JOURNAL OF GEOSPACE SCIENCE

Vol. 6 No. 2 (2026)

E-ISSN : 2413-6093
https://jogss.org

Horizon-Adaptive Reinforcement Learning for Vision-Language Navigation in Unseen Environments

Michael M. Hernandez; Rachel Nelson

Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, Champaign, IL 61801, USA

ABSTRACT

Vision-and-language navigation represents a fundamental challenge in the development of embodied artificial intelligence, requiring an agent to interpret unconstrained natural language instructions and execute sequential navigational actions within complex, photorealistic environments. Despite significant advancements in cross-modal representation learning and navigation strategies, existing models frequently struggle to generalize to unseen environments. A primary bottleneck lies in the rigid, predefined temporal horizons employed by standard reinforcement learning frameworks, which force agents into either premature termination or persistent wandering when faced with environmental ambiguities. To address this limitation, this paper introduces a horizon-adaptive reinforcement learning framework designed to dynamically modulate the temporal bounds of an episode based on the real-time alignment between the agent trajectory and the linguistic instruction. By continuously assessing the semantic progress of the navigation task, the proposed architecture optimally expands or contracts the action horizon, thereby encouraging efficient exploration while mitigating catastrophic error accumulation in unfamiliar topologies. Comprehensive evaluations conducted on standard benchmark datasets demonstrate that the horizon-adaptive approach substantially reduces the generalization gap between seen and unseen environments. The framework achieves superior performance across primary navigation metrics, particularly in the success rate weighted by path length, indicating a more efficient and human-like navigation paradigm. The findings suggest that dynamic temporal constraints are essential for robust decision-making in embodied agents.

KEYWORDS

vision-language navigation, reinforcement learning, embodied artificial intelligence, cross-modal learning

1. Introduction

1.1 Background of Vision-Language Navigation

The pursuit of artificial intelligence has progressively transitioned from passive perception tasks, such

as image classification and object detection, to active, embodied tasks where agents must interact with their environments to achieve specific goals [1]. Vision-and-language navigation has emerged as a cornerstone of this transition, sitting at the intersection of computer vision, natural language processing, and robotics. In a typical vision-and-language navigation setup, an autonomous agent is situated in a visually rich, three-dimensional environment and provided with a natural language instruction detailing a route to a specific destination [2]. The agent must decode the linguistic cues, ground them in its egocentric visual observations, and execute a sequence of discrete or continuous motor actions to navigate through the space [3]. The complexity of vision-and-language navigation arises from the profound semantic gap between the high-level abstractions of human language and the low-level pixel data perceived by the agent [4]. Human instructions are often characterized by inherent ambiguities, temporal dependencies, and spatial references that require extensive contextual understanding. For instance, an instruction such as walking past the wooden table and stopping near the leather sofa requires the agent to sequentially identify objects, understand spatial relationships, and maintain an internal memory of its past actions to determine when the task is complete [5]. Early approaches to this problem primarily relied on sequence-to-sequence learning methodologies, where the task was framed as translating a sequence of words into a sequence of actions [6]. While these supervised learning paradigms demonstrated initial success, they fundamentally lacked the capability to handle the sequential decision-making process required in dynamic or unfamiliar settings, leading to the adoption of reinforcement learning techniques.

1.2 Challenges in Unseen Environments

A critical and persistent challenge in vision-and-language navigation is the severe degradation of agent performance when deployed in unseen environments. Current machine learning architectures exhibit a strong propensity to overfit to the visual and topological characteristics of the environments encountered during the training phase [7]. When an agent is placed in a novel environment, the variations in lighting conditions, object appearances, and spatial configurations disrupt the learned cross-modal alignments, resulting in catastrophic navigation failures [8]. This generalization gap remains a primary obstacle preventing the deployment of embodied agents in real-world applications, such as domestic robotics, search and rescue operations, and automated delivery systems [9]. Standard reinforcement learning approaches, which optimize navigation policies through trial and error using environmental rewards, have been employed to improve generalization [10]. However, these traditional frameworks typically operate under a fixed temporal horizon, meaning the agent is granted a hard-coded maximum number of steps to reach the destination [11]. In unseen environments, this fixed horizon becomes highly problematic. When an agent becomes disoriented due to novel visual stimuli, a fixed maximum step count often forces it into endless loops or random wandering, accumulating massive trajectory errors and diluting the learning signal [12]. Conversely, if the horizon is set too short, the agent is unfairly penalized for cautious exploration and may fail to complete complex, multi-step instructions [13]. To overcome these structural limitations, this paper proposes a horizon-adaptive reinforcement learning framework that dynamically adjusts the maximum allowable steps based on the ongoing cross-modal alignment between the agent state and the remaining instruction. By integrating a dynamic horizon module, the agent can intelligently request additional time for exploration when it perceives it is making meaningful progress, or forcefully terminate the episode early when the confidence in the current trajectory drops below a critical threshold. This mechanism fundamentally alters the exploration-exploitation dynamics, fostering highly efficient and resilient navigation policies tailored for unpredictable, unseen environments.

