
DeeplabV3+Mob	88.966	88.534	**7 652 286 720**	**5 814 294**
DeeplabV3+Mob+SE	88.848	88.632	7 653 496 320	6 020 454
DeeplabV3+Mob+CA	**89.017**	**88.763**	7 665 852 160	6 124 294

由表 3 可知，DeeplabV3+Mob 的模型计算量及模型参数量相比原模型 DeeplabV3 显著减小，同时 DeeplabV3+Mob 相比于 DeeplabV3+Mob1 分割精度更高。虽然 DeeplabV3+Mob+CA 的模型计算量和模型参数量要比 DeeplabV3+Mob1 略有增加，但是其分割精度得到了提高，有利于机械臂的精确抓取。随机抽取 15 张图像对比两者的分割速度，DeeplabV3+Mob+CA 平均分割速度为 0.195 秒/张，原模型 DeeplabV3 平均分割速度为 0.293 秒/张，改进后的模型分割速度相较于原模型提升 33.45%。

DeeplabV3 + Mob + CA 对比 DeeplabV3 + Mob、DeeplabV3+Mob+SE 及原模型，在 BS = 2 时，MIOU 分别提升了 0.051%、0.169%和 0.625%。实际分割效果如图 13（a）～（b）所示，由图 13（a）可知，DeeplabV3+Mob + CA 相较原模型整体分割效果更好，特别是在细节上的处理，可明显看出 DeeplabV3+Mob +CA 能够检测出螺丝刀头部细小边界，由图 13（b）可知，而原网络对边界处理不够清晰。

(a) DeeplabV3+Mob+CA分割结果

(a) DeeplabV3+Mob+CA segmentation results

(b) DeeplabV3+分割结果

(b) DeeplabV3+segmentation results

图 13　实际分割效果对比

Fig. 13　Comparison of the actual segmentation effect

3.5　点云配准及机械臂抓取实验结果与分析

为了验证点云配准方法的准确性，对不同位姿的物体进行实验。以剪刀为例，利用机械臂完成 10 次不同位姿下剪刀的抓取，记录抓取时机械臂末端关节的旋转角作为真实值，以点云配准得到的旋转角作为测试值。其中，精配准算法均以粗配准得到的变换矩阵作为初始位姿，结果如表 4 所示。

表 4　抓取剪刀时的关节 6 旋转角估计值

Table 4　Estimate of the joint 6 rotation Angle when grasping scissors

真实旋转角/(°)	粗配准/(°)	点到面/(°)	点到点/(°)
−42	−40.576	−39.835	−40.147
−140	−145.079	−145.960	−145.790
25	25.601	24.944	25.092
90	93.638	90.908	92.069
60	63.900	61.921	62.199
−15	−17.197	−17.298	−17.248
−70	−71.921	−71.652	−71.740
−65	−63.848	−63.342	−63.998
15	17.076	16.842	16.989
70	72.631	72.910	72.812

结果表明，粗配准得到不同位姿下旋转角的平均误

差是 2.462°,点到面的精配准得到的旋转角平均误差是 2.137°,点到点的精配准得到的旋转角平均误差是 2.179°。对其他 4 种目标物体同样采取基于 FPFH 的粗配准以及点到面的精配准方式,得到钳子的旋转角平均误差为 1.879°,螺丝刀的旋转角平均误差为 1.678°,固体胶旋转角平均误差为 1.612°,胶带的旋转角平均误差为 2.475°。

抓取实验需要考虑夹持器尺寸和待测目标的形状及尺寸,夹持器最大夹持宽度为 34 mm,最小夹持宽度为 20 mm。以剪刀为例,剪刀的把手处最大宽度为 28 mm,最小宽度为 26 mm,因此可抓取位置较固定,对剪刀把手不同位置进行 30 次抓取实验,得到剪刀最优抓取位置如图 14 所示位置,记录此时机械臂末端对应的三维坐标 G_c = (− 13.75,192.76,715.27) 。

利用表 4 中得到剪刀不同位姿数据以及式(13)可以得到点云配准得到模板点云到目标点云的变换矩阵 $\boldsymbol{M}'$,模板点云中的抓取点通过变换矩阵 $\boldsymbol{M}'$ 可以得到在目标点云下的抓取点 G_c' ,如表 5 所示。

