

Grasp Generation in Dense Clutter for General Robotic Hands

Doctor of Philosophy in Electromechanical Engineering

2026

拓展中文摘要

随着机器人应用快速拓展至家庭和服务领域，机器人的操作能力及其通用性成为制约机器人普及的重要挑战之一。现有成熟的机器人操作算法缺乏有效的跨本体泛化能力，同时目前具有跨本体部署潜力的机械手操作算法面临严重的场景感知瓶颈。针对上述研究难点，本论文聚焦研究机器人抓取操作算法的跨机械手通用机理及现实可部署的接触空间建模方法，着力推动机器人抓取操作在二维和三维非结构化环境中部署至任意机械手及自主系统的目标。首先，针对基于学习的抓取生成框架难以泛化到多指机械手，而基于优化框架的抓取策略缺乏在非结构环境部署潜力的难点，本文首次提出适用于通用机械手的学习-优化混合的抓取规划框架。其次，为弥合场景感知与手部硬件之间的建模鸿沟，本文显式定义一种用于表征机器人-环境交互状态的接触特征，并基于该定义的接触特征构建了两个大规模与机械手无关的解耦接触特征数据库，以支持机械手无关的场景-接触推理网络的训练。进一步地，该接触特征网络的推理结果作为通用环境感知模型，输入到本文可更换机械手的抓取解析优化模块，结合构建的统一机械手描述以及多个与接触特征协同的抓取准则，实现在二维及三维空间中二指至五指通用机械手的鲁棒物体抓取姿态生成。

通过基于大规模的数据集评估以及在密集杂乱环境中的抓取实验，本文提出的抓取框架均超越了当前基于解析模型和基于学习方法构建的二维或三维通用抓取研究，取得了持续更优的抓取成功率（在三维领域，针对从二指到五指的七种机械手，本框架成功率达到 94.3%-98.0%）。此外，本文所建立的跨机械手抓取框架实现了零样本机械手泛化，即无需针对特定硬件进行任何重新训

练，可在真实世界中赋予任意关节式机械手以熟练的抓取能力。同时，多个拓展验证实验进一步展示了混合框架在密集杂乱场景下实现 20 Hz 动态抓取以及多手协同完成复杂任务的能力。以上实验结果证明，该混合建模范式通过共享所提出的接触交互特征将环境感知与机器人动作解耦，为实现跨机械体复用的操作智能提供了可行路径，是迈向通用机器人的关键一步。

杂乱环境中通用机械手的二维平面抓取姿态规划：

机器人平面抓取广泛应用于工业无序零件整理，物流服务分拣及其他自动化应用领域。此类应用场景的可利用传感设备有限，机械驱动自由度较低（或仅可完成平面平移运动），同时末端夹持器定制化程度高。因此，研究适用于此类应用场景的通用末端机械手平面机器人抓取算法时，需要同时兼顾算力消耗合理高效，高夹持器通用程度，以及传感设备成本低廉稳定的等实际需求。在现有单一机械手平面抓取算法研究中，适用于杂乱环境的抓取规划算法通常采用深度学习框架直接基于环境观测进行机械手位姿及抓取手型推理，该类推理策略具有直接应用于商业传感器观测下的非结构化环境的潜力，但其基于特

