Towards Autonomous Micromobility through Scalable Urban Simulation

Wayne Wu*, Honglin He*, Bolei Zhou
University of California, Los Angeles

Fig. 1: **Autonomous micromobility.** In public urban spaces, various mobile machines (circular images) are essential for short-distance travel. However, urban environments are complex and contain varied terrain and challenging situations (rectangular images). To bridge this gap, we present a scalable urban simulation solution to advance autonomous micromobility. Images are from our Urban-Tra-City data.

Abstract*—Micromobility, which utilizes lightweight mobile machines moving in urban public spaces - such as delivery robots and electric wheelchairs - emerges as a promising alternative to vehicular mobility. Current micromobility depends mostly on human manual operation (in-person or remote control), which raises safety and efficiency concerns when navigating busy urban environments full of unpredictable obstacles and pedestrians. Assisting humans with AI agents in maneuvering micromobility devices presents a viable solution for enhancing safety and efficiency. In this work, we present a scalable urban simulation solution to advance autonomous micromobility. First, we build `URBAN-SIM` – a high-performance robot learning platform for large-scale training of embodied agents in interactive urban scenes. `URBAN-SIM` contains three critical modules: Hierarchical Urban Generation pipeline, Interactive Dynamics Generation strategy, and Asynchronous Scene Sampling scheme, to improve the diversity, realism, and efficiency of robot learning in simulation. Then, we propose `URBAN-BENCH` – a suite of essential tasks and benchmarks to gauge various capabilities of the AI agents in achieving autonomous micromobility. `URBAN-BENCH` includes eight tasks based on three core skills of the agents: Urban Locomotion, Urban Navigation, and Urban Traverse. We evaluate four robots with heterogeneous embodiments, such as the wheeled and legged robots, across these tasks. Experiments on diverse terrains and urban structures reveal each robot's strengths and limitations.

I. Introduction

Micromobility becomes a promising urban transport way for short-distance travel [47, 3]. It includes a range of lightweight machines that have a mass of no more than 350 *kg* and operate at speeds not exceeding 45 *kph* [39] in public spaces. These machines encompass mobile robots with different forms, such as wheeled, quadruped, wheeled-legged, and humanoid robots, and assistive mobility devices for elderly and disabled people, such as electric wheelchairs and mobility scooters. They can accommodate various users' needs in individual travel and parcel delivery. The appeal of micromobility lies in its provision of a flexible, sustainable, cost-effective, and on-demand transport alternative, which enhances urban accessibility [55, 42] and reduces reliance on vehicles for short-distance trips [11, 62].

Current road designs predominantly cater to full-sized vehicles [19]. Micromobility machines thus have to move through intricate urban public spaces, such as sidewalks, alleys, and plazas, which contain unpredictable terrains, various obstacles, and dense pedestrian traffic. Traditional micromobility machines rely on either onboard control (like wheelchairs) or teleoperation by humans (like food delivery bots [1]) to navigate complex urban spaces. However, humans and their driven mobile machines face critical *safety* concerns from human fatigue and limited situational awareness. As reported by FARS [65], over 6,000 vulnerable road users died on U.S. streets in 2018, a 14% increase over 2015 and a 27% increase over 2014. Humans are prone to distractions that can lead to collisions with road hazards. On the other hand, human-driven machines have low operation *efficiency*, as they require high labor costs and have limited agility. For instance, in teleoperated systems for parcel delivery [1, 2], robots require continuous human monitoring, which limits the number of robots that can be operated simultaneously. Also, given the complexity of the urban environment, human teleoperators may find it challenging to move swiftly through a hustling street.

Autonomous micromobility harnesses embodied AI agents for decision-making and maneuvering, providing a viable way to improve safety and efficiency. Existing AI solutions are mainly targeted at specific abilities of robots, such as obstacle avoidance [60] and parkour [9]. However, micromobility tasks require agents to have versatile capabilities facing various complex and challenging terrains and situations (bottom row in Figure 1), *i.e.*, traversing varied terrains (stairs, slopes, and rough surfaces), moving on traversable paths in open spaces, and avoiding both static and dynamic obstacles. Current AI solutions, focused on *isolated* tasks, are thus incapable of conducting complex micromobility tasks. Apart from that, existing robot learning and simulation platforms are insufficient for agent training on micromobility. They either have simple training scenes with *no contextual environments* [36, 43] or have *low training performances* without environment parallelization on GPUs [16, 32, 71]. For example, IsaacGym [36] has superior performance but simple environments, while CARLA [16] provides rich town scenes but has low end-to-end training efficiency. However, for micromobility tasks, on the one hand, robots should learn situational awareness by interacting with large-scale scene contexts, such as urban facilities and pedestrians; on the other hand, robots need a high-performance training platform to sample diverse scenes to obtain strong generalizability. Yet, "large-scale training" with abundant diverse scenes and "high-performance training" are contradictory in the existing robot learning platforms. Current platforms can not balance these two demands and thus lack sufficient support for autonomous micromobility tasks.

In this work, we present a scalable urban simulation solution to advance autonomous micromobility. This solution consists of two critical components: a robot learning platform **`URBAN-SIM`**, and a suite of tasks and benchmarks **`URBAN-BENCH`**. It forges a path to autonomous micromobility by enabling large-scale training and evaluation of varied embodied AI agents in complex urban environments.

