



面向复杂环境的机器人触觉感知算法研究综述

高春艳, 刘琦, 李满宏, 刘璇

引用本文:

高春艳, 刘琦, 李满宏, 等. 面向复杂环境的机器人触觉感知算法研究综述[J]. 智能系统学报, 2026, 21(4): 834–848.

GAO Chunyan, LIU Qi, LI Manhong, et al. Review of robot tactile perception algorithms for complex environments[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 834–848.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202510008>

您可能感兴趣的其他文章

强化学习稀疏奖励算法研究——理论与实验

Survey of sparse reward algorithms in reinforcement learning — theory and experiment
智能系统学报. 2020, 15(5): 888–899 <https://dx.doi.org/10.11992/tis.202003031>

记忆神经网络在机器人导航领域的应用与研究进展

Research progress and application of memory neural network in robot navigation
智能系统学报. 2020, 15(5): 835–846 <https://dx.doi.org/10.11992/tis.202002020>

多智能体分层强化学习综述

A survey on multi-agent hierarchical reinforcement learning
智能系统学报. 2020, 15(4): 646–655 <https://dx.doi.org/10.11992/tis.201909027>

触觉手势情感识别的超限学习方法

Extreme learning machine for emotion recognition of tactile gestures
智能系统学报. 2019, 14(1): 127–133 <https://dx.doi.org/10.11992/tis.201804029>

基于卷积神经网络的盲文音乐识别研究

Research on braille music recognition based on convolutional neural networks
智能系统学报. 2019, 14(1): 186–193 <https://dx.doi.org/10.11992/tis.201805002>

关于深度学习的综述与讨论

Overview on deep learning
智能系统学报. 2019, 14(1): 1–19 <https://dx.doi.org/10.11992/tis.201808019>

DOI: 10.11992/tis.202510008

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260403.0843.002>

面向复杂环境的机器人触觉感知算法研究综述

高春艳, 刘琦, 李满宏, 刘璇

(河北工业大学 机械工程学院, 天津 300401)

摘要: 为应对复杂环境挑战并提升机器人触觉感知精度与鲁棒性, 本文系统综述了近 5 年基于机器学习、深度学习和强化学习的典型算法, 涵盖卷积神经网络 (convolutional neural network, CNN)、图神经网络 (graph neural network, GNN)、长短期记忆网络 (long short-term memory network, LSTM)、多模态融合与主动探索。结果表明, 现有方法针对噪声、非规则接触、动态时序等复杂特性已取得显著进展且各具专长, 但仍面临数据稀疏、公开数据集不足、跨模态弱对齐及仿真迁移困难等共性瓶颈。结论可为低监督学习、标准数据集构建、统一多模态建模及高保真迁移研究提供参考。

关键词: 触觉感知; 复杂环境; 人工智能; 机器人; 深度学习; 强化学习; 多模态融合; 机器学习

中图分类号: TP242.2 **文献标志码:** A **文章编号:** 1673-4785(2026)04-0834-15

中文引用格式: 高春艳, 刘琦, 李满宏, 等. 面向复杂环境的机器人触觉感知算法研究综述 [J]. 智能系统学报, 2026, 21(4): 834-848.

英文引用格式: GAO Chunyan, LIU Qi, LI Manhong, et al. Review of robot tactile perception algorithms for complex environments[J]. CAAI transactions on intelligent systems, 2026, 21(4): 834-848.

Review of robot tactile perception algorithms for complex environments

GAO Chunyan, LIU Qi, LI Manhong, LIU Xuan

(School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: To address the challenges posed by complex environments and improve the accuracy and robustness of robotic tactile perception, this study systematically reviews representative machine learning-, deep learning-, and reinforcement learning-based algorithms developed over the past five years, covering convolutional neural network(CNN), graph neural network(GNN), long short-term memory network(LSTM), multimodal fusion, and active exploration. The results indicate that existing methods have made significant progress in handling complex characteristics such as noise, irregular contact, and dynamic temporal patterns, while demonstrating distinct methodological advantages. Nevertheless, common bottlenecks remain, including data sparsity, limited publicly available datasets, weak cross-modal alignment, and challenges in simulation-to-real transfer. These findings provide valuable insights for future research on low-supervision learning, standardized dataset construction, unified multimodal modeling, and high-fidelity transfer in robotic tactile perception.

Keywords: tactile sensing; unstructured environments; artificial intelligence; robots; deep learning; reinforcement learning; machine learning; multimodal systems

实现融合多模态信息的具身感知已成为新一代智能机器人的关键发展趋势, 作为具身感知不可或缺的组成部分, 机器人触觉感知通过识别物体纹理、刚度、温度等物理属性^[1-3]有效弥补视觉感知的局限^[4], 为机器人提供可靠的环境信息。触觉感知技术已经广泛应用于工业、医疗等领域^[5-6], 尤其在视觉受限或复杂环境中, 能够为机器人提供更加稳定的感知能力。

机器人的应用场景逐步从结构化环境推广至半结构化和非结构化的复杂环境, 需面对极端挑战因素, 给机器人触觉感知带来了噪声和数据标注困难、动态变化引起的信号时效性差以及传感器干扰等问题。为应对复杂环境下的感知挑战, 当前研究正从算法架构、学习范式与多模态融合等维度系统性推动触觉感知技术的创新发展。

近年来, 随着人工智能技术的迅速发展, 特别是深度学习中的卷积神经网络 (convolutional neural network, CNN)、图神经网络 (graph neural network, GNN)、长短期记忆网络 (long short-term

收稿日期: 2025-10-10. 网络出版日期: 2026-04-03.

基金项目: 河北省中央引导地方科技发展资金项目 (226Z1801G).

通信作者: 高春艳. E-mail: gcy@hebut.edu.cn.

©《智能系统学报》编辑部版权所有

memory network, LSTM)及Transformer等模型^[7-9]在触觉感知任务中得到了广泛应用,触觉感知的精度和效率得到了大幅提升。然而,复杂环境要求模型具备在未知场景下主动获取信息的能力,这超出了深度学习机制的局限。因此,基于强化学习的主动感知方法应运而生。本文系统综述了面向复杂环境的机器人触觉感知关键算法,对比分析了不同方法的优势与局限,并对研究热点和趋势进行展望。

1 复杂环境特征及算法需求分析

复杂环境^[10-12]是指触觉信号采集显著受外界干扰的场景,触觉数据在空间结构、时序演化、噪声水平与接触稳定性等方面的不确定性大幅增加。典型情形包括弱视觉或无视觉条件、非结构化/可形变接触、信号动态时序变化以及未知环境探索等^[11-12]。在此类环境下,机器人在物体识别、操控稳定性评估、抓取规划与人机交互等任务中面临系统性挑战。

传统机器学习方法,如支持向量机(support vector machine, SVM)、K近邻(K-nearest neighbors, KNN)、决策树与随机森林(random forest, RF)等,已在物体识别、表面粗糙度评估和滑移检测等任务中取得较好效果^[13-17]。然而,这类方法通常依赖人工特征设计,在复杂环境下暴露出显著局限,如图1所示。1)特征工程流程繁琐且可迁移性弱,难以覆盖复杂接触中的多源扰动与非线性变化;2)在数据规模增大时识别精度易趋于饱和,难以持续受益于数据增长;3)多模态信息融合多采用浅层或后融合策略,跨模态互补性利用不足,导致融合效率与鲁棒性受限。

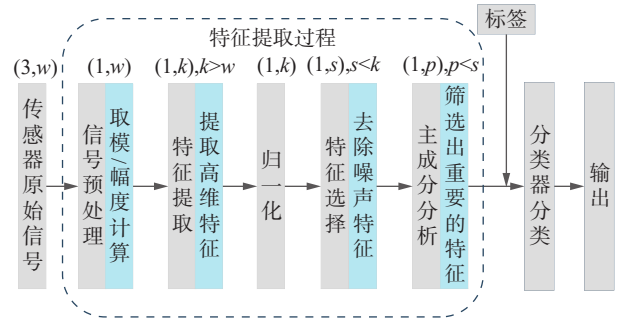


图1 人工特征提取流程

Fig. 1 Manual feature extraction flowchart

深度学习通过端到端表征学习^[18]能够从原始高维触觉数据中自动提取判别特征,减少对人工特征的依赖,从而更适合处理高维输入、复杂噪声以及时空依赖性。在复杂环境的关键问题上,不同深度模型基于其结构优势展现出针对性能力,CNN适用于组织为触觉时空张量的数据,可稳定提取多尺度局部模式;GNN能够显式建模触觉阵列中的空间关系与接触拓扑,强化对局部接触结构的理解;LSTM在接触过程的动态变化建模上表现突出,适合处理时序触觉信号;Transformer凭借自身机制展现出多模态融合的优势,能够在异构传感信息间实现更有效的全局关联建模。上述模型为复杂环境下的鲁棒感知提供了可行路径。

进一步地,基于强化学习的主动感知技术通过智能体与环境的闭环交互,使机器人能够主动选择探索动作以获取信息增益,同时在交互中实现策略自适应与自我调整。这一范式为复杂环境中提升感知的探索性、适应性与鲁棒性提供了新的技术方案。表1给出了针对不同复杂环境所提出的挑战及相应的算法类型。

表1 复杂环境挑战与代表算法

Table 1 Complex environmental challenges and representative algorithms

环境类型	难点	代表算法
非结构化/可形变接触	空间数据高噪声、稀疏且形变耦合	CNN、GNN
动态时序变化	时序非平稳、长短期依赖并存;瞬态信息对延迟敏感;	LSTM
弱视觉/需多模态融合	模态异构、时序不同步;融合噪声被放大	多模型协同、Transformer
未知环境/主动探索	信息依赖动作序列;奖励敏感;虚拟现实迁移技术难	Q学习(Q-learning, QL)、PPO(proximal policy optimization)等

