基于目标识别和FC-GQCNN网络的机械臂抓取检测技术研究

白杨凡[1],卞永明[1],杨继翔[2],杨　濛[1]

(1.同济大学 机械与能源工程学院,上海 201804; 2.宁波工程学院 机器人学院,浙江 宁波 315211)

摘要: 本文提出了一种基于目标识别和全卷积抓取质量网络(FC-GQCNN)的机械臂抓取检测技术。针对传统GQCNN在实际应用中存在的计算效率低、特征重复计算等问题,提出了一种改进的FC-GQCNN。该网络通过将GQCNN的全连接层替换为1×1卷积层,使其能够处理任意尺寸的输入图像。同时,将FC-GQCNN与YOLOv8目标识别算法相结合,构建了YOLOv8-FCGQCNN级联结构,有效解决了复杂环境下目标物体的识别和定位问题。实验结果表明:该方法在10类不同物体的抓取任务中有86%的抓取成功率,单帧平均检测时间仅为0.09 s,相比传统GQCNN的推理速度提升了22倍,显著提高了系统效率。该方法可以准确地检测感兴趣的物体的抓取位姿,并且较基准方法具有更高的可靠性。

关键词: 目标识别; 抓取位姿检测; 机械臂抓取系统; 算法融合

中图分类号: TP 241; TP 391.41　　**文献标志码:** A　　**文章编号:** 1672-5581(2025)02-0227-06

Research on robotic arm grasping detection technology based on target recognition and FC-GQCNN network

BAI Yangfan[1], *BIAN Yongming*[1], *YANG Jixiang*[2], *YANG Meng*[1]

(1.School of Mechanical Engineering, Tongji University, Shanghai 201804, China; 2.School of Robotics, Ningbo University of Technology, Ningbo 315211, Zhejiang, China)

Abstract: This paper proposes a robotic grasping technique based on object recognition and fully convolutional grasp quality convolutional neural network (FC-GQCNN). To address the limitations of traditional GQCNN, such as low computational efficiency and redundant feature calculations, an improved FC-GQCNN is developed. By replacing the fully connected layers in GQCNN with 1×1 convolutional layers, the proposed network can handle input images of arbitrary sizes. Furthermore, the integration of FC-GQCNN with the YOLOv8 object detection algorithm forms a YOLOv8-FCGQCNN cascade structure, effectively solving the challenges of object recognition and localization in complex environments. Experimental results demonstrate that this method achieves an 86% grasp success rate across 10 different objects, with an average detection time of 0.09 s per frame, which is 22 times faster than traditional GQCNN, significantly improving system efficiency. This method can accurately detect the grasping position of the object of interest and has higher reliability than the baseline method.

Key words: object recognition; grasp pose estimation; robotic arm grasping system; algorithm integration

基金项目:国家重点研发计划资助项目(2023YFC3806603);国家自然科学基金青年科学基金资助项目(52205279)

作者简介:白杨凡(1999—),男,硕士生。E-mail:2132643@tongji.edu.cn

通信作者:杨　濛(1990—),男,讲师,博士。E-mail:yangmeng@tongji.edu.cn

机器人可以代替人类在危险的环境中执行繁重的工作,而机械臂是工业机器人顺序运输物体的关键可执行部件[1]。操纵机械臂进行准确的抓取是机械臂研究领域的巨大挑战[2]。

机器人抓取系统通常由抓取检测、抓取规划和控制单元组成[3-4]。为了完成抓取任务,机器人必须提前检测到感兴趣的物体[5]。传统方法通常先检测并识别物体的位置,然后通过几何分析制定抓取计划[6],然而,几何分析过程中往往涉及大量计算。得益于深度学习在目标识别任务中的成功应用[7-8],可以设计一种深度学习模型,直接训练并生成抓取检测方案,从而实现从图像输入到抓取输出的端到端流程。Mahler等[9]提出基于深度图像的全卷积抓取质量网络(grasp quality convolutional neural network,GQCNN)方法,通过物体边缘检测生成抓取候选框,再对候选区域进行旋正处理后预测抓取成功率,实现了85.7%的分类准确率,但采样阶段耗时多且图像块间特征重复计算增加了计算量。Morrison等[10]提出基于语义分割的GGCNN,通过编码器-解码器结构实现像素级抓取框检测,生成抓取概率、角度及爪手宽度特征图,杂波环境中抓取成功率为81%。这些方法在识别物体方面存在一定局限性,难以有效减少无关物体对抓取检测的影响,同时在抓取视野中具有身份的预定义物体时表现不足。

