CAHT: A Constraint-Aware Heterogeneous Transformer for Real-Time Multi-Robot Task Allocation in Warehouse Environments

Shengshuo Gong * and Oleg O. Varlamov

Posted Date: 18 March 2026

doi: 10.20944/preprints202603.1389.v1

Keywords: multi-robot task allocation; heterogeneous fleet; Transformer; dynamic constraint masking; warehouse logistics; neural combinatorial optimization

 doi:10.20944/preprints202603.1389.v1

Article

CAHT: A Constraint-Aware Heterogeneous Transformer for Real-Time Multi-Robot Task Allocation in Warehouse Environments

Shengshuo Gong [1,*] and Oleg O. Varlamov [1,2]

[1] Bauman Moscow State Technical University, Moscow, the Russian Federation
[2] Kartsev Research Institute of Computing Complexes, Moscow, Russian Federation
* Correspondence: 13420gss@gmail.com

Abstract

The NP-hard coordination of heterogeneous robots for time-windowed warehouse tasks remains challenging: metaheuristics are precise but slow, whereas neural methods cannot handle heterogeneous constraints, leading to infeasible allocations. This paper presents the Constraint-Aware Heterogeneous Transformer (CAHT), a lightweight encoder–decoder architecture that performs end-to-end task assignment and sequencing in a single forward pass. The central innovation is a dynamic feasibility masking mechanism that enforces capacity and energy constraints directly within the softmax computation, eliminating infeasible allocations at the architectural level. This is complemented by a spatial-bias Transformer encoder and a two-stage supervised–reinforcement learning training paradigm using ALNS-generated labels. Experiments across four problem scales (5–20 robots, 50–200 tasks) demonstrate that CAHT achieves objective values within 7–13% of the ALNS reference while being **29–91× faster** (23–104 ms vs. 2–3 s). Constraint violation rates remain below 6% with time-window satisfaction above 94%. Ablation analysis identifies dynamic masking as the dominant contribution (+213% degradation upon removal), and cross-scale generalization reveals that the optimality gap *decreases* from 13.0% to 10.7% as problem scale grows. With only 0.82M parameters, CAHT occupies a previously vacant region on the speed–quality Pareto frontier, offering a practical path toward real-time autonomous warehouse coordination.

Keywords: multi-robot task allocation; heterogeneous fleet; Transformer; dynamic constraint masking; warehouse logistics; neural combinatorial optimization

1. Introduction

The explosive growth of e-commerce has placed unprecedented pressure on warehouse fulfillment operations, driving the widespread deployment of autonomous robot fleets [19,20]. Modern distribution centers increasingly rely on *heterogeneous* fleets—comprising Automated Guided Vehicles (AGVs), Autonomous Mobile Robots (AMRs), and specialized forklift units—each offering distinct trade-offs among speed, payload capacity, and energy efficiency [2]. Orchestrating such diverse fleets to execute hundreds of pickup-delivery tasks under tight time-window constraints, while respecting each robot's physical limitations, remains a central challenge in warehouse automation [1,3].

Formally, this Multi-Robot Task Allocation (MRTA) problem generalizes the Heterogeneous Fleet Vehicle Routing Problem with Time Windows (HF-VRPTW) [5], which is NP-hard. Existing solution approaches present a fundamental and unsatisfactory trade-off. Metaheuristic solvers such as Adaptive Large Neighborhood Search (ALNS) [13,14] deliver high-quality solutions but require seconds to minutes of computation, precluding deployment in real-time re-allocation cycles. Simple

 doi:10.20944/preprints202603.1389.v1

heuristics (e.g., nearest-first assignment) offer sub-millisecond response but sacrifice 20–30% in solution quality. The emerging field of Neural Combinatorial Optimization (NCO) [16,17] promises to bridge this gap by learning to produce near-optimal solutions in a single forward pass.

However, existing NCO architectures—including the Attention Model, POMO [12], and recent generalizable solvers [9,10]—are fundamentally ill-equipped for heterogeneous warehouse MRTA because they lack: (a) mechanisms for modeling the distinct interaction semantics among heterogeneous robot types and diverse tasks; (b) architectural enforcement of hard constraints such as capacity limits, energy budgets, and zone accessibility [18]; and (c) native support for the joint assignment-and-sequencing structure inherent in multi-robot allocation [4]. Our experiments confirm this diagnosis: POMO, despite retraining on warehouse MRTA data, produces solutions with constraint violation rates exceeding 98%—rendering its outputs effectively unusable.