First, we propose **`URBAN-SIM`** – a high-performance robot learning platform for autonomous micromobility. It can automatically construct infinite *diverse* and *realistic* interactive urban scenes for large-scale robot learning while providing more than 1,800 *fps* high *training performance* with large-scale parallelization in a single Nvidia L40S GPU. `URBAN-SIM` has three key designs: 1) The *Hierarchical Urban Generation* pipeline, which can construct an infinite number of static urban scenes, from street block to ground division to building and infrastructure placements to terrain generation. This pipeline remarkably enhances the *diversity* of training environments. 2) The *Interactive Dynamics Generation* strategy, which can provide rich dynamics of pedestrians and cyclists that are responsive to robots in real-time during training. This strategy highly improves the *realism* of dynamic agents while maintaining the performance in our large-scale, distributed robot learning workflows. 3) The *Asynchronous Scene Sampling* scheme, which can train robots on thousands of various urban scenes on GPUs in parallel. This scheme significantly enhances the *training performance*, especially for large-scale scenes, achieving more than 26.3% relative improvement compared to synchronous approaches with the same training steps. `URBAN-SIM` is built on top of Nvidia's Omniverse [45] and PhysX 5 [46] to provide realistic scene rendering and physics simulation.

Though the goal of autonomous micromobility is to move from point A to B in an urban environment, it requires the multifaceted capabilities of the agent. Thus, we construct **`URBAN-BENCH`** – a suite of essential tasks and benchmarks to train and evaluate different capabilities of an agent. We first construct a set of tasks for the agent to acquire two orthogonal skills in micromobility: *Urban Locomotion* and *Urban Navigation*. For urban locomotion, an agent must learn various movement skills to tackle different ground conditions, *i.e.*, flat surfaces, slopes, stairs, and rough terrain. We define four tasks for urban locomotion based on these ground conditions. For urban navigation, an agent needs to develop different operational skills to manage various scenarios, *i.e.*, unobstructed ground, static obstacles, and dynamic obstacles. We define three tasks for urban navigation based on these scene conditions. Furthermore, real-world micromobility often requires *kilometer-scale* navigation in complex urban spaces; it remains extremely challenging to tackle this problem. Thus, we define *Urban Traverse* as a new task with a substantially long time horizon, where a mobile robot needs to make tens of thousands of actions at a kilometer-scale distance. We further introduce a human-AI shared autonomous approach to tackle the task. It is designed with a flexible architecture that ranges from full human control to complete AI management of the workflow, allowing us to explore various labor division modes between humans and AI agents in the urban traverse task.

We construct comprehensive benchmarks across four robots with heterogeneous mechanical structures for all 8 defined tasks. Experimental results demonstrate that all `URBAN-BENCH` tasks are challenging for existing solutions. By presenting well-defined challenges beyond the capabilities of current solutions, URBAN-BENCH can serve as *a unified benchmark* that facilitates the future development of autonomous micromobility. Furthermore, through training in complex urban environments, qualitative results indicate that agents have developed interesting and surprising skills based on their mechanical structures. For instance, humanoid robots learn to maneuver through narrow spaces by sidestepping, while wheeled robots learn to navigate around stairs by detouring. Finally, we demonstrate our work's strong scale-up ability, which is essential for learning skills in autonomous micromobility.

II. Related Work

a) Micromobility.: Conventional mobility solutions [6], such as cars and buses, primarily operate on structured roadways, suited for *medium* to *long-distance* commutes. However, these systems often struggle with last-mile connectivity, where efficient transport is needed for the final leg of a journey, such as moving people from transit hubs to destinations or delivering parcels directly to recipients' doorsteps. Micromobility [47, 3], emerging in Europe and North America in the late 1900s [23, 40], offers a practical solution for *short-distance* travel in urban spaces. It relies on lightweight and low-speed devices, such as electric wheelchairs and e-mobility scooters for personal transport [35], or small robots for parcel delivery [14], providing flexible, sustainable, and cost-effective alternatives to private vehicles. This approach reduces emissions [56], alleviates congestion [38], and enhances accessibility [55], especially in densely populated areas.

Recently, a few AI-driven solutions [22, 70] have been introduced in micromobility, focusing on device-sharing systems [63] and scene understanding [73], including fleet management, demand prediction, as well as road change and hazard detection. While these improve operational efficiency, they do not tackle the core challenge of enabling autonomous travel from point A to B in urban spaces. Current solutions lack the *embodied intelligence* essential for real-time decision-making, which is crucial for tasks like assistive mobility and autonomous delivery.

b) Robot autonomy tasks.: Recent advances in robotics and embodied AI have significantly enhanced specific skills for robot autonomy, particularly in *locomotion* [26] and *navigation* [15]. In locomotion, the main goal is to enable robots to move efficiently across diverse terrains. Considerable progress has been achieved in tasks categorized by different mechanical structures (*e.g.*, bipedal [33], quadrupedal [4], multilegged [10]) or unique abilities (*e.g.*, parkour [9], whole-body control [34], jumping [59]). In navigation, the focus is on guiding robots to specific destinations while avoiding obstacles. Research has proposed various tasks categorized by goals and conditions, such as point navigation [5], object navigation [74], and social navigation [64]. However, these tasks address *isolated* skills and struggle to meet micromobility's demands, which require unique and versatile abilities for complex urban environments. A few pioneering studies have explored long-horizon outdoor navigation tasks, but they are limited to *case-specific* robots [41, 29] and scenarios [60, 54], lacking the generalizability needed for micromobility tasks.

c) Simulation platforms for robot learning.: Simulation platforms have rapidly advanced over the past decades, offering scalable training for embodied agents and robots, as well as safe evaluation before real-world deployment [30, 12, 13, 61, 49]. Existing platforms mainly focus on two types of environments: 1) *indoor* environments [48, 52], such as homes and offices, and 2) *driving* environments [24, 28], like roadways and highways. In indoor environments, platforms like AI2-THOR [25], Habitat [52], iGibson [57], OmniGibson [31], and ThreeDWorld [17] are tailored for tasks like indoor navigation, object rearrangement, and manipulation, which differ greatly from micromobility scenarios in complex urban spaces. In driving environments, platforms like GTA V [37], CARLA [16], DriverGym [27], Nuplan [8], and MetaDrive [32] support medium to long-distance driving tasks, focusing on vehicle-centric road scenarios rather than urban public spaces like sidewalks and alleys, which are crucial for micromobility tasks.