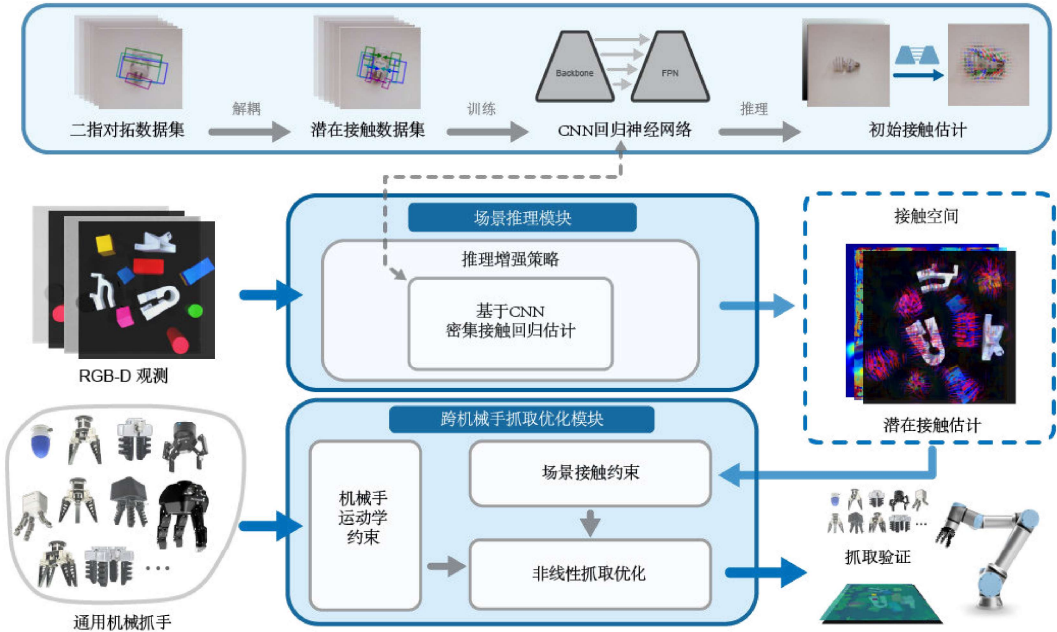


图 1 适用于通用机械手在杂乱环境中的平面抓取姿态规划

定机械手结构参数下显式确定的推理输出严重限制了此类模型可适用的机械手构型，且通常只适用于二指平行夹持器。针对拓展机械手通用性的挑战，近期代表性的研究尝试用神经网络编码器隐式编码不同机械手构型作为抓取评判网络输入的一部分，同时结合环境观测来预测机械手抓取点或对抓取构型进行打分评判。但实验数据表明，该类以神经网络建模机械手构型的方案仅能泛化至三指及以下的低自由度机械手，且应用于不同手型尺寸或新机械手时存在明显的泛化瓶颈。因此，需要研究设计出一种能够对高自由度机械手进行有效泛化，可用于密集杂乱的场景，且兼容低成本商业视觉传感器的平面抓取规划方法。

本研究提出的跨机械手二维平面抓取规划方法 HybridGen（图 1）首次结合环境感知学习推理框架和机械手解析优化框架进行抓取姿态的混合规划。该方案创新点在于定义了一种在平面空间中用于描述机械手与环境物体接触的状态特征形式，该平面特征简洁地描述了机器人与外部环境相互接触所必须的状态描述，即潜在接触的位置，姿态，以及置信值。该接触特征的新颖之处在于：其可行域遍布整个操作动作空间（而非为仅被抓取操作物体的表面）。这一关键点能够带来两方面的优势：1）所提出的接触特征与机械手构型无关，且可以直接从观测空间中通过神经网络推理获得；2）该接触特征不仅描述了与环境及被操作物体相关的交互点状态，同时可以被直接引入到解析优化框架中进行与机械手相关的抓取手部姿态优化。这些优势为后续的环境-接触推理模块和跨机械手解析优化模块的有效解耦及互相配合奠定核心基础。

基于所建立的平面接触特征模型，本研究进一步创新性地构建在杂乱环境中快速生成大规模接触数据集的方法。由于现存抓取数据集均以二指平行夹持器作为目标机械手，虽然该机械手型能够在处理不同的抓取情况，同时具备稳

定的抓取效果，但是面向二指夹持器的数据集难以拓展至其他机械手构型。为此，本文首次提出一种针对通用机械手（包括二指）抓取数据的解构方法，用于在操作动作空间中生成大规模与机械手无关的密集接触特征。首先，该方法利用面向杂乱环境的抓取数据集 Cornell Grasping Dataset 中的标注数据，将每一个二指标签的指尖位置解构为二个互不相关的区域，用以表征该位置的潜在接触方向以及置信度（参见图 1）。由于场景观测中存在多个标注的二指标签数据，且多个接触标签区域可能存在部分重叠，因此，整个抓取场景下的接触标签的集合可作为该场景所有已知可行的潜在接触分布。基于此原理及现存的密集二指抓取数据集，本方法实现了快速生成大规模潜在接触状态分布的数据库，且该数据库的环境观测数据包含了观测传感器结构化特点及原始噪声的分布，便于使用同类型视觉观测设备快速完成无偏的场景泛化。进一步地，该研究利用接触特征数据集训练基于 CNN 结构的深度神经网络（以 RGB-D 图像作为模型输入，与机械手无关的接触特征集合作为标签），该接触特征神经网络可对新的环境观测图像进行快速推理以得到其潜在接触分布，同时结合改进的推理增强策略，有效提高了通用机械手抓取框架的非结构化场景适应能力。

