

复杂场景下物料分拣动态目标检测评估指标体系构建

徐小辉

重庆工程职业技术学院 重庆 402260

【摘要】：针对复杂工业场景下动态目标检测通用指标失灵的问题，构建了一套涵盖技术性能、场景适配及工程经济效能的“三位一体”多维综合评估指标体系。通过深度解构光照波动、物料堆叠等特征，设计了加权动态平均精度、轨迹碎片化率等专项指标，并引入层次分析法（AHP）与客观数据融合的组合赋权机制。结果表明，该体系通过线性加权模型与雷达图可视化诊断，有效克服了传统评价的“非补偿性效应”，为物料分拣技术的标准化评估与落地应用提供了科学的方法论支撑。

【关键词】：复杂场景；物料分拣；动态目标检测；指标体系

DOI:10.12417/2705-0998.26.06.058

1 引言

在智能制造及工业 4.0 浪潮下，自动化分拣已成为智慧物流与精密制造领域提质增效的关键。基于深度学习的动态目标检测技术凭借卓越的特征捕捉能力，正逐步取代传统感知手段成为分拣系统的核心“大脑”[1]。然而，尽管各类检测算法在标准数据集上频破记录，但在真实工业产线落地中却普遍遭遇“水土不服”[2]。

学术表现与工程实效的显著背离，根源于现有评价标准的结构性缺陷。一般而言，学术研究更多聚焦于平均精度均值（Mean Average Precision, mAP）与每秒处理帧数（Frames Per Second, FPS）等通用指标，虽能衡量实验室环境下的算法上限，却难以量化工业现场剧烈光照波动、随机遮挡及运动模糊等复杂工况对系统鲁棒性的冲击[3]。这种评价失灵导致系统在实际部署时极易出现漏拣率波动或算力失控，陷入“学术跑分高、落地应用难”的窘境。

实际上，工业分拣系统遵循“木桶效应”，任一性能短板均会诱发整体崩溃。因此，构建兼顾技术先进性与业务实效性的专项评价体系，是打通算法落地“最后一公里”的核心前提。立足于此，本文深入解构复杂场景挑战，提出一套涵盖检测能力、运行质量与工程经济性的综合评估指标体系，并引入层次分析法（Analytic Hierarchy Process, AHP）实现科学赋权，旨在为工业生产领域物料分拣技术的标准化评估与示范应用提供方法论支撑。

2 复杂场景物料分拣特征与评估需求分析

2.1 工业分拣场景的复杂特征解析

工业现场环境与标准学术数据集存在本质差异，其复杂性

主要源于环境的不确定性与物料特征的非线性变化。首先，物理环境干扰影响。产线中光照波动的随机性、皮带高速运行引发的机械抖动及工业粉尘，极易诱发动态检测中的“运动模糊”，显著降低图像对比度。其次，物料形态耦合影响。区别于传统的单目标识别，工业物料常呈高密度堆叠与随机遮挡状态，这种几何层面的非线性重叠要求算法在极低特征信息下具备极强的边界勾勒与语义关联能力。再次，尺度演变挑战影响。分拣对象跨越多个尺度量级，既有占据大幅画面的构件，也有易被背景淹没的小微特征物[4]，传统标准对多尺度均衡性的忽视极易导致识别失衡。

2.2 动态目标检测应用的效能瓶颈

在工业生产复杂环境下，动态目标检测技术的应用价值不再仅局限于视觉层面的“检出率”，而在于其与执行机构协同产生的整体效能。其一，存在非补偿性“木桶效应”。智能分拣系统遵循感知、决策、执行的闭环逻辑。根据非补偿性评价原则，系统的整体可用度受限于性能最差的指标，即使 mAP 领先，若在特定遮挡下的鲁棒性存在短板，亦会导致分拣溃败。其二，时延与精度的实时博弈。动态分拣对响应时间有严苛约束，单帧推理时延必须在物料经过抓取位的极窄时间窗内完成[5]。算力受限与模型复杂度之间的矛盾是制约落地效率的关键瓶颈。其三，跟踪稳定性要求时间连续性。物料位移的实时性要求检测结果具备时间维度的平滑性，检测框抖动或 ID 频繁切换会直接导致机械手动作紊乱。

2.3 全流程业务评估需求的提炼

基于对复杂场景与效能瓶颈的洞察，本文提出从技术研发、现场落地到业务运营三个维度，提炼动态检测性能的全流程评估需求。具体如表 1 所示。

作者简介：徐小辉（1979-），男，中国福建三明人，教授，重庆工程职业技术学院智能制造与交通学院教师，主要研究方向为：传感器与智能检测、机器感知与机器视觉等。

基金项目：本文系重庆市教育委员会科学技术研究计划青年项目资助项目（项目编号：KJQN202403436）的阶段性研究成果。

表1 工业分拣场景特征与评估需求映射表

场景特征/瓶颈	评估需求维度	核心关注点
异形物料、多尺度	技术层面：精准导向	模型在干扰下的性能衰减率、分类准确性
光照波动、重叠遮挡	落地层面：鲁棒性验证	环境适配阈值、边缘场景失效率
实时博弈、能耗成本	运营层面：经济实用性	分拣效率、设备稼动率、单位能耗