Some recent works have constructed detailed urban spaces [18, 69, 72, 75]. However, these focus mainly on algorithm evaluation [18, 69] or scene generation [72, 75], and lack support for interactive robot training, which requires efficient scene sampling, physical simulation, and real-time dynamics. Recently, task-oriented robot learning platforms, such as IsaacGym [36], IsaacSim [44], and IsaacLab [43], built on Nvidia ecosystem, have shown impressive training efficiency with high visual and physical realism. However, these platforms are mainly suited for repetitive tasks in *uniform* environments, like locomotion and manipulation, and often neglect contextual scene simulation needed for complex, long-horizon micromobility tasks.

III. URBAN-SIM: A Robot Learning Platform for Autonomous Micromobility

To support robot learning in complex urban scenes, an ideal simulation platform needs to have two important features: **large-scale** – the platform should provide a vast array of *diverse* scenes with *realistic* interactions; and **high-performance** – the platform should support *high-efficiency* scene sampling for training. In this section, we introduce URBAN-SIM – a robot learning platform for autonomous micromobility, which can balance the contradiction between scale and performance. It supports infinite urban scene generation with arbitrary size and achieves high-performance training with more than 1,800 *fps* sampling rate in a single GPU. We highlight three key designs of URBAN-SIM: the **Hierarchical Urban Generation** pipeline (Section III-A), which ensures the *diversity* of static scenes on a large scale; the **Interactive Dynamics Generation** strategy (Section III-B), which ensures the *realism* of dynamics on a large scale; and the **Asynchronous Scene Sampling** scheme (Section III-C), which ensures *high-efficiency* training on complex urban environments.

A. Hierarchical Urban Generation

The *diversity* of simulation environments is essential for the robustness and generalizability of robot training, especially in

Fig. 2: **`URBAN-SIM`: a robot learning platform for autonomous micromobility.** (a) Hierarchical Urban Generation. It generates an infinite number of *diverse* scenes through four progressive stages. (b) Interactive Dynamics Generation. GPU-based generation of *realistic* agent-scene and agent-agent interactions on the fly. (c) Asynchronous Scene Sampling. An asynchronous sampling scheme to enable *high-efficiency* training on varied scenes with rich context information.

deep learning approaches. Following recent advancements in procedural generation in games [58], we introduce a hierarchical urban generation pipeline to procedurally create complex urban scenes, from macroscale street blocks to microscale terrains, enabling *infinite variations* of diverse scenes with *arbitrary sizes* (from a street corner to a city).

As shown in Figure 2 (a), this pipeline includes four progressive stages: 1) In street block connection, blocks (*e.g.*, straight, curve, roundabout, diverging, merging, intersection, and T-intersection) are sampled and connected to form diverse road networks. 2) In ground planning, we divide urban public areas into functional zones (*e.g.*, sidewalks, crosswalks, plazas, buildings, and vegetation) using randomized parameters for each area's dimensions. 3) In terrain generation, we employ the Wave Function Collapse (WFC) [21] algorithm to generate typical urban terrains - flat (*e.g.*, pathway on grass), stair (*e.g.*, front steps), slope (*e.g.*, assistive ramps), and rough (*e.g.*, cracked sidewalks) - each with adjustable parameters like step height or ramp angle, providing diverse ground conditions. 4) In object placement, static objects (*e.g.*, trees and bus stops) are placed adaptably within the functional areas according to their sizes, creating varied obstacle layouts. To ensure the coverage of objects, we have compiled a repository of over 15,000 high-quality 3D assets of urban objects. This pipeline enables the creation of enormous static urban scenes with diverse street layouts, functional divisions, obstacles, and terrains in a breeze[1].

B. Interactive Dynamics Generation

Beyond static scene diversity, the *realism* of dynamic agents, *i.e.*, vehicles, pedestrians, and other mobile machines, is crucial for simulated urban environments. To form realistic dynamics, the environmental agents should be interactive, with both the static scenes and other dynamic agents. A naive approach uses multi-agent path planning algorithms like

[1]Empowered by the UI of Omniverse [45], users can easily modify the scenes generated by our pipeline further, to cater to specific needs.

ORCA [68] to optimize agents' trajectories, avoiding collisions and deadlocks. However, these methods pre-compute trajectories, preventing real-time interaction with the trained agent, and run only on the CPU, causing inefficiencies when integrated with GPU-based platforms due to the frequent CPU-GPU data transfer during training.