本文提出了一种基于全卷积网络的改进抓取质量网络(fully convolutional grasp quality convolutional neural network,FC-GQCNN),并结合YOLOv8检测并定位感兴趣的目标区域(region of interest,ROI),有效减少了其他非抓取物体对抓取点生成的干扰。在真实场景下,通过对10种不同物体的抓取,评估了该算法与其他物体检测和抓取算法组合的性能。结果表明,该算法抓取成功率为86%,单帧平均检测时间为0.09 s,可以有选择、准确地抓取感兴趣的物体,并且比基准方法更可靠。

1 抓取检测技术概述

本文机械臂抓取方法如图1所示,主要包括3个模块:基于YOLOv8的目标识别与定位模块、基于FC-GQCNN的抓取位姿生成模块。

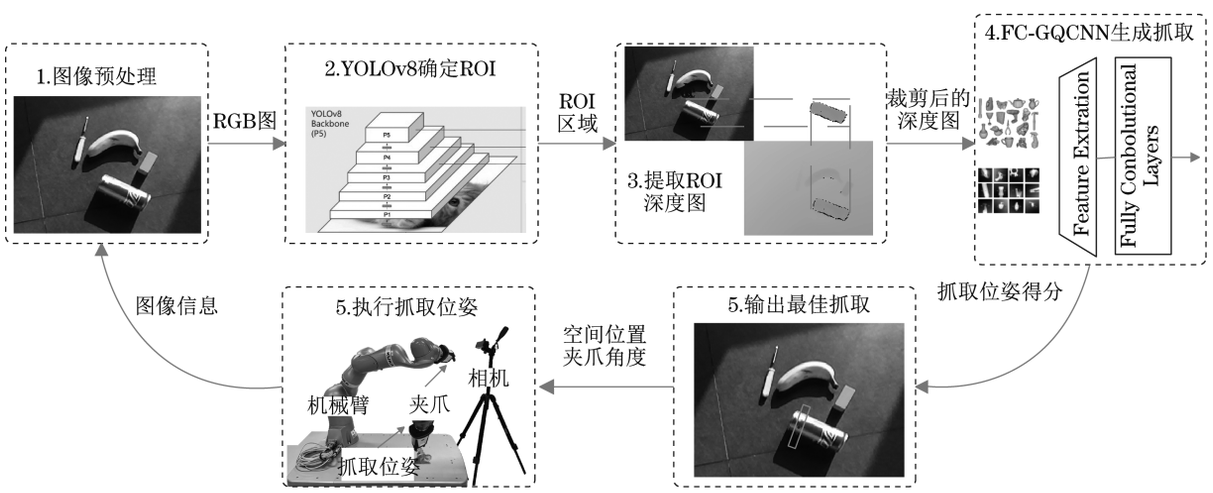

图1 基于目标识别和FC-GQCNN的抓取技术概述

Fig.1 Overview of target recognition and FC-GQCNN based crawling technique

首先,系统通过深度相机订阅场景的RGB-D图像。RGB图像输入至YOLOv8网络进行目标识别与定位,生成待抓取目标的边界框。边界框被映射至深度图像对应区域,提取局部物体深度特征。提取的深度特征输入到FC-GQCNN网络,该网络通过全卷积结构和优化的特征提取机制,实现了对任意尺寸输入的高效处理,输出最优抓取位姿候选。

2 FC-GQCNN结构

2.1 GQCNN结构

抓取质量网络(grasp quality convolutional neural network,GQ-CNN)是一种针对机器人抓取任务设计的深度学习模型,其核心目标是通过对传感器采集的深度图像与抓取姿态参数的联合建模,预测抓取动作的质量评分。该模型由卷积神经网络和全连接网络组成,首先通过卷积层提取深度图