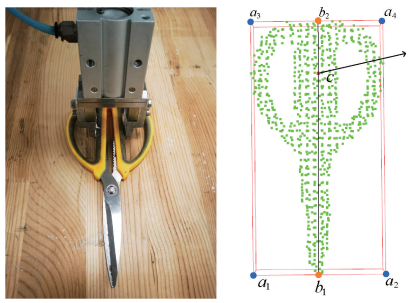


图 14　剪刀最优抓取点

Fig. 14　Optimal grasping point of scissors

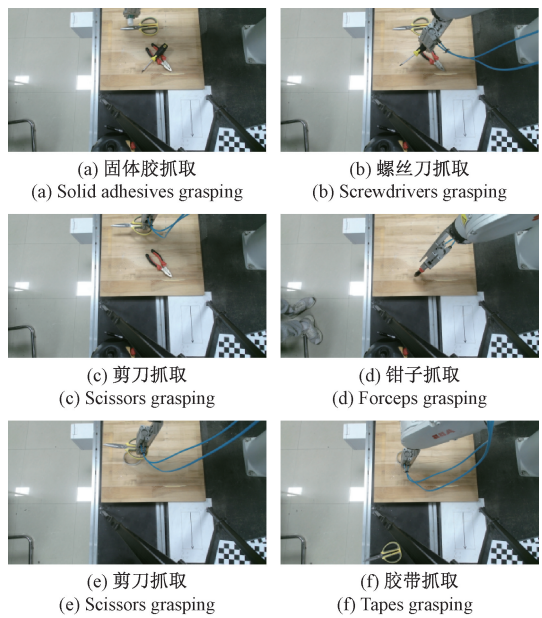
表 5　精确配准得到的三维坐标

Table 5　The three-dimensional coordinates by precise registration

真实旋转角/(°)	真实抓取点/mm	点到面抓取点/mm
−42	(−21.82,173.43,711.49)	(−23.02,172.46,711.02)
−140	(6.50,229.56,722.64)	(6.36,231.95,722.37)
25	(−5.12,165.23,711.30)	(−7.09,167.61,711.97)
90	(−6.47,165.61,711.37)	(−5.58,164.53,711.12)
60	(−9.18,165.65,711.37)	(−10.99,164.51,711.57)
−15	(1.12,253.53,733.44)	(1.30,253.72,733.58)
−70	(−3.51,172.40,712.81)	(−4.89,174.26,712.77)
−65	(−15.59,171.86,721.69)	(−18.08,174.03,721.56)
15	(−9.16,170.56,717.77)	(−10.42,169.41,717.48)
70	(1.59,152.97,708.89)	(−0.13,152.76,708.78)

由表 5 可知,点到面精配准得到对应抓取点与真实抓取点在 x,y,z 轴上的平均误差分别为 1.31 mm,1.35 mm,0.26 mm。对其他 4 种目标物体采取同样的配准方得到钳子 x,y,z 轴上的平均误差分别为 0.61 mm,0.67 mm,0.28 mm。螺丝刀在 x,y,z 轴上的平均误差为 1.86 mm,1.50 mm,0.24 mm。胶带在 x,y,z 轴上的平均误差为 0.34 mm,0.69 mm,0.44 mm。固体胶在 x,y,z 轴上的平均误差为 0.29 mm,0.49 mm,0.25 mm。

将三维坐标转化至机械臂坐标系下,完成实际抓取实验,其中部分抓取实验结果如图 15(a)~(f)所示。在多目标遮挡混合实验环境下进行 30 组实验得到各个目标物体的检测、分割与抓取实验结果如表 6 所示。由表 6 结果分析,螺丝刀与固体胶抓取成功率较高,因为其可抓取部位更多,且形状更规则。但螺丝刀分割准确率最低,原因是其头部较细,易与其他物体混合,导致分割不完整。钳子的抓取准确率最低是因其可抓取位置仅在头部,如图 15(d)所示,由于重心导致在抓取过程中易掉落,所以出现较多抓取失败的情况。



(a) 固体胶抓取　(a) Solid adhesives grasping
(b) 螺丝刀抓取　(b) Screwdrivers grasping
(c) 剪刀抓取　(c) Scissors grasping
(d) 钳子抓取　(d) Forceps grasping
(e) 剪刀抓取　(e) Scissors grasping
(f) 胶带抓取　(f) Tapes grasping

图 15　不同物体实验结果

Fig. 15　Experimental results on different objects

表 6　不同目标物体的检测、分割、抓取准确率

Table 6　Accuracy of detection, segmentation and capture of different target　(%)

目标物体	检测准确率	分割准确率	抓取准确率
钳子	96.00	91.33	81.63
剪刀	95.33	92.00	84.18
螺丝刀	94.00	88.00	86.60
固体胶	93.33	92.67	84.34
胶带	92.67	91.33	82.65