To address these issues, we follow Waymax [20] and Jax-MARL [51] in upgrading ORCA with JAX [7] for multi-agent path planning on GPUs without any CPU bottlenecks. This method enables parallelization across multiple environments for simultaneous collision avoidance with static and dynamic objects and interaction with the trained agent. Specifically, we first generate a 2D occupancy map labeling obstacles, roadways (for vehicles), and traversable areas (for pedestrians and mobile machines), then sample random start and end points for each agent. Using ORCA for initial trajectories, we adjust agents' positions in real-time based on proximity and relative velocity, all on GPUs. We illustrate the realistic interactions between agents and environments and other agents in Figure 2 (b). This strategy enables the creation of dynamic environments with realistic interactions on the fly in robot training.

C. Asynchronous Scene Sampling

So far, we can generate diverse scenes with realistic dynamics. However, the complexity of these scenes, with numerous objects and dense physical interactions, poses new challenges for the training performance, especially in learning long-horizon behaviors for robots with high degrees of freedom. Recent robot learning platforms like IsaacGym [36] and IsaacLab [43] achieve high performance through environment parallelization on GPUs. These platforms are designed for tasks that require extensive repetitive training in *uniform* environments with enormous trial and error, such as locomotion and manipulation. In micromobility tasks, however, rather than uniform environments, robots must make decisions based on *varied* environments with rich contextual information, such as ground paving, obstacles semantics, and pedestrian movements. Thus, existing *synchronous* scene sampling in [36, 43] will encounter huge barriers facing micromobility tasks, where *the essential is not the repetitive training in uniform environments but the multi-faceted training in enormous varied environments*.

To solve this problem, we propose an *asynchronous* scene sampling scheme, which can remarkably enhance training efficiency by training simultaneously on thousands of non-uniform environments with various static layouts, obstacles, dynamics, terrains, and episodes of agents. Specifically, as illustrated in Figure 3, all assets are initially loaded into a cache, from which environments randomly sample assets to create diverse settings simultaneously. Observations, rewards, and actions for each environment are fully vectorized on the GPU, enabling efficient parallel training of agents across multiple environments. Figure 2 (c) visualizes the parallel training on varied environments simultaneously with the asynchronous scene sampling scheme. This approach significantly accelerates model convergence and reduces training time, essential for context-aware micromobility tasks.

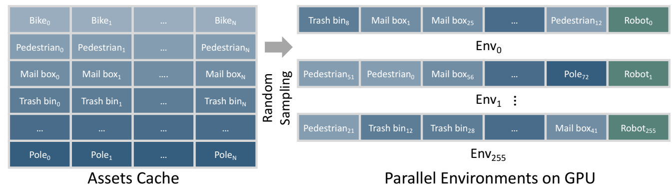

Fig. 3: **Scene sampling diagram.** (Left) Assets Cache that stores all assets in urban scenes. (Right) With a random sampling of assets, parallel environments can be constructed on GPU.

a) Performance benchmarking.: Using the asynchronous scene sampling scheme, we can enable parallelization with any number of unique environments, depending on the GPU used. On a single GPU, parallelized training can be conducted across 256 environments, achieving performance ranging from 1,800 to 2,600 *fps* with RGBD sensors, depending on the specific scenario. Note that, due to the scalable nature of our platform, the sampling rate can be continually increased by adding more GPUs.

IV. URBAN-BENCH: A SUITE OF ESSENTIAL TASKS FOR AUTONOMOUS MICROMOBILITY

In this section, we introduce URBAN-BENCH, a suite of essential tasks and benchmarks that capture high-frequency scenarios in autonomous micromobility. Based on the data from users of micromobility, we first summarize several key **Human Needs** (Section IV-A) as the basis of the task definition. The real-world demands for micromobility devices mainly ask for two primary skills: **Urban Locomotion** (Section IV-B) — moving stably across diverse terrains, including flat, slope, stair, and rough surfaces, and **Urban Navigation** (Section IV-C) — moving efficiently in spaces with varying conditions like unobstructed pathways, static, and dynamic obstacles. Furthermore, we define a long-horizon task, **Urban Traverse** (Section IV-D), where robots must navigate urban spaces at kilometer scales. To tackle this challenging task, we introduce a pilot approach - human-AI shared autonomy - leveraging the power of both humans and AI agents. We will present benchmark results for these tasks in Section V.

Fig. 4: **URBAN-BENCH: a suite of essential tasks for autonomous micromobility.** Simulation environments of eight essential tasks of (a) Urban Locomotion, (b) Urban Navigation, and (c) Urban Traverse.

A. Tasks Grounded in Human Needs

The selection of tasks in URBAN-BENCH is informed by urban mobility studies and infrastructure assessments, highlighting their practical importance. U.S. Department of

Transportation (DOT) reports [66] indicate the prevalence of diverse terrains like ramps, stairs, and uneven surfaces in public spaces, so it is necessary to have various locomotion capabilities, including *slope traversal*, *stair climbing*, and *rough terrain traversal*. Besides, the National Household Travel Survey (NHTS) [67] indicates that a significant portion of urban travel involves short trips on sidewalks and plazas, where micromobility devices must navigate both unobstructed pathways and crowded zones. This underscores the need for safe and efficient *clear pathway traversal*, and *static* and *dynamic obstacle avoidance*. Based on these scene conditions, we define a set of essential tasks of urban locomotion and navigation.

B. Urban Locomotion

In urban locomotion, the embodied AI agent controls the robot's locomotion, ensuring stable and efficient movement across various terrains such as flat surfaces, slopes, and stairs. We define four tasks for urban locomotion (Figure 4 (a)) based on different ground conditions as below:

LOC `LocoFlat` → Flat Terrain Traversal: Traversing stable, flat surfaces commonly found on sidewalks and pedestrian zones. This is necessary for basic mobility in city spaces designed for foot traffic.