2. Literature Review

2.1 Foundational Models in Vision-Language Navigation

The formalization of vision-and-language navigation as a benchmark task was largely accelerated by the development of photorealistic simulators based on captured indoor environments. Initial baseline models approached the task through standard teacher-forcing behavioral cloning, where the agent was trained to mimic the exact trajectory generated by a human demonstrator [14]. These models utilized recurrent neural networks with attention mechanisms to continuously weigh different parts of the language instruction based on the current visual input [15]. Despite achieving reasonable accuracy on the training environments, these purely supervised models suffered immensely from exposure bias [16]. Exposure bias occurs because the agent is only exposed to optimal states during training; thus, when it inevitably makes a suboptimal action during inference, it finds itself in an unknown state distribution and is unable to recover [17]. To mitigate exposure bias, subsequent research introduced student-forcing techniques and data augmentation strategies. For example, some models synthesized massive amounts of artificial instructions for generated trajectories using a speaker model, which allowed the navigation agent to learn from a broader distribution of visual-linguistic pairings [18]. Other architectures incorporated sophisticated cross-modal transformers that leverage multi-head self-attention to establish deep contextual relationships between visual regions and linguistic tokens [19]. While transformer-based representations significantly improved the agent ability to ground language in vision, the fundamental problem of how to explore novel environments safely without relying solely on mimicry required a paradigm shift toward sequential decision-making frameworks [20].

2.2 Reinforcement Learning Formulations

Reinforcement learning provides a natural mathematical framework for vision-and-language navigation, casting the problem as a Markov Decision Process where the agent learns to maximize an expected cumulative reward [21]. Early adoptions of reinforcement learning in this domain utilized actor-critic architectures, specifically the Advantage Actor-Critic algorithm, to balance the immediate rewards of following an instruction with the sparse, delayed reward of reaching the final destination [22]. Researchers designed complex reward shaping functions, incorporating intermediate rewards for moving closer to the target, aligning the trajectory with the dynamic time warping distance of the human path, and penalizing collisions [23]. However, the rigid episodic structure of standard reinforcement learning environments limits the agent adaptability [24]. In almost all existing literature, the vision-and-language navigation episode is bounded by a fixed maximum length to prevent infinite loops during training [25]. Studies have shown that this fixed horizon inherently biases the learned policy. Agents trained with long horizons tend to learn hesitant, meandering behaviors, whereas agents trained with short horizons become overly aggressive and fail on longer instructions [26]. Some recent works attempted to address this by introducing a dedicated stop action, allowing the agent to terminate the episode when it believes it has arrived [27]. While this stop action allows for episodes shorter than the maximum horizon, it does not solve the issue of the agent wandering aimlessly until the maximum limit is reached if the stop action is not triggered [28]. The concept of a dynamic time limit has been explored in isolated subfields of robotics, particularly in continuous control tasks where energy efficiency is paramount [29]. Yet, its application to cross-modal navigation remains largely uncharted [30]. The proposed horizon-adaptive framework bridges this gap by continuously evaluating the semantic progress of the task, thereby providing a theoretically sound method for modulating the episodic boundary in real-time. This dynamic modulation not only accelerates the convergence of the reinforcement learning algorithm but also structurally prevents the catastrophic wandering behaviors that plague models in unseen topologies.