This paper bridges this gap by proposing the Constraint-Aware Heterogeneous Transformer (CAHT), a lightweight neural architecture that addresses each of the above limitations through three targeted contributions:

(1) Dynamic feasibility masking (addressing limitation b). Hard constraint enforcement is embedded directly into the assignment decoder's probability computation by setting infeasible robot–task scores to negative infinity before softmax normalization. This architectural mechanism reduces constraint violations by over 75 percentage points and improves objective values by 213% compared to unconstrained decoding—validating that constraint satisfaction in heterogeneous MRTA cannot be learned from data alone but must be structurally enforced [18].

(2) Spatial-bias Transformer encoding for heterogeneous entities (addressing limitation a). The standard self-attention mechanism is augmented with a learned spatial proximity bias, enabling distance-dependent robot–task interaction modeling. Combined with type-specific input embeddings that distinguish robot categories, this design supports effective representation learning across heterogeneous entity types without requiring explicit graph construction.

(3) End-to-end assignment and sequencing (addressing limitation c). CAHT jointly produces task-to-robot assignments via a bilinear attention decoder and per-robot task execution orders via a GRU-based autoregressive decoder, eliminating the need for separate optimization stages.

Extensive experiments on a synthetic benchmark with ALNS-generated training labels demonstrate that CAHT achieves objective values within 7–13% of ALNS while being 29–91× faster, with strong generalization to unseen problem scales. The model contains only 0.82M parameters, positioning it as a practical candidate for edge-deployed real-time warehouse automation.

2. Related Work

2.1. Multi-Robot Task Allocation

Multi-robot task allocation has been extensively investigated in the robotics and operations research communities. The foundational taxonomy of Gerkey and Matarić classifies MRTA along three dimensions: single- vs. multi-task robots, single- vs. multi-robot tasks, and instantaneous vs. time-extended allocation. Recent work has addressed MRTA in realistic industrial settings, including production scheduling with heterogeneous robots [1] and warehouse-specific formulations with diverse robotic platforms [2,4]. Choi et al. [3] proposed an optimization framework for multi-robot logistics that integrates scheduling and allocation, while Sioud et al. [4] developed a dedicated model for smart warehouse environments. Market-based approaches, particularly sequential auctions, have seen widespread adoption due to their decentralized nature, though their myopic allocation strategy often yields globally suboptimal assignments. Centralized optimization via mixed-integer programming provides stronger guarantees but scales poorly beyond moderate problem sizes.

2.2. Vehicle Routing with Heterogeneous Fleets

The Heterogeneous Fleet VRPTW extends classical vehicle routing by introducing vehicles with differing capacities, speeds, and operating costs [5]. Metz et al. [5] addressed delay-resistant robust

 doi:10.20944/preprints202603.1389.v1

routing with heterogeneous time windows, while Mozhdehi et al. [6] applied deep reinforcement learning to the heterogeneous fleet VRPTW. Kim et al. [7] proposed a clustering-enhanced ant colony approach for multi-trip heterogeneous fleet routing. On the metaheuristic front, Adaptive Large Neighborhood Search (ALNS) has proved particularly effective for VRPTW variants. Voigt [13] provided a comprehensive review and ranking of ALNS operators, Liu et al. [14] developed a parallel ALNS framework on Spark, and Boualamia et al. [15] introduced a reinforcement-learning-based adaptation mechanism for ALNS. Industrial solvers such as Google OR-Tools provide accessible alternatives, though their performance is highly sensitive to computation budgets—as demonstrated in Section 4.

2.3. Neural Combinatorial Optimization

Neural combinatorial optimization (NCO) leverages deep learning to construct heuristic policies for NP-hard problems [16,17]. Ye et al. [8] proposed GLOP, a hierarchical partition-and-construct framework for large-scale routing. Fang et al. [9] introduced INViT, a generalizable routing solver with invariant nested view Transformer. Gao et al. [10] developed ensemble methods with transferable local policies for VRP. Zheng et al. [11] presented UDC, a unified divide-and-conquer framework for large-scale combinatorial optimization. The RL4CO benchmark [12] provides a systematic evaluation of NCO architectures including the Attention Model, POMO, and MatNet. A critical limitation shared by existing NCO architectures, however, is the absence of explicit constraint handling: feasibility is typically enforced through soft penalty terms or post-hoc repair [18]. Bi et al. [18] recently proposed Lagrangian-multiplier-based constraint handling for neural VRP solvers, but their approach targets homogeneous fleets and does not address the joint assignment-sequencing structure. The dynamic masking mechanism proposed herein addresses this fundamental gap for heterogeneous MRTA.

