E-Commerce Letters 电子商务评论, 2024, 13(4), 1887-1893
Published Online November 2024 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2024.1341345

Hans汉斯

基于机器视觉的工业机械臂自动抓取系统在电商物流中的应用研究

李熔鑫*，罗浩东#，李志文#，雷乔之#，曾以春#

贵州大学大数据与信息工程学院，贵州 贵阳

收稿日期：2024年8月6日；录用日期：2024年8月21日；发布日期：2024年11月7日

摘 要

随着电子商务的蓬勃发展，物流自动化特别是高效准确的物品分拣和抓取技术变得至关重要。针对传统机械臂在动态环境中适应性不足的问题，本文提出了一种基于机器视觉的工业机械臂自动抓取系统。该系统利用深度学习技术，通过实时处理图像数据，精确提取物品特征，并计算6D姿态信息，以指导机械臂执行准确的抓取动作。本研究的核心是一个深度神经网络，它结合了深度可分离卷积和U型网络的优势，显著提高了特征提取的效率和准确性，同时减少了模型的参数量，增强了系统的实时性。在Cornell抓取数据集上的实验结果表明，该系统达到了98.79%的准确率，证明了其在电商物流自动化领域的应用潜力。此外，系统还采用了高斯滤波和数据增强技术，进一步提升了模型的稳定性和泛化能力。对比实验显示，与其他现有技术相比，本系统在保持高准确率的同时，模型参数量更少，更适合实时应用。

关键词

工业机械臂，自动抓取系统，电商物流，深度学习

Application Research of Automatic Grasping System of Industrial Robot Arm Based on Machine Vision in E-Commerce Logistics

Rongxin Li*, Haodong Luo#, Zhiwen Li#, Qiaozhi Lei#, Yichun Zeng#

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: Aug. 6th, 2024; accepted: Aug. 21st, 2024; published: Nov. 7th, 2024

*第一作者。
#共同二作。

文章引用: 李熔鑫, 罗浩东, 李志文, 雷乔之, 曾以春. 基于机器视觉的工业机械臂自动抓取系统在电商物流中的应用研究[J]. 电子商务评论, 2024, 13(4): 1887-1893. DOI: 10.12677/ecl.2024.1341345

Abstract

With the booming development of e-commerce, logistics automation, especially efficient and accurate item sorting and grasping technology, has become crucial. Aiming at the problem of the lack of adaptability of traditional robot arm in dynamic environment, an automatic grasping system of industrial robot arm based on machine vision is proposed in this paper. The system uses deep learning technology to accurately extract item features by processing image data in real time, and calculates 6D pose information to guide the robot arm to perform accurate grasping actions. The core of this research is a deep neural network, which combines the advantages of deep separable Convolution and U-Net to significantly improve the efficiency and accuracy of feature extraction, while reducing the number of parameters in the model and enhancing the real-time performance of the system. The experimental results on the Cornell crawl data set show that the system achieves 98.79% accuracy, which proves its application potential in the field of e-commerce logistics automation. In addition, the system also adopts Gaussian filtering and data enhancement technology, which further improves the stability and generalization ability of the model. Comparative experiments show that compared with other prior art, the system has fewer model parameters while maintaining high accuracy, and is more suitable for real-time applications.

Keywords

Industrial Robot Arm, Automatic Grasping System, E-Commerce Logistics, Deep Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

随着互联网的快速发展，机械臂在制造业和物流业等多个领域的应用日益广泛，极大地推动了各行业的自动化进程。传统的机械臂抓取系统通常依赖示教方式，通过预设固定的关节角度来完成搬运任务。这种方法虽然在特定环境下能够实现一定程度的自动化，但在面对动态变化的仓储环境时显得力不从心。电子商务迅猛发展的今天，物流仓储行业对智能化和自动化的需求日益迫切，传统的机械臂抓取系统显然难以满足这一需求。电商物流的核心在于提高物品处理的效率和准确性。现有的机械臂抓取系统普遍存在灵活性不足、环境适应性差的问题，导致整个系统的鲁棒性不高[1] [2]。

机器视觉技术的飞速发展为机械臂的智能化提供了新的解决方案。通过摄像头实时捕获的图像数据，结合机器视觉技术提取物流包裹的形状、纹理、颜色等特征信息。根据这些特征信息可以得出物体在像素坐标系下的坐标信息，再结合位姿转换矩阵就能得到物体在世界坐标系下的位置和姿态。这样的系统能够提高抓取的精度，还可以增强机械臂对环境感知能力[3]。