2 针对复杂环境不同性能需求的深度学习触觉感知算法

2.1 基于卷积结构的触觉空间表征学习方法

CNN凭借其平移不变性、局部连接和参数共享等核心机制将触觉信号规整为规则张量表示,实现对高噪声、高维触觉数据的稳健特征提取与

高效推理^[7,19],展现出了好用、好训和好部署的优势。为了形式化刻画卷积结构在触觉张量上的局部连接与参数共享机制,可将1层卷积表示为

$$y_{ij}^{(k)} = \sigma \left(\sum_{c=1}^c \sum_{x=-\tau}^x \sum_{-r}^r w_{k,x}^{(k,k)} x_{1+\alpha,j+x}^{(k)} + b^{(k)} \right) \quad (1)$$

式中: $x_{i,j}^{(c)}$ 为规整后的触觉张量; $y_{i,j}^{(l,k)}$ 为第 l 层第

k 个特征图在空间位置的响应; $(u, v) \in [-r, r]^2$ 描述感受野的局部邻域, 体现局部连接; $\omega_{u,v}^{(l,k,c)}$ 和 $b^{(l,k)}$ 为卷积核权重和偏置, 在不同 (i, j) 位置共享, 体现参数共享; σ 为非线性激活函数。由式 (1) 可见, 卷积核 $\omega_{u,v}^{(l,k,c)}$ 在整个触觉阵列上复用, 使模型能够用有限参数捕捉局部压力/形变模式, 对高维、高噪声触觉观测具有良好的特征抽取与泛化能力。

针对多步接触中力/形变模式随时间演化的问题, Pastor 等^[20]将抓取过程中的二维压力图沿时间/力变化方向堆叠为 3D 触觉张量, 首次提出采用 3D CNN 网络架构, 该方法在保证网络结构简单、数据量有限的情况下, 较 2D CNN 获得更高的物体识别精度, 并且对不同抓取力度、不同接触轨迹具有更强的泛化能力, 图 2 对 3D CNN 模型进行了示意。

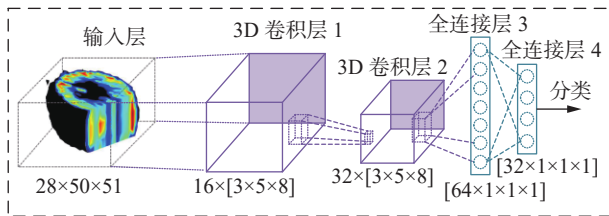


图 2 3D CNN 网络模型

Fig. 2 3D CNN network model diagram

Funabashi 等^[21]提出的形态特定卷积神经网络 MS-CNN(morphology specific convolutional neural networks)在输入 3D 数据的基础上, 依据传感器的物理拓扑结构分层组织卷积层, 有效避免了传统 CNN 将数据强制转换为矩形网格所带来的空间关系破坏问题, 该方法显著增强了对噪声的鲁棒性。在物体识别任务中准确率超过 95%, 并通过迁移学习后在新任务上仅需少量数据微调就保持较高识别率, 展现了对传感器损坏和新任务计算开销的容错性。另一方面, Park 等^[22]提出的 EIT-NN(electrical impedance tomography neural network)框架即使在传感器基底层受损的情况下, 模型仍保持 99.76%

的两点辨别性能, 定位误差仅为 (4.88 ± 1.63) mm, 也展现出对传感器损坏的高度容错能力。

在户外非结构化地形和高噪声环境中, Dog-Touch 框架^[23]利用安装于四足机器人足底的高密度触觉阵列, 将单步着地的压力分布编码为小尺寸触觉图像, 并采用单迭代时间仅为 12.6 ms 的轻量级 CNN 模型结合批量归一化(batch normalization)等正则化手段, 在仅 800 次自动踩踏采样的条件下, 对 8 种地面纹理取得 74.37% 的识别准确率, 其中线纹理识别率达 90%, 表明 CNN 能够从碎石、粉尘等复杂路面引起的局部压力纹理中有效提取判别性特征, 从而为复杂地形下的足端感知与自适应步态控制提供了关键依据。

Ma 等^[24]提出的双域(时间域和频率域)SE-CNN(squeeze-and-excitation CNN)触觉材料识别系统, 通过 3σ 异常点剔除(3σ criterion to eliminate outliers)、滑动平均滤波(moving average filter)和 Z-score 标准化(Z-score normalization)进一步提升了噪声鲁棒性。该系统基于 24 kHz 采样率与 14 通道信号融合, 为复杂环境中的大规模数据采集及半结构化接触问题提供了技术基础, 最终在 20 种材料识别任务中达到 99.375% 的精度。此外, 也有工作在保证识别精度的前提下对网络进行轻量化设计^[25], 使参数量降至原模型的约十分之一, 以适应嵌入式平台和在线感知的实时性要求。

综上, 研究者通过结合时空堆叠 3D 卷积、形态特定卷积以及通道重标定与轻量化等工程化设计, 在复杂接触环境中展现出各自的优势。对应地, 表 2 对本节典型 CNN 模型的输入信号尺寸、面对复杂环境的优势与不同实验任务性能评估结果进行了归纳, 由于各研究所用数据集、任务定义与评价指标不尽相同, 表 2 中的性能结果仅用于展示原文设定下的表现, 不作为严格横向对比依据。

表 2 CNN 模型的输入信号尺寸、优势及其性能评估结果

Table 2 Input signal dimensions, advantages, and performance evaluation results of CNN models

方案名称	输入信号尺寸	优势	性能评估
3D-CNN ^[20]	1400触点	噪声鲁棒性、小样本学习能力	物体识别准确率较2DCNN(TactNet6、VGG16-SVM等)提升约20%~35%
MS-CNN ^[21]	240触点+16个关节角度	传感器空间拓扑保持能力、迁移学习能力	物体与属性识别准确率较SVM、ResNet等模型提升约10%~20%
SE-CNN ^[24]	14通道时序信号	噪声鲁棒性、半结构化接触能力	材料类型识别准确率较单域CNN提升3%~5%, 参数量、响应时间、推理时间均优于单域CNN
EIT-NN ^[22]	重建图像分辨率24×24	传感器损坏抗性	单点压痕实验定位误差下降约30%~50%

同时需要指出的是, 这类方法普遍依赖将触觉阵列规整为规则栅格再进行卷积计算, 当传感单元拓扑高度不规则、接触区域持续变化或事件触觉信号呈现稀疏且强时序依赖时, 栅格化往往难以同时保留传感单元之间的真实连接结构, 从而在处理稀疏、噪声及动态不稳定接触模式时表现出建模能力不足, 这也正是引入 GNN 的关键动机。

2.2 基于图结构的触觉不规则拓扑建模方法

在复杂环境中, 外界扰动通常导致传感阵列的空间拓扑呈高度非规则性^[26], 且接触区域持续变化; 与此同时, 事件触觉信号在时间维度上具有高度稀疏性, 却仍表现出显著的时序依赖。传统基于 CNN 的方法依赖对数据进行规则栅格化输入, 这种操作往往无法同时保留传感单元之间真实的拓扑连接结构, 从而在处理稀疏、噪声及动态不稳定接触模式时表现出建模能力不足。相

比之下, 图结构触觉建模具有天然优势, 其将触觉传感器阵列表示为图结构, 并以每个传感单元作为节点 v , 为其赋予特征向量, 整个阵列的特征可表示为所有节点特征的拼接^[27]。

$$\mathbf{f}_G = [f_{v_1} \ f_{v_2} \ \cdots \ f_{v_N}] \in \mathbf{R}^{N \times D}$$

式中: f_{v_i} 为节点 v_i 的特征, N 为图中节点的个数。

节点更新基于邻域聚合机制, 其核心操作可为

$$f_{v^{(l)}} = \sum_{a^{(l)} \in N(v^{(l)})} \frac{1}{\sqrt{\deg(v^{(l)})} \cdot \sqrt{\deg(a^{(l)})}} (\mathbf{W} \cdot f_{a^{(l-1)}})$$

式中 $\deg(v^{(l)})$ 和 $\deg(a^{(l)})$ 分别为节点 v 及其邻居 a 的度。该操作先将层 $l-1$ 处的相邻节点 a 的特征 $f_a^{(l-1)}$ 用权重矩阵 $\mathbf{W}$ 进行变换, 然后在根据节点的度进行归一化之后进行线性组合。由此, 在保持传感器真实几何结构的同时, 模型能够自适应整合多尺度的局部与全局触觉信息, 从而提升在复杂环境中对稀疏、噪声和动态接触模式的表达能力。图3对图卷积神经网络进行了示例。

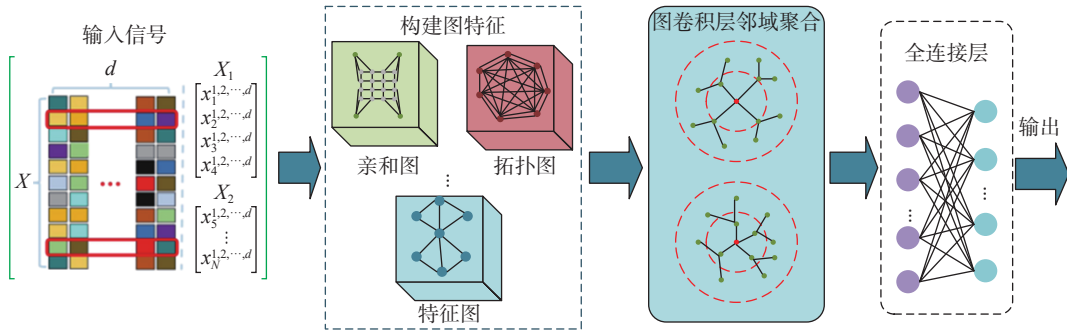


图3 图卷积网络流程

Fig. 3 Flowchart of graph convolutional network

围绕复杂环境下的结构不规则与信号多源异质性, MR-GCN(multireceptive graph convolutional network)^[28]通过同时构建拓扑图与亲和图, 并在多阶图卷积基础上引入注意力机制 (attention mechanism), 自适应融合不同图上的邻域信息, 从而在动态探索场景中增强对复杂触觉模式的表征能力。在物体识别和液体容器抓取稳定性检测任务中分别取得 92.05% 的准确率和 93.87% 的 F_1 分数。面向事件驱动触觉感知中数据高度稀疏、对噪声敏感的问题, GGT-SNN(graph learning and gaussian prior integrated spiking neural network)^[29]采用基于 M-tree(Minimum spanning tree) 与 Z-tree(Z score based tree) 的触觉图构建方式, 在保持图结构稀疏性的同时强化关键事件间的空间关系, 并在脉冲神经网络中引入高斯先验参数估计与高效的 LIF(leaky integrate and fire) 近似激活, 从统计角度缓解小样本和高噪声下的过拟合问题, 最终在