像的几何和边缘特征，然后将这些特征与抓取姿态参数的特征向量在特征融合层中结合，最终通过全连接层输出抓取质量评分。

GQCNN的网络结构如图2所示。GQCNN只能处理32×32的图像，且由于其全连接层限制了感受野大小，采样阶段耗时较多，且多个候选抓取框对应的图像块之间可能存在重叠区域，单独输入判别器网络会造成特征的重复计算，从而加大了计算量。这些原因以及可能的硬件差异导致了GQCNN输入条件严格，计算速度慢，内存使用效率低，因此，需要改进来使其获得更好的处理性能。

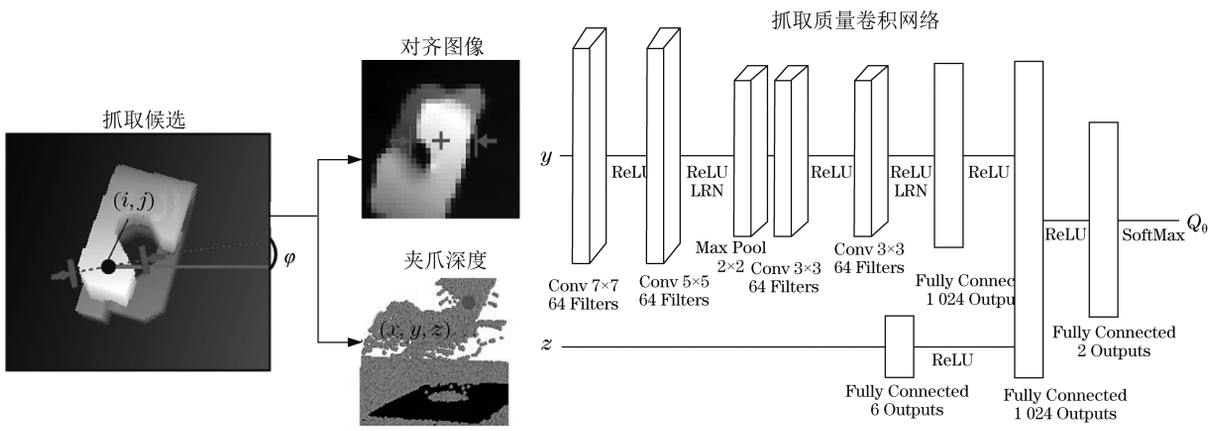

图2　GQCNN网络结构[9]

Fig.2　Diagram of GQCNN network structure[9]