本文提出了一种实时易于部署的基于机器视觉的自动抓取系统，该系统以被抓取物品的图像为输入，利用深度可分离卷积技术(Depthwise Separable Convolution, DWConv) [4]提取物品的特征信息，并计算其在平面上姿态。这些姿态信息被提供给机械臂的末端执行器，以实现对物品的准确抓取，解决了机械臂在仓储物流的非结构场景下的实时分拣问题。

2. 文献综述

机器视觉系统可以识别和分类不同类型的包裹，从而提高分拣效率和准确性。利用机器视觉技术，

机械臂能够准确定位和抓取物品，并按特定顺序进行装箱和码垛[5] [6]。机械臂在仓储物流的自动分拣任务的核心是机械臂抓取。通过引入机器视觉可以提高电商物流的自动化效率。机械臂的视觉抓取主要分为传统的基于分析的抓取以及目前主流的数据驱动的抓取方式[7]。

基于分析的抓取方式：主要是利用物体的几何信息、物理建模等方式进行分析抓取需要的力，设计专门的末端执行器尽可能地贴合被抓取物体的表面。A. T. Miller [8]等人提出了一种为机器人抓取设计的多功能模拟器，模拟器利用详细的几何和物理模型来评估不同的抓取技术，考虑各种因素，如接触点，摩擦和力分布。这种抓取方式的泛用性低不利于抓取的环境迁移，因为面对非结构化的场景很难找到一种适用于所有物品抓取的末端执行器结构。

基于数据驱动的抓取方式：主要是引入深度学习的办法来提取被抓取物品的特征信息，通过深度学习自动地得到抓取物品的位置信息，这种方式是目前最受欢迎的模式。Duan H [9]等人提出了端到端学习模型，利用点云数据直接输出抓取姿态，省略了中间步骤。该方法提高了机器人在动态环境中的灵活性和适应性，能够处理复杂的抓取任务。端到端模型通常需要大量数据进行训练，且在实际应用中可能面临数据不足的问题。S. Wang [10]等人提出了 TF-Grasp 模型，通过 Transformer 的编码器-解码器架构处理机械臂抓取任务。使用自注意力机制(Self-Attention)聚合图像的全局和局部信息，从而生成上下文相关的特征表示。通过 Swin-Transformer 的局部窗口注意力和跨窗口注意力机制，实现了全局和局部信息的平衡，从而提高了复杂场景中的抓取检测准确性。但 Transformer 的自注意力机制计算复杂度较高，处理大尺寸图像时可能导致计算开销增加。Transformer 模型通常需要大量的数据进行训练，以充分发挥其优势。对于数据量有限的应用场景，模型的泛化能力可能受限。

将视觉技术应用于机械臂的抓取，不仅能够提升电商物流的自动化水平，还推动了机器人通用性研究的发展。深度学习在视觉领域的成功为机器人感知研究提供了新的思路。相比于传统仓储物流，引入视觉抓取方式可以为物流系统带来更高的容错率。

3. 方法

3.1. 抓取定义

在现代仓储环境中，机械臂执行的挑拣抓取任务通常被视为一种二维平面内的作业。这种任务的自动化要求机械臂能够精确识别并抓取目标物品，而基于数据驱动的抓取技术在此过程中扮演着至关重要的角色。核心步骤之一是将摄像头采集的 RGB 图像数据实时传输给一个预先训练好的神经网络。

该神经网络经过精心设计和训练，旨在从图像中提取关键信息，包括被抓取物品的像素坐标、足式夹爪的对应打开宽度，以及末端执行器绕 Z 轴的旋转角度。这些信息通常以一种标准化的形式表示，被称为五元组抓取表征法[11]：

$$\{x, y, w, h, \theta\} \tag{1}$$

公式(1)中，G 代表抓取框的表征。其中，使用(x, y)表示物体在像素坐标系中的位置，后文中也用 Q 来表示。w 和 h 分别表示足式夹爪需要打开的宽度和允许的最小长度，这两个参数对于确保抓取的稳定性和准确性至关重要。θ 表示末端执行器相对于物体需要绕 Z 轴旋转的角度，后文中用 A 来表示，它对于机械臂正确对准和抓取目标物体起着决定性作用。

通过这种五元组的表征方法，神经网络能够为机械臂提供精确的抓取指令，从而实现高效、准确的自动化抓取。此外，这种方法的标准化和模块化特点也便于在不同的仓储环境和应用场景中进行扩展和定制。