EvTouch-Objects 与 EvTouch-Containers 数据集上分别取得 92.36% 与 75.00% 的识别准确率。

针对“空间拓扑不规则+时间依赖跨多帧”的联合难题 AM-SGCN(adaptive multichannel spiking graph convolutional network)^[30]与 SNN-Atten-Res-GCN(spiking neural network based attention residual GCN)^[31]则从图-脉冲神经网络联合建模的角度进一步提升模型能力。前者在 GCN 基础上引入自适应多通道图卷积与注意力机制, 分别在拓扑图与特征图上建模并进行自适应融合, 同时结合基于脉冲序列学习的 TSSL-BP(temporal spike sequence learning via backpropagation) 算法, 显式刻画 LIF 神经元的时空依赖关系, 在事件触觉分类任务中实现超过 91% 的准确率。后者通过引入残差结构的 ResGCN 与图注意力机制缓解深层图卷积的过度平滑问题, 在突出关键接触区域的同时保持全局结构一致性, 并利用多层 LIF-SNN 对

触觉图时间序列进行编码,在检测性能显著提升的同时基本不增加迭代开销。

面向非结构化接触环境中触觉数据的拓扑不规则、稀疏高噪声与接触区域动态变化等问题,基于图结构的方法以“节点-边”直接刻画传感单元关系,避免栅格化造成的邻接关系破坏,并通过注意力/残差、多图融合与多跳聚合等机制提升

深层训练稳定性与跨区域依赖建模能力。表 3 进一步对本节典型 GNN 模型的输入信号尺寸、数据集、优势及其物体识别任务性能评估结果进行了归纳与对比。这类方法对构图策略与超参较敏感,动态图与多跳传播也会带来额外计算开销,在计算效率、工程生态成熟度以及训练稳定性等方面仍普遍弱于传统卷积网络。

表 3 GNN 模型的输入信号尺寸、数据集、优势及其性能评估结果

Table 3 Input signal dimensions, datasets, advantages, and performance evaluation results of GNN models

方案名称	输入信号	数据集	优势	性能评估
MR-GCN ^[28]	16触点+4个光学近接单元	自建数据集	复杂接触场景鲁棒性强、稳定性检测性能高	准确率、稳定性检测、 F_1 分数优于传统模型
GGT-SNN ^[29]	39触点	EvTouch-Objects、EvTouch-Containers	噪声鲁棒性、事件触觉动态场景表现优秀	准确率较TactileSGNet提升2.92%、10.83%
AM-SGCN ^[30]	39触点	EvTouch-Objects	复杂时空结构建模能力强、图扰动鲁棒性	准确率、精度、 F_1 分数均优于Tactile-SGNet
SNN-Atten-ResGCN ^[31]	39触点	EvTouch-Objects、EvTouch-Containers	不规则触觉拓扑适应性高	准确率、精度、召回率、 F_1 分数较传统模型提升约2%~5%

2.3 基于序列建模的触觉时序特征学习方法

在触觉时序信号处理中,循环神经网络(recurrent neural network, RNN)虽可用于序列建模,但难以有效捕捉长期依赖,易出现梯度消失或爆炸等问题^[32]。LSTM通过门控机制显式控制历史信息的保留与遗忘,显著缓解了梯度退化,增强了对长时依赖与非平稳序列的建模能力^[33],更适用于噪声干扰强、接触模式随动作动态变化的触觉序列。

在接触条件快速变化、信号噪声较强的复杂触觉场景中,LSTM通过对时序依赖关系的建模展现出优于静态模型的感知能力。以颗粒介质中目标定位为例,Chen等^[34]将三维定位解耦为“距离回归+方位分类+埋藏深度分类”3个子任务,并用3个相互独立的LSTM子模型分别学习,从而降低单一模型的学习难度并便于对各维度性能独立评估。实验显示当目标进入约45 mm的有效检测范围后,方向与深度相关子任务宏平均 F_1 分数(macro F_1 score)在接近目标时可达约0.7~0.9,距离回归误差亦显著下降,验证了多子任务的时序建模策略在复杂介质中的有效性。

进一步地,在滑动接触、连续读取或纹理识别等任务中,判别信息常集中于序列的少量关键片段,单纯依赖循环结构的“平均记忆”容易被冗余扰动稀释,因此注意力机制成为强化关键时间片的重要途径。AM-LSTM^[35]在LSTM中引入注意力以自适应突出关键帧,并显示夹爪闭合速度提升会使触觉响应更平滑、随机干扰减弱,从而

带来识别精度提升。在速度为20 r/min条件下,4类小球样本的识别准确率可超过90%。同时,Yan等^[36]将弧形软磁触觉传感器与注意力双向LSTM耦合,直接以三轴磁通密度序列输入并自动聚焦区分纹理的关键信号片段,在连续盲文阅读与60种布料纹理识别中分别达到约97%与99%准确率,进一步说明“注意力+双向时序”对动态触觉纹理辨识具有显著增益。图4对“注意力机制+双向LSTM”机制进行了示意。

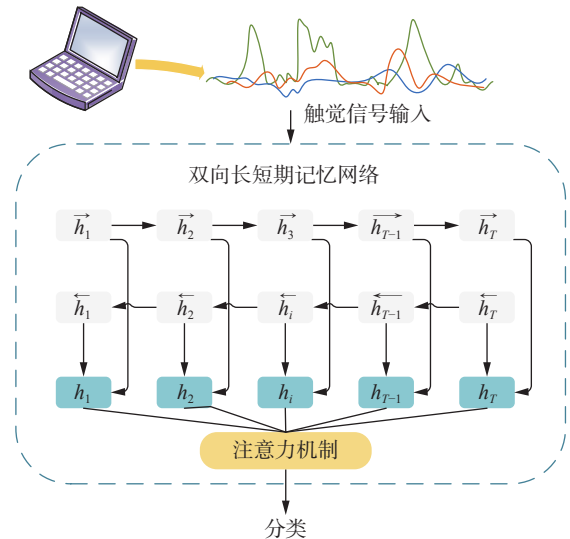


图 4 双向 LSTM 模型

Fig. 4 Bidirectional LSTM model

总体来看,基于序列建模的触觉时序特征学习主要面向复杂交互中的非平稳噪声与长短期依赖并存,LSTM通过门控机制提升长程依赖建模稳定性,多头/多子任务结构通过任务分解降低学

习难度,注意力与双向建模进一步实现对关键时间片段的聚焦并抑制冗余扰动。表4对典型方法的输入规模与采样频率、数据集及不同实验任务性能表现进行了归纳,由于各研究所用数据集、任务定义与评价指标不尽相同,表4中的性能结果

仅用于给出原文设定下的表现,不作为严格横向对比依据。同时也应注意,序列建模对切窗与采样敏感且单模态信息可能不完备,这促使后续方法进一步转向多模态融合,以在时间维度上联合利用低频趋势与高频细节,提高鲁棒性与泛化能力。

表4 LSTM模型的输入信号尺寸与采样频率、数据集及其性能评估结果

Table 4 Input signal dimensions, sampling frequencies, and performance evaluation results for the LSTM model

方案名称	输入信号尺寸+ 采样频率	数据集	性能评估
三头LSTM ^[34]	4触点×3轴+15 Hz	自建数据集(公开)	3D定位(距离回归、方向与深度分类)多触点三轴误差指标、宏平均 F_1 分数优于单轴输入
AM-LSTM ^[35]	25维×890帧+1000 Hz	自建OMT 13(未公开)	高速抓取触觉序列分类准确率可超过90%
双向LSTM ^[36]	3维×54帧+10 Hz	自建数据集(公开)	滑动触觉序列识别准确率优于传统方法(如DTW)

2.4 基于多模态融合的触觉综合感知方法

机器人触觉感知因其局部性与瞬时性而存在局限,在复杂环境中要实现稳定感知,亟需引入视觉、惯性、温度、磁场及语言等互补模态信息,图5为归纳的触觉多模态融合框架^[37-38]。通过多模态融合算法统筹利用多源信息,缓解模态异构、时间尺度不同步、环境与传感噪声放大以及多模态标注数据匮乏等^[39-41]落地瓶颈。基于深度学习的多模态融合方法可归纳为两类主流路径:一是以CNN、RNN及概率模型协同为代表的多模型协同路径;二是以Transformer为核心的统一多模态建模路径。

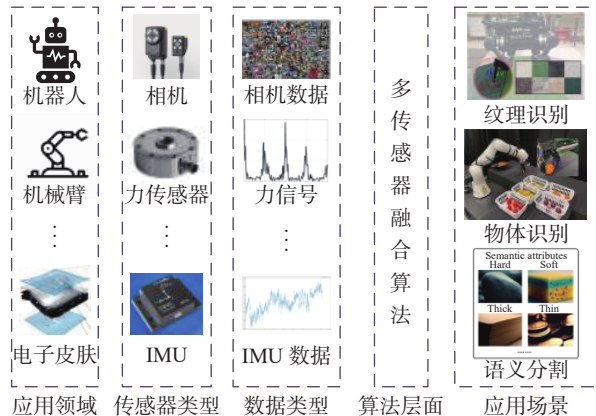


图5 多模态融合框架

Fig. 5 Multimodal fusion framework

2.4.1 多模型协同融合路径

复杂环境下的触觉多模型融合可按融合发生层级概括为3类:1)特征级融合(feature level fusion)在表征空间内显式建模跨模态交互以提升判别性^[40];2)决策级融合在后验/层级判定层组合多源信息以提高稳健性与可插拔性^[42];3)感知-决策级融合将对齐、不确定性建模与控制闭环耦合,以满足安全敏感与时延严苛场景的鲁棒需求^[43]。

