

# 基于“形态-感知-动作”仿生机理的机器人自适应力控抓取方法

赵洲<sup>1</sup> 耿明强<sup>1</sup> 何秋实<sup>2</sup> 何赞鑫<sup>1</sup> 蔡明达<sup>3,4</sup> 周翔宇<sup>3,4</sup> 罗晶<sup>3,4</sup>

**摘要** 随着机器人技术快速发展,其对精细感知能力需求日益增长.然而,现有机器人仍难以具备如人类般灵活的操作能力.在精细抓取任务中,机器人恒力抓取策略存在局限性:抓取力过大易损伤物体,抓取力过小则导致抓取不稳.为应对上述问题,提出一种基于视觉与触觉融合的机器人自适应力控抓取方法.该方法由视觉模块、触觉模块和抓取策略组成:视觉模块用于预测目标抓取位置;在接触阶段,触觉模块借助视触觉传感器恢复触觉深度并估算接触面积与法向力;随后,通过最大深度变化率和帧间均方差进行形变判定,并触发抓取力调整策略,从而实现“渐进增力-形变检测-力回退”的仿生反馈抓取机制.实验结果表明,该方法将多种日常物体的整体抓取成功率由 87.50% 提升至 98.75%,在易碎物体抓取中实现零损坏.

**关键词** 机器人抓取;视触觉传感器;多模态融合;抓取力

**引用格式** 赵洲,耿明强,何秋实,何赞鑫,蔡明达,周翔宇,罗晶.基于“形态-感知-动作”仿生机理的机器人自适应力控抓取方法.自动化学报,2026,52(5):1116-1127

**DOI** 10.16383/j.aas.c250453 **CSTR** 32138.14.j.aas.c250453

## Robot Adaptive Force Control Grasping Method Based on Bionic Mechanism of “Shape-Perception-Action”

ZHAO Zhou<sup>1</sup> GENG Ming-Qiang<sup>1</sup> HE Qiu-Shi<sup>2</sup> HE Yun-Xin<sup>1</sup>  
CAI Ming-Da<sup>3,4</sup> ZHOU Xiang-Yu<sup>3,4</sup> LUO Jing<sup>3,4</sup>

**Abstract** With the rapid development of robot technology, its demand for fine sensing ability is increasing. However, it is still difficult for existing robots to have the flexible operation ability as human beings. In the fine grasping task, the robot's constant force grasping strategy has limitations: Too much grasping force is easy to damage the object, and too little grasping force leads to unstable grasping. In order to solve the above problems, this paper proposes a robot adaptive force control grasping method based on the fusion of vision and touch. The method consists of a visual module, a tactile module and a grasping strategy: The vision module is used for predicting the grab position of the target; In the contact stage, the tactile module recovers the tactile depth with the help of the visual tactile sensor and estimates the contact area and normal force; Then, the deformation is judged by the maximum depth change rate and the mean square error between frames, and the grasping force adjustment strategy is triggered. Thus, the bionic feedback grasping mechanism of “gradual force increase-deformation detection-force retreat” is realized. The experimental results show that this method improves the overall success rate of grasping various daily objects from 87.50% to 98.75%, and achieves zero damage in grasping fragile objects.

**Keywords** robot grasping; visual tactile sensor; multimodal fusion; grasping force

**Citation** Zhao Zhou, Geng Ming-Qiang, He Qiu-Shi, He Yun-Xin, Cai Ming-Da, Zhou Xiang-Yu, Luo Jing. Robot adaptive force control grasping method based on bionic mechanism of “shape-perception-action”. *Acta Automatica Sinica*, 2026, 52(5): 1116-1127

收稿日期 2025-09-04 录用日期 2025-12-24

Manuscript received September 4, 2025; accepted December 24, 2025

国家自然科学基金(62203341), 鹭江创新实验室自主部署科技项目(25FV0CZZ03), 湖北省自然科学基金(2024AFB245, 2024AFB614), 中央高校基本科研业务费专项资金(CCNU25ai023)资助

Supported by National Natural Science Foundation of China (62203341), Fujian Ocean Innovation Center (25FV0CZZ03), Hubei Provincial Natural Science Foundation (2024AFB245, 2024AFB614), and Fundamental Research Funds for the Central Universities (CCNU25ai023)

本文责任编辑 刘志杰

Recommended by Associate Editor LIU Zhi-Jie

1. 华中师范大学计算机学院 武汉 430079 2. 中国人民武装警察部队警官学院 成都 610213 3. 鹭江创新实验室 厦门 361102

生物在自然演化过程中逐渐形成了“形态-感知-动作”协同机制,以适应环境变化和完成多样化任务<sup>[1]</sup>.这种机制在人类的日常动作学习中体现得尤为明显.例如,婴幼儿在成长过程中通过不断尝试与探索逐步掌握抓取技巧:先依靠视觉获取目标的大致位置和形态信息,然后在手部与物体接触的

4. 武汉理工大学自动化学院 武汉 430070

1. School of Computer Science, Central China Normal University, Wuhan 430079 2. Officers College of PAP, Chengdu 610213 3. Fujian Ocean Innovation Center, Xiamen 361102 4. School of Automation, Wuhan University of Technology, Wuhan 430070

过程中, 通过皮肤和深层组织中的机械感受器感知法向力、剪切力和形变<sup>[2]</sup>. 这些触觉信息经神经系统迅速反馈至大脑和运动神经元, 实现抓取力与手部姿态的细微调整, 从而在避免对柔软或易碎物体施加过大压力的同时, 确保抓取动作的稳定和可靠.

近年来, 机器人研究在感知模式、运动控制和智能决策逻辑上重现了这种自然界的优势机制<sup>[3-4]</sup>. 随着传感器和人工智能技术的快速发展, 在机器人抓取中使用视觉传感器(如 RGB 相机、深度相机等)能够实现远程的目标识别、姿态估计与环境建模<sup>[5-6]</sup>; 利用触觉传感器(如电阻式、压电式以及视触觉(见图 1)等)则可在近距离物理交互中, 直接获取物体表面形态、接触力分布等信息<sup>[7-10]</sup>.

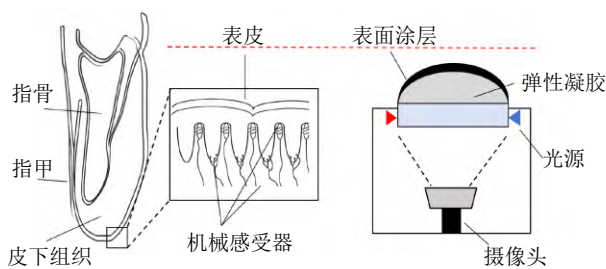


图 1 人类手指与视触觉传感器

Fig.1 Human finger and visual tactile sensor

在目标抓取检测方面, 早期方法多依赖几何建模与传统图像处理技术<sup>[11]</sup>, 通过边缘、角点和轮廓特征推算目标的候选抓取位姿. 然而这类方法泛化能力不足, 难以应对复杂环境. 伴随着深度学习的发展, 基于数据驱动的抓取检测框架得到广泛应用, 如 GGCNN<sup>[12]</sup>、TF-Grasp<sup>[13]</sup>、GraspNet<sup>[14]</sup> 等. 通过神经网络直接从图像中预测目标抓取位置、角度和宽度, 这些方法在未见过的物体上仍表现出较高的成功率. 在触觉感知方面, 传统触觉传感器(如力矩传感器、压电传感器)虽然在某些应用中具有较高的空间分辨率, 但在复杂物体的抓取中, 尤其是需要同时捕捉物体形状与受力分布时, 仍可能面临一定的局限性.

视触觉传感器(如 GelSight<sup>[9]</sup>、DIGIT<sup>[10]</sup>)结合柔性凝胶与高分辨率相机, 通过光度立体技术重建接触面形变, 可获得亚毫米级的深度与力学分布信息. 这为机器人抓取提供更丰富的感知信息.

然而, 单一模态在复杂环境中仍存在显著局限性: 单视觉方法易受光照、遮挡、纹理缺失等影响, 对透明、反光或柔性物体的抓取效果较差; 单触觉方法缺乏全局空间信息, 无法独立完成目标定位与路径规划, 并且需要机械臂先行接触才能获取信息, 限制了反应速度<sup>[15-16]</sup>. 为此, 视觉与触觉的融合逐渐

成为机器人抓取领域的重要发展方向. 视觉在远程阶段完成抓取位置与姿态的粗略规划, 触觉在近距离交互阶段实现对抓取力和接触状态的精确闭环调节, 这种多模态融合方法不仅提升了抓取的成功率, 也增强了机器人对非结构化、动态环境的适应能力<sup>[17-18]</sup>.

