Reinforcement Learning for Optimal Polycube Packing under Physical Constraints: Application to Robotic Handling of Sequentially Arriving Luggage

Proposed by:
Lama Tarsissi (SUAD, lama.tarsissi@sorbonne.ae)
Damien Jamet (HDR, Univ. Lorraine/Loria, damien.jamet@loria.fr)
Laurent Ciarletta (Univ. Lorraine/Loria, laurent.ciarletta@loria.fr)

Submission for the SAFIR 2025 call

Abstract

Automated packing of 3D volumes, like luggage containers (ULDs), is a major complex real-world logistics challenge, complicated by sequential item arrival and the need to consider heterogeneous physical properties (mass, centre of gravity, fragility, shape) for stability and integrity, aspects often overlooked by traditional optimisation or simplified AI models in real-time scenarios. This project investigates using deep reinforcement learning (Deep RL) to develop optimal, adaptive robotic packing strategies for items (approximated as polycubes) within this dynamic, physically constrained environment. The primary objective is to design an RL agent learning sophisticated policies covering item selection, strategic placement mode choice (precise direct placement vs. efficient dynamic 'gentle throws'), and optimal pose determination, tackling key RL challenges like sparse rewards and safe exploration. The agent aims to maximise packing density while adhering to complex physical constraints, preventing damage and maintaining stack stability based on support polygons and load distribution, a key challenge for physics-informed RL. We will develop an advanced simulation environment incorporating realistic physics modelling, including accurate multi-contact and motion dynamics for throwing actions. The agent learns via extensive simulated interactions, targeting robust, generalisable policies for diverse shapes/scenarios. This research pushes boundaries in combinatorial optimisation and autonomous robotic manipulation for complex real-world tasks, offering significant impact for logistics automation, enhancing overall operational efficiency and safety. Expected outcomes include evolutions of physics-aware RL algorithms, a validated, reusable simulation platform, and a compelling proof-of-concept validated through simulation and ideally including experiments executed on real Cobot and Humanoid robot platforms, demonstrating the feasibility of this approach.

Research Question and Objectives

Main Research Question: How can a Reinforcement Learning agent be designed and trained to learn effective, robust, and physically consistent strategies for sequential packing (online 3D packing) of heterogeneous items (polycubes with varying physical properties like mass distribution, friction, fragility) by a robotic manipulator (ideally including cobots or humanoid robots), while mastering complex dynamic placement actions ('gentle throws') and potentially incorporating advanced behaviours like mobility, buffering, and rearrangement to collaboratively optimise packing density, throughput, and overall physical safety?

Scientific and Technical Objectives:

1. **Develop a high-fidelity physics simulation environment:** Integrating a detailed robot model (e.g., cobots, humanoids), a configurable stream of polycube items with diverse geometric/physical parameters, a target container, and reliably modelling multi-contact dynamics (including frictional contacts and restitution), impacts, and trajectories for 'gentle throws' for realistic interaction results and sim-to-real

transfer potential. Include realistic sensor simulation (vision, force/torque). Evaluate physics engines (e.g., PyBullet, Isaac Gym) for accuracy/speed trade-offs, possibly calibrated with real data if available.

2. **Design effective state and action representations:** Suitable for deep RL, capturing container state (e.g., occupancy grid, object graph), item properties, robot configuration, potentially sensor data. Action space must accommodate hybrid decisions (discrete: pick, place/throw; continuous: pose, throw velocity) and potentially navigation/rearrangement primitives. Explore GNNs or VoxelNets for efficient geometric and relational reasoning between objects.
3. **Implement, adapt, and advance Deep RL algorithms:** Select/enhance methods (e.g., PPO, SAC, hierarchical RL, model-based RL) for this challenging sequential problem, investigating trade-offs between sample efficiency and final performance. Design and test multi-objective reward functions and constraint-handling techniques balancing density, throughput, strict constraint adherence (stability, fragility), and efficiency, addressing potential sparse reward and safe exploration challenges.
4. **Integrate and evaluate learning of dynamic placement ('gentle throws'):** Investigate the agent's capability to learn *when* dynamic placement is beneficial and *how* (control parameters) for accurate, stable landing, considering item physics and environment state. Analyse speed/accuracy/risk trade-offs in detail using statistical methods and robustness tests.
5. **Rigorously evaluate performance and generalisation:** Assess agent performance via metrics (density, throughput, constraint violations, stability, robustness, efficiency, scalability). Compare against relevant baselines (heuristics, other RL). Evaluate policy's ability to generalise to unseen items/containers/sequences via diverse benchmark scenarios, assessing zero-shot and few-shot adaptation capabilities, a critical factor for real-world deployment.
6. **Analyse and interpret learned policies:** Apply interpretability methods (e.g., attention map visualisation, saliency analysis, causal analysis) to understand emerging complex strategies. Analyse failure modes and decision factors (e.g., place vs. throw choice) to build trust, ensure safety/predictability, and potentially extract human-understandable heuristics.
7. **Explore advanced packing strategies (Exploratory Goal):** Investigate, potentially in later stages, integrating more complex behaviours if beneficial. This includes:
   (a) *robot base mobility* for navigation in larger workspaces, potentially involving path planning integration,
   (b) *strategic buffering* involving temporary item placement for improved long-term packing, and
   (c) *local rearrangement* using pushing actions. Evaluate complexity vs. performance gains.