3.2. 网络结构

为了进一步提升抓取任务的性能和鲁棒性，我们利用深度学习进行更复杂的场景理解和抓取策略优

化。本文提出的网络架构如图 1 所示，其设计由四种核心模块构成，共同协作以实现高效的特征处理和抓取决策。

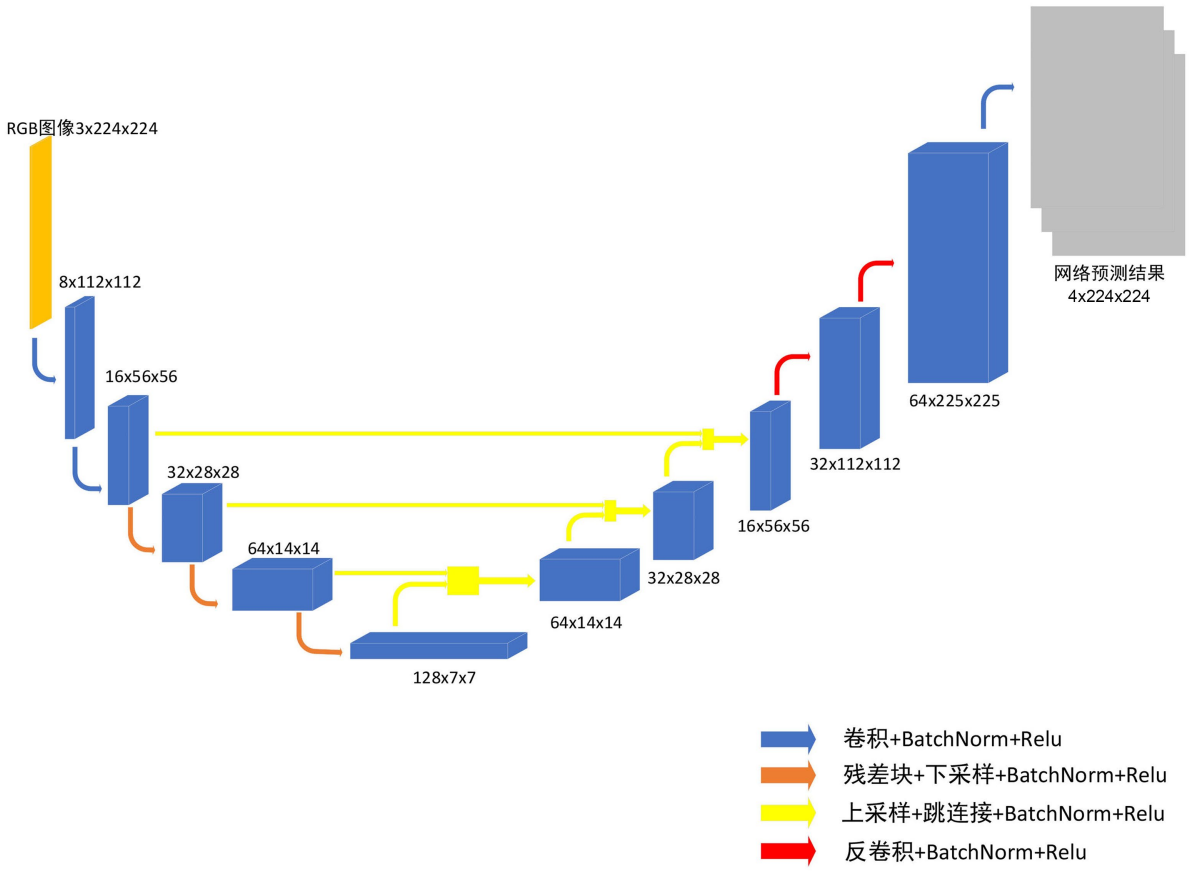

Figure 1. Model structure
图 1. 模型结构

网络的初始阶段采用两个卷积层进行特征提取。编码器部分由三个残差块(Residual Networks, ResNet)和一个 1 × 1 的卷积层组成，这些残差块通过 DWConv 替代了传统的卷积操作。这种设计选择显著减少了模型的参数数量，同时保持了特征表达的能力，从而提高了网络的计算效率和实时性。

在网络的解码阶段，通过三个双线性插值上采样模块(UpsamplingBilinear2d)逐步恢复特征图的空间分辨率。这些模块利用了 U 型网络(U-Net) [12]架构中的跳跃连接来实现特征的融合，确保了解码过程中信息的完整性和丰富性。最后，通过两个转置卷积模块(Transposed Convolution Modules, TCM)来细化特征并生成最终的抓取结果。