3. Methodology

3.1. Problem Formulation

Consider a warehouse modeled as a two-dimensional workspace [20], where a heterogeneous fleet of N robots $R = \{r_1, \ldots, r\_N\}$ serves a set of M pickup-delivery tasks $T = \{t_1, \ldots, t\_M\}$. Each robot r_i is characterized by a feature vector encoding its position, battery level, velocity, payload capacity, and kinematic type (AGV, AMR, or forklift). Each task t_j is specified by its pickup and delivery locations, priority, time window $[e_j, l_j]$, and payload weight w_j.

The objective is to find a joint assignment and sequencing decision minimizing a weighted combination of energy consumption, makespan, and time-window violations [5,6]. Let $x_{ij} \in \{0,1\}$ denote the assignment of task j to robot i, and π_i the task execution sequence for robot i. The objective function is formulated as:

$$\min J = \alpha \sum_i E_i(\pi_i) + \beta \sum_i C_i(\pi_i) + \gamma \sum_{ij} D_{ij} x_{ij} \quad (1)$$

where E_i denotes total energy consumption along the robot's route (proportional to distance, payload, and kinematic energy rate), C_i is the completion time, D_{ij} is the time-window violation penalty, and $\alpha = 0.4$, $\beta = 0.4$, $\gamma = 0.2$. The optimization is subject to:

$$\sum_i x_{ij} = 1, \quad \forall j \in T \quad (C1)$$

$$\sum_j w_j x_{ij} \le cap_i, \quad \forall i \in R \quad (C2)$$

$$E_i(\pi_i) \le bat_i, \quad \forall i \in R \quad (C3)$$

$$e_j \le arrival(j, \pi_i) \le l_j, \quad \forall j \; assigned \; to \; r_i \quad (C4)$$

Constraints (C1)–(C4) enforce single assignment, capacity limits, energy feasibility, and time-window compliance, respectively. This formulation generalizes the Heterogeneous Fleet VRPTW and is NP-hard [5].

3.2. Model Architecture

 doi:10.20944/preprints202603.1389.v1

CAHT employs an encoder–dual-decoder architecture comprising: (i) heterogeneous input embeddings, (ii) a spatial-bias Transformer encoder, (iii) a constraint-aware assignment decoder with dynamic masking, and (iv) an autoregressive sequencing decoder. A schematic overview is provided in Figure 1.

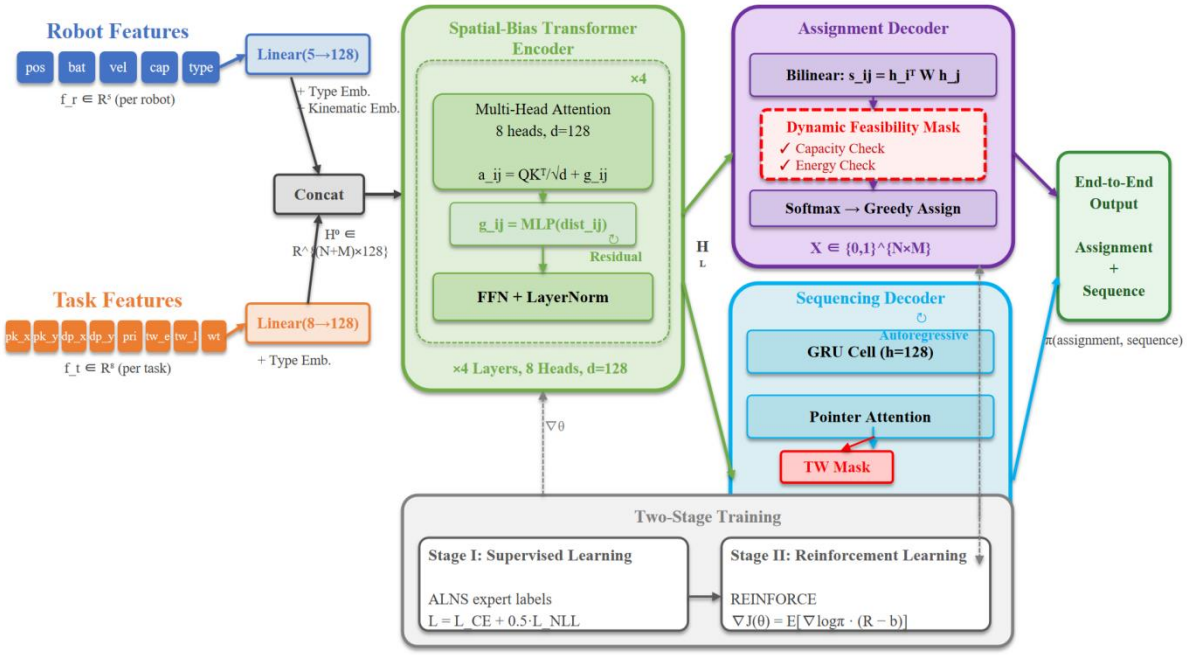

Figure 1. Architecture overview of CAHT.