LOC `LocoSlope` → Incline Ascent and Descent: Moving up and down ramps and inclined surfaces with varying slope angles. This is essential in urban areas where slopes and accessibility ramps are common.

LOC `LocoStair` → Stair Ascent and Descent: Ascending and descending stairs with varying heights. This is critical in urban spaces where ramps are unavailable, allowing access to multi-level areas.

LOC `LocoRough` → Uneven Terrain Traversal: Maintaining stability on uneven surfaces like cobblestones or damaged sidewalks. This is important for robust movement in urban environments with irregular, worn-down paths.

C. Urban Navigation

In urban navigation, the embodied AI agent handles local navigation, determining how the robot should move to stay within traversable areas while avoiding obstacles and pedestrians. We define three tasks for urban navigation (Figure 4 (b)) based on different scene conditions as below:

NAV `NavClear` → Traversable Area Finding: Moving across open, unobstructed ground, avoiding non-walkable areas like mud or bushes. This is essential for efficient navigation on open plazas and trails on lawns.

NAV `NavStatic` → Static Obstacle Avoidance: Navigating around stationary urban obstacles such as benches, trash bins, and signposts. This is vital for safely maneuvering in crowded city environments with fixed structures.

NAV `NavDynamic` → Dynamic Obstacle Avoidance: Adjusting paths to avoid moving obstacles like pedestrians and cyclists. This is crucial in urban spaces with high foot traffic, ensuring safe interactions with moving entities.

D. Urban Traverse

In kilometer-scale urban traverse, the embodied AI agent's goal is to reach the target point as efficiently as possible, minimizing travel time while ensuring safety in the journey. We define the urban traverse task (Figure 4 (c)) as below:

TRA `Traverse` → Urban Traverse: Moving from point A to point B with a distance of more than 1 km within a complex urban environment safely and efficiently. A challenging real-world setting for micromobility.

a) Human-AI shared autonomous approach.: We propose a human-AI shared autonomous approach as a pilot study to address this task, combining AI capabilities with human interventions. In this approach, we structure the robot control into three layers: high-level decision-making, mid-level navigation, and low-level locomotion. With the layered architecture, we decompose the complex urban traverse task into a series of subtasks, with AI managing mid-level and low-level routine tasks, and humans making high-level decisions and intervening in risky situations. This approach allows a flexible transition between human and AI control. Humans can manage the entire process if needed, while AI can manage the entire operation using an extra rule-based/AI-based decision model to direct the dispatch of urban navigation and locomotion models. We evaluate these control variants to study micromobility performance at the kilometer scale in Section V. Please refer to the Appendix for a detailed discussion of this approach.

V. Benchmarks

We benchmark four tasks in urban locomotion, three tasks in urban navigation, and one long-horizon task in urban traverse. We describe the benchmarks below regarding the **Settings** (Section V-A) of robots, data, and models, as well as the analysis of the **Results** (Section V-B) of benchmarks. These benchmarks will be maintained and updated as time goes on to cover more robots, tasks, and models, as we aim to build a standard evaluation platform to facilitate research in autonomous micromobility and robot learning in urban spaces.

A. Settings

a) Robots.: We evaluate four representative robots, each with distinct mechanical structures, to gain insights and demonstrate the general applicability of the proposed platform. The robots selected for this study include a wheeled robot (COCO Robotics' delivery robot), a quadruped robot (Unitree Go2), a wheeled-legged robot (Unitree B2-W), and a humanoid robot (Unitree G1) [2].

b) Data.: We construct 4 datasets in `URBAN-SIM`: Urban-Nav is used for the train and test of urban navigation; Urban-Loc is used for the train and test of urban locomotion; Urban-Tra-Standard and Urban-Tra-City are used for the test of urban traverse.

[2]It is simple to import new robots in `URBAN-SIM`.

c) Models.: For the urban navigation and locomotion task, we formulate it as a Markov Decision Process (MDP) [50], where the AI learns to optimize its navigation or locomotion policy using the reinforcement learning algorithm Proximal Policy Optimization (PPO) [53]. For each robot, we train and test three models for urban navigation tasks on Urban-Nav and four models for urban locomotion on Urban-Loc (except wheeled devices), which form a *24-model matrix*. For the urban traverse task, we construct 4 control modes, spanning from the full human to full AI: Human – a full human control mode; Human-AI-Mode-1 – a human AI shared control mode with the dispatch of foundational navigation and locomotion models; Human-AI-Mode-2 – a human AI shared control mode with the dispatch of foundational navigation models and a general locomotion model; AI – a full AI control model.

B. Results

a) Urban locomotion benchmark.: Table I brings the following insights: *1) Quadruped robot achieves optimal smoothness*: The quadruped robot consistently demonstrates the best Smoothness scores across all terrains, highlighting its stability and controlled movement, even on challenging surfaces like stairs and rough ground. *2) Wheeled-legged robot excels in versatility*: Leveraging its hybrid leg-wheel design, the wheeled-legged robot leads in both distance traversal (X-displacement and Time to Fall) and keeping Balance, enabling it to cover diverse urban terrains efficiently. *3) Humanoid robot shows stability on even surfaces*: The Humanoid robot achieves the best Balance performance on both flat and inclined ground, indicating its capability for navigation in even urban environments.

TABLE I: **Urban Locomotion benchmark.** Different colors indicate the best performance of different metrics among three robots: Balance; X-displacement; Time to Fall (TTF); Smoothness.