基于上述 CNN 神经网络推理得到的接触概率分布，本研究进一步提出跨机械手的解析优化抓取策略模块。该模块结合提出的以数学解析式描述的机械手模型，对任意机械手进行平面抓取构型优化。作为本工作重要创新点之一，该模型梳理并总结了在二维平面抓取背景下通用机械手的基本关节运动类别及其解析描述形式，其中涉及的关节分类标准包含了关节类型（平行关节和旋转关节）和关节相对方向（即关节轴垂直或平行于抓取平面）。进一步地，该关节解析方法成功应用于十个不同机械手（从二指到四指，包含软体抓持器）的快速建模，在保证机械手建模精度的同时，其简洁的数学解析形式有利于更改结构

参数以及对定制化夹爪进行快速建模。平面抓取优化模块以机械手的候选接触姿态作为状态变量，以描述机械手构型的解析方程为优化约束，基于场景推理得到的潜在接触分布建立场景接触约束，同时以抓取稳定性和优化速度构建目标函数以优化机械手手型。该优化模块利用数学优化工具 CasADi 构建。

该研究通过多层次的消融实验成功验证了所提出的接触推理及机械手优化策略模块的有效性。首先，针对接触推理模块中改进的集成学习策略，本文分别验证了其以深度图像和 RGB-D 图像作为模型输入的接触预测性能，结果表明其有效提高回归推理模型在仅深度图输入情况下的预测能力，即系统在观测条件不足情况下的鲁棒适应能力。在二指抓取数据集夹爪检测实验以及真实世界下的多样化物体的抓取实验中，平面混合抓取优化框架均显著优于现有研究方法，达到 92.5% 以上的二指抓取成功率。研究进一步建立面向多个通用机械手的数据集验证实验，其聚焦三大泛化指标：机械手对场景物体的泛化程度、抓取构型的多样化程度、以及抓取构型生成的稳定性。实验结果表明提出的混合抓取框架成功根据不同的机械手构型优化手部关节位置，所生成的抓取构型覆盖了十个被验证机械手的灵巧抓取范围，且算法生成抓手姿态的速度及稳定性能够满足现实场景需求。进一步地，本研究配置十个真实机械手进行在多物体杂乱环境的抓取实验，在跨机械手泛化表现以及的抓取成功率等指标上，本文混合抓取框架的抓取性能表现远优于现有跨抓手算法（AdaGrasp）。

基于以上构建的算法框架及综合实验验证，本文提出的跨机械手混合平面抓取框架具备以下显著优越性能：1) 对机器人部署的非结构场景适应能力强，且在极端环境观测条件下表现鲁棒；2) 具备对跨多关节多手指的机械手实现零样本泛化的能力，无需对系统模块作微调或后训练即可立刻部署到新机械手；

3) 场景推理神经网络及抓取优化策略的结构精简高效, 且面对多抓手同步抓取时可复用同一场景推理模块, 进一步提高系统计算效率。这些性能表现极大拓展了机器人操作硬件的适用性以及面向非结构化环境时的算法适应能力。

杂乱环境中通用机械手的三维空间抓取姿态规划

以形封闭和力封闭分析为核心的机器人抓取理论经过数十年的研究已趋于完善, 但相关抓取理论在真实环境中的部署始终面临严峻挑战。其中最突出的部署难点在于, 机器人难以获取精确的待抓物体几何模型, 因而无法进行有效的力封闭计算以及碰撞检测等优化策略。在非结构化环境中抓取未建模的多样化物体时, 基于低成本视觉传感器的三维重建技术在常规单相机视角的观测条件下无法得到物体的完整精确几何模型。尽管部分抓取研究着力引入观测预处理算法来尝试获得物体完整的几何表面(如场景物体分割算法, 基于概率函数的物体表面估计, 基于编码器-解码器结构的物体重建), 其效果仍然显著依赖于物体的先验信息以及受限于场景的复杂度, 而无法良好地部署于多物体杂乱环境中。另一方面, 现有基于深度学习-微分优化的跨机械手抓取算法面临着类似的场景适应挑战。该类研究着重针对悬浮且具有完整几何表面的单物体(即已知完好建模的物体几何网格或点云)进行多机械手手型优化, 缺乏考虑当前机器人环境观测的可行性及条件限制, 从而严重限制其实际部署的潜力。