1) 特征级融合

特征级融合通常对应早期或中期融合,在共享表示空间内显式建模跨模态相关性,并通过特征拼接、双线性/张量积或注意力等机制实现互补信息耦合,从而提升判别表征的可分性。其优势往往体现在“单模态证据不完备但仍存在可对齐线索”的场景,但前提是时空对齐与尺度归一较可靠。一旦存在异步采样、接触点随机性或噪声非平稳,早期融合更容易学到偶然共现的伪相关,导致跨场景泛化不稳,图6(a)对多模型特征级融合机制进行了示意。在同源多模态触觉中,Lü等^[44]将低频应力与高频率序列作为互补输入,在SA-LSTM^[44]中引入自注意力以突出时序关键片段,抑制机械臂引入的额外振动干扰、提升判别鲁棒性。该方法在粗糙度与硬度组合构成的6类材料上达到93.61%准确率。去除任一模态会显著降性能,验证了双模态融合的必要性。对于压力-摩擦强耦合的指尖触觉,Liu等^[45]以CNN抽取空间接触模式、LSTM建模时序演化,并以注意力强调关键时间片,实现压力与摩擦的高精度回归($R^2 > 0.95$)及细粒度类别识别,体现了同源信号中通过时空特征学习与注意力重加权实现多物理量解耦与下游识别的有效性。当模态来自异源传感器时,Ruan等^[46]提出的RFCT(ResNet50-IDTP-CBAM-TCN based visual-tactile fusion model)用于弱刚度目标的抓取状态检测,用ResNet50与触觉CNN分别提取视、触特征,采用增维张量积构造初始融合特征,同时编码模态固有属性与跨模态相关。考虑到外积会显著增加通道/空间冗余,进一步引入CBAM(convolutional block attention module)在通道与空间维进行选择抑制与强化。该方法在未见目标测试上达到精度82.89%、召回率82.07%、 F_1 分数81.65,表明“高阶交互+注意力抑冗余”有助于弱刚度目标的可分性与跨目标泛化。

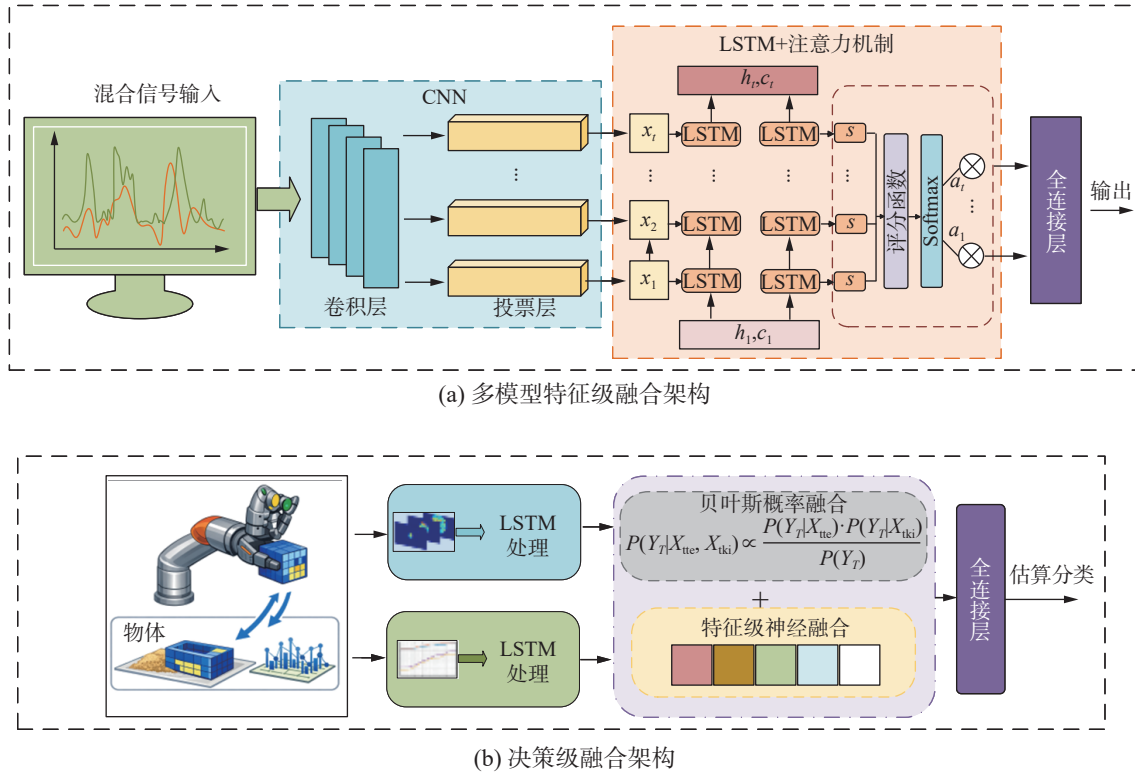


图 6 多模型特征级融合与决策级融合架构对比

Fig. 6 Comparison of multi-model feature-level fusion and decision-level fusion architectures

2) 决策级融合

决策级融合属于后融合,先由各模态独立推断并输出概率/置信度,再以投票、贝叶斯或层级级联综合证据,具备模块化与对缺失或失效模态的容错优势,适用于数据稀缺与分布漂移场景。但可能会出现模态间细粒度互补被延后利用、单模态系统性偏差会被融合继承等问题,图 6(b)对决策级融合进行了示意。代表工作有彭欢等^[47]将磁性触觉层与柔性压阻阵列集成为单一感知层以实现时空天然对齐,并采用 CNN-SVM-MLP 级联:CNN 表征压力分布、SVM 划分磁场特征、MLP 融合判别;对 24 类(含磁性/非磁性)物体识别精度达 99.42%,且在外界磁场干扰下保持稳定,体现出“结构一体化+级联判别”对电磁复杂环境的鲁棒性。Mao 等^[10]通过热-力复合传感获取多种材料与接触属性的时序信号,并以由粗到细的级联 LSTM 实现层级分类,在生活垃圾抓取中对未见物体识别率达 98.85%,说明层级推理可缓解姿态与接触点不确定导致的类内变化。Pastor 等^[48]进一步在后验概率层融合触觉与运动觉分类器,经过贝叶斯概率(Bayesian probability)融合后平均识别率 86.2%,高于单模态与神经融合,并呈现更低的训练划分敏感性,表明决策级后验融合在触觉数据稀缺时具有稳定优势。

3) 感知-决策级融合

感知-决策级融合将融合过程显式嵌入状态估计与控制闭环,通常包含时空对齐、不确定性评估、在线调权与控制输出以应对复杂环境中观测可靠性随时变化的情形。其优势在于可将融合收益直接转化为系统安全与控制精度,但链路长、工程耦合强,且对对齐与噪声建模误差敏感,误差累积可能降低闭环稳定裕度。Chi 等^[49]提出 HHFA-DBFL-PNN(hierarchical heterogeneous feature alignment-dynamic bayesian fusion layer-probabilistic neural network) 框架,HHFA 将多源触觉的时间尺度误差压缩至亚毫秒级,DBFL 基于熵-方差耦合不确定性在线调权,PNN 输出控制指令。在微创手术等场景实现约 0.15 mm 定位误差与约 96% 操作成功率,并在高湿与机械应变下维持约 40 dB 信噪比,体现了不确定性驱动融合对极端复杂环境的稳定性价值。与此相呼应,滤波式视触融合方法^[50]通过学习异方差噪声实现动态调权,并可结合主动信息增益提高数据效率,说明将融合嵌入状态估计与交互策略可进一步增强环境适应性。

表 5 对多模型协调方法的融合层级、代表模型性能评估及局限性进行了归纳。总体上,多模型协同融合沿“表征耦合(特征级)-后验集成(决策级)-闭环融合(感知-决策级)”演进,分别侧重提升判别性、增强迁移稳健性与保障闭环精度,

在复杂环境的不确定性与模态失效条件下形成互补。但是该方法仍受缺乏统一框架(需为不同模态分别设计特征提取、对齐与融合点)、采样率/时

延不一致与弱配对会放大对齐误差(早融合尤其易产生伪相关)等瓶颈制约。基于此,研究逐步转向以 Transformer 为代表的统一建模路径。

表5 多模型协同的融合层级、代表模型性能评估、局限性

Table 5 Integration levels, representative model performance evaluation, and limitations of multi-model collaboration

融合层级	代表性能评估	局限性
多模型特征级融合	RFCT ^[46] 较最优基线模型精度、召回率、 F_1 分数提升了6.72%、10.04%、9.47%	对齐与尺度归一性要求高,易产生伪相关。
多模型决策级融合	LSTM-Haptic-Fusion ^[48] 较触觉模型、神经模型准确率提升了5.6%、4.7%	单模态偏差可能被继承,细粒度互补延迟。
多模型感知-决策级融合	HHFA-DBFL-PNN ^[49] 的定位误差较最优基线下降了46.4%	对噪声误差敏感,闭环稳定性受限。

2.4.2 基于 Transformer 的多模态融合方法

针对多模型融合在统一表示与长程关联建模方面的局限,Transformer 通过统一的 token 化接口与自注意力机制(self attention)/交叉注意力(cross attention)为核心^[51],将“对齐-相关性建模-冗余抑制”纳入同一计算框架,其基本思路是各模态输入 $X^{(m)}$ 经模态投影映射到统一维度的 token 序列,并叠加位置与模态编码。

$$\tilde{H}^{(m)} = X^{(m)} W_{\text{emb}}^{(m)} + P^{(m)} + E^{(m)}$$

$$Z = [\tilde{H}^{(v)} \quad \tilde{H}^{(t)} \quad \tilde{H}^{(l)}]$$

式中: $W_{\text{emb}}^{(m)}$ 为模态投影; $P^{(m)}$ 、 $E^{(m)}$ 分别为位置与模态编码,用于保持时间结构与模态差异性。这一步在一定程度上缓解了多模型融合中需要为每一模态单独设计特征空间的问题。

