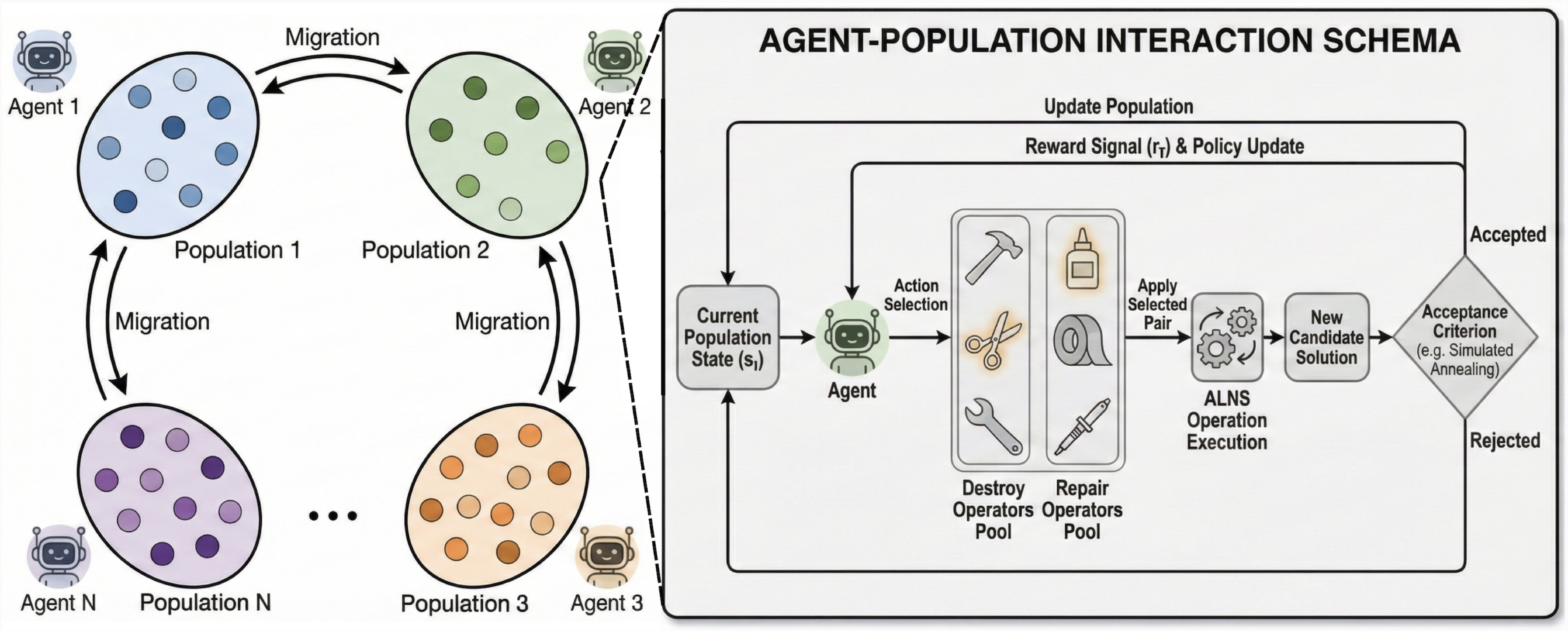
Island-DR-ALNS: A Diversity-Driven Multi-Agent Framework for Deep Reinforcement Learning-based Combinatorial Optimization

Adam Żychowski[1], Xin Yao[2], Jacek Mańdziuk[1,3]

[1]Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland

[2]School of Data Science, Lingnan University, Hong Kong SAR

[3]Faculty of Computer Science, AGH University of Krakow, Poland



Introduction

- **Context:** The approach to combinatorial optimization has evolved from manually engineered heuristics toward Reinforcement Learning-based Combinatorial Optimization (RLCO). Methods like DR-ALNS replace handcrafted rules with learned search policies.
- **Problem:** Current RLCO approaches rely on a monolithic learning agent. While these neural models excel at optimizing a specific trajectory, they are prone to converging toward deterministic, dominant policies, leading to a lack of behavioral diversity.
- **Proposed Solution: Island-DR-ALNS** fills this gap by integrating Deep Reinforcement Learning with a decentralized island model. It deploys a population of heterogeneous agents cooperatively optimizing the Adaptive Large Neighborhood Search (ALNS) parameters.

Framework & State Space

Multi-Agent Architecture & Action Space

- Each agent controls its local ALNS process via a Markov Decision Process $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$.
- Action space $\mathcal{A}$ involves 4 choices: Destroy Operator, Repair Operator, Destroy Severity, and Acceptance Parameter.
- Agent is modeled as a Multi-Layer Perceptron with two hidden layers of 64 units.

Socially-Aware State Representation

The state space extends standard local search dynamics with asynchronous *social features*, enabling agents to contextualize their performance relative to the global population:

Feature	Description
	Island Search Dynamics
Best Improved	Has the best global solution improved?
Current Accepted	Has the last candidate solution been accepted?
Current Improved	Is the current solution better than the previous one?
Is Current Best	Is the current solution the best found so far?
Cost Difference	Relative gap between current and best solution objective.
Stagnation Count	Number of iterations without improvement.
	Social Context (New)
Island Rank (Best)	Rank of this island's best solution within the population.
Island Rank (Avg)	Rank of this island's average population fitness.
Island Distance	Normalized distance to the closest neighboring island.
Migration Time	Iterations since the last migration event.

Training & Migration

Negatively Correlated Search (NCS)

NCS explicitly penalizes similarity between agent policies to promote diversity. The loss for agent i is:

$$L(\theta_i) = L_{PPO}(\theta_i) - \lambda \mathbb{E}_{s \sim \rho}[C_i(s)]$$

where C_i computes the distance between its policy and the policies of all other agents.