为克服单一感知模态在复杂抓取场景下的局限, 研究者提出多种视觉-触觉融合策略, 主要可按融合层次分为信息级(特征级)融合<sup>[19-20]</sup>、决策级融合<sup>[21-22]</sup>. Calandra 等<sup>[20]</sup>提出的端到端动作条件模型使用高分辨率 GelSight 触觉图像与视觉图像联合训练卷积网络, 尝试将视觉与触觉的原始或中间特征映射到同一表示空间, 学习预测候选微调动作, 并在迭代选择策略中执行重抓或微调动作, 从而提高在接触后调整抓取的效率与安全性, 但是该方法需要大量交互数据并面临训练时的样本效率问题. Li 等<sup>[21]</sup>关于透明物体抓取的工作, 采取先用视觉方法在复杂背景中定位抓取点, 再在接触阶段用触觉进行微调 and 类别校验的策略, 从而解决了透明或反光目标在视觉感知中鲁棒性不足的问题. 在更复杂或接触丰富的任务(例如装配、插销对位)中, 研究者将视觉与触觉作为策略学习的观测输入, 采用深度强化学习的方法学习感知到动作的映射<sup>[22]</sup>, 该方法在不确定环境下能显著提升装配鲁棒性与收敛性. Lee 等<sup>[23]</sup>融合视觉-触觉并加入交叉模态注意力机制, 在强化学习框架下实现对柔性物体的高效抓取, 尤其对遮挡或反光物体具有强泛化能力.

然而, 现有视觉-触觉融合抓取研究仍存在显著不足: 一方面, 多数方法采用固定抓取力或简单阈值策略, 很少对接触形变与受力演化进行显式建模和在线细粒度调节, 这导致抓取力偏大时易损伤薄壁、可压缩或脆性目标, 偏小时则难以在扰动下保持稳固握持, 尤其在真实场景中难以平衡“稳定抓取”与“避免损伤”的双目标; 另一方面, 现有视觉-触觉融合研究的核心目标多集中于抓取位姿的优化与预测, 触觉信息通常仅用于辅助视觉定位、位姿微调<sup>[15, 24-26]</sup>、对象类别校验<sup>[21]</sup>或融入强化学习提升复杂环境适应性<sup>[23, 27]</sup>, 却未对抓取力的动态闭环控制给予足够重视. 此外, 深度融合视觉与触觉特征的方法还常面临多模态数据对齐困难、模型训练复杂且依赖大量标注数据等挑战, 进一步限制了其在未知物体抓取场景中的适用性.

针对上述问题, 本文提出一种基于“形态-感知-动作”仿生机理的机器人自适应力控抓取方法(如图 2), 核心创新在于构建解耦的“视觉位姿预测-触觉力控闭环”架构. 与现有方法将触觉用于辅助位

姿规划或依赖固定力/简单规则调整不同, 本文方法明确划分视觉与触觉的职能边界: 视觉模块借鉴人类“全局感知”模式, 独立完成抓取位姿的预测; 触觉模块则聚焦“局部精控”, 在接触阶段独立运行“渐进增力-形变检测-力回退”的闭环策略, 通过视觉触觉传感器实时恢复形变深度、估算接触面积与法向力, 基于最大深度变化率与帧间均方差实现形变判定, 进而动态调节抓取力. 这种架构既规避了复杂跨模态融合的弊端, 降低了对数据量和计算资源的依赖, 又实现基于触觉反馈的抓取力精确控制, 能实时响应物体力学特性, 在多种物体及复杂环境下均能实现稳定、安全地抓取, 更本质地模拟人类抓取的仿生机理.

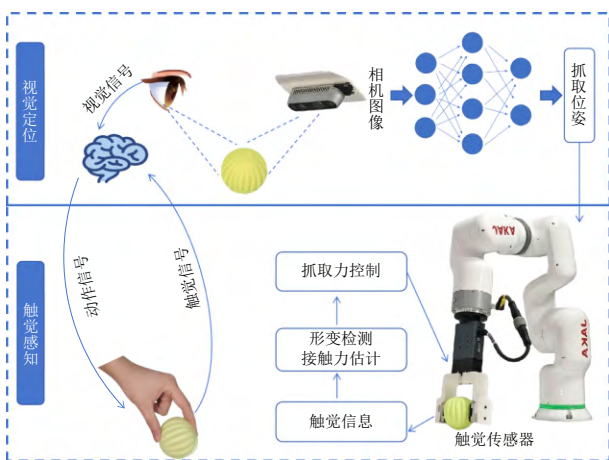


图 2 基于“形态-感知-动作”仿生机理的机器人自适应力控抓取方法

Fig.2 Robot adaptive force control grasping method based on bionic mechanism of “shape-perception-action”

本文的主要贡献如下:

1) 为解决远程感知的问题, 本文提出一种基于 DINOv3 的目标抓取检测模型, 在多个抓取检测数据集上取得良好的检测效果, 实现物体抓取位姿的准确检测.

2) 为解决近端感知的问题, 本文基于对视触觉传感器 (GelSight Mini) 的形变深度恢复, 构建触觉接触面积、形变检测和抓取力估计的计算方法, 为接触信息分析提供数据来源.

3) 为解决灵活稳定的物体抓取问题, 本文提出基于触觉反馈的抓取力调整策略. 相比传统的固定力抓取方式, 该方法在抓取过程中动态感知物体的接触状态与力学响应并实时调整抓取力, 系统在保持抓取稳定性的同时, 有效避免了过大抓取力造成的物体损伤, 体现出良好的通用性与鲁棒性.

## 1 方法

### 1.1 视觉模块

视觉模块对应于生物的远程感知功能, 负责在接触物体之前完成目标位置与抓取姿态的规划. 本方法主要目标是确定抓取中心位置  $(x, y)$ 、夹爪旋转角度  $\theta$  和张开宽度  $w$ .

#### 1.1.1 抓取位姿参数化

一般俯视平面抓取位姿用四元组表示为:

$$\mathbf{g} = (x, y, \theta, w) \quad (1)$$

其中,  $x, y$  为抓取中心在图像坐标系中的像素位置;  $\theta$  为平行夹爪垂直平面的旋转角度 (考虑对称性, 取值范围  $[-\pi/2, \pi/2]$ );  $w$  表示夹爪张开宽度, 归一化至  $[0, 1]$  表示在夹爪最大开度中的比例, 如图 3 所示.

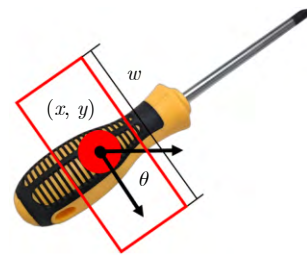


图 3 平面抓取位姿表示

Fig.3 Plane grasping pose representation

#### 1.1.2 网络结构与预测过程

为实现抓取位姿的预测, 本文构建一个以 DINOv3<sup>[28]</sup> 模型为编码器的深度神经网络 (具体型号为 ViT-B/16), 然后经过四层反卷积层进行特征上采样逐步恢复特征图分辨率, 最终通过一层卷积层输出逐像素预测结果, 网络结构如图 4 所示. 该网络输入视觉传感器获取的 RGB 图像并输出四个与输入图像相同尺寸的预测图: 抓取置信度图  $Q \in [0, 1]^{m \times n}$ , 其中  $m$  为图像高度,  $n$  为图像宽度; 抓取二倍角正弦图  $\sin 2\theta$  与二倍角余弦图  $\cos 2\theta$ , 保证角度预测的有界性与连续性; 抓取宽度图  $W \in [0, 1]^{m \times n}$ , 其值与最大夹爪宽度归一化. 最后我们通过式 (2) ~ (4) 从神经网络输出的结果中构建抓取位姿四元组.