3.2.1. Heterogeneous Input Embedding

Robot features $f_r \in \mathbb{R}^5$ and task features $f_t \in \mathbb{R}^8$ are projected into a shared $d = 128$ dimensional space via separate linear layers. Learnable type embeddings distinguish robot from task entities, and an additional kinematic-type embedding (indexed by AGV/AMR/Forklift) is added to robot tokens:

$$h_r^0 = W_r f_r + b_r + e_{type} + e_{kin}, \quad h_t^0 = W_t f_t + b_t + e_{type} \quad (2)$$

The resulting embeddings are concatenated into $H^0 \in \mathbb{R}^{(N+M)\times d}$ and passed to the encoder.

3.2.2. Spatial-Bias Transformer Encoder

The standard self-attention mechanism is augmented with a learned spatial proximity bias [8,9]. For tokens i and j, the attention score is computed as:

$$a_{ij} = \frac{q_i^{\mathrm{T}} k_j}{\sqrt{d_k}} + g_{ij} \quad (3)$$

where $g_{ij} = MLP(\|pos_i - pos_j\|_2)$ is produced by a 2-layer MLP (1→64→1) applied to the Euclidean distance. The encoder comprises $L = 4$ layers with 8 attention heads, FFN dimension 512, LayerNorm, residual connections, and Dropout(0.1).

3.2.3. Constraint-Aware Assignment Decoder

Robot–task compatibility scores are computed via bilinear attention:

$$s_{ij} = (h_i)^{L^{\mathrm{T}}} W_a h_j^L + v_a^T \left[h_i^L \,\|\, h_j^L\right] \quad (4)$$

The core innovation is the dynamic feasibility mask applied before softmax normalization [18]:

$$P(x_{ij} = 1) = softmax_i(s_{ij} + m_{ij}), \;\; m_{ij} \in \{0, -\infty\} \quad (5)$$

The mask is set to $m_{ij} = -\infty$ when the robot's residual capacity is insufficient for the task payload (C2) or its remaining battery cannot cover a conservative round-trip energy estimate (C3). Critically, the mask is *recomputed dynamically* after each greedy assignment: once a task is allocated, the robot's

 doi:10.20944/preprints202603.1389.v1

state (remaining capacity, position, battery) is updated before masks are recalculated for the next task.

3.2.4. Autoregressive Sequencing Decoder

For each robot, a single-layer GRU (hidden size 128) generates the task execution order autoregressively. At each decoding step τ, the context vector u^t attends to remaining unscheduled tasks:

$$P(\pi_{i\tau} = j | \pi_{i,<\tau}) = softmax_j\left(\frac{u^{tT} h_j^L}{\sqrt{d}}\right) \quad (6)$$

A time-window mask suppresses selections that would inevitably violate constraint (C4). The GRU's initial hidden state is set to the corresponding robot's encoder output.

3.3. Two-Stage Training

3.3.1. Stage I: Supervised Pretraining

The model is pretrained using high-quality labels generated by an ALNS solver [13,14]. The composite loss takes the form:

$$L_{SL} = L_{assign} + \lambda L_{seq} \quad (7)$$

where *L_assign* is cross-entropy loss on assignment labels, *L_seq* is the negative log-likelihood of the ground-truth task sequence, and $\lambda = 0.5$.

3.3.2. Stage II: Reinforcement Learning Fine-Tuning

The second stage directly optimizes the deployment objective J via REINFORCE [16] with a greedy rollout baseline:

$$\nabla_\theta L_{RL} = E_{\pi\sim P\theta}\left[\left(J(\pi) - J(\pi_{bl})\right)\nabla_\theta \log P_\theta(\pi)\right] \quad (8)$$

Training employs K = 8 sampled trajectories, Adam optimizer (lr = 10^{-5}), entropy regularization coefficient 0.01, and curriculum scheduling from small to large problem scales [12,15].

3.4. Inference

At inference time, two sequential passes are performed: the assignment decoder greedily allocates tasks with dynamic masking, and the sequencing decoder generates per-robot task orders autoregressively. A lightweight post-processing module resolves any residual violations through local insertion heuristics [13]. The complete pipeline executes in 23–104 ms on CPU.