Metrics	LocoFlat	LocoSlope	LocoStair	LocoRough
Quadruped Robot				
Balance (%) ↑	100.00 ± 0.00	90.56 ± 3.13	91.89 ± 2.07	72.18 ± 4.76
X-dis. (m) ↑	19.58 ± 0.41	4.63 ± 0.23	9.20 ± 0.36	4.88 ± 0.14
TTF (s) ↑	20.00 ± 0.00	19.50 ± 0.44	19.58 ± 0.39	18.31 ± 0.25
Smooth. ↓	7.85 ± 0.04	5.18 ± 0.07	8.11 ± 0.12	10.02 ± 0.09
Wheeled-Legged Robot				
Balance (%) ↑	100.00 ± 0.00	95.57 ± 3.31	83.01 ± 2.37	85.04 ± 2.16
X-dis. (m) ↑	19.62 ± 0.15	12.54 ± 0.34	16.73 ± 0.27	18.24 ± 0.22
TTF (s) ↑	20.00 ± 0.00	19.95 ± 0.02	19.07 ± 0.17	19.13 ± 0.11
Smooth. ↓	210.43 ± 0.07	253.24 ± 0.28	236.52 ± 0.18	231.96 ± 0.14
Humanoid Robot				
Balance (%) ↑	100.00 ± 0.00	95.67 ± 2.24	80.98 ± 4.32	82.45 ± 3.15
X-dis. (m) ↑	16.61 ± 0.50	7.16 ± 0.22	13.99 ± 0.27	16.28 ± 0.31
TTF (s) ↑	20.00 ± 0.00	19.91 ± 0.03	19.03 ± 0.36	19.02 ± 0.33
Smooth. ↓	40.94 ± 0.15	57.69 ± 0.31	42.36 ± 0.19	53.67 ± 0.24

b) Urban navigation benchmark.: Table II brings the following insights. *1) Wheeled robot excels in clear pathway navigation*: The wheeled robot achieves the highest Success Rate (97.60%) and Route Completion (98.61%) in the `NavClear` task, highlighting its suitability for open, predictable urban environments. *2) Quadruped robot leads in safety metrics*: The quadruped robot outperforms others in tasks with obstacles, achieving the lowest Collision rates (0.08 in `NavSta` and 0.13 in `NavDyn`) and the highest percentage On Walkable Regions. This demonstrates its stability in complex, obstacle-rich environments. *3) Humanoid robot performs best in complex scenarios*: The humanoid robot shows the highest Success Rates and Route Completion in tasks with static and dynamic obstacles, indicating its flexibility in navigating crowded urban spaces.

TABLE II: **Urban navigation benchmark.** Different colors indicate the best performance of different metrics among four robots: Success Rate; Route Completion; On Walkable Region; SPL; Collision.

Metrics	NavClear	NavStatic	NavDynamic
Wheeled Robot			
Success Rate (%) ↑	97.60 ± 0.92	51.95 ± 2.63	48.82 ± 3.26
Route Completion (%) ↑	98.61 ± 1.28	53.11 ± 2.92	50.04 ± 3.02
On Walkable Region (%) ↑	74.38 ± 0.99	81.88 ± 1.00	84.82 ± 1.49
SPL ↑	0.48 ± 0.05	0.24 ± 0.04	0.23 ± 0.01
Collision ↓	-	0.31 ± 0.09	0.35 ± 0.04
Quadruped Robot			
Success Rate (%) ↑	90.29 ± 3.25	76.13 ± 3.07	77.14 ± 2.57
Route Completion (%) ↑	94.28 ± 2.16	77.47 ± 2.99	77.63 ± 2.12
On Walkable Region (%) ↑	93.96 ± 3.38	85.81 ± 1.67	88.20 ± 2.17
SPL ↑	0.37 ± 0.05	0.36 ± 0.04	0.36 ± 0.05
Collision ↓	-	0.08 ± 0.02	0.13 ± 0.02
Wheeled-Legged Robot			
Success Rate (%) ↑	79.94 ± 3.06	42.97 ± 4.14	31.06 ± 3.77
Route Completion (%) ↑	80.44 ± 2.97	44.33 ± 3.74	33.95 ± 3.21
On Walkable Region (%) ↑	67.93 ± 0.85	62.17 ± 2.95	63.29 ± 2.71
SPL ↑	0.37 ± 0.03	0.19 ± 0.02	0.14 ± 0.02
Collision ↓	-	0.15 ± 0.04	0.19 ± 0.02
Humanoid Robot			
Success Rate (%) ↑	80.47 ± 2.29	77.86 ± 3.54	79.23 ± 2.71
Route Completion (%) ↑	80.92 ± 1.36	79.72 ± 2.76	80.26 ± 2.92
On Walkable Region (%) ↑	65.86 ± 1.56	86.89 ± 1.73	65.85 ± 1.94
SPL ↑	0.37 ± 0.01	0.37 ± 0.03	0.38 ± 0.03
Collision ↓	-	0.13 ± 0.03	0.15 ± 0.04

c) Urban traverse benchmark.: We evaluate a quadruped robot on a kilometer-scale urban traverse task using the Urban-Tra-Standard dataset with three control modes. As shown in Figure 6, the AI mode achieves the lowest human intervention but exhibits the poorest completeness and safety. Conversely, the Human mode achieves the highest completeness and safety but at a significantly higher labor cost. The two human-AI shared autonomy modes balance completeness and cost while maintaining moderate safety. Future research in urban traverse should aim to move the dot closer to the origin with minimal dot size, indicating optimized completeness, cost, and safety.

d) Emerging robot behaviors.: Through large-scale training in diverse urban environments, different robots obtain movement skills that exploit their unique mechanical structures, as shown in Figure 5: quadruped robots, known to be proficient at stair climbing, can traverse challenging terrain directly to reach the goal; wheeled robots prefer detouring over even surfaces to reduce the risk of getting stuck, despite the longer path; Wheeled-legged robots benefit from their hybrid design and show the ability to partially descend on slopes and stairs simultaneously; The humanoid robot, with greater degrees of freedom, can sidestep through narrow spaces efficiently.