抓取点选择公式:

$$(x^*, y^*) = \arg \max_{(x, y)} Q(x, y) \quad (2)$$

角度恢复公式:

$$\theta^* = \frac{1}{2} \text{atan2}(\sin 2\theta, \cos 2\theta) \quad (3)$$

宽度计算公式:

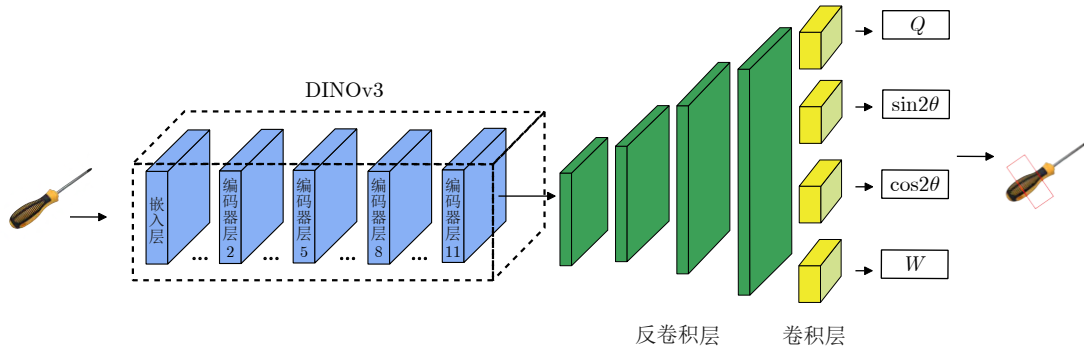


图4 视觉模块神经网络结构

Fig.4 Neural network structure of visual module

$$w^* = W(x^*, y^*) \quad (4)$$

## 1.2 触觉模块

触觉模块对应于生物的近端感知功能, 通过 GelSight Mini 传感器实现对接触区域的形变与力感知. GelSight Mini 输出为 RGB 图像, 不包含直接的深度信息, 因此需经过触觉深度恢复、高度标定、接触面积估计的处理获取有效的触觉信息.

### 1.2.1 深度恢复

首先, 利用法向估计网络<sup>[9]</sup>对 GelSight Mini 输出的接触面 RGB 图像  $I(x, y)$  进行处理, 得到接触表面的法向量  $\mathbf{n}(x, y) = (n_x(x, y), n_y(x, y), n_z(x, y))$ . 根据几何约束, 表面高度函数的梯度可由法向量表示为:

$$\nabla h(x, y) = \left( -\frac{n_x(x, y)}{n_z(x, y)}, -\frac{n_y(x, y)}{n_z(x, y)} \right) \quad (5)$$

在此基础上, 通过泊松方程对梯度场进行积分, 得到相对高度分布  $H_{\text{rel}}(x, y)$ :

$$\nabla^2 H_{\text{rel}}(x, y) = \frac{\partial}{\partial x} \left( -\frac{n_x}{n_z} \right) + \frac{\partial}{\partial y} \left( -\frac{n_y}{n_z} \right) \quad (6)$$

最后, 结合实验标定得到的比例系数  $\lambda$ , 将相对高度转化为真实形变深度:

$$H_{\text{real}}(x, y) = \lambda \times H_{\text{rel}}(x, y) \quad (7)$$

其中, 比例系数  $\lambda$  由已知高度的标准物体压痕实验确定.

### 1.2.2 接触面积求解

接触面积  $A$  定义为深度值超过阈值  $\delta$  的像素区域面积:

$$A = \sum_{H_{\text{real}}(x, y) > \delta} A_{\text{pixel}} \quad (8)$$

其中,  $A_{\text{pixel}}$  为单像素对应的实际面积.

### 1.2.3 形变检测

在抓取接触过程中, 发生形变的物体产生的瞬时最大触觉深度值  $d_{\text{max}}(t)$  随时间先增大后减小. 当变化率为负且接触区域与上一帧的深度分布均方差 (mean squared error, MSE) 超过阈值  $\eta$  时, 判定发生形变.

$$\frac{d}{dt} d_{\text{max}}(t) < 0 \wedge \text{MSE}(t, t-1) > \eta \quad (9)$$

其中

$$\text{MSE}(t, t-1) = \frac{1}{N} \sum_{i=1}^N (H_{\text{real}}^t(i) - H_{\text{real}}^{t-1}(i))^2 \quad (10)$$

### 1.2.4 抓取力估计

GelSight Mini 传感器表面凝胶可视各向同性弹性体, 因此抓取时的法向力  $F_{\text{est}}$  可由胡克定律近似估算.

$$F_{\text{est}} = E \times \frac{\Delta d}{d_0} \times A \quad (11)$$

其中,  $E$  为凝胶的杨氏模量;  $\Delta d$  为形变量;  $d_0$  为初始厚度.

## 1.3 抓取策略

基于获取的形变估计和抓取力的触觉信息, 本文在抓取阶段设计基于触觉反馈的自适应力控抓取策略. 本节介绍策略的状态定义、实时估计与判定量以及抓取力调整方法.

### 1.3.1 状态定义

本文将抓取过程抽象为如表 1 的状态. 状态转移基于触觉形变检测: 如果在  $FORCE\_BUILDUP$  中满足形变判定条件 (见式 (9)), 转入  $ADJUST$ ; 若在  $ADJUST$  后力稳定且无进一步形变, 转入  $HOLD$ ; 若调整无效或检测到异常 (超阈值滑脱/超载), 进入  $ABORT$ . 策略的状态转移如图 5 所示.

表 1 抓取过程状态表  
Table 1 Grasping process state table

状态	描述
<i>APPROACH</i>	按视觉规划, 机械臂靠近目标位姿
<i>CONTACT_DETECT</i>	夹爪逐步闭合, 直至检测到触觉传感器接触
<i>FORCE_BUILDUP</i>	继续逐步闭合夹爪并实时估计抓取力 $F_{est}$ 与形变指标
<i>ADJUST</i>	若检测到形变或力异常, 执行力调整 (降力/保持) 并必要时调整抓姿或撤回
<i>HOLD</i>	达到稳定抓取 ( $F_{est} \approx F_d$ 且无形变) 后进入保持态, 执行搬运任务
<i>RELEASE/ABORT</i>	完成任务或发生异常时释放或放弃抓取

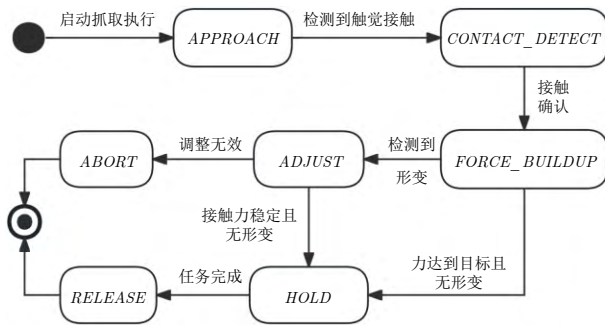


图 5 基于触觉反馈的抓取策略状态转移

Fig. 5 Grasping strategy state transition based on tactile feedback

### 1.3.2 实时估计与判定量

在抓取过程中, 触觉传感器输出需经过特征提取与物理量估计, 以作为状态转移与控制律的依据. 首先, 单像素对应的实际面积  $A_{\text{pixel}}$  由标定过程得到, 并作为接触面积计算的尺度因子. 接触区域的面积  $A(t)$  可由式 (8) 计算, 其值随抓取力的变化反映接触面的扩展情况. 同时, 通过式 (7) 进行触觉深度恢复并计算最大形变量  $d_{\text{max}}(t)$ .

为检测动态形变, 还需计算相邻帧的均方差 (MSE), 用于刻画接触区域在时间序列上的变化特性. 若满足判定条件 (见式 (9)), 则认为发生形变事件.

在抓取力估计方面, 本文采用弹性近似模型, 将最大形变量与接触面积结合, 用于实时估计法向力  $F_{\text{est}}(t)$ , 其计算方式见式 (11). 该估计值不仅为抓取力控制律提供参考, 也作为判断系统进入稳定抓取状态的重要依据.

### 1.3.3 抓取力调整

本文采用弱比例闭环与速率限制的抓取力指令生成方法, 能抵抗估计噪声并避免突增导致形变. 给定固定力  $F_d$ , 定义抓取力输出更新:

$$u(t) = u(t - \Delta t) + K_p \times \text{sat}(F_d - F_{\text{est}}(t), \Delta F_{\text{max}}) \quad (12)$$

其中,  $u(t)$  为当前夹爪力命令;  $\Delta t$  为力更新周期;

$\text{sat}(\cdot, \Delta F_{\text{max}})$  表示对力差的速率限制 (避免一次性大幅闭合);  $\Delta F_{\text{max}}$  为限制值;  $K_p$  为比例增益.