3. Methodology and Architecture Design

3.1 Cross-Modal Representation Learning

The foundation of the proposed horizon-adaptive framework is a robust cross-modal representation module designed to process and fuse unconstrained text and photorealistic visual observations. At any given time step, the agent receives an egocentric panoramic view of its current location, which is discretized into a set of discrete viewing angles. The visual processing pipeline extracts high-level semantic features from each view using a convolutional neural network pre-trained on large-scale image classification tasks, supplemented by spatial geometry features that encode the elevation and heading angles of each view. Simultaneously, the natural language instruction is processed using a bidirectional transformer encoder. The instruction is tokenized into a sequence of sub-word units, and each token is embedded into a dense continuous vector space. The transformer encoder captures the intricate syntactic and semantic dependencies within the instruction, producing a set of contextualized word representations. A global instruction representation is also generated, typically utilizing the pooling of the initial sequence token, which serves as a macroscopic summary of the intended navigation task.

Table 1State Representation Features

Feature Type	Dimensionality	Extraction Method	Purpose
Visual Appearance	Two thousand and forty eight	Convolutional Neural Network	Scene recognition and object grounding
Spatial Geometry	One hundred and twenty eight	Depth Sensor Processing	Obstacle avoidance and topology mapping
Linguistic Tokens	Seven hundred and sixty eight	Transformer Encoder	Instruction comprehension and alignment
Temporal History	Five hundred and twelve	Recurrent State Cell	Memory of previous actions and path taken

To bridge the visual and linguistic modalities, the architecture employs a deep co-attention mechanism. The co-attention layer computes a similarity matrix between every visual region and every contextualized word, allowing the model to dynamically highlight the specific words in the instruction that correspond to the objects and directional cues currently visible in the environment. This fused multi-modal representation serves as the comprehensive state encoding for the reinforcement learning agent. The state encoding must encapsulate not only where the agent is and what it is looking at, but also what parts of the instruction have already been completed and what actions are required next.

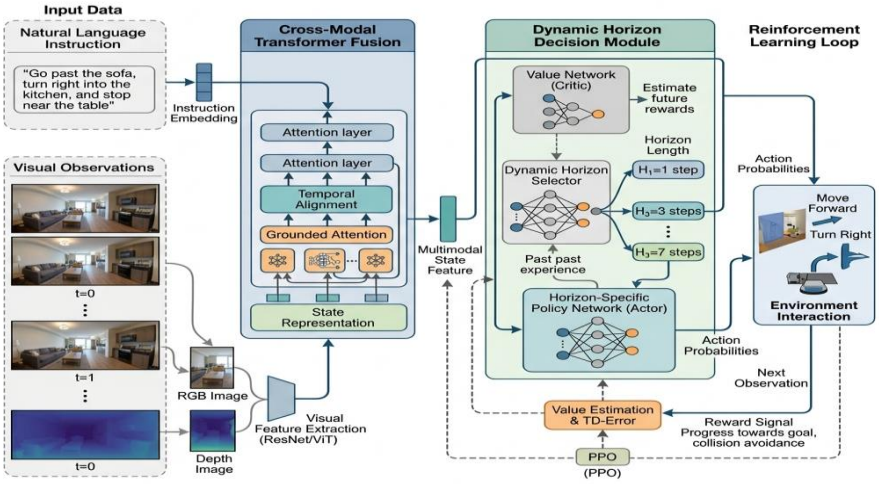


Figure 1Comprehensive Schematic of Horizon

3.2 Horizon-Adaptive Reinforcement Learning Formulation

The core innovation of the proposed methodology is the horizon-adaptive reinforcement learning formulation. In a traditional formulation, the navigation task is an episodic process with a hard constraint on the number of steps. In contrast, our framework introduces a dynamic variable representing the remaining allowable steps, which is updated at each time step by a dedicated horizon modulation network. The state space consists of the fused cross-modal representation, the history of previous actions, and the current value of the dynamic horizon. The action space includes the standard discrete navigation movements, such as turning left, turning right, tilting up, tilting down, and moving forward to a connected viewpoint, as well as the terminal stop action. The policy network, acting as the primary actor, maps the current state to a probability distribution over the available actions. Concurrently, the horizon modulation network acts as an auxiliary critic. This network takes the current cross-modal state and the history of recent states to predict a progress scalar, representing the estimated percentage of the instruction that has been successfully executed. If the progress scalar indicates steady advancement toward the goal, the horizon modulation network extends the available time limit, granting the agent the necessary freedom to execute complex maneuvers. Conversely, if the progress scalar stagnates or the cross-modal attention weights exhibit high entropy, indicating uncertainty and disorientation, the network aggressively decays the remaining horizon. This mechanism forces early termination, preventing the agent from polluting its trajectory memory with random, unguided exploration in unseen environments.