之后在统一表示 Z 上施加跨模态自注意力、交叉注意力,可自适应模态间关联与重要性,其本质是以数据驱动动态权重替代人工预设或静态融合策略,从而在噪声或模态失效时抑制不可靠信息、增强关键模式与上下文表征。

$$Z' = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

1) 触觉模态内融合

在复杂接触过程中,触觉信号往往同时包含高频瞬态与低频稳态信息,且多变量在不同时间尺度上耦合。因子化多模态时空 Transformer(如 FMT(factorized multimodal transformer)^[52]、SVP-Transformer(shape-level variable-position Transformer)^[53])将触觉序列组织为统一的时空 token,并在自注意力中显式建模跨时间窗口的依赖关系,从而更稳定地捕捉接触阶段结构与长程关联。在 FBG(fiber bragg grating) 触觉弯曲传感场景中,该类方法在材料硬度-粗糙度识别与摩擦表面分类等任务上取得约 87%~93% 的平均准确率。同时模型体量可控制在约 10~12 MB,更适合嵌入式部署。该结果表明,Transformer 的时空注

意在“模态内”即可有效缓解复杂环境中触觉多时间尺度与长程依赖带来的表征瓶颈。图 7(a) 对因子化多模态时空 Transformer 进行了示意。

2) 跨模态融合

跨模态融合的关键困难在于模态异构与证据不完备,视觉可能受遮挡/反光影响,触觉受接触点随机性影响,而语言往往以类别或属性先验补足不可观测信息。DT-Transformer(double T: tactile and text)^[54] 将低分辨率阵列触觉通过插值与伪彩色编码构造图像 token,并引入由 BERT(bidirectional encoder representations from Transformers) 编码的属性词序列,通过解码器 cross-attention 将语义先验与触觉证据对齐。在 STAG 与 AUBO 数据集上,识别准确率分别达到 86% 和 95%。说明语言先验能够在弱监督与信息缺失条件下提升触觉识别的鲁棒性与泛化。进一步地,在视-触融合中,基于 TimeSformer/ViViT 的框架^[55] 将视频 RGB(red, green, blue) 与触觉等时序信号统一为时空 patch 序列,通过因子化时空自注意力分别建模空间结构与时间演化;在滑移检测与可变形目标抓取结果预测等复杂交互任务中,滑移安全损伤预测精度可达到 90% 以上,且对未见过的可变形物体的泛化性能优于 CNN+LSTM 等多模型基线。面向更强互补的 3 模态设置,TVT-Transformer^[56] 在视觉、触觉与文本 3 模态上采用“模态独立编码+Q/K/V 重组的跨模态自注意力”策略,将 3 模态统一到共享词元空间,在 MSDO 与 ObjectFolder2.0 这 3 模态数据集上的最高分类准确率分别达到约 99.7% 和 99.6%,显著高于任一单模态及任意双模态子网络。并通过 Adapter^[40] 与 LoRA^[57] 等参数高效微调技术,在标注样本仅占完整数据集数个百分点时仍能保持接近全参数微调的性能,部分缓解了复杂环境下多模态标注稀缺与分布迁移带来的工程瓶颈。图 7(b) 对跨模态 Transformer 进行了示意。

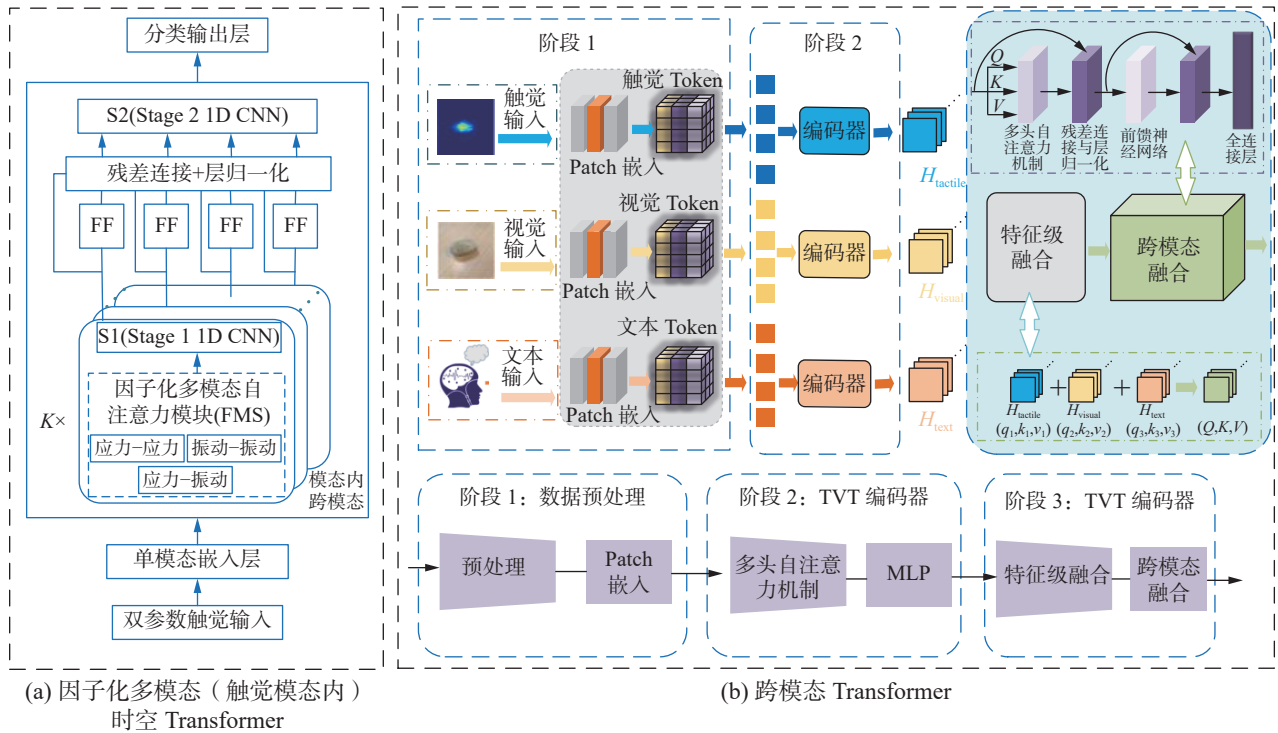


图 7 因子化多模态（触觉模态内）时空 Transformer 与跨模态 Transformer 架构对比

Fig. 7 Comparison of factorized multimodal (within the tactile modality) spatio-temporal transformer and cross-modal transformer architectures

表 6 对本节典型 Transformer 模型的融合模态、数据集和不同任务下的性能评估进行了归纳, 由于各研究所用数据集、任务定义与评价指标不尽相同, 表 6 中的性能结果仅用于展示原文设定下的表现, 不作为严格横向对比依据。综上, Transformer 统一建模路线结合跨模态注意力、时空建模与参数高效微调等工程化设计, 实现了在复杂

接触环境中的动态信息选择与可靠性驱动融合。相较于传统多模型拼装式框架, 该路线在一定程度上体现出对弱配对与异步采样的软对齐能力、对长时序交互与多阶段接触的依赖建模能力、对噪声非平稳与遮挡或失接触等模态退化的鲁棒性增强以及在小样本、跨场景迁移条件下借助预训练与轻量化适配提升泛化与部署可行性的潜力。

表 6 Transformer 的融合模态、数据集及性能评估

Table 6 Fusion modalities, datasets, and performance evaluation of Transformers

方案名称	融合模态	数据集	性能评估
FMT	应力、振动	自建数据集(未公开)	纹理识别准确率较MFN、CNN-TEMMA分别提升了5.79%、4.68%
TimeSformer/ ViViT	视觉、触觉	Slip detection	TimeSformer、ViViT较CNN + LSTM滑坡检测准确率分别提升了4.3%、2.0%
		自建Fruit grasping dataset (公开)	TimeSformer、ViViT较CNN + LSTM物体识别准确率分别提升了 5.2%、1.0%
DT-Transformer	触觉、文本	STAG	物体识别准确率较Transformer(仅触觉)、ResNet分别提升了16.49%、24.62%
		AUBO	物体识别准确率较CNN提升了8.03%
TVT-Transformer	视觉、触觉、文本	自建MSDO(公开)	物体识别准确率较现有最优方法提升了4.12%
		ObjectFolder2.0	物体识别准确率较现有最优方法提升了 2.16%

尽管多模态融合提升了复杂环境下的触觉感知能力, 但多模态算法仍受到为覆盖多接触状态与环境扰动需大量多模态数据, 且注意力或多分支结构带来较高计算与存储开销。而且, Trans-

former 统一 token 化并不是完全解决了异步采样、空间错位等弱配对问题, 缺乏显式物理约束导致跨场景稳定性不足。并且模态失效与长尾事件下常以启发式调权或注意力权重近似可靠性, 即使

能给出不确定性估计,也往往难以保证其可靠校准,从而难以在控制环中形成可执行的风险边界。

3 基于强化学习的触觉主动探索技术

在实际应用中,机器人常被要求在视觉受限、几何先验欠缺甚至完全未知的环境中执行抓取、检测和结构探查等任务,例如管道检测、狭窄通道内操作以及黑暗或被遮挡场景下的操作^[58-59]。此类场景中,触觉信息具有显著的局部性和动作

依赖性,感知结果不仅取决于传感位置,还高度依赖于接触方式与时序。因此,如何在缺乏精确几何模型条件下,通过主动规划触觉动作并依据反馈自适应调整策略,成为触觉感知领域的核心问题之一。将触觉探索建模为马尔可夫决策过程(Markov decision process, MDP),利用强化学习在“状态-动作-回报”闭环中优化策略,为构建统一的触觉主动感知框架提供了自然途径^[60-61]。图8给出了以Q-learning为代表的典型强化学习流程。