综上所述，构建一套兼顾技术性能与业务实效、覆盖全流程的动态目标检测性能评估指标体系，已成为突破工业分拣落地瓶颈的关键路径。

3 物料分拣检测评估指标体系的多维构建

3.1 指标体系构建原则与框架设计

构建面向工业现场的评估体系，必须彻底摒弃“唯跑分论”的单一视角，确立以“解决产线实际问题”为最终检验标准的核心价值导向。本文在指标体系设计上，严格遵循 S.M.A.R.T 原则（具体性、可度量性、可达成性、相关性、有时限性），确保评价标准的科学性与工程可执行性。基于工业场景的业务特性，提出一种“技术性能—场景适配—业务与工程经济”三位一体的综合评估维度模型。该模型自下而上贯通了从底层算法基准测试到顶层商业价值决策的全流程逻辑，形成了一个层级严密、互为支撑的指标矩阵。

3.2 技术性能指标：面向算法基础能力的评估

技术性能层主要评估目标检测算法在理想或标准工况下的基础能力，是对模型精度、速度及资源消耗的量化考察。针对动态分拣场景，本文对传统通用指标进行了改良设计，具体指标包括：

(1) 加权动态平均精度 (Weighted Dynamic mAP)。传统目标检测采用的 mAP 指标忽略了物料在高速运动下产生的畸变对精度的削弱。为此，本文引入基于传送带带速 v 的动态惩罚权重函数 $\mu(v)$ ，构建动态平均精度度量公式：

$$mAP_{dyn} = \frac{1}{C} \sum_{i=1}^C \left(\int_0^1 P_i(R) \cdot \mu(v) dR \right)$$

式中， C 为物料类别总数， $P_i(R)$ 为类别 i 在召回率 R 下的准确率曲线。 $\mu(v)$ 的引入使该指标能更真实地反映算法在不同皮带速度下的实际检出一致性。

(2) 端到端推理时延 (End-to-End Inference Time)。与仅关注前向传播时间的研究不同，工业评估强调包含图像预处理、模型推理与后处理（如 NMS 非极大值抑制）的全局耗时。在实时分拣系统中，单帧推理时延必须严格控制在机械手运动补偿的时间窗内。

(3) 算力资源占用波动率。为评估系统长期运行的稳定性，引入显存及计算资源占用的波动率指标。通过计算系统在连续作业周期内的峰值资源占比与平均占比的方差，识别算法是否存在内存泄漏或潜在的计算瓶颈。

3.3 场景适配指标：面向工业复杂工况的评估

场景适配层专项评估模型应对工业复杂物理环境干扰的能力，是填补工业分拣专项评价空白的关键。具体指标包括：

(1) 鲁棒性衰减率 (Robustness Decay Rate)。用于量化系统在遮挡、半透明物料或极端光照等边缘场景 (Corner Cases) 下的失效程度。通过在标准光照与极端光照（或无遮挡与高密度堆叠）状态下分别测试抓取成功率或召回率，计算其性能衰减幅度：

$$\Delta R_{decay} = \frac{R_{standard} - R_{degraded}}{R_{standard}} \times 100\%$$

式中， $R_{standard}$ 为标准环境下的召回率， $R_{degraded}$ 为复杂干扰环境下的召回率。该值越小，说明系统对抗场景波动的鲁棒性越强。

(2) 轨迹碎片化率 (Trajectory Fragmentation Rate, TFR)。

在动态分拣中，检测框的闪烁或频繁 ID 切换 (Switch) 会导致执行机构动作紊乱。本文通过 TFR 评估跟踪稳定性：

$$TFR = \frac{N_{frag}}{N_{traj}}$$

其中， N_{frag} 为由于检测失效导致的轨迹中断次数， N_{traj} 为物料经过检测区的实际总轨迹数。高 TFR 直接预示着机械臂“抓空”或“重复抓取”风险的增加。

3.4 业务实效与工程经济指标：面向决策价值的评估

该维度旨在将技术参数转化为管理者关注的业务产出与商业决策依据。具体指标包括：(1) 业务实效指标：以产线最终收益为导向，涵盖分拣准确率、单位小时吞吐量以及分拣成本分析。通过构建分类混淆矩阵，量化不同类别物料错拣、漏拣对生产成本的具体影响。(2) 系统可用度 (System Availability)：结合平均无故障工作时间 (MTBF)，计算系统在规定生产周期内的实际有效作业时长占比，评估其作为工业级装备的可靠性。(3) 柔性化指标 (换产时间)：评估系统在不同品类物料任务切换时的微调时长。换产时间越短，代表系统的柔性化水平与部署便捷性越高。(4) 单位分拣综合能耗：将视觉处理单元与执行机构的总能耗平摊至单件物料，形成评估生产经济性与环保效益的交叉指标。