基于前述的平面混合抓取框架, 将跨机械手抓取从二维空间拓展至三维空间同样面临若干算法难点: 1) 环境和机械手的建模复杂度以及抓取姿态的计算难度显著增加; 2) 提出适用于三维空间的与抓手无关的接触数据集; 3) 基于三维离散场景特征的抓取力封闭验证及碰撞检测。这些难点主要涉及基于微分优化过程的计算复杂度以及算法的稳定性问题。针对这些挑战, 本文将平面

混合抓取框架重构，加入多个与三维潜在接触空间匹配的抓取优化策略，提出适用于杂乱环境的跨机械手三维空间抓取规划框架 SpaHybGen（见图 2）。该架构由三个模块构成：接触估计模块，通用机械手构建模块，以及抓取优化模块。由于同样采用与机械手无关的场景-接触估计策略和面向通用机械手的抓取优化框架，SpaHybGen 需要在三维空间中，建立适用于接触特征推理同时可结合空间抓取优化的潜在接触特征形式。此外，面向三维空间所增加的机械手复杂度对建立统一机械手描述模型提出了新的挑战。

为在三维空间中高效推理得到潜在接触特征，特征推理网络以 3D 体素作为环境观测输入，并以相同空间分辨率输出对应的潜在接触估计，包括潜在接触的位置，方向，以及置信度。此外，针对在三维杂乱环境中物体质心难以定位