2.2　FC-GQCNN改进点

基于GQCNN改进的FC-GQCNN结构如图3所示。在架构设计上，FC-GQCNN用1×1的卷积层替换全连接层，使网络可以处理任意尺寸的输入图像。

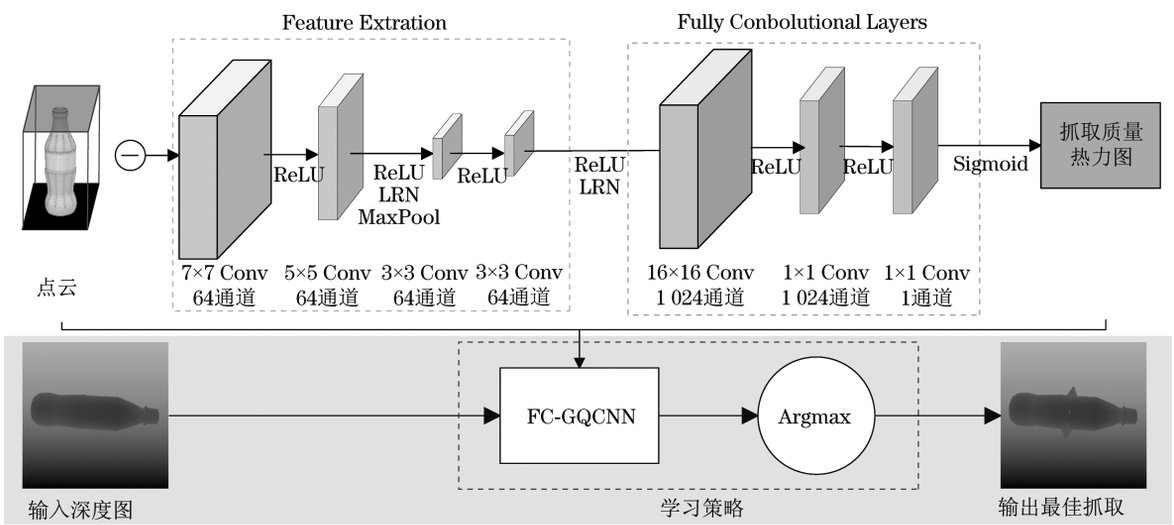

图3　FC-GQCNN网络结构

Fig.3　Diagram of FC-GQCNN network structure

1×1卷积是一种特殊的卷积操作，其卷积核大小为1×1，仅对每个像素位置的通道维度进行线性变换。假设输入特征图大小为$N\times M\times C$，其中，N和M为空间尺寸，C为通道数。1×1卷积输出特征图的大小为

$$O\in\mathbb{R}^{N\times M\times C'} \tag{1}$$

式中：C'为输出通道数。由于1×1卷积不依赖于输入图像的空间维度N和M，其只对通道维度进行操作，因此能够处理任意尺寸的输入图像。

基准GQCNN使用滑动窗口策略评估抓取质量。假设输入图像大小为$H\times W$，每次滑动窗口提取局部区域大小为$P\times P$。需要对每个$P\times P$区域单独进行前向传播，计算复杂度为

$$\mathcal{O}\left(\frac{H\cdot W}{P^2}\cdot C\right) \tag{2}$$

式中：C为单次前向传播的计算复杂度。滑动窗口方法会导致特征重复计算，效率较低。

FC-GQCNN使用全卷积网络直接生成整幅图

像的抓取质量热力图。每个像素位置(i,j)的抓取质量评分$Y(i,j)$由局部感受野的卷积计算得到。其热力图生成公式为

$$Y(i,j)=f[X(i-k:i+k,j-k:j+k] \quad (3)$$

式中:f为卷积操作;$(i-k:i+k,j-k:j+k)$为卷积核的感受野。

通过全卷积方式,一次前向传播可以高效计算整个输入特征图的抓取质量热力图,其复杂度为

$$\mathcal{O}(H\cdot W\cdot K\cdot C) \quad (4)$$

式中:K为卷积核的大小,与式(2)相比复杂度相比显著降低。

FC-GQCNN移除了夹持器与目标物体的相对高度的影响,扩展到4自由度(3D位置和平面方向),比之前的3自由度实现提供了更丰富的抓取策略空间。

3 YOLOv8-FCGQCNN级联流程

FC-GQCNN可以为每个32×32图像块生成一个最佳抓取位置,并在每个位置生成多个抓取角度统合为一个抓取候选,对每个抓取候选提取特征后输出抓取概率(0~1之间的一个分数),最后以全部图像块得分最高的抓取分数作为最终输出。因此,其能抓取任何形状规则或不规则的物体。但不能根据实际环境中的各种要求选择预定义的对象。例如,当一些感兴趣的物体与其他物体混在一组,但都在摄像机的同一视野中时,FC-GQCNN无法准确识别预定义对象。

YOLO是一种单阶段目标识别算法,利用卷积神经网络直接预测目标的位置和类别概率。作为YOLO系列的最新版本,YOLOv8在继承其前代高效实时检测能力的基础上,进一步优化了网络架构[11],能够在复杂背景中高效识别和定位感兴趣的目标,这正是传统FC-GQCNN算法的不足之处。因此,本文提出了一种两步级联结构YOLOv8-FCGQCNN。该方法利用YOLOv8生成的边界框剔除图像中无关区域,并将感兴趣区域裁剪为深度图像,再输入至FC-GQCNN,以确定目标物体的最佳抓取位姿。YOLO-GQCNN的线性级联设计有效结合了YOLOv8的目标识别能力和FC-GQCNN的抓取优化能力,其级联流程如图4所示。