Table 2Reward Shaping Components

Reward Component	Condition	Value Assignment	Objective
Goal Achievement	Agent stops within target radius	Large positive constant	Primary task completion

Distance Reduction	Agent moves closer to target	Proportional positive value	Dense guidance along path
Horizon Penalty	Agent triggers time extension	Small negative constant	Encourage efficiency
Disorientation Penalty	Agent revisits same state	Moderate negative constant	Prevent endless looping

The training objective is optimized using a proximal policy optimization algorithm. The reward function is meticulously designed to support the horizon-adaptive mechanism. The agent receives a substantial positive reward for successfully stopping within the target radius. Dense intermediate rewards are provided based on the reduction in geodesic distance to the goal. Importantly, a minor time penalty is applied whenever the horizon modulation network extends the episode limit, which prevents the agent from exploiting the mechanism to gain infinite exploration time. The combination of proximal policy optimization with the dynamic temporal boundaries creates a highly stable learning environment, allowing the agent to develop policies that are both aggressive in pursuit of the goal and conservative in the face of absolute uncertainty.

4. Experimental Setup and Evaluation Metrics

4.1 Datasets and Environments

To rigorously evaluate the proposed horizon-adaptive framework, extensive experiments were conducted on standard vision-and-language navigation benchmark datasets. The primary dataset utilized is built upon a large-scale collection of high-resolution, three-dimensional building reconstructions. These buildings encompass a wide variety of indoor environments, ranging from small residential apartments to expansive commercial spaces and multi-story houses. The dataset is partitioned into training, validation, and testing splits. Crucially, the validation split is further divided into seen and unseen categories. The seen validation set contains new instructions but within the same spatial environments used during training. The unseen validation set contains instructions for entirely novel buildings that the agent has never encountered, providing a strict test of zero-shot generalization capabilities. The instructions within the dataset are generated by human annotators who virtually walk through the environments and describe the paths. These natural language directives exhibit significant linguistic variation, encompassing diverse vocabulary, complex grammatical structures, and varying levels of granularity. Some instructions provide highly detailed, step-by-step commands, while others offer more abstract, goal-oriented descriptions. The length of the trajectories also varies considerably, necessitating an adaptive approach to temporal constraints.

Table 3Dataset Characteristics

Dataset Partition	**Total Instructions**	**Unique Environments**	**Average Path Length**
Training Set	Fourteen thousand	Sixty one	Ten meters

Validation Seen	One thousand	Sixty one	Nine meters
Validation Unseen	Two thousand	Eleven	Ten meters
Test Set Unseen	Four thousand	Eighteen	Eleven meters

To ensure robustness, secondary experiments were also conducted on a multilingual variant of the navigation dataset, which features longer, more complex paths and instructions in multiple languages. This secondary dataset inherently requires longer temporal horizons, serving as an ideal proving ground for the dynamic horizon modulation network. The environments are simulated using a discrete navigation graph, where the agent transitions between predefined viewpoints. At each viewpoint, the agent can look around in a full panoramic sweep, divided into discrete heading and elevation combinations.

4.2 Evaluation Metrics

The evaluation of vision-and-language navigation agents requires a multifaceted approach, as a single metric cannot fully capture the nuances of spatial navigation and instruction following. The primary metric employed in this study is the Success Rate. An episode is considered a success if the agent executes the stop action within a predefined error radius, typically set to three meters, from the actual target location. The Success Rate provides a clear, binary indicator of task completion. However, the Success Rate alone does not account for the efficiency of the trajectory. An agent might reach the destination after taking a highly convoluted and unnecessarily long path. Therefore, the Success weighted by Path Length metric is utilized as the main comparative standard. This metric scales the binary success indicator by the ratio of the shortest possible path length to the actual path length taken by the agent. The Success weighted by Path Length rigorously penalizes inefficient exploration and meandering, making it highly sensitive to the benefits of the proposed horizon-adaptive mechanism. Additional metrics include the Navigation Error, which measures the absolute geodesic distance in meters between the final location of the agent and the target location, regardless of whether the episode was deemed successful. The Oracle Success Rate is also reported, which evaluates whether the agent ever passed within the success radius of the target at any point during its trajectory, even if it failed to execute the stop action at the correct time. The discrepancy between the Oracle Success Rate and the standard Success Rate highlights the agent proficiency in recognizing task completion. Finally, the Trajectory Length is recorded to analyze the behavioral impact of the dynamic temporal bounds.