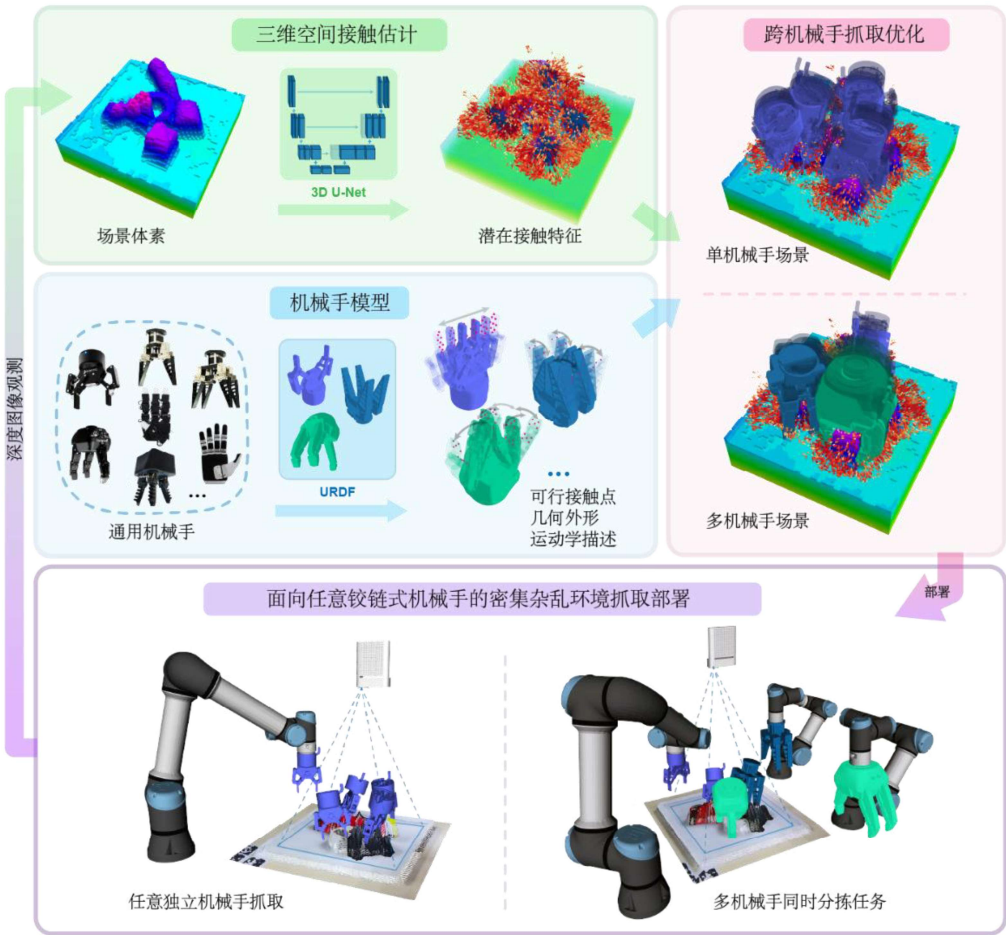


图 2 适用于通用机械手在杂乱环境中的平面抓取姿态规划

的问题，特征推理模型同时预测当前场景下物体的可行抓取中心。为提高推理得出的接触特征的鲁棒性以及便于后续跨机械手优化，该三维抓取框架探究了多种环境编码以及特征表征的格式，并进一步设计多种特征优化聚类策略，以得到更精简有效的潜在接触特征分布。本研究采用的三维空间机械手模型以 URDF 文件作为通用描述格式，同时根据不同手型引入候选接触区域（点）以引导三维微分优化建立有效的交互接触。基于 1) 深度传感器观测的环境几何信息；2) 推理提炼得到的空间潜在接触特征；以及 3) 构建的机械手模型，本研究的抓取微分优化模块同时优化多个抓取目标：机械手候选接触与相关场景潜在接触之间的位姿偏差，机械手当前抓取中心与场景可行抓取中心的距离，机械手-场景物体的碰撞惩罚，机械手自定义运动学约束的偏离程度，以及当前机械手手型的力封闭评估。该五类优化目标共同实现对任意铰链式机械手的抓取手型优化，且能够有效保留机械手的灵巧性及发挥被优化机械手的可用抓取构型。不同于经典的抓取优化理论，本文涉及的抓取优化策略均非基于单个物体的几何表面，而是直接建立在学习推理得到的空间潜在接触特征上，该特点有效解决了在观测条件严重受限且环境杂乱的情况下无法获取准确物体信息的难题。同时，该场景建模方式也给机械手抓取优化理论的重构实现提出多重挑战。具体地，本算法着重重构了机械手力封闭计算的依据，通过有效建立接触状态表征，将力封闭的评判函数与当前各潜在接触的置信程度相关，从而实现将已有抓取理论引入所构建的潜在接触空间的目标。进一步地，为应对更严峻的场景物体杂乱程度以及增强跨机械手的算法通用性，本文改进了一系列机械手批量初始化以及优化策略，从而使得所提出的混合抓取系统在杂乱多样场景下对任意铰链式机械手实现零样本的直接部署，充分保证在多机械手上实现灵巧、鲁棒、高成功率的抓取效果。

为评估三维跨机械手抓取框架在二指数据集的基本抓取检测性能，本文将 SpaHybGen 与基于解析模型的二指抓取方法，基于推理学习的二指抓取方法，以及自身框架的消融实验进行对比，结果表明该框架达到了当前（仅）二指抓取研究算法同等优良的检测水平，而且在未训练的抓取场景中表现出更好的泛化性能。此外，除了二指抓取检测，SpaHybGen 还具备面向通用机械手的抓取生成能力。针对跨机械手性能，本文测试算法在七个显著不同的机械手（二指至五指，两个手指关节至十六个手指关节）面向单物体及轻微杂乱环境（包含五个物体，测试环境同平面跨机械手抓取实验）的抓取成功率，实验结果表明本研究三维跨抓手算法显著优于前述二维框架以及测试基线算法，达到了 92.6% 至 100.0% 的抓取成功率。由于本文所提出的二阶段混合抓取框架解构了环境感知推理步骤和跨机械手抓取优化步骤，在面向自定义操作任务，特别是针对严重视觉遮挡的密集杂乱场景以及多机械手在场景中同时部署的抓取难题，其具备灵活的框架重构自由度。首先，为了改善单视角观测在复杂密集环境中的遮挡情况，本研究进一步设计了一个集成手内深度视觉传感器的抓取系统，成功实现 20 Hz 的动态抓取伺服。该系统通过实时更新深度图像的观测，对抓取目标周围的场景几何细节进行实时更新，进而完成机械手的在线伺服控制。基于 64 个丰富物体近千次的密集抓取结果表明，该动态伺服系统可有效提高三指及四指机械手抓取成功率及物体的单一抓取率。此外，本研究进一步开发了在单一视觉传感器系统中部署多个不同机械手的拓展算法框架，该拓展框架通过复用同一个场景-接触推理模块，同时为多个独立机械手抓取优化模块提供潜在接触特征，以实现更高的场景理解效率。上述两个拓展抓取框架充分展示了 SpaHybGen 具备优于基于解析模型或学习模型抓取算法的框架重构能力，为实现广泛的自定义机器人操作任务提供了新的架构及途径。