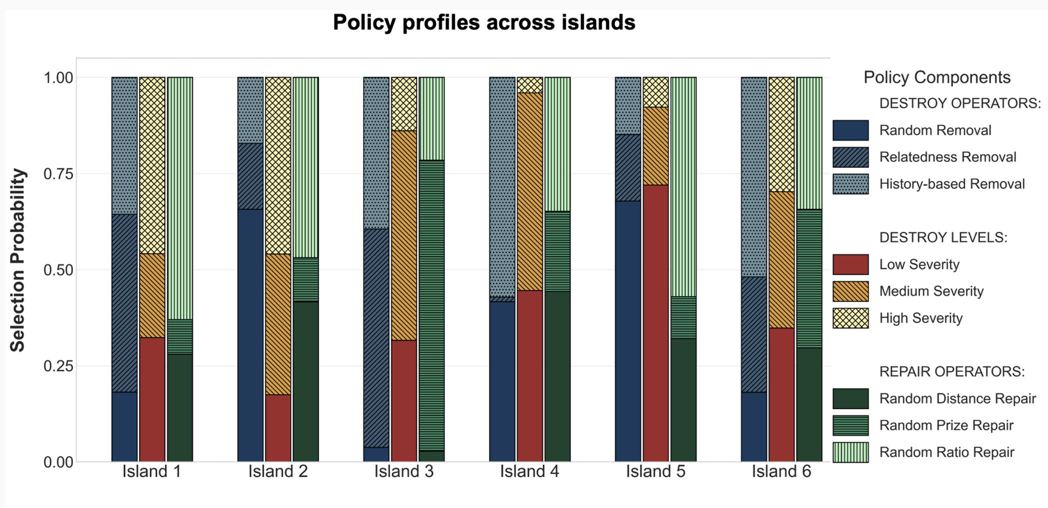
- **Training:** Proximal Policy Optimization (PPO) over 300,000 search steps.
- **Reward system:** +1 for individual improvement, +5 for a new local island best, and +10 for a new global best solution.

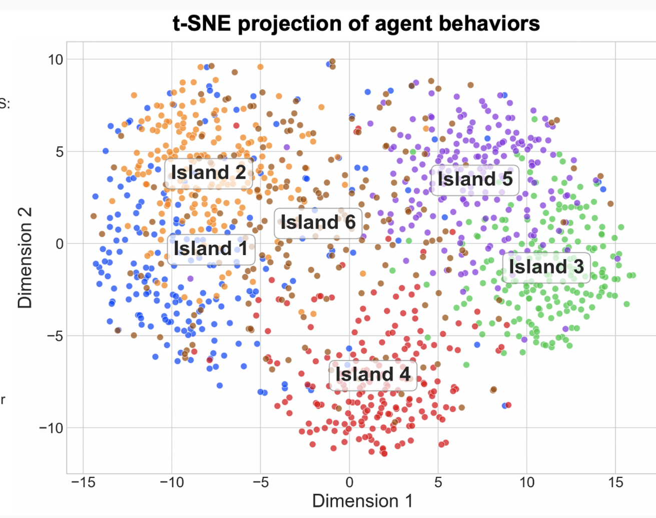
Diversity-Driven Migration

- **Topology:** Ring (migration strictly between immediate neighbors).
- **Timing:** Asynchronous, triggered by local stagnation.
- **Content:** Selects solutions maximizing a weighted sum of fitness and contribution to target island diversity.

Agent Specialization

Agents spontaneously adopt distinct search roles (e.g., aggressive exploration vs. fine-grained exploitation) without manual engineering.





Main Performance Comparison

Island-DR-ALNS achieves the lowest average optimality gap (1.22%) compared to the single-agent baseline (2.07%) within the same 10^6 evaluations budget.

Method	Optimality Gap (%) Small	Medium	Large	Computation time [s]	Training time [min]
ALNS-Vanilla	4.15	5.12	6.42	31.2	0.0
ALNS-BO	3.55	4.20	5.21	34.5	265.5
DRLH	2.10	3.15	4.50	73.3	1472.2
DR-ALNS	1.45	2.05	2.71	63.4	147.6
Island-DR-ALNS	**0.85**	**1.18**	**1.63**	78.4	180.3

Generalization & Transferability

Within-Problem Generalization

Island-DR-ALNS exhibits a significantly lower performance decay when scaling to unseen problem sizes (both Upward and Downward) compared to the monolithic agent.

Upward: Trained on Small				Downward: Trained on Large			
Method	Test: S	Test: L	Decay	Method	Test: L	Test: S	Decay
DR-ALNS	1.45	3.85	+1.14	DR-ALNS	2.71	1.85	+0.40
Island-DR-ALNS	**0.88**	**2.15**	**+0.52**	Island-DR-ALNS	**1.63**	**0.94**	**+0.06**

Zero-Shot Cross-Problem Transfer

The multi-agent framework learns universal search heuristics, achieving competitive zero-shot results on structurally distinct problems (TSP and mTSP).

Configuration	TSP gap	mTSP gap
Island-DR-ALNS (Trained on target)	*0.05*	*0.42*
DR-ALNS (Transferred from CVRP)	1.12	2.35
Island-DR-ALNS (Transferred from CVRP)	**0.35**	**1.15**

Conclusions

- NCS and socially-aware states successfully mitigate policy homogenization, driving parallel exploration of disjoint heuristic spaces.
- Island-DR-ALNS acts as a robust solver, ensuring highly competitive scaling and transfer capabilities without retraining.

19th International Conference on Parallel Problem Solving From Nature 2026, Trento, Italy