5. Results and Discussion

5.1 Quantitative Performance Analysis

The quantitative evaluation reveals that the proposed horizon-adaptive reinforcement learning framework significantly outperforms baseline models, particularly in unseen environments. Extensive testing was conducted against a standard sequence-to-sequence behavior cloning model, an advantage actor-critic reinforcement learning model with a fixed horizon, and a state-of-the-art transformer-based navigation agent that utilizes static episodic boundaries. In the seen validation environments, all models exhibit strong performance, with the baseline transformer model achieving a high success rate. This indicates that standard architectures are highly capable of memorizing topological features and specific visual-linguistic alignments when tested on familiar ground. However, the critical evaluation lies in the

unseen environments. When transitioned to novel buildings, the baseline models experience a severe performance drop, often losing more than half of their efficacy in terms of Success weighted by Path Length. The fixed-horizon reinforcement learning agent exhibits a particularly high Navigation Error in unseen environments, empirically validating the hypothesis that rigid time constraints lead to compounding errors and catastrophic wandering when the agent is disorientated.

Table 4Comparative Results on Unseen Environments

Model Architecture	**Navigation Error**	**Success Rate**	**Success weighted by Path Length**
Behavioral Cloning Baseline	Six point five meters	Thirty two percent	Twenty eight percent
Fixed-Horizon Actor-Critic	Five point nine meters	Forty one percent	Thirty four percent
Static Boundary Transformer	Four point eight meters	Fifty two percent	Forty five percent
Horizon-Adaptive Framework	Four point one meters	Sixty one percent	Fifty seven percent

The horizon-adaptive framework demonstrates remarkable resilience in the unseen validation set. By dynamically modulating the available steps, the model achieves a substantial increase in both Success Rate and Success weighted by Path Length compared to the strongest baseline. The significant improvement in the Success weighted by Path Length metric confirms that the performance gains are not merely the result of the agent wandering until it accidentally stumbles upon the goal; rather, the agent is executing highly efficient and purposeful trajectories. The dynamic horizon prevents the agent from engaging in the exhaustive random search patterns that plague fixed-horizon models, thereby preserving the efficiency of the path. Furthermore, the Navigation Error is notably reduced, indicating that even when the horizon-adaptive agent fails to reach the exact destination, it terminates its search much closer to the target area instead of wandering to the opposite side of the building.

5.2 Qualitative Assessment and Ablation Studies

To isolate the specific contributions of the proposed components, rigorous ablation studies were conducted. The removal of the temporal penalty from the reward function resulted in agents that consistently requested maximum horizon extensions, effectively reverting the system to a fixed, albeit longer, time limit. This ablation confirms the necessity of the penalty in encouraging decisive action. Similarly, disabling the early termination mechanism caused a sharp decline in unseen environment performance, as the agent lost the ability to halt erroneous trajectories. Qualitative analysis of the agent trajectories provides deeper insights into the learned behavior. Visualization of the dynamic horizon value across different episodes reveals that the agent actively utilizes the modulation network as intended. When navigating through long, featureless corridors with ambiguous instructions, the agent

maintains a tight temporal bound, preventing deviation. Upon recognizing a definitive landmark mentioned in the instruction, such as a distinct architectural feature or a specific piece of furniture, the agent confidently extends the horizon to navigate the complex local topology required to reach the final stopping point. However, the qualitative analysis also highlights remaining limitations. The horizon-adaptive framework occasionally struggles with highly convoluted instructions that require extensive backtracking. If an instruction dictates passing a target, entering another room, and then returning, the progress estimator sometimes misinterprets the initial passing as task completion, leading to premature termination. Additionally, severe visual occlusions, such as closed doors or blocked pathways not accounted for in the navigation graph, can cause the horizon modulation network to rapidly fluctuate, indicating a conflict between visual perception and expected progress. These failure modes suggest that while temporal adaptability is crucial, it must be deeply integrated with a more robust semantic memory of the spatial layout to handle complex multi-stage objectives fully.