若在  $t$  时刻判断为发生形变 (见式 (9)), 则执行将力回退并锁定不再增加.

$$u(t) \leftarrow u(t - \Delta t) \quad (13)$$

若连续  $N_{\text{abort}}$  帧仍然检测到形变或力不能稳定, 则进入 *ABORT*: 回退至安全位并放弃抓取以保护物体与传感器.

综上, 本策略 (见算法 1、图 6) 构造了抓取中的“渐进增力-形变检测-力回退”模式, 使机械臂在抓取柔性或易碎物体时能保持稳定性与低损伤性.

### 算法 1. 基于触觉反馈的抓取策略

输入. 抓取位姿  $g$ , 固定力  $F_d$ .

输出. 抓取成功或抓取失败.

- 1) 机械臂移动至预抓取位姿 (*APPROACH*);
- 2) 夹爪逐渐关闭至检测到触觉接触 ( $H_{\text{real}} > \delta$ ) 进入 *CONTACT\_DETECT* 状态;
- 3) 初始化  $u = u_0$  (初始闭合量),  $u_{\text{prev}} = u$ , 时刻  $t = 0$ ;
- 4) **While** 抓取任务未结束
- 5) 获取  $H_{\text{real}}(t)$ , 计算  $d_{\text{max}}(t)$ 、 $\Delta d(t)$  与接触面积  $A(t)$ ;
- 6) 对  $d_{\text{max}}$ 、 $\Delta d$  和  $A$  进行滑动窗口平滑处理;
- 7) 计算法向力估计:  $F_{\text{est}}(t) = E \times \frac{\Delta d(t)}{d_0} \times A(t)$
- 8) 计算帧间均方差:
- 9)  $\text{MSE} = \text{MSE}(H_{\text{real}}(t), H_{\text{real}}(t - 1))$
- 10) **if**  $\frac{d}{dt}d_{\text{max}}(t) < 0 \wedge \text{MSE}(t, t - 1) > \eta$
- 11)  $u \leftarrow u_{\text{prev}}$ ;
- 12)  $\text{counter}_{\text{deform}} \leftarrow \text{counter}_{\text{deform}} + 1$ ;
- 13) **if**  $\text{counter}_{\text{deform}} > N_{\text{abort}}$
- 14) 转入 *ABORT* 状态;
- 15) **else**
- 16)  $\Delta \leftarrow \text{sat}(F_d - F_{\text{est}}(t), \Delta F_{\text{max}})$ ,  $u_{\text{prev}} \leftarrow u$ ;
- 17)  $u \leftarrow u_{\text{prev}} + K_p \times \Delta$ ;
- 18) **if**  $|F_{\text{est}}(t) - F_d| < \epsilon$  且稳定保持时间  $\geq$  阈值  $T_{\text{hold}}$
- 19) 转入 *HOLD* 状态;
- 20)  $t \leftarrow t + 1$ ;

- 21) *HOLD*: 保持  $u$  执行搬运任务, 直至进入 *RELEASE* 状态;
- 22) *RELEASE*: 张开夹爪, 报告成功;
- 23) *ABORT*: 张开夹爪, 报告失败.

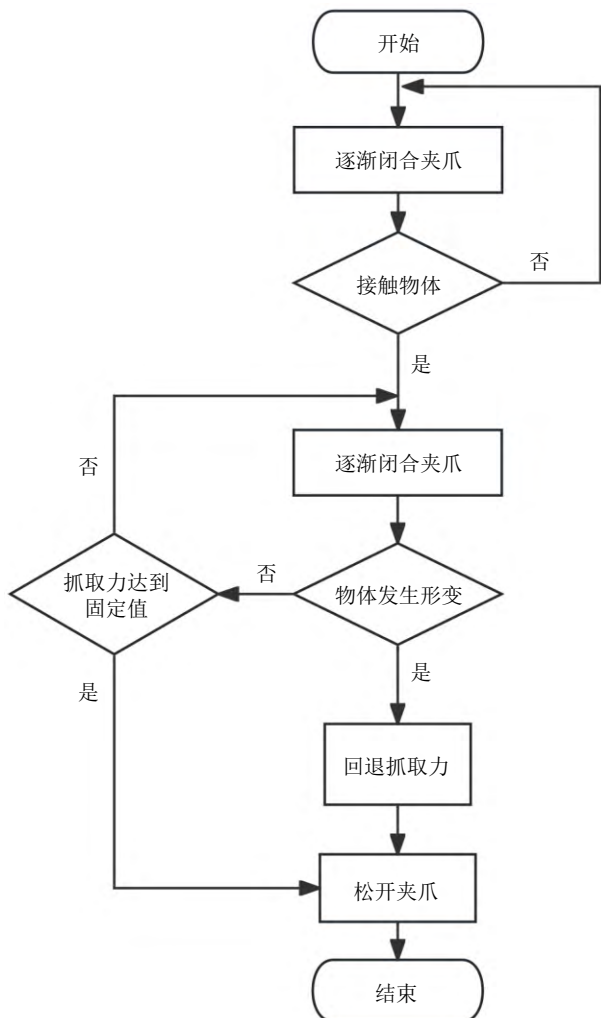


图 6 基于触觉反馈的自适应力控抓取策略流程图  
Fig.6 Flow chart of adaptive force control grasping strategy based on tactile feedback

## 2 实验与结果

### 2.1 实验平台

实验平台如图 7 所示, 其核心组件是一台 6 自由度机械臂 (节卡机器人的 JAKA MiniCobo). 通过末端执行器的法兰盘安装一个平行双指夹爪 (大寰机器人的 PGE-5-26), 该夹爪的行程范围为 26 mm, 最大抓取力为 5 N. 夹爪配备两对定制的 3D 打印手指 (PLA 材料), 手指张开宽度不同, 以适应不同尺寸的物体. 每个手指的内表面都集成一个高分辨率触觉传感器 (GelSight Mini). 这些传感器的感应

面积为 26 mm × 20 mm, 能以 640 × 480 像素的分辨率提供接触几何数据, 采样率为 25 Hz. 视觉感知系统采用一台 Intel RealSense D435i RGB-D 相机, 该相机直接安装在机械臂工作空间的正上方. 触觉传感器和 RGB-D 相机均与一台基于 Linux 系统的计算机相连, 所有采集到的图像数据均在此计算机上进行处理和分析.

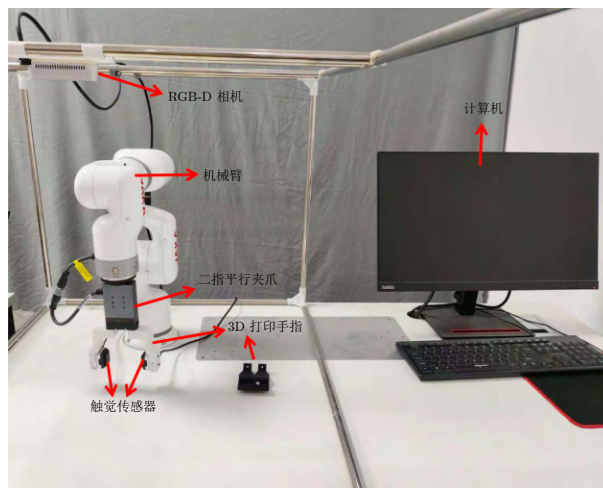


图 7 实验平台  
Fig.7 Experimental platform

### 2.2 视觉模块评估

#### 2.2.1 数据集

我们在三个数据集上评估视觉模块的性能.

1) 康奈尔 (Cornell) 抓取数据集<sup>[29]</sup>. 该数据集包含 885 张孤立物体的 RGB-D 图像, 这些图像带有人工标注的平行夹爪抓取矩形框.

2) RC\_49 抓取数据集<sup>[30]</sup>. 该数据集包含 15 000 张通过半全局匹配生成的合成深度图像, 用于模拟真实立体相机的成像缺陷, 包括深度不连续和未匹配像素. 它包含 49 个物体, 采用随机相机位姿和基于物理的场景生成方式, 提供自动标注的抓取信息, 无需人工预处理.

3) Jacquard 抓取数据集<sup>[31]</sup>. 该数据集由 54 485 个不同对象的 11 619 个不同场景组成, 共有 4967454 个抓取注释, 每个场景都提供渲染的 RGB 图像、分割蒙版、两个深度图像和抓取注释.