先预存无物体的深度图像作为背景参考;再使用YOLOv8检测RGB图像中的目标物体,生成对应的边界框;然后将边界框的索引和位置信息编码,并映射至背景深度图中的相应区域;替换深度图中边界框对应的区域以获取目标物体的深度信息;最后截取包含目标物体的深度图片段,并调整为符合FC-GQCNN输入尺寸。通过训练好的FC-GQCNN,即可生成抓取质量最高的抓取位姿所对应的机械臂末端执行器位置和角度。

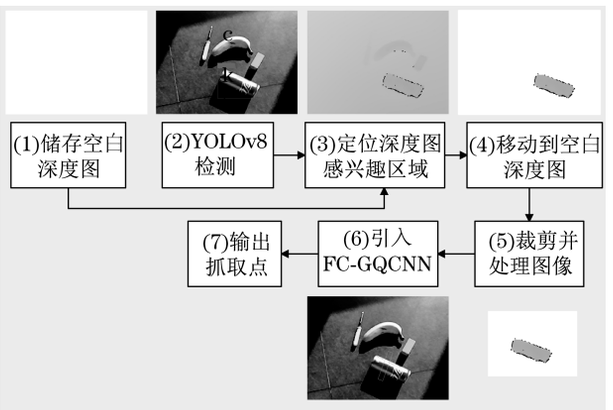

图4 YOLOv8-FCGQCNN级联流程

Fig.4 YOLOv8-FCGQCNN cascade process

4 实验验证

4.1 实验环境

FC-GQCNN的训练基于PyTorch 1.11.0搭建深度学习框架,训练环境基于软硬件环境进行,操作系统为乌班图20.04,Python 3.8,Cuda 11.3,GPU NVIDIA Geforce RTX 4060,CPU Intel Core i9-13900H,内存32 GB。图片输入尺寸为300像素×300像素,batchsize为256,训练轮数为60轮。

为了测试改进后的FC-GQCNN的能力,数据集选择与原始GQCNN相同的Dex-net2.0数据集,这是一个包含高质量三维物体模型的多模态数据集。数据集中的物体模型来源于多个公开的3D模型数据库,并通过仿真生成了大量用于抓取评估的样本,具有较强的较高的跨域泛化能力。其中,训练集与测试集的分配比例为5∶1。

使用了由KUKA KMR 200 iiwa 14 R82七轴机械臂、ZED2i深度摄像头和中控计算机组成的抓取实验平台测试了算法的抓取效果。

4.2 训练结果与分析

图像分类中的准确率是一种常见的评价指标,但是Dex-net2.0数据集中正负样本比例约为1∶5,准确率无法正确反映模型的识别效果。为此,引入精确率P和召回率R作为评价指标,其定义为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}} \quad (5)$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}} \tag{6}$$

式中：N_{TP}为正确预测为正类的样本数；N_{FP}为错误预测为正类的负样本数；N_{FN}为错误预测为负类的正样本数。

通过精确率-召回率曲线（*P-R* Curve），如图5所示，可观察FC-GQCNN和GQCNN不同阈值下对两者的权衡。曲线下的面积越大，说明模型同时具有较高的精确率和召回率。FC-GQCNN既能准确预测，又能有效识别正类样本的能力要好于GQCNN，如图6所示。

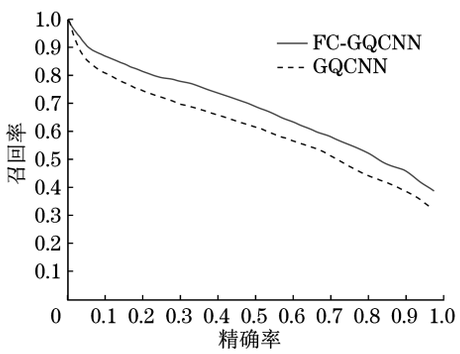

图5　FC-GQCNN和GQCNN的*P-R*曲线

Fig.5　*P-R* curves for FC-GQCNN and GQCNN

此外，模型推理速度是抓取模型在实际应用中的关键性能指标之一，尤其在实时性要求较高的任务中具有重要意义。如图6所示：GQ-CNN可实现每秒13.5次推理；FC-GQ-CNN可实现每秒540次推理，速度提高了22倍，效率有了明显的提升。

