

具身智能框架中的感知—决策耦合机制建模

胡江洪^{1,2}, 曹彬^{1,2*}, 陈立名^{1,2}, 刘瑞芳^{1,2}

1. 天津市工业 AI 视觉检测技术重点实验室（筹），天津 300000

2. 菲特（天津）检测技术有限公司，天津 300000

DOI: 10.61369/TACS.2025110003

摘 要： 具身智能系统在复杂工业环境中的应用受限于感知精度不足与决策响应滞后问题，其本质是感知模块与决策模块的解耦架构导致信息传递断层与实时性缺失，针对这一问题构建了基于多模态数据融合的感知—决策耦合机制。通过三维视觉与多光源协同感知技术获取目标物体的空间几何特征与表面缺陷信息，结合大规模数据处理与行业垂类决策模型实现智能决策生成，在软硬件协同平台上设计实时数据传输通道与多模态特征融合算法，形成感知数据采集，特征提取，决策指令生成与执行反馈的闭环耦合架构。工业场景验证表明该耦合机制在汽车零部件、整车检测以及医药化工行业检测场景等应用中实现了微米级精度与毫秒级响应，证明了耦合建模在工业智能系统中的有效性。

关 键 词： 具身智能；感知—决策耦合；多模态融合；工业视觉；实时优化

Modeling of the perception-decision coupling mechanism in the embodied intelligence framework

Hu Jianghong^{1,2}, Cao Bin^{1,2*}, Chen Liming^{1,2}, Liu Ruifang^{1,2}

1. Tianjin Key Laboratory of Industrial AI Visual Inspection Technology, Tianjin 300000

2. Fitt (Tianjin) Testing Technology Co., LTD, Tianjin 300000

Abstract： The application of embodied intelligent systems in complex industrial environments is limited by insufficient perception accuracy and lagging decision-making response. The essence of this problem lies in the decoupling architecture of the perception module and the decision-making module, which leads to information transmission discontinuity and real-time loss. To address this issue, a perception-decision-making coupling mechanism based on multimodal data fusion has been constructed. The spatial geometric features and surface defect information of the target object are obtained through 3D vision and multi-light source collaborative perception technology. Intelligent decision generation is achieved by combining large-scale data processing and industry vertical decision-making models. Real-time data transmission channels and multi-modal feature fusion algorithms are designed on the software and hardware collaborative platform to form perception data collection and feature extraction. A closed-loop coupled architecture of decision instruction generation and execution feedback. Industrial scenario verification shows that this coupling mechanism has achieved micron-level accuracy and millisecond-level response in applications such as automotive parts, vehicle inspection, and inspection scenarios in the pharmaceutical and chemical industries, proving the effectiveness of coupling modeling in industrial intelligent systems.

Keywords： embodied intelligence; perception-decision coupling; multimodal fusion; industrial vision; real-time optimization

引言

工业智能系统要求环境感知与决策控制的高度协同，传统分离式架构因数据传输延迟与特征信息损失制约了系统性能，感知模块需实时获取物体三维形态与表面缺陷等多维度特征，决策模块需基于感知数据生成优化控制指令，二者的耦合程度直接决定系统的响应速度与控制精度。感知—决策耦合机制建模的核心在于打通数据流通过程，建立特征表示映射与设计协同优化策略，通过构建多模态感知技术体系捕获环境信息，基于数据驱动方法训练决策模型^[1]，在软硬件协同平台上实现感知与决策的实时交互与深度融合，可有效提升具身智能系统在高精度与强实时工业场景中的适应能力与作业性能。

一、高精度多模态感知技术体系

（一）三维视觉与多光源协同感知

工业场景中的高反射率金属表面与复杂几何形态对视觉感知系统提出了严苛要求，多角度光学融合技术通过构建多视角同步采集架构获取目标物体的完整空间信息，使得感知模块能够突破单一视角的遮挡限制。针对镜面材料的微米级缺陷检测采用条纹反射原理建立间接感知机制，光路逆向追踪算法通过定制化条纹投影与反射光路重构实现缺陷深度的三维复原^[2]，多光源 AI 融合技术整合不同波长与入射角度的光源组合生成多模态图像集，反射率图，曲率图与 3D 融合图经深度神经网络提取的特征向量在决策层实现语义级融合，因而构建了从光学物理层到语义抽象层的完整感知链路。该技术在头部汽车厂商的铝压铸件检测中实现了缩孔与纤维等微观缺陷的精准定位，感知处理延迟被控制在 110ms 以内，故而为后续决策模块提供了高置信度与低延迟的输入数据（见表1）。

表1 多光源融合图像特征维度与性能指标

图像模态	特征维度	缺陷检出率 (%)	误检率 (%)	处理时延 (ms)
反射率图	512×512×3	92.3	5.7	35
曲率图	512×512×1	87.6	8.2	28
3D 融合图	512×512×6	96.8	2.1	47