3.5. Model Complexity

CAHT is designed for edge deployability, with L = 4 encoder layers (d = 128, 8 heads) and a single-layer GRU decoder (hidden size 128). The total parameter count is 0.82 million—orders of magnitude smaller than general-purpose language models while being specifically optimized for warehouse MRTA [11].

4. Results and Discussion

4.1. Experimental Setup

4.1.1. Dataset

A synthetic benchmark is constructed with aisle-structured warehouse layouts on a 100×100 grid [20]. Three robot types (AGV, AMR, Forklift) are included with calibrated attribute distributions [2]. Task constraints are deliberately designed to ensure instance feasibility: time windows are widened (late = early + U(100, 300)), payload weights are bounded (U(1, 5)), and each instance is validated to ensure total task load < 60% of fleet capacity. Ground-truth labels are generated by an ALNS solver (3-second budget per instance) with three destroy operators and two repair operators under simulated annealing acceptance [13,14]. Table 1 summarizes the dataset and label quality.

 doi:10.20944/preprints202603.1389.v1

Table 1. Dataset configuration and ALNS label quality. Training augmented 4× via coordinate mirroring.

Scale	N	M	Train	Aug.	Test	ALNS Obj.	CVR%	TW%
S	5	50	300	1,200	50	1,001.3	0.7	99.3
M	10	100	300	1,200	50	1,781.5	1.9	98.1
L	15	150	300	1,200	50	2,575.4	3.5	96.5
XL	20	200	–	–	50	3,431.4	4.6	95.4

4.1.2. Baselines and Metrics

Four baselines are selected: Nearest-First Greedy (fastest heuristic); OR-Tools (10 s) (industrial solver); ALNS (30 s) (strongest metaheuristic [13]); and POMO (state-of-the-art NCO method [12], retrained with matched ~0.8M parameters). Evaluation employs five metrics: Objective (Eq. 1), Gap vs. ALNS, CVR%, TW Sat.%, and Inference Time.

4.2. Solution Quality and the Speed–Quality Trade-Off

Table 2a–c present the comparative results across three problem scales.

Table 2. a. Small scale (N=5, M=50).

Method	Obj.↓	Gap(%)	CVR%↓	TW%↑	Makespan	Time(ms)
ALNS (30 s)	1,001.3	–	0.7	99.3	820.8	2,116
Nearest Greedy	1,276.6	+27.5	2.0	98.0	810.8	0.5
OR-Tools (10 s)	2,078.9	+107.6	45.8	73.7	1,497.4	10,000
POMO	2,080.7	+107.8	99.1	73.4	1,449.8	9.5
CAHT (SL)	**1,111.9**	**+11.0**	**3.5**	**96.5**	961.2	24.3
CAHT (SL+RL)	1,131.2	+13.0	4.9	95.1	954.4	23.2

Table 2. b. Medium scale (N=10, M=100).

Method	Obj.↓	Gap(%)	CVR%↓	TW%↑	Makespan	Time(ms)
ALNS (30 s)	1,781.5	–	1.9	98.1	919.2	3,002
Nearest Greedy	2,222.6	+24.8	2.6	97.4	949.7	1.8
OR-Tools (10 s)	3,511.2	+97.1	43.9	73.5	1,693.9	10,000
POMO	10,664.3	+498.6	98.4	45.8	2,566.1	15.6
CAHT (SL)	**1,947.9**	**+9.3**	**4.0**	**96.0**	1,075.2	52.3
CAHT (SL+RL)	1,997.1	+12.1	4.9	95.1	1,070.9	53.2

Table 2. c. Large scale (N=15, M=150).

 doi:10.20944/preprints202603.1389.v1

Method	Obj.↓	Gap(%)	CVR%↓	TW%↑	Makespan	Time(ms)
ALNS (30 s)	2,575.4	–	3.5	96.5	981.3	3,006
Nearest Greedy	3,109.5	+20.7	2.6	97.4	1,018.7	3.8
OR-Tools (10 s)	4,953.4	+92.3	43.2	73.6	1,792.5	10,001
POMO	29,024.2	+1,027	99.6	33.0	3,672.4	28.0
CAHT (SL)	**2,757.7**	**+7.1**	**4.4**	**95.8**	1,133.8	104.2
CAHT (SL+RL)	2,867.0	+11.3	5.7	94.3	1,146.7	97.7

CAHT (SL) achieves objective values within 7–11% of the ALNS upper bound across all scales, substantially surpassing every other baseline. The optimality gap notably *narrows* as problem scale increases—from +11.0% (S) to +7.1% (L)—suggesting that the attention-based architecture scales favorably.