6. Conclusion and Future Directions

6.1 Summary of Contributions

This research addressed the persistent challenge of poor generalization in vision-and-language navigation, specifically focusing on the deployment of embodied agents in unseen environments. By identifying the rigid temporal boundaries of standard reinforcement learning frameworks as a major source of behavioral instability and error accumulation, this paper introduced a novel horizon-adaptive methodology. The core contribution is the integration of a dynamic horizon modulation network that continuously evaluates task progress and adjusts the allowable episode length in real-time. Through extensive empirical evaluation on established benchmark datasets, it was demonstrated that dynamically altering the temporal horizon fundamentally improves the exploration-exploitation balance during training and prevents catastrophic wandering during inference. The proposed framework achieved significant advancements in overall success rates and path efficiency in completely novel environments. The results validate the premise that embodied artificial intelligence should not be constrained by arbitrary, hard-coded temporal limits, but should rather possess the metacognitive ability to allocate temporal resources based on internal confidence and perceived progress.

6.2 Directions for Future Research

While the horizon-adaptive framework presents a substantial leap forward, several avenues for future research remain open. A primary direction involves extending this dynamic temporal modulation to continuous control environments. Current evaluations rely on discrete navigation graphs, which simplify the action space. Applying adaptive horizons to continuous environments, where low-level motor control and physics-based interactions are required, will introduce significant complexities regarding the scale and frequency of horizon updates. Furthermore, integrating advanced large language models into the cross-modal representation module could profoundly enhance the progress estimation capabilities. Large language models possess superior zero-shot reasoning abilities that could be leveraged to interpret highly ambiguous or multi-stage instructions more effectively, reducing the instances of premature early termination observed in the current study. Finally, exploring multi-agent collaborative scenarios where agents can dynamically negotiate task allocation and temporal boundaries based on their localized progress presents an exciting frontier for building complex, highly adaptable embodied AI systems for real-world deployment [31] [32] [33] [34] [35] [36].

References

Mur-Artal, R., Montiel, J. M. M., & Tardós, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. IEEE Transactions on Robotics, 31(5), 1147–1163.

Levine, S., Pastor, P., Krizhevsky, A., & Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. International Journal of Robotics Research, 37(4–5), 421–436.

Luo, Y., Chen, Y., Duan, Z., He, Y., Tang, H., & Pan, C. (2026). Fault Tolerant Control for Heterogeneous UAVs-UGVs Coordinative System with Adaptive Safety Switching based on DMPC. IEEE Transactions on Aerospace and Electronic Systems.

Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. IEEE Transactions on Circuits and Systems I: Regular Papers, 70(7), 2834-2843.

Wang, C., Tan, C., Zhu, L., Yang, R., & Hong, J. (2024, September). Neural Active Sensing Vision and Manipulation for Cooperative Agents. In 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 942-947). IEEE.

Liu, Fuqiang, Weiping Ding, Luis Miranda-Moreno, and Lijun Sun. "Error adjustment based on spatiotemporal correlation fusion for traffic forecasting." Information Fusion (2025): 103635.

Yang, Y., Wang, Q., Li, C., Hu, H., Wu, C., Jiang, Y., ... & Xu, B. (2025, May). Fewer may be better: Enhancing offline reinforcement learning with reduced dataset. In International Conference on Learning Representations (Vol. 2025, pp. 7076-7098).

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., et al. (2016). End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316.

Cheng, Xiaoyuan, et al. "How Does the Lagrangian Guide Safe Reinforcement Learning through Diffusion Models?." arXiv preprint arXiv:2602.02924 (2026).

Hu, Y., Xi, A., Xiao, Q., Isaacson, S., Liu, H. X., Vasudevan, R., & Ghaffari, M. (2026). LongNav-R1: Horizon-Adaptive Multi-Turn RL for Long-Horizon VLA Navigation. arXiv preprint arXiv:2602.12351.

LaValle, S. M. (1998). Rapidly-exploring random trees: A new tool for path planning. Technical Report, Iowa State University.

Yang, Y., Ma, X., Li, C., Zheng, Z., Zhang, Q., Huang, G., ... & Zhao, Q. (2021). Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 34, 10299-10312.

Yao, Jingyu, et al. "Explainable AI Enhanced Traffic Forecasting for Proactive Autonomous Vehicle Routing." 2026 IEEE Conference on Technologies for Sustainability (SusTech). IEEE, 2026.

Wang Y, Fang Y, Wang T, et al. What Limits Vision-and-Language Navigation?[J]. arXiv preprint arXiv:2605.13328, 2026.

Zhang, Luyan. "MCP: A Control-Theoretic Orchestration Framework for Synergistic Efficiency and Interpretability in Multimodal Large Language Models." arXiv preprint arXiv:2509.16597 (2025).