整个网络的设计不仅考虑了特征的高效提取和处理，还特别强调了模型在实际应用中的实时性和泛化能力。通过在各种模拟环境中的测试，验证了模型对于不同形状、大小和颜色物品的识别和抓取决策的准确性和可靠性。

ResNet 可以有效地进行深层次特征提取。残差块通过短路连接解决了深层网络中的梯度消失问题，从而使得训练更深的神经网络成为可能。在抓取任务中，ResNet 能够提取图像中的高层次特征，帮助识别并定位目标物体。U-Net 是一种典型的卷积神经网络架构，主要用于图像分割任务。它通过编码器－解码器结构对图像进行逐步下采样和上采样，从而精确地预测像素级的物体边界。由于机械臂抓取的定义

和图像分割类似，因此本文借鉴 U-Net 的结构思想。在抓取任务中，U-Net 可以用于精确地分割出目标物体的轮廓，为抓取位置的确定提供高精度的信息。DWConv 将标准卷积分解为深度卷积和逐点卷积两个步骤，大大减少了计算量和参数量。在抓取任务中，DWConv 能够提高模型的计算效率，使得网络在实时应用中更加高效。

通过结合这些深度学习技术，可以构建一个高效且准确的抓取神经网络。该网络可以处理输入的 RGB 图像数据，输出物体的像素坐标、夹爪开口宽度以及末端旋转角度，从而实现精确的抓取操作。

4. 实验

本节深入探讨了本文提出的模型在性能方面的评估，并展示了其在实际环境中的推理结果。我们的模型在广泛认可的 Cornell 数据集上进行了训练，该数据集由 240 种不同对象的 885 张图像组成，分辨率为 640 × 480 像素。数据集中包含 5110 次正样本抓取和 2909 次负样本抓取，是抓握检测领域内一个公认的训练集。在训练过程中，我们仅使用了正样本抓取数据，以确保模型专注于学习有效的抓取模式。

为了评估模型性能，我们采用了标准的 IoU (交并比)阈值大于等于 25%，并结合抓取角度 A 与标签的误差小于等于 30˚作为合格推理的标准。此外，为了增强模型对不同环境条件的适应性，我们对数据集进行了包括旋转、随机裁剪和缩放在内的一系列增强操作，以模拟实际应用中可能遇到的多样化情况。

在训练过程中，我们采用了 Adam 优化器，进行了 50 个 epoch 的训练，batch size 设置为 8，同时引入了 dropout 机制，概率为 0.05，以减少过拟合的风险。为了进一步提升模型推理结果的稳定性和可靠性，我们在测试阶段对预测结果应用了高斯滤波处理。

模型的训练和测试均在高性能的硬件平台上进行，包括 NVIDIA RTX 4070S GPU、Intel Core i5-12600KF CPU，以及 Ubuntu 18.04 操作系统的 ROS (Robot Operating System) melodic。这一平台的配置确保了模型训练和推理过程的高效性和稳定性。

通过这些综合的实验设置和评估方法，我们证明了所提出模型在抓握检测任务上的有效性和鲁棒性，为实际应用中的自动化抓取提供了坚实的技术基础。

4.1. 对比实验

此部分，我们对比了近几年提出的模型和本文提出模型的性能对比，分别是 GG-CNN [13]、TF-Grasp。如表 1 网络性能对比表所示，我们把上述的两种模型放到了和我们相同的硬件和软件环境中，在同一个数据集 Cornell 数据集中做了准确率和模型参数量的对比。我们仅将 RGB 图像作为网络的输入，可以看到本文提出的模型在准确率和模型参数量上均取得了不错的效果。

Table 1. Network performance comparison table
表 1. 网络性能对比表

模型	参数量(大约)	准确率
GG-CNN [13]	62 k	74%
TF-Grasp [10]	800 k	97.99%
本文模型	220 k	98.79%

本文提出的模型融合了 ResNet 的优势，该网络通过引入残差块连接来有效提取图像特征，同时跳连接解决了深层网络中的梯度消失问题。此外，残差网络中的卷积使用了 DWConv 进行替换，与传统卷积相比，它具有更少的参数，从而降低了模型的复杂度。结合 U-Net 的结构，该模型通过跳跃连接将编码器捕获的上下文信息与解码器相结合，有效传递了位置信息，使得解码器能够保留图像的颜色、形状等