（二）动态目标特征提取与表示

生产线上的移动工件要求感知系统具备动态目标的实时捕获与特征稳定表达能力，3D 超动态融合技术采用快速连续图像采集机制结合点云融合算法完成动态场景的三维重建，时序图像帧通过深度学习网络映射为点云集合，使得目标物体的空间几何特征得以量化保存。点云拼接算法基于迭代最近点原理完成多帧数据的配准，配准误差控制在 0.03mm 以内，特征向量的构建采用多层次几何描述符体系实现从局部到全局的表征映射^[3]，局部表面法向量直方图捕获点云邻域的微观几何特性，全局形状签名则基于主成分分析提取目标物体的宏观形态参数。特征向量采用 8 位定点化编码降低存储开销，单帧特征数据压缩至 2.3KB，从而满足了实时数据传输通道的带宽约束，该流程在轮毂毛坯检测中实现了 40 秒内完成单件工件的完整特征提取，关键尺寸测量精度达到 ± 0.05mm，可耦合特征表示通过标准化接口向决策模块输出结构化数据。

二、数据驱动的智能决策建模

（一）大规模数据处理与模型训练

工业决策模型的泛化能力依赖大规模异构数据处理与高效训练架构，基于 K8S 构建的分布式集群通过容器化部署实现计算资源动态分配，多节点训练任务可依负载自适应迁移。GPU 利用率优化采用 dlRover 框架，通过监控各节点显存占用率与算力使用率动态调整批处理大小，因而将 GPU 平均利用率从 62% 提升至 89%，针对工业场景异常数据稀缺问题，少样本学习算法结合对

抗生成网络合成缺陷样本与元学习策略，实现模型在 5-10 个标注样本下的快速收敛。表 2 展示了不同样本规模下的模型泛化性能指标，所述算法在食品包装异物检测与化工容器裂纹识别任务中实现了 92% 以上的检出率，故而验证了少样本决策机制在数据受限环境下的有效性。

表2 不同样本规模下的决策模型性能指标

样本数量	训练时长 (h)	检出率 (%)	误检率 (%)	泛化损失
5 样本	0.3	85.7	12.4	0.187
10 样本	0.8	92.1	7.6	0.142
50 样本	1.5	96.3	3.2	0.098
全量数据	2.7	97.8	2.1	0.073

（二）行业适配的智能决策生成机制

行业垂类决策需深度融合领域知识与实时响应能力，汽车行业垂类大模型构建涵盖生产数据，高精度图像与运营文本的多模态数据集，打通生产平台与 AI 中台数据通路，采用预训练 - 微调范式将冲压与焊装等六大领域工艺规范嵌入模型参数空间，提示学习技术通过少量示例引导模型理解缺陷语义^[4]，使得决策输出直接对应工艺优化指令。同时整合 M-RAG 检索增强与 A-agent 多智能体机制，形成轻量化边缘模型实时响应与云端大模型复杂推理的分层架构，响应延迟控制在 150ms 内，为满足毫秒级响应要求，模型量化采用 TensorRT 框架将浮点参数转为 INT8 定点表示，量化公式为：

$$Q(x)=(x-z)\cdot s$$

其中 z 为零点偏移而 s 为缩放因子。结构化剪枝在精度损失 < 1.2% 前提下将参数压缩至 38%，卷积层与批归一化层融合缩短前向传播时间 23%，嵌入式部署结合 3D 相机与机械臂标定（误差 < 0.8mm），构建从图像输入到动作指令输出的端到端低延迟通路，单次决策周期稳定在 45ms。

三、感知决策协同耦合机制

（一）软硬件协同耦合平台

智能魔方平台通过硬件层与软件层的深度集成构建了感知 - 决策 - 执行的全流程耦合载体，硬件层整合算力服务器，3D 相机与边缘计算设备形成分布式算力网络，软件层则由光学算法，AI 决策引擎与工业控制逻辑组成模块化软件栈。光学模块接口采用标准化数据帧格式输出点云与纹理信息，数据帧头部包含时间戳，坐标系标识与质量评估参数，使得感知结果能够被 AI 模块无缝接收。AI 模块通过 RESTful API 与光学模块建立异步通信机制，决策网络在接收到完整特征向量后触发推理流程并将决策结果封装为 JSON 格式指令。工业控制模块解析决策指令后转换为 PLC 可识别的 Modbus 协议或 EtherCAT 实时总线信号驱动机械臂执行动作，执行反馈通过状态寄存器回传至 AI 模块形成闭环校正机制，该平台在汽车焊装与化工分拣场景中实现了多模式按需组合，光学检测模式，AI 识别模式与机器人抓取模式可根据生产节拍动态切换^[5]。