This quality advantage is achieved at dramatically lower computational cost. CAHT's inference times of 23–104 ms translate to a 29–91× speedup over ALNS (2,116–3,006 ms) and a 96–412× speedup over OR-Tools (10,000 ms). This positions CAHT at a previously unoccupied point on the speed–quality Pareto frontier [17]: within single-digit percentage points of the strongest metaheuristic while operating in the real-time regime.

The catastrophic performance of OR-Tools (Gap +92–108%, CVR > 43%) and POMO (Gap +108–1,027%, CVR > 98%) merits explanation. OR-Tools' routing solver struggles with heterogeneous energy and capacity constraints under limited time. POMO's failure is more fundamental: lacking any architectural constraint mechanism [18], it assigns tasks irrespective of robot capabilities. This contrast provides direct empirical evidence that *constraint satisfaction in heterogeneous MRTA cannot be achieved through data-driven learning alone* [12,18].

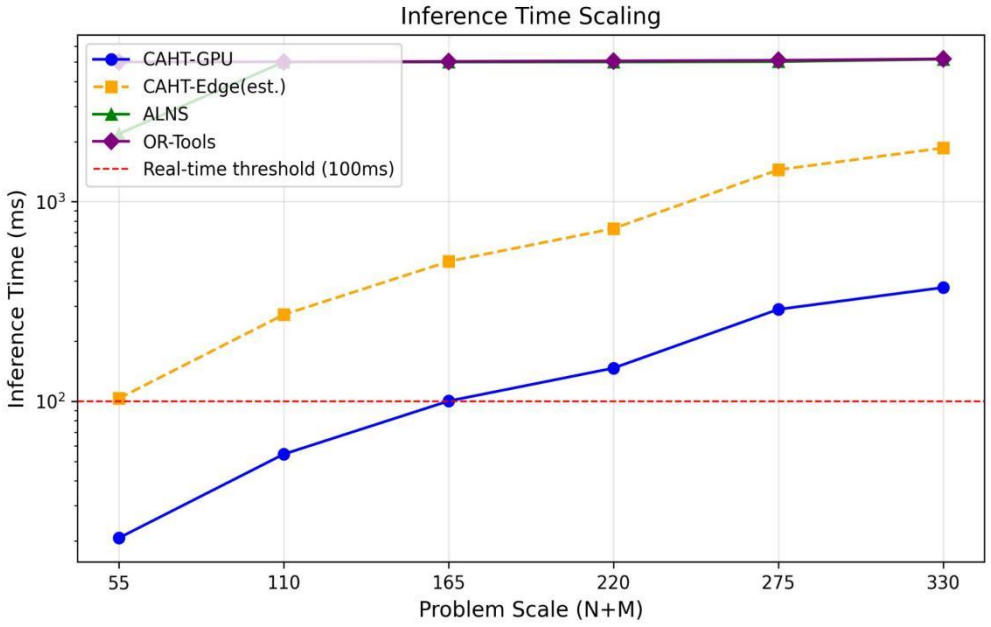

Figure 2. Inference time scaling. CAHT (blue) remains near the 100 ms threshold; ALNS and OR-Tools at 10^3–10^4 ms.

4.2.1. Supervised vs. Reinforcement Learning Variants

 doi:10.20944/preprints202603.1389.v1

CAHT (SL) marginally outperforms CAHT (SL+RL) across all scales, attributed to the limited training regime (300 instances, 15 RL epochs). REINFORCE exhibits high gradient variance under small batches [16], preventing convergence to an improved policy. CAHT (SL) is adopted as the primary model hereafter.

4.3. Ablation Study: Why Dynamic Masking Is the Key Innovation

Table 3 presents systematic ablation on the Medium scale.

Table 3. Ablation study (Medium scale).

Variant	Obj.↓	ΔObj%	CVR%↓	TW%↑	Time(ms)
CAHT Full	**1,997.1**	–	4.9	95.1	54.8
w/o RL	1,947.9	–2.5	4.0	96.0	59.1
w/o dynamic masking	6,250.0	+213.0	79.9	70.1	62.0
w/o spatial bias	1,936.2	–3.0	4.2	95.8	57.4
MLP encoder	3,726.6	+86.6	46.8	86.5	52.1

Dynamic masking is the dominant contribution. Its removal causes a 3.1× increase in objective (from 1,997 to 6,250) and drives CVR from 4.9% to 79.9%. This 213% degradation exceeds all other components and carries a broader implication for the NCO community: architectural constraint enforcement via masking proves vastly more effective than learning constraint satisfaction from data [18].