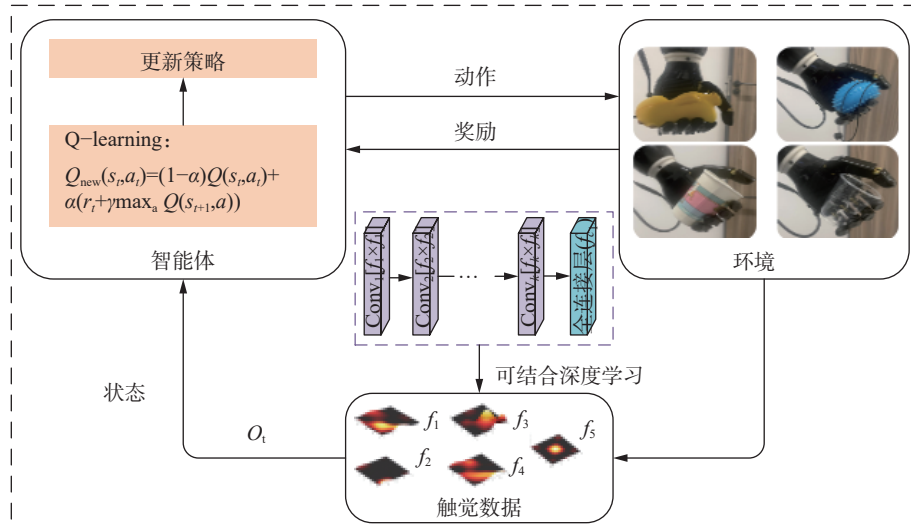


图8 基于Q-learning的强化学习框架

Fig. 8 Reinforcement learning framework based on Q-Learning

早期工作 Fler 等^[62]提出的 HAM (haptic attention model), 可以视作触觉主动感知在强化学习框架下的典型示例。其将一次“触觉一瞥”视为一个动作, 状态由若干次触觉观测的历史信息编码而成, 通过策略梯度 (REINFORCE) 直接优化“下一次触摸的位姿选择”。在受控仿真环境中, 该方法在只允许少量接触的条件下即可获得约 90% 的物体类别识别率, 随着触觉次数增加, 准确率接近 100%, 明显优于随机探索与无记忆的基线策略。这一结果说明, 即便触觉观测具有强局部性, 通过强化学习仍可以学习到有效的动作序列, 用极少的探索完成识别任务。不过, HAM 主要针对简单刚体和有限类别的分类问题, 其对复杂三维结构和更高层操作任务的适用性仍有限。Shahidzadeh 等^[58]提出的 AcTEExplore 则面向三维几何未知的表面探索, 该方法将单指触觉在物体表面上的微小位姿变化直接作为动作空间, 状态由当前及近期的触觉观测构成, 奖励函数同时考虑“保持稳定接触”与“尽量覆盖未探索区域”。在此设定下, 策略通过 PPO (proximal policy optimization) 学习如何在有限步长内最大化表面覆盖率。实验表明, 即

便仅在简单几何体上进行训练, 该策略仍能在多种未 YCB (yale-CMU-berkeley) 物体上实现约 96% 的表面覆盖 IoU (intersection over union), 表现出一定的泛化能力。与 HAM 相比, AcTEExplore 更关注“为后续任务建立几何先验”的能力, 但同样主要在单物体、刚体以及仿真环境中验证。

与上述“以几何探索为主”的方法不同, Duan 等^[63]提出的 FSTR-Net 将强化学习与 sim-to-real 迁移结合, 直接面向线缆穿管等狭窄空间操作任务。首先通过特征级域适应, 将仿真与真实的触觉信号对齐, 再将对齐后的触觉与视觉信息统一作为强化学习的状态输入; 动作则对应线缆推进与姿态调整等操作; 奖励函数反映穿管成功与否及过程中的接触质量。在此基础上, 采用多阶段 SAC (soft actor critic) 进行策略学习, 使策略能够在复杂接触条件下保持稳定插入。实验显示, 经由这种“域适应+强化学习”的流程训练得到的策略, 以约 95.8% 的成功率完成真实线缆穿管任务, 说明强化学习在处理带有显著接触不确定性的精细操作问题上具有现实应用潜力。在典型工业装配场景中, Tang 等^[64]进一步将多模态感知与

强化学习结合,用于不确定环境下的插销孔任务。其系统在 Gazebo 中构建 UR10 装配平台,综合利用 RGB 图像、深度图、末端姿态以及力/力矩信息。首先通过自编码器结构对视觉与位姿数据进行融合压缩,获得低维多模态潜在表示;随后将该潜在向量与六维力/力矩信号拼接,作为 SAC 策略网络的状态输入,动作为末端位置增量,奖励函数则综合孔口距离与插入深度。仿真结果表明,在存在较大初始姿态偏差和视觉/力觉噪声的条件下,所学策略在训练场景(方孔)中能保持近乎 100% 的装配成功率,并能在圆孔、三角孔和六边孔等未见几何上保持较高成功率,体现出多模态融合与连续控制强化学习在典型插销孔装配任务中的鲁棒性和一定几何泛化能力。

多指抓取则进一步将问题扩展到多接触点、多关节以及软硬不同材料的综合作用。Qi 等^[65]提出的 RLMP (RL-based multimodal perception) 框

架,重点探索在此类场景下如何构造合适的状态空间及学习稳定策略。该系统先通过数据手套采集人手示教轨迹,用于初始化机械手的抓取姿态;在此基础上,将指尖压力阵列与关节本体感知组合成离散化状态,动作对应各指关节的调整方式,奖励同时考虑物体位移、滑落和握持稳定性等因素。其分别采用 Q-learning 和 SARSA(state action reward state action) 对策略进行更新,并对比分析了两种方法在不同折扣因子下的表现:前者具有更强的探索性,适合需要频繁调整握力的柔软物体;后者更新更保守,在刚性物体抓取中表现出更好的稳定性。此外,RLMP 还使用深度网络对触觉矩阵进行物体识别,识别准确率约 98.5%,为在视觉受限条件下的抓取提供了有益补充。为便于梳理强化学习在触觉主动探索中的应用情况,表 7 对典型方法的仿真环境、算法类别及主要特点进行了归纳比较。

表 7 强化学习案例的仿真环境、算法类别及背景意义

Table 7 Simulation environments, algorithm categories, and contextual significance in reinforcement learning case studies

方案名称	仿真环境	算法类别	算法背景与特点
HAM ^[62]	Gazebo	策略梯度(REINFORCE)	直接优化策略处理连续/复杂动作空间
AcTEExplore ^[58]	Isaac Gym	改进PPO	结合策略梯度与价值函数; Actor负责执行策略, Critic负责评价
FSTR-Net ^[63]	MuJoCo	SAC	最大熵强化学习、适合连续动作控制
RLMP ^[65]	Unity3D	时序差分(SARSA、Q-learning)	无模型, 在线高效学习

4 问题与展望

4.1 问题

尽管深度学习、强化学习与多模态建模推动了触觉感知的进展,但在复杂环境约束下,现有方法仍存在以下关键局限:

1) 触觉数据标注困难导致样本稀疏与分布失衡。在复杂环境对触觉数据的标注质量(如细粒度语义、动态状态划分)提出了更高要求,而这又显著加剧了标注难度。触觉信号本身具有时序非平稳、高噪声与局部性强等特点,其精准标注严重依赖领域专家对多模态信息的协同判读,成本极高。同时,关键状态(如微滑移、隐蔽碰撞)在自然交互中呈现长尾分布,在有限标注预算下难以被充分覆盖,最终导致训练样本稀缺且分布失衡,严重制约模型的泛化与鲁棒性。

2) 公开数据集较少且泛化场景覆盖有限,现有数据集规模难以满足复杂环境需求。当前触觉公开数据集不仅在数量上稀缺,更关键的是其覆盖的场景、任务与条件较为单一,多基于受控实验室环境构建。这导致现有数据难以充分反映真

实复杂环境中存在的多源干扰、动态变化及任务多样性,模型仅能在受限场景下取得有限性能,无法直接推广至开放环境下的实际应用。

3) 跨模态融合种类有限,且仍受弱配对、异步与尺度差异制约。现有跨模态触觉研究多集中于视觉、力觉等相近模态,对声觉、嗅觉等高差异模态的融合仍显薄弱。不同模态间在采样率、延迟与语义层级上的不一致性,导致时序对齐困难;触觉信号的局部高噪特性与高层语义模态差异显著,在弱配对或无对齐标注的情况下,模型易学到虚假关联,难以兼顾跨模态语义一致性与触觉表征的物理合理性。

4) 强化学习主动感知仿真-现实差异显著。主动感知算法研究存在仿真建模理想化的问题^[66-67],触觉仿真在接触力学、材料非线性及传感器噪声等方面存在简化进而导致仿真数据与真实信号在统计分布上存在差异,导致策略迁移到现实世界性能下降。

4.2 展望

面向上述问题,未来面向复杂环境触觉感知的研究可从以下方向入手:

1)发展面向触觉特性的低监督表征学习框架。通过自监督学习(如对比学习^[68]、掩码建模^[69])从大规模未标注触觉数据中提取鲁棒表征,结合半监督与弱监督方法(如一致性正则化^[70]、伪标签学习),在少量标注下提升对长尾关键状态的学习效率,缓解标注稀疏与分布失衡问题。

2)推动多机构协作建立面向复杂环境的公开数据集,并建立跨场景、跨硬件的标准化数据集与评测协议,明确泛化评估设置,为算法研究与性能比较提供可靠基础。

3)扩展跨模态对齐机制并增强触觉中心统一表征。在现有多模态融合基础上,引入语言、声音等高层次模态,通过生成对抗网络^[71-72]、稀疏编码^[73-74]与哈希算法^[75-76]等技术缓解模态间结构差异;结合参数高效微调方法,提升模型在引入新模态时的训练与部署效率。

4)提升触觉仿真保真度并推动可迁移强化学习方法。通过高精度材料建模、接触动力学仿真及传感器噪声注入,增强触觉仿真的物理真实性。研究重点正转向开发更高效的域适应方法,例如特征级域适应与对比学习,以降低对大量真实标注数据的依赖。