4　结　论

针对遮挡环境中机器人抓取存在的问题,提出了一

种改进的遮挡目标检测与抓取算法。在目标检测方面，对 YOLOv5 模型进行改进，添加 CA 模块，考虑检测框匹配方向的问题，并对检测部分进行解耦，提高对目标的检测准确率。在姿态获取方向，首先通过 DeeplabV3+解决遮挡干扰问题，用 MobileNetV2 替换其主干网络，减小模型复杂度，在 ASPP 结构中添加 CA 模块融合像素坐标信息提高分割精度。完成语义分割后，将目标二维坐标信息转向三维坐标，利用点云配准得到目标姿态相对于模板姿态的末端旋转角及最优抓取点。相较数据驱动法，该方法无需对抓取框进行评估且利用语义分割解决遮挡干扰后，可以达到较高精度抓取。最后，实现遮挡目标的抓取，通过消融实验，验证了该方法的有效性。在未来如何更进一步的提高目标分割精度，以及更高效获取目标抓取位姿，或者是对遮挡目标进行复原等方向都值得深入探究。

参考文献

[1] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.

[2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[3] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[5] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

[6] DAI J F, LI Y, HE K M, et al. R-FCN: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29, DOI:10.48550/arXiv.1605.06409.

[7] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[8] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561-5569.

[9] NING CH CH, ZHOU H J, SONG Y, et al. Inception single shot multibox detector for object detection[C]. 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2017: 549-554.

[10] DENG T M, LIU X H, WANG L. Occluded vehicle detection via multi-scale hybrid attention mechanism in the road scene[J]. Electronics, 2022, 11(17): 2709-2723.

[11] 吴文涛，何赟泽，杜旭，等. 融合相机与激光雷达的目标检测与尺寸测量[J]. 电子测量与仪器学报，2023，37(6)：169-177.
WU W T, HE Y Z, DU X, et al. Fusing camera and Lidar for object detection and dimensional measurement[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(6): 169-177.

[12] LI Y J, LI SH SH, DU H H, et al. YOLO-ACN: Focusing on small target and occluded object detection[J]. IEEE Access, 2020, 8: 227288-227303.

[13] 乔贵方，蒋欣怡，高春晖，等. 基于多目标优化的工业机器人位置与姿态精度提升方法[J]. 仪器仪表学报，2023，44(12)：217-224.
QIAO G F, JIANG X Y, GAO CH H, et al. Method for improving position and attitude accuracy of industrial robots based on multi-objective optimization[J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 217-224.

[14] CHOI C, SCHWARTING W, DELPRETO J, et al. Learning object grasping for soft robot hands[J]. IEEE Robotics & Automation Letters, 2018, 3(3): 2370-2377.

[15] ZENG A, SONG SH R, YU K T, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching[J]. The International Journal of Robotics Research, 2022, 41(7): 690-705.

[16] MORRISON D, CORKE P, LEITNER J. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach[J]. ArXiv preprint arXiv: 1804.05172, 2018.

[17] CHENG B, WU W N, TAO D P, et al. Random cropping ensemble neural network for image classification in a robotic arm grasping system[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(9): 6795-6806.

[18] WANG ZH T, XU Y T, HE Q, et al. Grasping pose estimation for SCARA robot based on deep learning of point cloud[J]. The International Journal of Advanced Manufacturing Technology, 2020, 108: 1217-1231.

[19] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[20] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[J]. ArXiv preprint arXxiv: 2205.12740, 2022.

[21] SONG G L, LIU Y, WANG X G. Revisiting the sibling head in object detector[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11563-11572.

作者简介

林哲,2020 年于中国计量大学获得学士学位,2023 年于福州大学获得硕士学位,现为澳门理工大学博士研究生,主要研究方向为机器人多模态融合感知技术。

E-mail: Linzhe0208@outlook.com

Lin Zhe received his B. Sc. degree from China Jiliang University in 2020, M. Sc. degree from Fuzhou University in 2023. Now he is a Ph. D. student at the Macao Polytechnic University. His main research interest includes robot multimodal fusion sensing technology.

陈丹(通信作者),2011 年于中国科学院沈阳自动化研究所获得博士学位,现为福州大学电气工程与自动化学院副教授,主要研究方向为机器视觉、机械臂路径规划等。

E-mail: 632151807@qq.com

Chen Dan (Corresponding author) received her Ph. D. degree from Shenyang Institute of Automation Chinese Academy of Sciences in 2011. Now she is an associate professor at the Fuzhou University. Her main research interests include robot vision, robotic arm path planning, etc.