#### 2.2.2 评估指标

在评估视觉模块的性能时, 我们采用 Jacquard 指数 (J@N)<sup>[31-32]</sup> 作为核心评估指标. J@N 通过量化模型在排名前  $N$  的高质量抓取候选中包含至少一个有效抓取的能力, 来评估抓取预测模型的性能. 具体而言, 对于单张输入图像, 模型生成的抓取候

选会按质量从高到低排序.  $J@N$  指标计算测试集中“前  $N$  个预测结果包含正确抓取”的图像所占比例. “正确性”需同时满足两个关键条件:

1) 预测抓取矩形  $\hat{G}R$  与真实抓取矩形  $GR$  的交并比 (intersection over union, IoU) 大于 0.25, 即

$$IoU = \frac{|\hat{G}R \cap GR|}{|\hat{G}R \cup GR|} > 0.25 \quad (14)$$

2) 预测抓取角度  $\hat{\phi}$  与真实抓取角度  $\phi$  之间的角度偏差 ( $\Delta\phi$ ) 在  $30^\circ$  以内, 即

$$\Delta\phi = |\hat{\phi} - \phi| < 30^\circ \quad (15)$$

### 2.2.3 训练细节

对于第 2.2.1 节中提到的三个数据集, 我们将 90% 的数据用于训练, 10% 用于测试. 在视觉模块中, 我们采用 DINOv3 作为骨干网络, 输入图像尺寸为  $224 \times 224$  像素. 训练过程共进行 50 个 epoch, 批处理大小设为 32, 采用均方误差 (MSE) 作为损失函数, Adam 优化器的学习率为 0.001. 所有实验均在配备 NVIDIA GeForce RTX 4090 GPU 的服务器上完成.

### 2.2.4 性能分析

视觉模块在三个数据集上的评估结果如表 2 所示. 实验结果表明, 该模块在三个不同数据集上均表现出优异的抓取预测性能. 在 Cornell 数据集上的表现最为突出,  $J@1$  和  $J@5$  分别达到 93.26% 和 96.77%, 表明该模块对相关场景中物体的抓取位置预测具有极高的准确率. 在 RC\_49 数据集上,  $J@1$  和  $J@5$  分别为 90.32% 和 93.54%, 虽略低于 Cornell 数据集, 但仍保持较高水平, 验证了该模块对合成深度图像中物体的适应性. 在 Jacquard 数据集上,  $J@1$  为 87.10%,  $J@5$  为 93.55%, 这一结果表明, 当允许系统输出多个候选抓取位置时, 预测成功率可得到显著提升. 三个数据集的  $J@5$  指标均超过 90%, 充分证明了基于 DINOv3 的视觉特征提取和基于反卷积上采样方法的有效性.

表 2 不同数据集下视觉模块的性能比较 (%)  
Table 2 Performance comparison of vision module under different datasets (%)

指标	Cornell	RC_49	Jacquard
$J@1$	93.26	90.32	87.10
$J@5$	96.77	93.54	93.55

## 2.3 触觉模块评估

我们设计的触觉模块通过分析视触觉传感器 (GelSight Mini) 采集的接触图像, 恢复传感器表面

的微观形变信息, 并基于深度和接触面积计算法向力, 实现对物体接触状态的精细估计.

为了评估触觉模块在不同条件下的性能, 我们在抓取实验中分别设定 5 N 和 2 N 两种固定抓取力 (考虑到夹爪的施力范围为 1 N 至 5 N, 因此选择 5 N 作为较大抓取力的代表, 2 N 作为较小抓取力的代表), 系统性分析在不同力作用和物体特性下的触觉反馈图像与法向力估算表现. 实验对象选取八类常见物体 (见表 3), 在几何形状和材质特性上均具有代表性, 并根据形状规则性分为几何规则与不规则两类.

表 3 实验采用的八类不同物体  
Table 3 Eight different objects used in the experiment

几何规则		几何不规则	
			
纸杯	塑料瓶	易拉罐	弹力球
			
玻璃瓶	橡皮	胶棒	螺丝刀

图 8 展示几何规则物体的抓取结果, 其中第 2 行为抓取力为 2 N 时的触觉图像, 第 3 行为抓取力为 5 N 时的触觉图像. 触觉模块估计的法向力标注在每张触觉图像的上方. 实验采用两个不同的 3D 打印手指, 其内侧表面均安装有触觉传感器. 由于抓取过程中物体直接与传感器接触, 更换 3D 打印手指不会影响触觉反馈图像. 在未发生显著形变的情况下, 触觉图像中的接触面积相对稳定, 估算的法向力通常接近或略高于设定抓取力, 差异主要来源于局部压痕与受力集中效应. 例如, 橡皮在实验条件下几乎不产生明显形变, 其估计结果表现出与设定抓取力较高的一致性. 而当物体在抓取过程中发生显著形变时, 情况则不同: 接触面积会急剧增大, 即便深度值增长不显著, 估算的法向力也会大幅上升, 远高于设定抓取力. 图 8(a) 中的纸杯便是典型案例, 高抓取力作用下纸杯被压迫形变并与传感器表面紧密贴合, 导致接触面积和估算法向力同步显著增加.

表 4 进一步对比不同物体在相同抓取力下的最大深度值与接触面积, 结果显示, 未发生形变的物

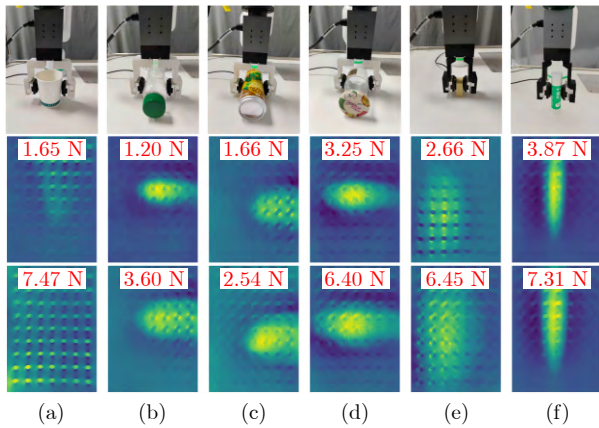


图 8 以固定力抓取几何规则物体的触觉图像 ((a) 纸杯; (b) 塑料瓶; (c) 易拉罐; (d) 玻璃瓶; (e) 橡皮; (f) 胶棒)

Fig.8 Grasping tactile images of geometrically regular objects with the fixed force ((a) Paper cup; (b) Plastic bottle; (c) Can; (d) Glass bottle; (e) Rubber; (f) Glue stick)

表 4 抓取几何规则物体的触觉最大深度值与接触面积  
Table 4 Maximum tactile depth value and contact area of grasping geometrically regular objects

	(a)	(b)	(c)	(d)	(e)	(f)
<i>depth</i> <sub>2</sub>	1.11	1.30	1.21	1.53	1.21	1.53
<i>depth</i> <sub>5</sub>	1.17	2.04	1.50	2.14	1.37	2.14
<i>area</i> <sub>2</sub>	11 612	7 234	10 718	16 573	17 123	14 842
<i>area</i> <sub>5</sub>	49 591	13 789	13 195	23 268	36 630	19 572

注: 对应图 8 中的触觉图像, “*depth/area*<sub>2/5</sub>”中, *depth* 代表有效接触区域的深度平均值, *area* 代表接触面积 (以深度值大于阈值的像素数量表示), 2/5 代表抓握力 (单位: N).

体主要表现为深度值主导的力估算, 而发生显著形变的物体则以接触面积的扩展为主导因素. 以塑料瓶与玻璃瓶为例 (如图 9, p/g 表示瓶子材料 (p: 塑料, g: 玻璃); 5/2 表示抓取力大小 (5 N 或 2 N); d/f 表示曲线类型 (d: 深度值, f: 估计法向力)), 两者外形相似, 但玻璃瓶在相同抓取力下产生更大的压痕和接触面积, 因此估计法向力远高于设定值; 塑料瓶由于低于形变阈值, 无法完全贴合传感器表面, 估算值则低于设定力.