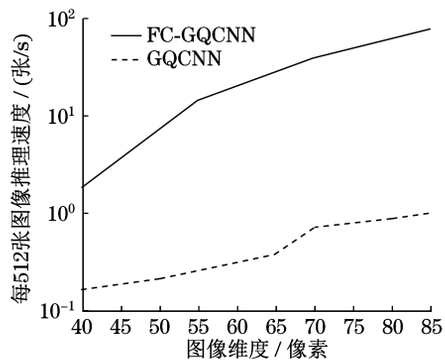

图6　FC-GQCNN和GQCNN的推理速度

Fig.6　Inference speed of FC-GQCNN and GQCNN

4.3　实验结果与分析

机器人在被预设了不同的抓取对象，同时场景中存在其他的干扰物体来执行抓取动作，如图7所示。由图7可知，单独引用FC-QGCNN的情况下只能生成整个深度图中的最佳抓取位置，这个位置是和物体类别无关的。所以预定义的抓取对象无法被准确抓取。

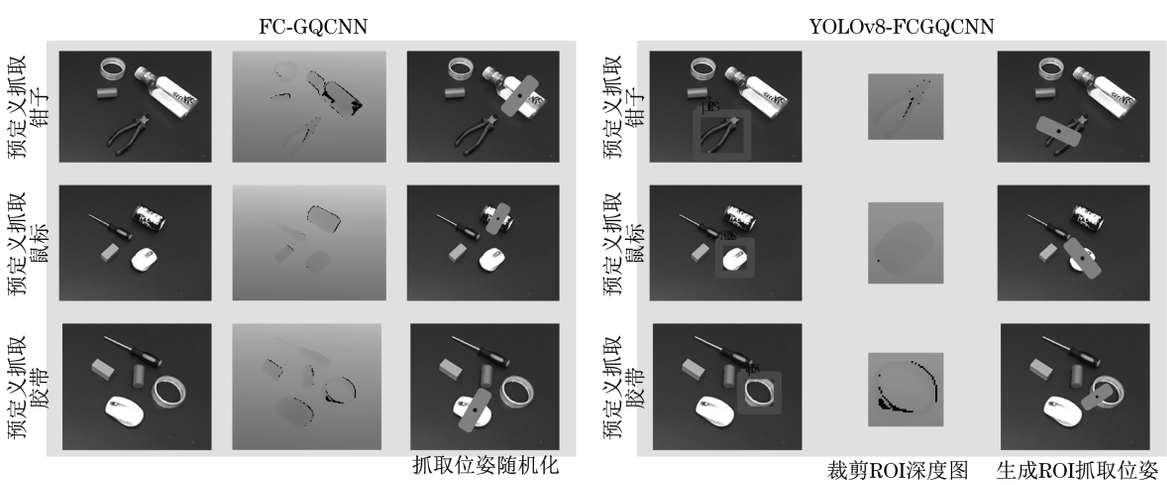

图7　FC-GQCNN与YOLO-GQCNN的检测比较

Fig.7　Comparison of FC-GQCNN and YOLO-GQCNN detection

而在使用YOLOv8-FCGQCN的情况下准确定位到了目标位置抓取区域，这是因为YOLOv8的识别和定位能力有助于FC-GQCNN在目标区域生成抓取点，从而提高机器人在干扰较大的环境中的抓取识别率。

使用不同的算法组合对10类物体分别进行了100次抓取任务，并在每种抓取场景中都放置了1种目标物体和3～4种干扰物体，各类物体的抓取时间和抓取统计结果见表1。其中，FDT为单帧平均检测时间，SR为成功率，是抓取成功的次数占总抓取次数的百分比。结果表明，YOLOv8-FCGQCNN抓取成功率为86%，单帧平均检测时间为0.09 s，在大多数情况下都可以成功预测物体的抓取方案，并且效果好于其他算法组合。