5 结束语

本文围绕复杂环境下机器人触觉感知这一关键问题,对近年来基于机器学习、深度学习及强化学习的触觉感知算法进行了系统梳理与分析。通过从复杂环境特征出发,重点总结了卷积神经网络、图神经网络、时序建模方法、多模态融合技术以及主动触觉感知策略在提升感知精度、鲁棒性与泛化能力方面的研究进展,并对不同方法在实际应用中的优势与局限进行了对比讨论。综述结果表明,随着深度学习模型结构的不断演进以及多模态与主动感知范式的引入,机器人触觉感知在复杂、非结构化环境中的应用潜力正持续增强。

同时也应看到,受限于触觉数据获取与标注成本高、公开数据集规模和场景覆盖不足、跨模态协同与仿真到现实迁移困难等因素,现有方法距离大规模工程化应用仍存在一定差距。未来研究有必要在低监督学习、大规模标准化数据集构建、高效多模态统一建模以及高保真触觉仿真与可迁移强化学习等方向进一步深入探索。通过算法、数据与系统层面的协同发展,有望推动机器人触觉感知技术在真实复杂环境中的可靠落地,为新一代具身智能机器人的感知与决策能力提供坚实支撑。

参考文献:

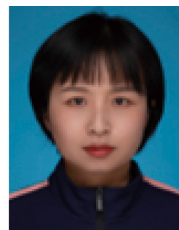
- [1] 孙世政, 何江, 秦鸿宇, 等. 基于 FBG 的机器手指尖触觉感知研究[J]. *仪器仪表学报*, 2024, 45(5): 72-81.
SUN Shizheng, HE Jiang, Qin Hongyu, et al. Research on tactile perception of machine fingertip based on FBG[J]. *Chinese journal of scientific instrument*, 2024, 45(5): 72-81.
- [2] YI Zhengkun, ZHANG Yilei, PETERS J. Bioinspired tactile sensor for surface roughness discrimination[J]. *Sensors and actuators A: physical*, 2017, 255: 46-53.
- [3] SUN Shizheng, HE Shenggang, HAN Yu, et al. Soft bionic skin with touch-slip and thermal perception using double-layer distributed FBG sensing array[J]. *IEEE sensors journal*, 2022, 22(5): 4096-4105.
- [4] 赵恒, 陈清森, 李成钢, 等. 融合机器视觉和触觉的机器人自动接线方法研究[J]. *计算技术与自动化*, 2022, 41(4): 84-90.
ZHAO Heng, CHEN Qingmiao, LI Chenggang, et al. Research on robot automatic wiring method integrating machine vision and tactile sense[J]. *Computing technology and automation*, 2022, 41(4): 84-90.
- [5] SUN Jia, ZHANG Qifeng, LU Yu, et al. A review of touching-based underwater robotic perception and manipulation[J]. *Machines*, 2025, 13(1): 41.
- [6] MAZZEO A, AGUZZI J, CALISTI M, et al. Marine robotics for deep-sea specimen collection: a systematic review of underwater grippers[J]. *Sensors*, 2022, 22(2): 648.
- [7] GANDARIAS J M, GARCIA-CEREZO A J, GOMEZ-DE-GABRIEL J M. CNN-based methods for object recognition with high-resolution tactile sensors[J]. *IEEE sensors journal*, 2019, 19(16): 6872-6882.
- [8] CONTE ALCARAZ J, MOGHADDAMNIA S, PEISSIG J. Efficiency of deep neural networks for joint angle modeling in digital gait assessment[J]. *EURASIP journal on advances in signal processing*, 2021(1): 10.
- [9] LIU Chongyu, LIU Hong, CHEN Hu, et al. Touchformer: a transformer-based two-tower architecture for tactile temporal signal classification[J]. *IEEE transactions on haptics*, 2024, 17(3): 396-404.
- [10] MAO Qian, ZHU Rong. Enhanced robotic tactile perception with spatiotemporal sensing and logical reasoning for robust object recognition[J]. *Applied physics reviews*, 2024, 11(2): 021424.
- [11] HUANG Jinrong, GUO Yuchen, JIANG Yongchang, et al. Recent advances and future prospects in tactile sensors for normal and shear force detection, decoupling, and applications[J]. *Journal of semiconductors*, 2024, 45(12): 121601.
- [12] PARK M, BOK B G, AHN J H, et al. Recent advances in

- tactile sensing technology[J]. *Micromachines*, 2018, 9(7): 321.
- [13] 张宪民, 王浩楠, 黄沿江. 机器人抓取对象硬度触觉感知研究[J]. *机械工程学报*, 2021, 57(23): 12–20.
ZHANG Xianmin, WANG Haonan, HUANG Yanjiang. Research on robotic grasping object hardness perception based on tactile sensing[J]. *Journal of mechanical engineering*, 2021, 57(23): 12–20.
- [14] 朱家辰. 基于强化学习的模块化软硬结合的仿生象鼻抓取器设计与控制研究[D]. 上海: 上海师范大学, 2025.
ZHU Jiachen. Research on the design and control of a modular soft-hard hybrid bionic elephant trunk gripper based on reinforcement learning[D]. Shanghai: Shanghai Normal University, 2025.
- [15] LI Tianliang, ZHANG Ao, DU Mingchang, et al. A fingertip optical fiber composite sensor with conformal design for robotic perception of tactile force[J]. *ASME transactions on mechatronics*, 2025, 30(2): 1523–1535.
- [16] LECHUZ-SIERRA J G, MARTIN A E H, SUNDARAM A M, et al. Bayesian optimization for robust robotic grasping using a sensorized compliant hand[J]. *IEEE robotics and automation letters*, 2024, 9(11): 10503–10510.
- [17] ZHANG Xingxing, LI Shaobo, YANG Jing, et al. Tactile perception object recognition based on an improved support vector machine[J]. *Micromachines*, 2022, 13(9): 1538.
- [18] RUSK N. Deep learning[J]. *Nature methods*, 2016, 13(1): 35.
- [19] CAO Lele, SUN Fuchun, LIU Xiaolong, et al. End-to-end convNet for tactile recognition using residual orthogonal tiling and pyramid convolution ensemble[J]. *Cognitive computation*, 2018, 10(5): 718–736.
- [20] PASTOR F, GANDARIAS J M, GARCÍA-CEREZO A J, et al. Using 3D convolutional neural networks for tactile object recognition with robotic palpation[J]. *Sensors*, 2019, 19(24): 5356.
- [21] FUNABASHI S, GANG Yan, FEI Hongyi, et al. Tactile transfer learning and object recognition with a multifingered hand using morphology specific convolutional neural networks[J]. *IEEE transactions on neural networks and learning systems*, 2024, 35(6): 7587–7601.
- [22] PARK H, PARK K, MO S, et al. Deep neural network based electrical impedance tomographic sensing methodology for large-area robotic tactile sensing[J]. *IEEE transactions on robotics*, 2021, 37(5): 1570–1583.
- [23] WEERAKKODI MUDALIGE N D, NAZAROVA E, BABATAEV I, et al. DogTouch: CNN-based recognition of surface textures by quadruped robot with high density tactile sensors[C]//2022 IEEE 95th Vehicular Technology Conference. Helsinki: IEEE, 2022.
- [24] MA Feihong, LI Yuliang, CHEN Meng, et al. A data-driven robotic tactile material recognition system based on electrode array bionic finger sensors[J]. *Sensors and actuators A: physical*, 2023, 363: 114727.
- [25] LI Qian, DU Zhijiang, LIU Feng, et al. Tactile perception for surgical status recognition in robot-assisted laminectomy[J]. *IEEE transactions on industrial electronics*, 2022, 69(11): 11425–11435.
- [26] GU Fuqiang, SNG W, TAUNYAZOV T, et al. TactileSGNet: a spiking graph neural network for event-based tactile object recognition[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas: IEEE, 2020.
- [27] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016–09–09)[2025–09–16]. <https://arxiv.org/abs/1609.02907>.
- [28] LIAO Junjie, XIONG Pengwen, LIU P X, et al. Enhancing robotic tactile exploration with multireceptive graph convolutional networks[J]. *IEEE transactions on industrial electronics*, 2024, 71(8): 9297–9308.
- [29] YANG Jing, YU Zukun, LI Shaobo, et al. GGT-SNN: graph learning and Gaussian prior integrated spiking graph neural network for event-driven tactile object recognition[J]. *Information sciences*, 2024, 677: 120998.
- [30] YANG Jing, LIU Tingqing, REN Yaping, et al. AM-SGCN: tactile object recognition for adaptive multichannel spiking graph convolutional neural networks[J]. *IEEE sensors journal*, 2023, 23(24): 30805–30820.
- [31] 吴培良, 林为梁, 毛秉毅, 等. 一种基于触觉信息的脉冲图注意力残差卷积物体检测算法[J]. *控制与决策*, 2023, 38(9): 2537–2544.
WU Peiliang, LIN Weiliang, MAO Bingyi, et al. An object detection algorithm based on convolution of attention residuals of pulse graph based on tactile information [J]. *Control and decision*, 2023, 38(9): 2537–2544.
- [32] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE transactions on neural networks*, 1994, 5(2): 157–166.
- [33] SHERSTINSKY A. Fundamentals of recurrent neural network(RNN) and long short-term memory(LSTM) network[J]. *Physica D: nonlinear phenomena*, 2020, 404: 132306.
- [34] CHEN Zhengqi, VERSACE E, JAMONE L. 3D localization of objects buried within granular material using a distributed 3-axis tactile sensor[C]//2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. Abu Dhabi: IEEE, 2024.
- [35] XU Zhe, YI Wei, CHEN Muxin, et al. Robotic tactile recognition system based on AM-LSTM model[C]//2021 IEEE International Conference on Robotics and Biomimetics. Sanya: IEEE, 2021.