细节特征。这三者的有效结合保留了彼此的优势，因此比 GG-CNN 这种基于传统的全卷积网络在准确率上高很多而又比基于 Transformer 的 TF-Grasp 网络的参数量小。

4.2. 可视化展示

本小节主要展示本模型的推理性能。我们分别提供了单件物品的推理结果图、相应的热图，以及多件物品的推理效果和热图。这些物品图像均源自现实生活场景，且在模型训练阶段未被包含，以检验模型的泛化能力和鲁棒性。图 2 模型推理效果图，其中详细标注了抓取质量分数(Q)和所需抓取角度(Angle)，并在 RGB 图像上进行了相应的标注，结果也证明本文提出的模型在未见物品上的推理能力和鲁棒性。

此外，为了验证模型的实时性，我们将模型与机器人操作系统(Robot Operating System, ROS)集成，并在 Gazebo 仿真环境中进行了机械臂仿真。在仿真中，我们设计了不同尺寸和颜色的方块和球体，以模拟实际仓储环境中可能遇到的各种包裹。图 3 所示，仿真结果表明，在保持实时性的同时，本模型对于不同形状和颜色的物品也展现出了高度的识别准确性。

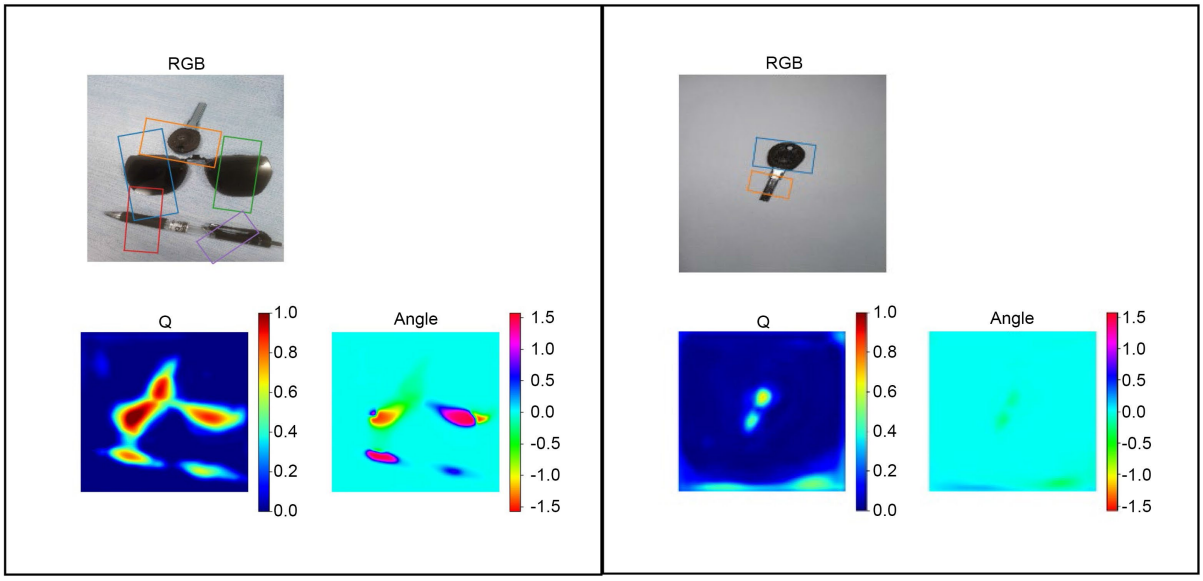

Figure 2. Model inference renderings
图 2. 模型推理效果图

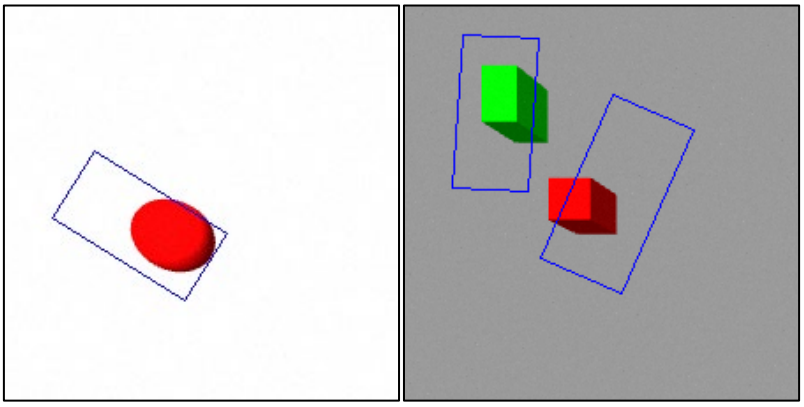

Figure 3. Diagram of simulation results
图 3. 仿真结果图

5. 总结

本文研究了一种基于机器视觉的工业机械臂自动抓取系统，并将其应用于电商物流领域。随着电子商务的快速发展，物流自动化的需求日益增长，而传统的机械臂抓取系统在动态和非结构化环境中的适应性和灵活性不足。为了提高系统的鲁棒性，本文提出了一种结合深度学习技术的自动抓取系统，该系统能够实时处理图像数据，准确提取物品特征，并计算其 6D 姿态信息，指导机械臂完成精确抓取。

系统的核心是一个深度神经网络，它利用 DWConv 和 U-Net 结构，通过残差块提高特征提取能力，并减少参数量以提高实时性。在 Cornell 抓取数据集上，该系统达到了 98.79%的准确率，展现了出色的性能。此外，系统还采用了高斯滤波和数据增强技术，进一步提高了模型的稳定性和泛化能力。实验结果表明，该系统不仅在准确率上优于现有技术，而且在参数量上更为经济，满足了电商物流领域对实时性和自动化的高要求。

参考文献

[1] 赵源麒, 王清珍. 基于视觉协同的双机械臂物品动态抓取系统设计与实现[J]. 仪器与设备, 2024, 12(2): 107-116.

[2] 杜学丹, 蔡莹皓, 鲁涛, 王硕, 闫哲. 一种基于深度学习的机械臂抓取方法[J]. 机器人, 2017, 39(6): 820-828, 837.

[3] Du, G., Wang, K., Lian, S. and Zhao, K.Y. (2019) Vision-Based Robotic Grasping from Object Localization, Pose Estimation, Grasp Detection to Motion Planning: A Review. arXiv: 1905.06658.

[4] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/cvpr.2017.195

[5] Rey, R., Corzetto, M., Cobano, J.A., Merino, L. and Caballero, F. (2019) Human-Robot Co-Working System for Warehouse Automation. 2019 24*th IEEE International Conference on Emerging Technologies and Factory Automation* (*ETFA*), Zaragoza, 10-13 September 2019, 278-585. https://doi.org/10.1109/etfa.2019.8869178