表1 不同算法组合对不同感兴趣对象的抓取结果

Tab.1 Grabbing results of different combinations of algorithms for different objects of interest

物体	抓取次数	YOLOv5-GQCNN		YOLOv8-GQCNN		YOLOv5-FCGQNNN		YOLOv8-FCGQCNN	
		FDT/s	SR/%	FDT/s	SR/%	FDT/s	SR/5	FDT/s	SR/%
可乐罐	15	1.72	93.33	1.37	93.33	0.10	93.33	0.08	93.33
圆柱体	15	1.75	93.33	1.41	86.67	0.12	86.67	0.10	93.33
长方体	15	1.73	93.33	1.38	93.33	0.13	93.33	0.09	93.33
螺丝刀	15	1.76	53.33	1.42	86.67	0.09	86.67	0.11	73.33
香蕉	15	1.71	93.33	1.36	93.33	0.12	93.33	0.07	93.33
胶带卷	15	1.77	53.33	1.43	53.33	0.11	66.67	0.10	86.67
钳子	15	1.74	86.67	1.39	93.33	0.10	88.67	0.09	73.33
饮料瓶	15	1.73	93.33	1.38	93.33	0.12	93.33	0.08	93.33
锤子	15	1.75	86.67	1.40	93.33	0.11	93.33	0.10	86.67
鼠标	15	1.74	93.33	1.37	93.33	0.11	93.33	0.09	93.33
平均值		1.74	84.00	1.39	85.00	0.11	84.00	0.09	86.00

5 结论

针对机械臂抓取系统中存在的计算效率低下、特征重复计算以及复杂环境下目标识别困难等问题，提出了以下改进方案：

（1）设计了基于全卷积网络的改进抓取质量网络（FC-GQCNN），通过替换全连接层为1×1卷积层，实现了对任意尺寸输入图像的处理能力，使系统在保持性能的同时大幅提升了计算效率。

（2）将FC-GQCNN与YOLOv8目标识别算法进行级联，构建了YOLOv8-FCGQCNN结构，有效解决了复杂环境下的目标识别和定位问题。

（3）通过实验验证，该方法在各类物体的抓取任务中表现出优异的性能，抓取成功率达到86%，检测速度相比传统方法明显提升。

研究结果表明：该方法在实际应用中具有较强的实用性和可靠性，为机械臂抓取系统的改进提供了新的思路。

参考文献：

[1] WANG D, KOHLER C, TEN PAS A, et al. Towards assistive robotic pick and place in open world environments[C]// International Symposium of Robotics Research. Cham, Switzerland: Springer, 2019: 360-375.

[2] RAMALEPA L P, JAMISOLA R S. A review on cooperative robotic arms with mobile or drones bases[J]. International Journal of Automation and Computing, 2021, 18(4): 536-555.

[3] DU G, WANG K, LIAN S, et al. Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review[J]. Artificial Intelligence Review, 2021, 54(3): 1677-1734.

[4] CALDERA S, RASSAU A, CHAI D. Review of deep learning methods in robotic grasp detection[J]. Multimodal Technologies and Interaction, 2018, 2(3): 57.

[5] ZHANG J, LI M, FENG Y, et al. Robotic grasp detection based on image processing and random forest[J]. Multimedia Tools and Applications, 2020, 79(3): 2427-2446.

[6] CAO H, CHEN G, LI Z J, et al. Lightweight convolutional neural network with Gaussian-based grasping representation for robotic grasping detection[EB/OL]. (2021-01-25)[2024-10-09]. https://arxiv. org/abs/2101. 10226.

[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] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.

[9] MAHLER J, LIANG J, NIYAZ S, et al. Dex-Net 2. 0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics[EB/OL]. (2017-03-27)[2024-10-09]. https://arxiv. org/abs/1703. 09312.

[10] LENZ I, LEE H, SAXENA A. Deep learning for detecting robotic grasps[J]. The International Journal of Robotics Research, 2015, 34: 705-724.

[11] JOCHER G, QIU J, CHAURASIA A. Ultralytics YOLO[EB/OL]. (2023-01-10)[2024-10-10]. https://github. com/ultralytics/ultralytics.