在几何不规则物体实验 (见图 10, 其中第 2 行为抓取力为 2 N 时的触觉图像, 第 3 行为抓取力为 5 N 时的触觉图像, 触觉模块估计的法向力标注在每张触觉图像的上方) 中, 法向力估计随抓取位置呈现较大波动. 以螺丝刀为例, 不同抓取位置的接触面积差异明显, 当接触区域较小或局部压痕极深时, 估算力显著偏高. 柔性物体如弹力球则展现出典型的“阈值效应”: 在低抓取力作用下, 估算力低于设定值; 一旦抓取力超过临界点, 物体发生明显形变,

接触面积急剧扩展, 导致估算力快速升高 (见表 5).

总体来看, 法向力的估算主要受深度值与接触面积的共同影响: 在未发生明显形变时, 深度是主

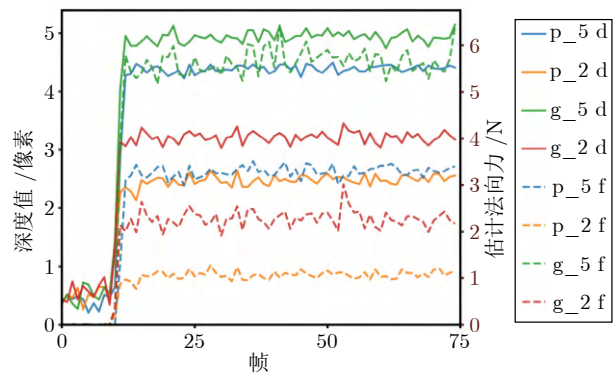


图 9 不同材料瓶子在不同抓取力下的深度值与估计法向力

Fig.9 Depth value and estimated normal forces of bottles made of different materials under different grasping forces

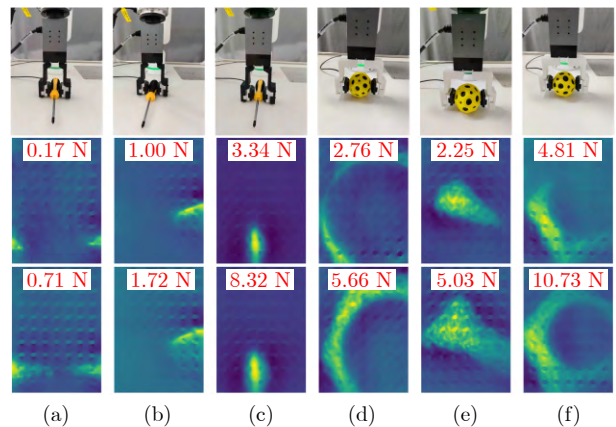


图 10 以固定力抓取几何不规则物体的触觉图像 ((a) ~ (c) 从三个位置抓取螺丝刀; (d) ~ (f) 从三个位置抓取弹力球)

Fig.10 Grasping tactile images of geometrically irregular objects with the fixed force ((a) ~ (c) Grasp the screwdriver at three positions; (d) ~ (f) Grasp the elastic ball at three positions)

表 5 抓取几何不规则物体的触觉最大深度值与接触面积  
Table 5 Maximum tactile depth value and contact area of grasping geometrically irregular objects

	(a)	(b)	(c)	(d)	(e)	(f)
<i>depth</i> <sub>2</sub>	1.00	1.50	3.86	0.91	1.65	1.04
<i>depth</i> <sub>5</sub>	1.34	1.65	2.53	1.71	1.74	1.45
<i>area</i> <sub>2</sub>	1 366	5 208	6 734	23 658	10 629	36 063
<i>area</i> <sub>5</sub>	4 105	8 121	25 567	25 840	22 529	57 648

注: 对应图 10 中的触觉图像.

导因素;当物体产生显著形变时,接触面积的增长成为决定性因素.通过对比设定抓取力与估算力,揭示了接触过程中的受力分布与传感器反馈特性,为后续抓取力的合理调节提供了实验依据.

在抓取力的估算中,我们采用基于胡克定律的简化模型,该模型假设材料为线性弹性,而未计入材料的非线性、凝胶的迟滞效应与粘弹特性,因此可能导致估算误差.为评估该模型在不同抓取力及物体刚度条件下对法向力的估算偏差,我们设计误差分析实验,通过与按压式压力传感器的对比进行验证.实验选取表 3 所列的 8 种物体,分别施加 9 种逐级增大的抓取力进行抓取.每次抓取过程中,同步记录触觉传感器与按压式压力传感器所测得的法向力数据,进而计算两者之间的绝对误差,并进行统计分析.

图 11 以热力图形式呈现不同抓取力与物体刚度条件下的误差分布,其中横轴为抓取力,纵轴为物体类别,色彩深浅反映误差大小.实验结果显示,对于刚性物体,误差整体较小.对于柔性物体,在抓取力较小时误差也较为有限;然而,当抓取力较大时,因物体形变导致法向力估算值出现显著偏差.值得指出的是,本文所采用的抓取策略在检测到物体形变时会自动回退抓取力,因而在实际抓取操作中,此类误差并不会发生.

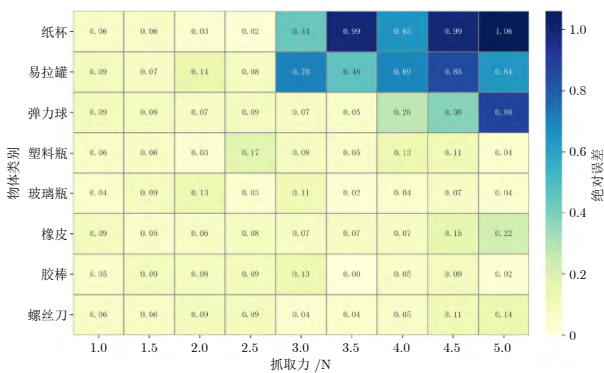


图 11 误差热力图

Fig. 11 Error heatmap

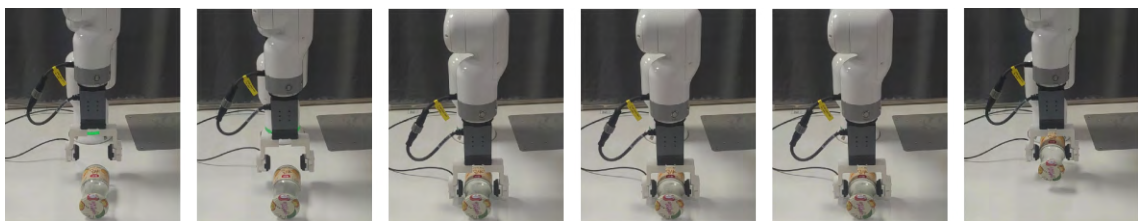


图 12 实物抓取流程

Fig. 12 Physical grasping process

## 2.4 抓取策略评估

如图 12,为验证所提方法的有效性,我们开展物理抓取实验,并与采用固定抓取力的方法进行对比.具体而言,选取与第 2.3 节相同的 8 个常见物体作为抓取目标,每个物体进行 10 次抓取尝试.我们重点关注四个评估指标:成功率、最终施加的抓取力、损坏率和滑动率.一次成功的抓取操作必须满足两个标准:一是不发生滑动,二是物体未受损坏(对于柔性物体尤为重要).我们通过视觉检查物体是否发生严重形变、损坏或滑动来判断抓取操作是否成功.

由于本文方法和基线方法均使用视觉模块进行预测,因此未将预测精度纳入评估指标.我们仅在结果中呈现该模块在物理抓取过程中的表现,并以预测失败率对其进行评估.此外,由于我们不对比预测精度,对每个物体进行的抓取实验均在假设预测准确的前提下进行,结果如表 6 所示.

实验结果表明,本文方法在涉及 8 个常见物体的抓取任务中展现出显著优势.与仅使用视觉模块并采用固定抓取力的基线方法相比,本文方法将整体成功率从 87.50% 提升至 98.75%,其中在抓取纸杯等易碎物体时成功率的提升尤为显著.在力控制方面,本文方法实现抓取时的动态力调节,将抓取力从固定的 4 N 优化为适应物体特性的合理范围.这有效避免基线方法因统一施加较大抓取力而导致的物体损坏,实现了零损坏.值得注意的是,基线方法在物体抓取过程中虽实现零滑动,但这是以施加过大抓取力为代价的.而本文方法在有效降低抓取力的同时,将滑动率控制在仅 1.25%.在预测性能方面,视觉模块总体表现良好,平均预测失败率为 11.25%.具体而言,对于弹性球等可形变物体,本文方法通过动态力调节,完全避免了基线方法常见的挤压形变问题.