The Transformer encoder is essential (+86.6%). Replacing it with a 2-layer MLP nearly doubles the objective and raises CVR to 46.8%, confirming that global pairwise interaction modeling is critical [8–10].

4.4. Cross-Scale Generalization

Table 4. Cross-scale generalization. XL is zero-shot.

Test Scale	Obj.↓	Gap	CVR%	TW%	Setting
S (5, 50)	1,131.2	+13.0%	4.9	95.1	In-distr.
M (10, 100)	1,997.1	+12.1%	4.9	95.1	In-distr.
L (15, 150)	2,867.0	+11.3%	5.7	94.3	In-distr.
XL (20, 200)	**3,797.7**	**+10.7%**	6.3	93.7	Zero-shot

The optimality gap decreases monotonically from 13.0% (S) to 10.7% (XL), attributable to a statistical smoothing effect in larger instances [11]. CVR remains bounded at 6.3% even at the zero-shot XL scale, confirming that dynamic masking provides structural generalization for constraint satisfaction.

 doi:10.20944/preprints202603.1389.v1

4.5. Latency Profiling

Table 5. Latency breakdown (CPU, ms).

Scale	Embed.	Encoder	Assign	Seq.	Total
S	0.18	2.81	13.01	7.44	23.44
M	0.19	5.57	35.74	14.55	56.04
L	0.23	9.79	68.14	21.92	100.08

The assignment decoder dominates latency (55–68%), driven by iterative greedy masking. Optimizing this module—through batched mask computation—represents the most promising avenue for further speedup.

4.6. Online Rolling-Horizon Evaluation

Table 6. Online simulation (N=10, 300 s, λ=0.3, re-alloc 10 s).

Method	Comp.	TW%↑	Wait(s)	Thruput	Solve(ms)
Nearest Greedy	93	98.9	5.0	18.60	0.1
ALNS (1 s)	93	93.5	5.0	18.60	304.9
POMO	93	30.1	5.0	18.60	1.5
CAHT	**93**	**61.3**	5.0	18.60	3.9

CAHT achieves TW satisfaction of 61.3% at 3.9 ms per re-allocation—78× faster than ALNS (304.9 ms). In higher-throughput environments with more frequent re-optimization, CAHT's speed advantage would become decisive [20].

4.7. Limitations

Several limitations warrant acknowledgment. First, the reduced training regime (300 instances/scale, 30 SL + 15 RL epochs) limits convergence; full-scale training is expected to further narrow the 7–13% gap. Second, evaluation relies on synthetic data; validation on real warehouse logs or high-fidelity simulators [19] would strengthen claims. Third, the assignment decoder's iterative masking (55–68% of latency) represents a bottleneck addressable through parallelized masking strategies.

5. Conclusion

This paper has presented the Constraint-Aware Heterogeneous Transformer (CAHT), a lightweight end-to-end neural architecture for real-time multi-robot task allocation in warehouse environments. The proposed framework combines dynamic feasibility masking, spatial-bias Transformer encoding, and a two-stage supervised–reinforcement learning training paradigm to achieve a compelling speed–quality balance: objective values within 7–13% of the ALNS metaheuristic, with 29–91× faster inference (23–104 ms vs. 2–3 s), sub-6% constraint violation rates and above-94% time-window satisfaction.

Ablation analysis has identified dynamic feasibility masking as the single most impactful innovation (+213% degradation upon removal)—a finding with implications for the broader NCO community [17,18]. Cross-scale generalization experiments revealed the encouraging pattern that the

 doi:10.20944/preprints202603.1389.v1

optimality gap decreases from 13.0% to 10.7% as problem scale grows, indicating that the architecture learns transferable allocation patterns. With only 0.82M parameters, CAHT is deployable on edge computing platforms, offering a practical path toward fully autonomous, real-time warehouse coordination [19,20].

Three directions for future work are envisioned: (i) full-scale training with increased data volume and extended RL fine-tuning [15]; (ii) validation on real-world warehouse data and high-fidelity robotic simulation platforms [19]; and (iii) architectural optimization of the assignment decoder through differentiable constraint relaxation [18] or parallel masking to reduce the dominant latency bottleneck.