综上所述，本论文分别在二维及三维空间中建立了适用于跨机械手的学习-推理混合机器人抓取算法框架。该框架不仅实现了复杂场景驱动的手型无关数据集构建与学习方法，同时提出统一的通用机械手模型描述与跨机械手抓取优化策略，从而在复杂、噪声干扰严重的真实环境中实现最优的抓取通用性与成功率。此外，本研究成果首次实现了在深度杂乱环境中跨机械手的零样本泛化，并在平面与空间域中均取得了当前最优机械手泛化性能。在面向多个自定义场景的拓展验证中，所构建的混合抓取架构展现出在通用操作任务上灵活的结构可塑性。因此，本文研究成果能够以低成本传感器和极高的部署潜力，快速赋能商用及定制化机械手在密集杂乱场景中的智能操作，实现通用机器人手的多样化部署与快速推广应用。

Abstract

A long-standing goal in robotics is to create general-purpose agents that can learn foundational skills applicable across diverse hardware and environments. Centered on the field of prehensile manipulation, this vision is hindered by a fundamental mismatch: deep learning policies are data-efficient but hardware-specific, while analytical planners are hardware-adaptable but fail under perceptual uncertainty. The diverse hand designs consistently emerged for industrial and household use pose further challenges in designing grasp synthesis algorithms applicable for different hand hardware with varied joint types and actuation systems. This thesis aims to tackle these challenges by proposing a novel learning-optimization hybrid methodology capable of hand-generalizable grasp generation in dense clutter. The hybrid grasping paradigm is comprehensively developed across both planar (two-dimensional) and spatial (three-dimensional) domains, culminating in two grasping frameworks in the planar (HybridGen) and spatial (SpaHybGen) spaces. Both proposed architectures integrate hand-agnostic scene inference with hand-adaptive numerical optimization to enable grasping in densely cluttered environments, applicable to a wide range of customized robotic hands. In each grasping framework, to bridge the modeling divide between scene perception and hand hardware, an exclusive formulation of robot-environment interaction state, i.e., contact features, is explicitly formed as the shared motion representation between the applied scene and the adopted hardware. Given the established concept of contact states in both domains, two large-scale contact-feature databases are individually generated from existing grasping data to facilitate hand-agnostic scene-contact inferences, which are further integrated with hand-changeable grasp optimization

using the formulated universal hand models described with analytical formulation (planar) and unified robot description format (spatial). In the construction of the grasp optimization schemes, multiple conventional grasp criteria are re-established in coordination with the newly defined contact feature to enable successful and robust object gripping. Eventually, throughout extensive dataset-based and physical assessments, HybridGen and SpaHybGen advanced the state-of-the-art performance of contemporary planar and spatial grasping approaches constructed with model-based and learning-based methodologies, achieving consistent superior grasping success rates (94.3-98.0% across seven robotic hands ranging from 2 to 5 fingers in the spatial domain). Notably, the established hybrid designs enable zero-shot hand generalization: without any hardware-specific retraining, the systems successfully empowered distinct robotic hands with proficient grasping capabilities in the real world. Advanced validations further verified the framework’s capability of dynamic grasping over 20 Hz in dense clutter and multi-hand coordination for complex tasks. The validated hybrid modeling paradigm, decoupling sensing perception from robotic action through a shared interaction interface, provides a pathway toward reusable and adaptable manipulation intelligence, a key step for general-purpose robotics.