如图 13 所示,为验证本方法在非结构化环境中的普适性与长期稳定性,我们设计多物体连续抓取实验.实验在模拟杂乱物体布局的场景中进行,包含多种刚性、柔性及不规则形状物体.机械臂需依次抓取所有物体并移栽至目标位置,共执行 43

表 6 自适应力控与固定力抓取策略的实验结果对比

Table 6 Comparison of experimental results between adaptive force control and fixed force grasping strategy

物体	成功率 (%)		施加力 (N)		损坏率 (%)		滑动率 (%)		视觉模块预测失败率 (%)
	自适应力	固定力	自适应力	固定力	自适应力	固定力	自适应力	固定力	
橡皮	100	100	4	4	0	0	0	0	30
胶棒	100	100	4	4	0	0	0	0	0
弹力球	100	100	1	4	0	0	0	0	0
纸杯	100	0	1	—	0	100	0	0	10
玻璃瓶	100	100	4	4	0	0	0	0	10
塑料瓶	100	100	2.58	4	0	0	0	0	10
易拉罐	90	100	2.44	4	0	0	10	0	20
螺丝刀	100	100	4	4	0	0	0	0	10
总计	98.75	87.50	—	—	0	12.50	1.25	0	11.25

注: 表中“—”表示由于固定力方法对纸杯的抓取完全失败, 未产生有效的施加力数据; 自适应力指自适应力控抓取方法, 固定力指使用夹爪默认抓取力; 所有实验中每个物体均进行 10 次抓取尝试, 统计结果以百分比形式呈现。

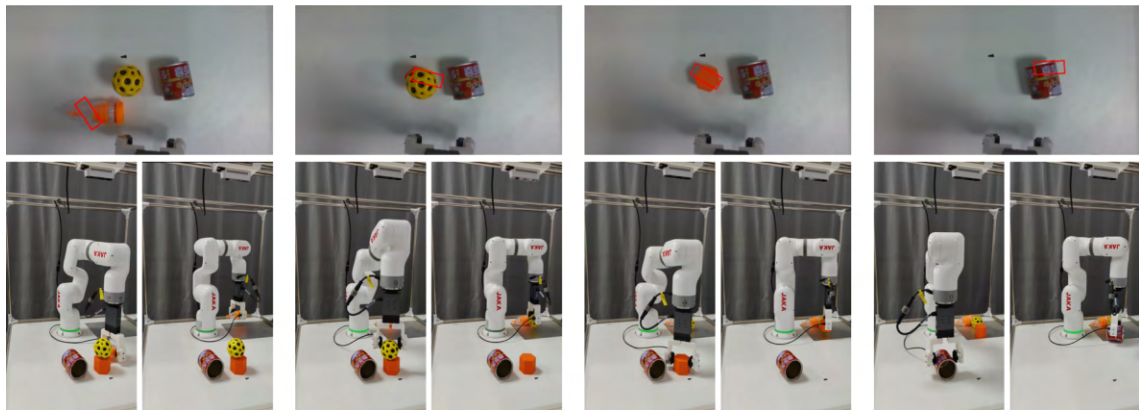


图 13 多物体连续抓取

Fig. 13 Multi-object continuous grasping

次连续抓取任务, 成功 36 次, 整体成功率为 83.72%。

在长期连续操作过程中, 该方法展现出良好的抗误差积累能力。具体而言, 触觉深度恢复与法向力估计在连续 40 余次操作中未出现显著偏移, 且未呈现随时间单调增长的漂移趋势。这一性能主要源于 GelSight Mini 传感器基于光度立体的工作原理, 其输出为图像域数据, 不涉及应变片或压阻式传感器常见的电信号漂移; 同时, 抓取过程中的触觉深度在每次接触时均重新计算, 从而避免将上一轮误差传递至后续任务。这些机制共同保障了系统在长期运行中的稳定性与重复性。

进一步地, 该方法在杂乱堆叠等非结构化场景中也表现出较强的适应能力。尽管物体与夹爪之间的接触方式和受力路径具有不确定性, 但本文所提出的触觉反馈机制能够在接触阶段直接反映实际力学状态, 并基于深度变化及其趋势触发形变判定与力回退, 从而在接触模式变化时维持安全的施力范围。该机制不依赖物体类别或刚度先验, 即可在刚

性物体上自然建立稳定抓取力, 在柔性或薄壁物体上及时停止进一步闭合以避免损伤。实验结果表明, 即使在接触区域随机变化、物体材质混杂的复杂环境中, 该力控策略仍能维持较低损坏率和稳定的抓取成功率。

针对抓取任务中的失败案例, 我们进行分析。结果显示, 失败主要源于两类因素: 环境碰撞 (6 次, 13.95%) 与视觉预测偏差 (1 次, 2.32%), 而抓取力控制本身未导致失败。环境碰撞多发生于机械臂末端进入杂乱区域时, 因末端与其他物体发生碰撞而中断任务; 视觉偏差则主要出现在透明或纹理稀疏物体上, 其抓取位姿预测易受光照或反射干扰。

综上所述, 尽管在复杂环境中执行多次连续抓取, 本方法仍保持较高的操作稳定性与重复性, 误差累积对抓取力的影响较小。失败主要来源于环境碰撞干扰与视觉预测误差, 而非抓取力控制本身, 进一步验证了所提抓取策略在非结构化环境中的有效性与长期可靠性。

这些结果验证了我们的方法在实现安全、自适应抓取方面的有效性,其基于触觉反馈的抓取策略为具有不同材料属性的物体提供了一种更稳健的抓取方法。

### 3 结束语

本文提出一种基于“形态-感知-动作”仿生机理的机器人自适应力控抓取方法,旨在解决恒力抓取时容易导致物体滑动或受损的问题。本方法以基于 DINOv3 的视觉模块完成抓取位姿预测,通过在接触阶段恢复视触觉传感器 (GelSight Mini) 的触觉深度并估算接触面积与法向力,基于深度变化率与帧间 MSE 的形变判定触发抓取力调节策略,构建“渐进增力-形变检测-力回退”的抓取机制。本方法在提高抓取成功率、降低物体损伤和实现能效抓取方面具有明显优势,为服务机器人中的多模态感知与操作提供了可复现的实验平台与方法论依据。