[6] Desaraju, V.R. and How, J.P. (2011) Decentralized Path Planning for Multi-Agent Teams in Complex Environments Using Rapidly-Exploring Random Trees. 2011 *IEEE International Conference on Robotics and Automation*, Shanghai, 9-13 May 2011, 4956-4961. https://doi.org/10.1109/icra.2011.5980392

[7] Morrison, D., Corke, P. and Leitner, J. (2019) Learning Robust, Real-Time, Reactive Robotic Grasping. *The International Journal of Robotics Research*, **39**, 183-201. https://doi.org/10.1177/0278364919859066

[8] Miller, A.T. and Allen, P.K. (2004) Graspit! A Versatile Simulator for Robotic Grasping. *IEEE Robotics & Automation Magazine*, **11**, 110-122. https://doi.org/10.1109/mra.2004.1371616

[9] Duan, H., Wang, P., Huang, Y., Xu, G., Wei, W. and Shen, X. (2021) Robotics Dexterous Grasping: The Methods Based on Point Cloud and Deep Learning. *Frontiers in Neurorobotics*, **15**, Article 658280. https://doi.org/10.3389/fnbot.2021.658280

[10] Wang, S., Zhou, Z. and Kan, Z. (2022) When Transformer Meets Robotic Grasping: Exploits Context for Efficient Grasp Detection. *IEEE Robotics and Automation Letters*, **7**, 8170-8177. https://doi.org/10.1109/lra.2022.3187261

[11] Jiang, Y., Moseson, S. and Saxena, A. (2011) Efficient Grasping from RGBD Images: Learning Using a New Rectangle Representation. 2011 *IEEE International Conference on Robotics and Automation*, Shanghai, 9-13 May 2011, 3304-3311. https://doi.org/10.1109/icra.2011.5980145

[12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv: 1505.04597. https://api.semanticscholar.org/CorpusID:3719281

[13] Morrison, D., Corke, P. and Leitner, J. (2019) Learning Robust, Real-Time, Reactive Robotic Grasping. *The International Journal of Robotics Research*, **39**, 183-201. https://doi.org/10.1177/0278364919859066