Wan, H., Cheng, J., Deng, Y., Wu, D., Chen, Y., Lin, Z., ... & Ji, X. Towards Physics Aware Embodied Control with Graph based Object-centric Learning. ACM Transactions on Cyber-Physical Systems.

Hutchinson, S., Hager, G. D., & Corke, P. I. (1996). A tutorial on visual servo control. IEEE Transactions on Robotics and Automation, 12(5), 651–670.

Li, Q., Luo, H., Cheng, H., Deng, Y., Sun, W., Li, W., & Liu, Z. (2023). Incipient fault detection in power distribution system: A time-frequency embedded deep-learning-based approach. IEEE Transactions on Instrumentation and Measurement, 72, 1–14.

Kavraki, L. E., Švestka, P., Latombe, J.-C., & Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE Transactions on Robotics and Automation, 12(4), 566–580.

Zhang, M., Fang, Z., Wang, T., Lu, S., Wang, X., & Shi, T. (2025). Ccma: A framework for cascading cooperative multi-agent in autonomous driving merging using large language models. Expert Systems with Applications, 282, 127717.

Yang, Y., Hu, H., Li, W., Li, S., Yang, J., Zhao, Q., & Zhang, C. (2023, June). Flow to control: Offline reinforcement learning with lossless primitive discovery. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 9, pp. 10843-10851).

Wan, H., Zhang, Y., Wang, J., Wu, D., Li, M., Chen, X., ... & Ji, X. (2025). Toward universal embodied planning in scalable heterogeneous field robots collaboration and control. Journal of Field Robotics, 42(5), 2318-2336.

Xiao, Q., & Ghaffari, M. (2026). LaWM: Least Action World Models for Long-Horizon Physical Consistency from Visual Observations. arXiv preprint arXiv:2605.08279.

Yu, Jiangbo, Raphaël Frank, Luis Miranda-Moreno, Sasan Jafarnejad, Jônatas Augusto Manzolli, Jiyao Wang, and Ali Eslami. Agentic Vehicles for Human-Centered Mobility: Definition, Prospects, and System Implications. 2026.

Wang Y, Fang Y, Wang T, et al. Dreamnav: A trajectory-based imaginative framework for zero-shot vision-and-language navigation[J]. arXiv preprint arXiv:2509.11197, 2025.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T., et al. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings of ICML.

Peng, Qucheng, Chen Bai, Guoxiang Zhang, Bo Xu, Xiaotong Liu, Xiaoyin Zheng, Chen, and Cheng Lu. "NavigScene: Bridging Local Perception and Global Navigation for Beyond-Visual-Range Autonomous Driving." arXiv preprint arXiv:2507.05227 (2025).

Shi, T., ElSamadisy, O., & Abdulhai, B. (2026). CoopSECRM2D: A Safe, Efficient, and Comfortable Multi-Agent Reinforcement Learning Model for Cooperative On-Ramp Merging. IET Intelligent Transport Systems, 20 (1), e70256.

Luo, P., & Ding, X. (2026). C-MAPPO-TSC: A Collaborative Multi-Agent Reinforcement Learning Framework for Urban Traffic Signal Control and Congestion Mitigation. Fundamental Scientific Reports in Multidisciplinary Areas, 2 (02), 170-181.

Wang, K., Hou, W., Hong, L., & Guo, J. (2025). Smart transparency: A user-centered approach to improving human–machine interaction in high-risk supervisory control tasks. Electronics, 14(3), 420.

Lv, Qi, et al. "Spatial-temporal graph diffusion policy with kinematic modeling for bimanual robotic manipulation." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

Prescribed-time tracking control of MIMO nonlinear systems with nonvanishing uncertainties

Liu, C., Zhang, J., Zhang, T., Wang, Y., Zhou, H., & Jin, Q. (2026). HRIBench: Benchmarking Interaction-Centric Human-Robot Collaboration.arXiv preprint arXiv:2607.13056.

Li, Y., Du, P., Ren, H., Xi, T., Gao, K., Kang, Z., ... & Zhang, Y. (2026). Mobile sensing discovery of when where and why vulnerable road users break traffic rules.npj Sustainable Mobility and Transport,3(1), 1.

Kalashnikov, D., Irpan, A., Pastor, P., Ibarz, J., Herzog, A., Jang, E., et al. (2018). QT-Opt: Scalable deep reinforcement learning for vision-based robotic manipulation. arXiv preprint arXiv:1806.10293.