### 参考文献

- Liu Hua-Ping, Guo Di, Sun Fu-Chun, Zhang Xin-Yu. Morphology-based embodied intelligence: Historical retrospect and research progress. *Acta Automatica Sinica*, 2023, **49**(6): 1131-1154 (刘华平, 郭迪, 孙富春, 张新钰. 基于形态的具身智能研究: 历史回顾与前沿进展. *自动化学报*, 2023, **49**(6): 1131-1154)
- Liu Z, Hu X N, Bo R H, Yang Y Z, Cheng X, Pang W B, et al. A three-dimensionally architected electronic skin mimicking human mechanosensation. *Science*, 2024, **384**(6699): 987-994
- Li J F, Xu Z, Zhu D J, Dong K, Yan T, Zeng Z, et al. Bio-inspired intelligence with applications to robotics: A survey. *Intelligence & Robotics*, 2021, **1**(1): 58-83
- Chatzivasiliadis D, Zhong S, Christopoulos V, Karydis K. Adaptive environment-aware robotic arm reaching based on a bio-inspired neurodynamical computational framework. In: Proceedings of the 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN). Pasadena, CA, USA: IEEE, 2024. 510-515
- An B S, Geng Y R, Chen K, Li X Q, Dou Q, Dong H. RGB-Manip: Monocular image-based robotic manipulation through active object pose estimation. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE, 2024. 7748-7755
- Fang H S, Wang C X, Fang H J, Gou M H, Liu J R, Yan H X, et al. AnyGrasp: Robust and efficient grasp perception in spatial and temporal domains. *IEEE Transactions on Robotics*, 2023, **39**(5): 3929-3945
- Zhu L F, Wang Y C, Mei D Q, Wu X. Highly sensitive and flexible tactile sensor based on porous graphene sponges for distributed tactile sensing in monitoring human motions. *Journal of Microelectromechanical Systems*, 2019, **28**(1): 154-163
- Xiang G Y, Wang X X, Cheng N Y, Hu L, Zhang H W, Liu H H. A flexible piezoelectric-based tactile sensor for dynamic force measurement. In: Proceedings of the 2022 International Conference on High Performance Big Data and Intelligent Systems (HDIS). Tianjin, China: IEEE, 2022. 207-211
- Yuan W Z, Dong S Y, Adelson E H. GelSight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 2017, **17**(12): 2762-2782
- Lambeta M, Chou P W, Tian S, Yang B, Maloon B, Most V R, et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters*, 2020, **5**(3): 3838-3845
- Saxena A, Driemeyer J, Kearns J, Ng A Y. Robotic grasping of novel objects. In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2007. 1209-1216
- Morrison D, Corke P, Leitner J. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach. In: Proceedings of Robotics: Science and Systems. Pittsburgh, USA: MIT Press, 2018.
- Wang S C, Zhou Z L, Kan Z. When Transformer meets robotic grasping: Exploits context for efficient grasp detection. *IEEE Robotics and Automation Letters*, 2022, **7**(3): 8170-8177
- Fang H S, Gou M H, Wang C X, Lu C W. Robust grasping across diverse sensor qualities: The GraspNet-1Billion dataset. *The International Journal of Robotics Research*, 2023, **42**(12): 1094-1103
- Jiang J Q, Cao G Q, Butterworth A, Do T T, Luo S. Where shall I touch? Vision-guided tactile poking for transparent object grasping. *IEEE/ASME Transactions on Mechatronics*, 2022, **28**(1): 233-244
- Zhang B Y, Andrussov I, Zell A, Martius G. The role of tactile sensing for learning reach and grasp. In: Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA). Atlanta, USA: IEEE, 2025. 11817-11824
- Xue Teng, Liu Wen-Hai, Pan Zhen-Yu, Wang Wei-Ming. Stable robotic grasping based on visual perception and prior tactile knowledge learning. *Robot*, 2021, **43**(1): 1-8 (薛腾, 刘文海, 潘震宇, 王伟明. 基于视觉感知和触觉先验知识学习的机器人稳定抓取. *机器人*, 2021, **43**(1): 1-8)
- Wang Rui-Dong, Wang Rui, Zhang Tian-Dong, Wang Shuo. A survey of research on robotic brain-inspired intelligence. *Acta Automatica Sinica*, 2024, **50**(8): 1485-1501 (王瑞东, 王睿, 张天栋, 王硕. 机器人人类脑智能研究综述. *自动化学报*, 2024, **50**(8): 1485-1501)
- Cui S W, Wang R, Wei J H, Li F R, Wang S. Grasp state assessment of deformable objects using visual-tactile fusion perception. In: Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE, 2020. 538-544
- Calandra R, Owens A, Jayaraman D, Lin J, Yuan W Z, Malik J, et al. More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robotics and Automation Letters*, 2018, **3**(4): 3300-3307
- Li S J, Yu H X, Ding W B, Liu H D, Ye L Q, Xia C K, et al. Visual-tactile fusion for transparent object grasping in complex backgrounds. *IEEE Transactions on Robotics*, 2023, **39**(5): 3838-3856
- Tang J X, Yuan X G, Li S D. Visual-tactile fusion and sac-based learning for robot peg-in-hole assembly in uncertain environments. *Machines*, 2025.
- Lee Y, Hong S, Kim M G, Kim G, Nam C. Grasping deformable objects via reinforcement learning with cross-modal attention to visuo-tactile inputs. arXiv preprint arXiv: 2504.15595, 2025.
- Fu L, Huang H, Berscheid L, Li H, Goldberg K, Chitta S. Safe self-supervised learning in real of visuo-tactile feedback policies for industrial insertion. In: Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). London, United Kingdom: IEEE, 2023. 10380-10386
- Matak M, Hermans T. Planning visual-tactile precision grasps via complementary use of vision and touch. *IEEE Robotics and Automation Letters*, 2022, **8**(2): 768-775
- Peng Z C, Cui T, Chen G Y, Lu H Y, Yang Y, Yue Y F. High-precision object pose estimation using visual-tactile information for dynamic interactions in robotic grasping. In: Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA). Atlanta, USA: IEEE, 2025. 14799-14805
- Galaiya V R, de Oliveira T E A, Jiang X T, da Fonseca V P. Grasp approach under positional uncertainty using compliant tactile sensing modules and reinforcement learning. In: Proceedings of the 2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). Kingston, Canada: IEEE, 2024. 424-428
- Siméoni O, Vo H V, Seitzer M, Baldassarre F, Oquab M, Jose

C, et al. Dinov3. arXiv preprint arXiv: 2508.10104, 2025.

- 29 Jiang Y, Moseson S, Saxena A. Efficient grasping from RGBD images: Learning using a new rectangle representation. In: Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai, China: IEEE, 2011. 3304–3311
- 30 Drögemüller J, Garcia C X, Gambaro E, Suppa M, Steil J, Roa M A. Automatic generation of realistic training data for learning parallel-jaw grasping from synthetic stereo images. In: Proceedings of the 20th International Conference on Advanced Robotics (ICAR). Ljubljana, Slovenia: IEEE, 2021. 730–737
- 31 Depierre A, Dellandréa E, Chen L M, Jacquard: A large scale dataset for robotic grasp detection. In: Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain: IEEE, 2018. 3511–3516
- 32 Ainetter S, Fraundorfer F. End-to-end trainable deep neural network for robotic grasp detection and semantic segmentation from RGB. In: Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE, 2021. 13452–13458



**赵 洲** 华中师范大学计算机学院副教授。2023 年获得法国巴黎索邦大学计算机科学与技术专业博士学位。主要研究方向为机器人智能感知。

E-mail: zhaozhou@ccnu.edu.cn

**(ZHAO Zhou** Associate professor at the School of Computer Science,

Central China Normal University. He received his Ph.D. degree in computer science and technology from Sorbonne University in Paris, France in 2023. His main research interest is robot intelligent perception.)



**耿明强** 华中师范大学计算机学院硕士研究生。2025 年获得华中师范大学软件工程专业学士学位。主要研究方向为机器人技术。

E-mail: gengmq@mails.ccnu.edu.cn

**(GENG Ming-Qiang** Master student at the School of Computer Science,

Central China Normal University. He received his bachelor degree in software engineering from Central China Normal University in 2025. His main research interest is robotic technology.)



**何秋实** 中国人民武装警察部队警官学院教员。2025 年获得四川大学管理学专业硕士学位。主要研究方向为军事指挥, 管理科学与工程, 无人系统智能感知与决策。

E-mail: 15680879000@163.com

**(HE Qiu-Shi** Instructor at the Of-

ficers College of PAP. He received his master degree in management from Sichuan University in 2025. His research interests include military command, management science and engineering, intelligent perception and decision-making of unmanned systems.)



**何贇鑫** 华中师范大学计算机学院硕士研究生。2023 年获得华东交通大学计算机科学与技术专业学士学位。主要研究方向为机器人技术。E-mail: heyunxin@mails.ccnu.edu.cn

**(HE Yun-Xin** Master student at the School of Computer Science,

Central China Normal University. He received his bachelor degree in computer science and technology from East China Jiaotong University in 2023. His main research interest is robotic technology.)



**蔡明达** 武汉理工大学自动化学院硕士研究生。2024 年获得长沙理工大学测控技术与仪器专业学士学位。主要研究方向为人机协作。

E-mail: 19307496164@163.com

**(CAI Ming-Da** Master student at the School of Automation, Wuhan

University of Technology. He received his bachelor degree in measurement and control technology and instrument from Changsha University of Science and Technology in 2024. His main research interest is human-robot collaboration.)



**周翔宇** 武汉理工大学自动化学院硕士研究生。2021 年获得武汉理工大学电气工程及其自动化专业学士学位。主要研究方向为人机协作。

E-mail: xiangyuzhou@whut.edu.cn

**(ZHOU Xiang-Yu** Master student at the School of Automation,

Wuhan University of Technology. He received his bachelor degree in electrical engineering and automation from Wuhan University of Technology in 2021. His main research interest is human-robot collaboration.)



**罗 晶** 武汉理工大学自动化学院教授。2020 年获得中国广州华南理工大学和英国伦敦帝国理工学院联合博士学位。主要研究方向为机器人, 远程操作, 可穿戴设备和人机交互。本文通信作者。

E-mail: ljing\_ac@whut.edu.cn

**(LUO Jing** Professor at the School of Automation, Wuhan University of Technology. He received his joint Ph.D. degree from South China University of Technology in Guangzhou, China and Imperial College London, UK in 2020. His research interests include robotics, teleoperation, wearable devices, and human-machine interaction. Corresponding author of this paper.)