4 评价指标赋权与综合评估得分模型

4.1 异构指标的无量纲化处理策略

智能分拣系统的评价体系包含众多性质迥异的指标，如检测精度（通常用百分比表示）、推理时延（以毫秒计）及部署

成本（以元计）等。为消除量纲与量级的差异，必须对原始数据进行无量纲化与归一化处理。

根据指标对系统效能的贡献方向，本文将其划分为“效益型指标”（如 mAP_{dyn} 、系统可用度；数值越大越好）与“成本型指标”（如端到端推理时延、轨迹碎片化率；数值越小越好）。随后，采用 Min-Max 极值标准化方法将指标映射至[0,1]区间。

对于效益型指标，归一化公式为：

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

对于成本型指标，归一化公式为：

$$x'_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)}$$

式中， x_{ij} 为第*i*个方案在第*j*项指标下的原始数据； x'_{ij} 为归一化后的标准值； $\max(x_j)$ 与 $\min(x_j)$ 分别为所有参评方案中该指标的极值。通过此步骤，所有指标均转化为无量纲的纯数值，为后续加权计算奠定基础。

4.2 基于 AHP 与客观数据的组合赋权机制

不同指标在系统整体效能中的重要度存在显著差异，单一的主观或客观赋权法均存在局限性。为此，本文提出主客观相融合的组合赋权机制：

(1) 主观权重的确定（AHP 法）：采用 AHP 法将业务专家的先验知识进行量化，得出主观权重向量 α_j 。为确保专家判断逻辑的自洽性，需对判断矩阵进行一致性检验；仅当一致性比率 $CR < 0.1$ 时，判断矩阵与对应权重方为有效。在实际应用中（如精密零件分拣与大宗废料初筛场景），主观权重分配具备动态调整的灵活性。

(2) 客观权重的确定（熵权法或 CRITIC 法）：为克服人为偏差，依据实测数据的变异程度计算得出客观权重向量 β_j 。若某项指标在多个方案中得分差异极小，客观赋权法将自动削弱其权重，凸显真正拉开方案差距的关键指标。

(3) 组合权重的合成：

通过乘法合成逻辑确定最终组合权重 w_j ，使评估模型兼顾产线管理者的业务侧重与客观测试数据的分布规律：

$$w_j = \frac{\alpha_j \beta_j}{\sum_{k=1}^n \alpha_k \beta_k}$$

4.3 综合得分模型与可视化诊断框架

在完成归一化与赋权后，需要提出一种综合评价数学模型计算出系统总得分，以便于对不同技术方案进行优劣排序。

(1) 综合得分决策模型：本文采用线性加权求和法（WSA），计算方案*i*的综合得分 S_i ：

$$S_i = \sum_{j=1}^n w_j x'_{ij}$$

该总分可直观呈现不同技术方案的综合优劣排序。

(2) 弱项可视化诊断：针对工业应用中的“非补偿性效应”，单看总分可能掩盖系统的致命短板。本文设计了结合雷达图的多维效能诊断报告形式。例如，通过对比雷达图的不同轴线，产线工程师可快速识别某方案是否存在“精度极高但抗光照干扰能力崩溃”的偏科现象。

(3) 分级认证体系：基于加权总分与短板分布，构建“分级认证方案”供行业参考。例如，总分高于 90 分且各项维度均衡的方案可认证为“复杂场景高可靠级（L3）”，而存在明显受限的方案则评定为“受限场景适用级（L1）”。该体系有助于企业根据预算与产线特性实现精准采购。

5 结论

本文针对复杂工业分拣场景，深入剖析了通用评估指标在应对光照、遮挡及动态模糊等极端工况时的适配性缺陷，构建了一套涵盖技术性能、场景适配及工程经济效能的“三位一体”多维综合评估体系。研究通过引入加权动态平均精度、轨迹碎片化率等专项指标，成功克服了传统标准存在的“非补偿性效应”，实现了评估维度从单一学术跑分向真实产线效能的跨越，为打通算法落地“最后一公里”提供了科学的“度量衡”。

在方法论层面，本文提出主客观融合的组合赋权机制，通过 AHP 层次分析法量化业务先验知识，并辅以一致性检验和客观数据修正，确保了权重分配的自洽性与科学性。该体系及其配套的可视化诊断框架，不仅能辅助工程师精准定位算法短板，更为实际工业场景应用效果评价提供了标准化工具。尽管目前在 3D 视觉领域的验证尚存局限，但研究思路为未来向多模态融合及边缘计算评估方向延伸奠定了坚实基础。

参考文献：

- [1] 张阳婷,黄德启,王东伟,等.基于深度学习的目标检测算法研究与应用综述[J].计算机工程与应用,2023,59(18):1-13.
- [2] 闫明,陶大鹏,普园媛.面向工业零件分拣系统的低纹理目标检测[J].中国图象图形学报,2022,27(08):2418-2429.
- [3] 明阳.基于目标检测的散乱工件定位及分拣技术研究[D].沈阳理工大学,2025.
- [4] 曹学鹏,李鑫,冯艳丽,等.面向无序分拣场景的工件 6D 位姿检测方法[J].工程科学与技术,2025,57(05):298-308.
- [5] 于灵鑫,陈艺博,曲浩君,等.基于 GraspNet 的多阶段无序混装抓取方法[J].计算机科学,2026,53(04):318-325.