Fig. 5: **Emerging behaviors.** The results of evaluating different robots in the same environment. After training in diverse urban scenes, robots with distinct structures have developed their unique movement skills.

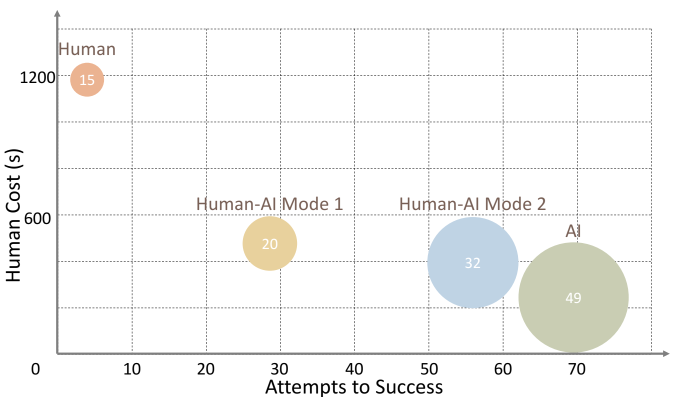

Fig. 6: **Comparison of different control modes in urban traverse.** X-axis: Attempts to Success – the number of failures before reaching the goal points (completion ability). Y-axis: Human Cost – time of human takeover of the control (labor cost). Size of circle: Collision Times to obstacles and pedestrians (safety property). indicate four control modes.

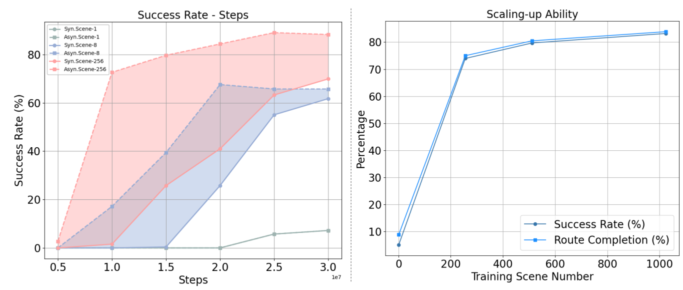

Fig. 7: **Effectiveness of scalable urban simulation.** (Left) Comparison between synchronous and synchronous scene sampling. X-axis: training steps; Y-axis: Success Rate. Different colors indicate training scene numbers – 1, 8, or 256. (Right) Scaling-up ability. X-axis: training scene number; Y-axis: Success Rate and Route Completion.

VI. Evaluation of Scalability

We try to address a fundamental question underlying the strengths demonstrated in this work: *How does the scalability of our urban simulation contribute to autonomous micromobility?*

The proposed asynchronous scene sampling scheme in URBAN-SIM enables high-performance, large-scale robot training in diverse urban environments with realistic interactions. We compare it to synchronous sampling, as used in IsaacLab [43], where all scenes in a batch are identical. In our asynchronous approach, however, each scene in a batch is unique. Furthermore, to assess the impact of large-scale training, we vary the number of training scenes from 1 to 1,024 and observe performance changes. All experiments are conducted using the NavStatic task.

As shown in Figure 7 (Left), asynchronous sampling performs the same as synchronous sampling with only one scene. However, as unique training scenes increase from 8 to 256, a substantial performance gap (the colored areas) emerges, showing the strong scalability of our platform for diverse scene training. Further, as seen in Figure 7 (Right), the performance remarkably improves as the number of training scenes increases from 1 to 1,024, rising from 5.1% to 83.2% (Success Rate). The result highlights the importance of large-scale training on a greater variety of scenes.

VII. Conclusion

We introduce a scalable urban simulation solution to advance research in autonomous micromobility. This solution comprises a high-performance robot learning platform, URBAN-SIM, and a suite of essential tasks and benchmarks, URBAN-BENCH. Through experiments, we evaluate various capabilities of AI agents across different tasks and demonstrate the platform's scalability for large-scale training in urban environments. Looking ahead, we plan to support real-world deployments of models trained on our platform. Our strategy includes building a sim-to-real pipeline based on ROS2 and enabling an integrated workflow for model training, evaluation, and deployment.