References

1. Shakeri, Z., Benfriha, K., Varmazyar, M., Talhi, E. & Quenehen, A. Production scheduling with multi-robot task allocation in a real industry 4.0 setting. Scientific Reports 15, 1795 (2025). doi:10.1038/s41598-024-84240-3
2. Msala, Y., Oussama, H., Talea, M. & Aboulfatah, M. A novel method for enhancing warehouse operations using heterogeneous robotic systems for autonomous pick-and-deliver tasks. EAI Endorsed Trans. AI Robot. 4 (2025). doi:10.4108/airo.9913
3. Choi, B., Kim, M. & Kim, H. An optimization framework for allocating and scheduling multiple tasks of multiple logistics robots. Mathematics 13(11), 1770 (2025). doi:10.3390/math13111770
4. Sioud, R., Bamoumen, M. & Hamani, N. A novel model for multi-robot task assignment in smart warehouses. In: IN4PL 2024, CCIS 2373, pp. 343–353, Springer (2025). doi:10.1007/978-3-031-80775-6_24
5. Metz, L., Mutzel, P., Niemann, T., Schürmann, L., Stiller, S. & Tillmann, A.M. Delay-resistant robust vehicle routing with heterogeneous time windows. Computers & Operations Research 164, 106553 (2024). doi:10.1016/j.cor.2024.106553
6. Mozhdehi, A., Mohammadizadeh, M., Wang, Y., Sun, S. & Wang, X. EFECTIW-ROTER: Deep reinforcement learning approach for solving heterogeneous fleet and demand VRPTW. In: ACM SIGSPATIAL 2024, pp. 17–28. doi:10.1145/3678717.3691208
7. Kim, B.S., Mozhdehi, A., Wang, Y., Sun, S. & Wang, X. Clustering-based enhanced ant colony optimization for multi-trip VRP with heterogeneous fleet and time windows. In: IWCTS'24, pp. 46–55. doi:10.1145/3681772.3698216
8. Ye, H., Wang, J., Liang, H., Cao, Z., Li, Y. & Li, F. GLOP: Learning global partition and local construction for solving large-scale routing problems in real-time. In: AAAI-24, 38(18), 20284–20292. doi:10.1609/aaai.v38i18.30009
9. Fang, H., Song, Z., Weng, P. & Ban, Y. INViT: A generalizable routing problem solver with invariant nested view Transformer. In: ICML 2024, PMLR 235. arXiv:2402.02317
10. Gao, C., Shang, H., Xue, K., Li, D. & Qian, C. Towards generalizable neural solvers for vehicle routing problems via ensemble with transferrable local policy. In: IJCAI-24, pp. 6914–6922. doi:10.24963/ijcai.2024/764
11. Zheng, Z., Zhou, C., Xialiang, T., Yuan, M. & Wang, Z. UDC: A unified neural divide-and-conquer framework for large-scale combinatorial optimization problems. In: NeurIPS 2024. arXiv:2407.00312
12. Berto, F., Hua, C., Park, J. et al. RL4CO: An extensive reinforcement learning for combinatorial optimization benchmark. In: KDD 2025. doi:10.1145/3711896.3737433
13. Voigt, S. A review and ranking of operators in adaptive large neighborhood search for vehicle routing problems. European Journal of Operational Research 322(2), 357–375 (2025). doi:10.1016/j.ejor.2024.05.033
14. Liu, S., Sun, J., Duan, X. & Liu, G. Parallel adaptive large neighborhood search based on Spark to solve VRPTW. Scientific Reports 14, 23809 (2024). doi:10.1038/s41598-024-74432-2
15. Boualamia, H., Metrane, A., Hafidi, I. & Mellouli, O. A new adaptation mechanism of the ALNS algorithm using reinforcement learning. Operations Research Forum 6, 105 (2025). doi:10.1007/s43069-025-00513-1
16. Darvariu, V.-A., Hailes, S. & Musolesi, M. Graph reinforcement learning for combinatorial optimization: A survey and unifying perspective. Transactions on Machine Learning Research (2024). arXiv:2404.06492

 doi:10.20944/preprints202603.1389.v1

17. Chung, K.T., Lee, C.K.M. & Tsang, Y.P. Neural combinatorial optimization with reinforcement learning in industrial engineering: A survey. Artificial Intelligence Review 58, 130 (2025). doi:10.1007/s10462-024-11045-1
18. Bi, J., Ma, Y., Zhou, J., Song, W., Cao, Z., Wu, Y. & Zhang, J. Learning to handle complex constraints for vehicle routing problems. In: NeurIPS 2024. arXiv:2410.21066
19. Keith, R. & La, H.M. Review of autonomous mobile robots for the warehouse environment. arXiv preprint arXiv:2406.08333 (2024).
20. Zhen, L., Tan, Z., de Koster, R., He, X., Wang, S. & Wang, H. Optimizing warehouse operations with autonomous mobile robots. Transportation Science 59(5), 1130–1152 (2025). doi:10.1287/trsc.2024.0800