（二）多模态感知与决策的动态交互机制

多模态感知与决策的动态交互机制围绕数据传输与特征融合两大核心构建，确保感知信息能精准、实时转化为决策指令，数据传输采用共享内存与环形缓冲区设计，共享内存通过进程间映射避免数据拷贝，将点云传输时延压缩至 2.7ms；环形缓冲区以 16 帧容量实现无锁并发，保障峰值采集速率下的数据完整性，同时结合 3D 相机与机械臂标定技术，基于标准球棒构建旋转轴约束模型，通过最小二乘法求解齐次变换矩阵：

$$T = \arg \min \sum \| X_{camera} T - X_{robot} \|^2$$

其中 X_{camera} 为相机坐标系坐标， X_{robot} 为机械臂坐标系坐标，将标定误差控制在 0.8mm 内，特征融合采用双流编码器架构，分别提取图像特征向量 f_{img} 与文本嵌入向量 f_{text} ，通过对比学习优化相似度函数

$$\text{sim}(f_{img}, f_{text}) = \cos(W_{img} \cdot f_{img}, W_{text} \cdot f_{text})$$

其中 W 为可学习的投影矩阵。多级筛选策略在融合特征空间中构建粗粒度与精细化的层级判别机制，粗筛层基于全局特征快速排除正常样本，精筛层则针对疑似缺陷区域进行局部注意力增强与边界精修，少样本缺陷生成技术通过扩散模型合成多角度高维图像扩充训练集，生成样本在纹理一致性与几何真实性方面达到了与真实数据等效的质量水平，该机制在汽车涂装缺陷检测中实现 96.8% 的检出率，误检率仅 2.1%，验证了多模态交互对决策精度的提升价值。

四、工业场景验证与性能分析

（一）轮毂毛坯检测实验与评估

轮毂毛坯在线检测实验通过感知-决策耦合机制实现了高精度与高节拍的协同优化，光学系统设计采用多相机环形布局结合旋转台运动控制完成360度全方位扫描，每个相机视场角覆盖45度空间范围，使得八视角点云数据能够无缝拼接重构完整三维模型。基于标准球棒的旋转轴标定技术通过测量球心轨迹拟合旋转轴方程，标定算法求解最小化目标函数：

$$\min \sum (d_i - r)^2$$

其中为第 i 个球心到拟合轴的距离， r 为标准球半径，该方法

将轴线定位误差控制在0.05mm以内，因而消除了机械装配误差对测量精度的影响，重叠点云去除算法在拼接过程中识别相邻视角的冗余区域并执行体素滤波，点云密度从原始的 2.3×10^6 点/帧降至 1.5×10^6 点/帧，故而将特征提取时间缩短至18秒。耦合机制在轮毂检测中的关键性能指标里，检测节拍达到40秒/件，关键尺寸精度 $\pm 0.05\text{mm}$ 与重复性 $\pm 0.03\text{mm}$ 均满足汽车行业质量标准。

（二）多行业应用案例分析

感知-决策耦合机制在多行业部署验证了跨场景适配能力，汽车领域覆盖冲压件表面缺陷检测、铝压铸缩孔识别与焊装质量评估，多光源 AI 融合技术精准捕获微米级缺陷，保障零部件质检精度。化纤行业生产线中，该机制实现丝锭表面微观缺陷实时检测，单件处理耗时 ≤ 2.5 秒，检测效率较传统人工提升3倍以上，航空航天领域依托太赫兹成像与深度学习耦合，突破传统无损检测穿透深度限制，实现金属浅表层微米级裂纹可靠识别，解决关键零部件质量管控难题。钢铁行业热态钢管检测系统集成动态环境自适应算法与多光谱成像，在1000℃高温与粉尘雾气干扰下仍能完成直径跨度超300mm钢管的实时扫描，精准识别表面裂纹与壁厚偏差，检测效率较传统方法提升5倍，食品医药领域采用少样本决策算法，仅需10个标注样本即实现包装异物92.1%的检出率，累计完成20余项工业级技术部署。

五、结束语

感知-决策耦合机制建模为具身智能在工业场景的深度应用奠定了理论与技术基础，通过三维视觉与多光源协同感知技术实现了微米级的检测精度与全方位的环境感知能力，通过大规模数据处理与行业垂类决策模型保障了算法在复杂场景下的泛化性与适配性，通过软硬件协同平台的实时数据传输与多模态特征融合算法构建了感知与决策的深度耦合机制。多行业应用案例验证了耦合机制在不同材质与不同缺陷类型场景中的有效性与可靠性，为具身智能系统在智能制造领域的规模化推广提供了可行路径。

参考文献

- [1] 张昌盛. 从具身智能到具身智能体 [J]. 北京工业大学学报 (社会科学版), 2024, 24(06): 154-165.
- [2] 何俊; 张彩庆; 李小珍; 张德海. 面向深度学习的多模态融合技术研究综述 [J]. 计算机工程, 2020(05).
- [3] 张丽媛, 戈维峰, 崔晋. 具身智能的科学内涵与研究范式分析 [J]. 华东科技, 2024, (10): 36-41.
- [4] 孙长银, 穆朝絮, 柳文章, 等. 自主无人系统的具身认知智能框架 [J]. 科技导报, 2024, 42(12): 157-166.
- [5] 魏圣九. 媒介化的身体感知：基于智能穿戴用户使用经验的研究 [D]. 中南民族大学, 2023.