Research Impact and Significance

Scientific Significance:

- **Advancement in Combinatorial Optimisation and AI:** Addresses online 3D-BPP with realistic physics and dynamics. A high-performing RL solution advances beyond heuristics/simplified RL, handling complex constraints and dynamic actions requiring new methods. Contributes significantly to physics-aware RL, sample efficiency in complex simulations, safe exploration techniques for robotics, hybrid action space learning, and constrained optimisation under significant uncertainty, establishing new benchmarks for complex sequential decision-making.
- **Contribution to Autonomous Robotics:** Learning physics-aware manipulation, including dynamic placement for optimisation, advances robotic dexterity, control, and planning for contact-rich tasks in unstructured, dynamic environments, moving towards more versatile and physically grounded AI robots capable of complex logistical tasks.
- **Synergy between Disciplines:** Integrates AI (Deep RL), Robotics, Physics Simulation, Optimisation Theory, and Discrete Geometry, fostering interdisciplinary approaches and potentially novel insights applicable across these key fields, bridging theoretical models with practical robotic implementation.

Potential Societal and Economic Impact:

- **Logistics Automation:** Automating luggage handling offers benefits: improved ergonomics (reducing repetitive strain injuries), reduced costs, faster turnarounds in time-sensitive airport operations, potentially less damage through consistent handling. Directly addresses labour shortages and improves supply chain resilience and sustainability. Creates demand for skilled workforce in robotics and AI maintenance/supervision.

- **Broader Applicability:** Core methods transferable to warehouse automation, e-commerce fulfilment, manufacturing (bin picking/assembly), construction (automated placement), waste management (sorting/compaction), optimising resource utilisation and complex assembly.
- **Strategic Alignment:** Aligns with AI/robotics/logistics priorities. Contributes to SUAD/SAFIR goals in cutting-edge research, enhancing regional competitiveness in advanced automation technologies and Industry 4.0. Fosters international collaboration.

Novelty and Originality: Lies in the *holistic integration* via deep RL of: (i) online 3D packing, (ii) realistic, heterogeneous object physics as hard constraints, and (iii) advanced robotic manipulation including strategically learned dynamic placement. Existing research typically addresses subsets with significant simplifications, lacking this integrated learning approach.

Administrative and Financial Framework (SAFIR 2025 Call)

The following information is extracted from the SAFIR 2025 call for proposals and outlines the conditions of the doctoral contract associated with this project.

Project Duration and Timeline:

- **Duration:** The project is funded for a duration of three years for a doctoral contract.
- **Project Start Date:** The selected candidate must start no later than September 30th, 2025.

Conditions for the PhD Candidate:

- **Residency:** The selected candidate must reside in the United Arab Emirates (UAE) for the duration of the project and actively participate in the scientific and academic life of the institution.
- **Enrollment and Affiliation:** The doctoral candidate will be enrolled in a French doctoral school and will be affiliated with the SAFIR laboratory at Sorbonne University Abu Dhabi.
- **Recruitment Process:** Following the approval of this research proposal, an open call for applications will be published. Interested candidates will apply, and the project leaders will select one applicant. The final award is subject to an audition (10-minute presentation) of the candidate before the SUAD Research Council.

Stipend and Benefits:

- **Monthly Allowance:** The doctoral candidate will receive a monthly allowance of **15,000 AED**.
- **Travel:** A round-trip ticket between the candidate's country of origin and Abu Dhabi is provided.
- **Coverage:** Visa and health insurance coverage are provided by Sorbonne University Abu Dhabi for the entire duration of the grant.

Intellectual Property:

- The SUAD Intellectual Property Policy applies. Any IP generated solely by SUAD personnel or students will be assigned to SUAD. In the case of jointly generated IP with an external entity, the IP will be jointly assigned under conditions specified in a formal agreement.