- [36] YAN Youcan, HU Zhe, SHEN Yajing, et al. Surface texture recognition by deep learning-enhanced tactile sensing[J]. *Advanced intelligent systems*, 2022, 4: 2100076.
- [37] 陆军, 赵颖然, 鲁林超. 基于多模态融合的三维目标检测方法研究[J]. *智能系统学报*, 2025, 20(5): 1167–1177.
- LU Jun, ZHAO Haoran, LU Linchao. Research on 3D object detection based on multi-modal fusion[J]. *CAAI transactions on intelligent systems*, 2025, 20(5): 1167–1177.
- [38] 王路遥, 王凤随, 闫涛, 等. 结合多尺度特征与混淆学习的跨模态行人重识别[J]. *智能系统学报*, 2024, 19(4): 898–908.
- WANG Luyao, WANG Fengsui, YAN Tao, et al. Cross-modal person re-identification combining multi-scale features and confusion learning[J]. *CAAI transactions on intelligent systems*, 2024, 19(4): 898–908.
- [39] WU Renjie, WANG Hu, CHEN H T, et al. Deep multimodal learning with missing modality: a survey [EB/OL]. (2024–09–12)[2025–09–16]. <https://arxiv.org/abs/2409.07825>.
- [40] YANG Fengyu, FENG Chao, CHEN Ziyang, et al. Binding touch to everything: learning unified multimodal tactile representations[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024.
- [41] DING Zihao, CHEN Guodong, WANG Zhenhua, et al. Adaptive visual–tactile fusion recognition for robotic operation of multi-material system[J]. *Frontiers in neurorobotics*, 2023, 17: 1181383.
- [42] YI Xiongfeng, WANG Furui, ZUO Wenyu, et al. Robotics assisted smart-touch pipeline inspection[J]. *International journal of intelligent robotics and applications*, 2021, 5(3): 326–336.
- [43] MURALI P K, DUTTA A, GENTNER M, et al. Active visuo-tactile interactive robotic perception for accurate object pose estimation in dense clutter[J]. *IEEE robotics and automation letters*, 2022, 7(2): 4686–4693.
- [44] LÜ Chengang, DAI J, DU Yanxia, et al. Tactile sensing system based on FBG and SA-LSTM[J]. *Sensors and actuators A: physical*, 2025, 388: 116483.
- [45] LIU Xueqian, LI Yuanyue, LI Yiang, et al. Bionic fingerprint tactile sensor with deep learning-decoupled multimodal perception for simultaneous pressure-friction mapping[J]. *Advanced functional materials*, 2025, 35(49): e06158.
- [46] RUAN Wenjun, ZHU Wenbo, ZHAO Zhijia, et al. RFCT: multimodal sensing enhances grasping state detection for weak-stiffness targets[J]. *Mathematics*, 2023, 11(18): 3969.
- [47] 彭欢, 熊鹏文, 张宇, 等. 面向机器人交互识别的抗干扰多模态柔性电子皮肤[J]. *机器人*, 2025, 47(1): 55–63.
- PENG Huan, XIONG Pengwen, ZHANG Yu, et al. Interference-resistant multimodal flexible electronic skin for robotic interactive recognition[J]. *Robot*, 2025, 47(1): 55–63.
- [48] PASTOR F, GARCIA-GONZALEZ J, GANDARIAS J M, et al. Bayesian and neural inference on LSTM-based object recognition from tactile and kinesthetic information[J]. *IEEE robotics and automation letters*, 2021, 6(1): 231–238.
- [49] CHI Xinrui, GUO Zhanbin, CHENG Fu. A probabilistic neural network-based bimanual control method with multimodal haptic perception fusion[J]. *Alexandria engineering journal*, 2025, 127: 892–919.
- [50] DUTTA A, BURDET E, KABOLI M. Predictive visuo-tactile interactive perception framework for object properties inference[J]. *IEEE transactions on robotics*, 2025, 41: 1386–1403.
- [51] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach: NIPS Foundation, 2017.
- [52] LÜ Chengang, DAI J, DU Yanxia. Dual-parameter bionic tactile sensing system[J]. *IEEE internet of things journal*, 2025, 12(11): 16777–16788.
- [53] LÜ Chengang, LI Lin, ZHANG Ze, et al. FBG tactile sensing system based on SVP-transformer for material classification[J]. *IEEE sensors journal*, 2025, 25(3): 5370–5379.
- [54] QIU Shengjie, LI Baojiang, WANG Xichao, et al. DT-transformer: a text-tactile fusion network for object recognition[J]. *IEEE transactions on haptics*, 2025, 18(1): 164–174.
- [55] HAN Yunhai, YU Kelin, BATRA R, et al. Learning generalizable vision-tactile robotic grasping strategy for deformable objects via transformer[J]. *ASME transactions on mechatronics*, 2025, 30(1): 554–566.
- [56] LI Baojiang, LI Liang, WANG Haiyan, et al. TVT-transformer: a tactile-visual-textual fusion network for object recognition[J]. *Information fusion*, 2025, 118: 102943.
- [57] HU E J, SHEN Yelong, WALLIS P, et al. LoRA: low-rank adaptation of large language models[EB/OL]. (2021–06–17)[2025–09–16]. <https://arxiv.org/abs/2106.09685>.
- [58] SHAHIDZADEH A H, YOO S J, MANTRIPRAGADA P, et al. AcTEExplore: active tactile exploration on unknown objects[C]//2024 IEEE International Conference on Robotics and Automation. Yokohama: IEEE, 2024.
- [59] CHEN Rui, WANG Huigang, WANG Haoji, et al. A motion-sensing integrated soft robot with triboelectric nanogenerator for pipeline inspection[J]. *Advanced intelligent systems*, 2025, 7(6): 2400643.

- [60] 沈健, 宋智功. 基于深度学习的双臂系统协同控制综述[J/OL]. 控制工程, (2025-09-09)[2025-10-03]. <https://doi.org/10.14107/j.cnki.kzgc.20250283>.
SHEN Jian, SONG Zhigong. A review of cooperative control for dual-arm systems based on deep learning control[J/OL]. Engineering of China, (2025-09-09)[2025-10-03]. <https://doi.org/10.14107/j.cnki.kzgc.20250283>.
- [61] OLIFF H, LIU Ying, KUMAR M, et al. Reinforcement learning for facilitating human-robot-interaction in manufacturing[J]. *Journal of manufacturing systems*, 2020, 56: 326–340.
- [62] FLEER S, MORINGEN A, KLATZKY R L, et al. Learning efficient haptic shape exploration with a rigid tactile sensor array[J]. *PLoS one*, 2020, 15(1): e0226880.
- [63] DUAN Boyi, QIAN Kun, LIU Aohua, et al. Visual-tactile learning of robotic cable-in-duct installation skills[J]. *Automation in construction*, 2025, 170: 105905.
- [64] TANG Jiaxian, YUAN Xiaogang, LI Shaodong. Visual-tactile fusion and SAC-based learning for robot peg-in-hole assembly in uncertain environments[J]. *Machines*, 2025, 13(7): 605.
- [65] QI Wen, FAN Haoyu, ZHENG Cankun, et al. Human-like dexterous grasping through reinforcement learning and multimodal perception[J]. *Biomimetics*, 2025, 10(3): 186.
- [66] WANG Shaoxiong, LAMBETA M, CHOU Powei, et al. TACTO: a fast, flexible, and open-source simulator for high-resolution vision-based tactile sensors[J]. *IEEE robotics and automation letters*, 2022, 7(2): 3930–3937.
- [67] JIANU T, GOMES D F, LUO Shan. Reducing tactile Sim2Real domain gaps *via* deep texture generation networks[C]//2022 International Conference on Robotics and Automation. Philadelphia: IEEE, 2022.
- [68] KERR J, HUANG H, WILCOX A, et al. Learning self-supervised representations from vision and touch for active sliding perception of deformable surfaces[EB/OL]. (2022-09-26)[2025-09-10]. <https://arxiv.org/abs/2209.13042v1>.
- [69] HIGUERA C, SHARMA A, BODDULURI C K, et al. Sparsh: self-supervised touch representations for vision-based tactile sensing[EB/OL]. (2024-10-31)[2025-09-10]. <https://arxiv.org/abs/2410.24090>.
- [70] MIYATO T, MAEDA S I, KOYAMA M, et al. Virtual adversarial training: a regularization method for supervised and semi-supervised learning[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 41(8): 1979–1993.
- [71] ZHENG Wendong, LIU Huaping, WANG Bowen, et al. Cross-modal surface material retrieval using discriminant adversarial learning[J]. *IEEE transactions on industrial informatics*, 2019, 15(9): 4978–4987.
- [72] ZHENG Wendong, LIU Huaping, WANG Bowen, et al. Cross-modal material perception for novel objects: a deep adversarial learning method[J]. *IEEE transactions on automation science and engineering*, 2020, 17(2): 697–707.
- [73] LIU Huaping, WANG Feng, SUN Fuchun, et al. Surface material retrieval using weakly paired cross-modal learning[J]. *IEEE transactions on automation science and engineering*, 2019, 16(2): 781–791.
- [74] LIU Huaping, SUN Fuchun, FANG Bin, et al. Multimodal measurements fusion for surface material categorization[J]. *IEEE transactions on instrumentation and measurement*, 2018, 67(2): 246–256.
- [75] ZHAO Wenying, XU Qian, WANG Haoyu, et al. Cross-modal hashing for material surface properties fusion [C]//2023 International Wireless Communications and Mobile Computing. Marrakesh: IEEE, 2023.
- [76] LI Fengling, WANG Tong, ZHU Lei, et al. Task-adaptive asymmetric deep cross-modal hashing[J]. *Knowledge-based systems*, 2021, 219: 106851.

作者简介:



高春艳, 教授, 博士, 研究方向为智能机器人技术。获省部级科技奖励 4 项, 发表学术论文 50 余篇。E-mail: gcy@hebut.edu.cn。



刘琦, 硕士研究生, 研究方向为机器人触觉感知技术。E-mail: 2894570097@qq.com。



李满宏, 教授, 博士生导师, 研究方向为特种机器人技术与控制。主持国家自然科学基金等项目 20 余项, 获省部级科技奖励 10 项, 发表学术论文 60 余篇。E-mail: lmh9181219@163.com。