References

[1] Coco robotics. https://www.cocodelivery.com/. 2

[2] Kiwibot. https://www.kiwibot.com/. 2

[3] Rusul L Abduljabbar, Sohani Liyanage, and Hussein Dia. The role of micro-mobility in shaping sustainable cities: A systematic literature review. *Transportation research part D: transport and environment*, 2021. 1, 3

[4] Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Legged locomotion in challenging terrains using egocentric vision. In *CoRL*, 2023. 3

[5] Peter Anderson, Angel X. Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, and Amir R. Zamir. On evaluation of embodied navigation agents. *arXiv preprint arXiv:1807.06757*, 2018. 3

[6] David Banister. The sustainable mobility paradigm. *Transport policy*, 2008. 3

[7] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, et al. Jax: composable transformations of python+ numpy programs. 2018. 5

[8] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021. 3

[9] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. In *ICRA*, 2024. 2, 3

[10] Baxi Chong, Juntao He, Daniel Soto, Tianyu Wang, Daniel Irvine, Grigoriy Blekherman, and Daniel I Goldman. Multilegged matter transport: A framework for locomotion on noisy landscapes. *Science*, 2023. 3

[11] Regina R Clewlow. The micro-mobility revolution: the introduction and adoption of electric scooters in the united states. Technical report, 2019. 1

[12] Matt Deitke, Winson Han, Alvaro Herrasti, Aniruddha Kembhavi, Eric Kolve, Roozbeh Mottaghi, Jordi Salvador, Dustin Schwenk, Eli VanderBilt, Matthew Wallingford, Luca Weihs, Mark Yatskar, and Ali Farhadi. Robothor: An open simulation-to-real embodied AI platform. In *CVPR*, 2020. 3

[13] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Kiana Ehsani, Jordi Salvador, Winson Han, Eric Kolve, Aniruddha Kembhavi, and Roozbeh Mottaghi. Procthor: Large-scale embodied ai using procedural generation. *NeuIPS*, 2022. 3

[14] Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, 2009. 3

[15] Guilherme N DeSouza and Avinash C Kak. Vision for mobile robot navigation: A survey. *TPAMI*, 2002. 3

[16] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *CoRL*, 2017. 2, 3

[17] Chuang Gan, Jeremy Schwartz, Seth Alter, Damian Mrowca, Martin Schrimpf, James Traer, Julian De Freitas, Jonas Kubilius, Abhishek Bhandwaldar, Nick Haber, Megumi Sano, Kuno Kim, Elias Wang, Michael Lingelbach, Aidan Curtis, Kevin T. Feigelis, Daniel Bear, Dan Gutfreund, David D. Cox, Antonio Torralba, James J. DiCarlo, Josh Tenenbaum, Josh H. McDermott, and Dan Yamins. Threedworld: A platform for interactive multimodal physical simulation. In *NeurIPS Datasets and Benchmarks*, 2021. 3

[18] Chen Gao, Baining Zhao, Weichen Zhang, Jinzhu Mao, Jun Zhang, Zhiheng Zheng, Fanhang Man, Jianjie Fang, Zile Zhou, Jinqiang Cui, Xinlei Chen, and Yong Li. Embodiedcity: A benchmark platform for embodied agent in real-world city environment. *arXiv preprint arXiv:2410.09604*, 2024. 3

[19] Jan Gehl. Life between buildings. 2011. 1

[20] Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, et al. Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research. *NeurIPS*, 2024. 5

[21] Maxim Gumin. Wave function collapse algorithm. https://github.com/mxgmn/, 2016. 4

[22] Suining He and Kang G Shin. Dynamic flow distribution prediction for urban dockless e-scooter sharing reconfiguration. In *Proceedings of the web conference*, 2020. 3

[23] Stewart Home. The assault on culture: utopian currents from lettrisme to class war. 1991. 3

[24] Saman Kazemkhani, Aarav Pandya, Daphne Cornelisse, Brennan Shacklett, and Eugene Vinitsky. Gpudrive: Data-driven, multi-agent driving simulation at 1 million fps. *arXiv preprint arXiv:2408.01584*, 2024. 3

[25] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana Ehsani, Daniel Gordon, Yuke Zhu, Kembhavi Aniruddha, Gupta Abhinav, and Farhadi Ali. Ai2-thor: An interactive 3d environment for visual ai. *arXiv preprint arXiv:1712.05474*, 2017. 3

[26] Swapnil Saha Kotha, Nipa Akter, Sarafat Hussain Abhi, Sajal Kumar Das, Md Robiul Islam, Md Firoj Ali, Md Hafiz Ahamed, Md Manirul Islam, Subrata Kumar Sarker, Md Faisal Rahman Badal, et al. Next generation legged robot locomotion: A review on control techniques. *Heliyon*, 2024. 3

[27] Parth Kothari, Christian Perone, Luca Bergamini, Alexandre Alahi, and Peter Ondruska. Drivergym: Democratising reinforcement learning for autonomous driving. *arXiv preprint arXiv:2111.06889*, 2021. 3

[28] Daniel Krajzewicz, Georg Hertkorn, Christian Rössel, and Peter Wagner. Sumo (simulation of urban mobility)-an open-source traffic simulation. In *MESM*, 2002. 3

[29] Joonho Lee, Marko Bjelonic, Alexander Reske, Lorenz Wellhausen, Takahiro Miki, and Marco Hutter. Learning robust autonomous navigation and locomotion for wheeled-legged robots. *Science Robotics*, 2024. 3

[30] Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Elliott Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, Andrey Kurenkov, C. Karen Liu, Hyowon Gweon, Jiajun Wu, Li Fei-Fei, and Silvio Savarese. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. In *CoRL*, 2021. 3

[31] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Wensi Ai, Benjamin Martinez, Hang Yin, Michael Lingelbach, Minjune Hwang, Ayano Hiranaka, Sujay Garlanka, Arman Aydin, Sharon Lee, Jiankai Sun, Mona Anvari, Manasi Sharma, Dhruva Bansal, Samuel Hunter, Kyu-Young Kim, Alan Lou,

Caleb R. Matthews, Ivan Villa-Renteria, Jerry Huayang Tang, Claire Tang, Fei Xia, Yunzhu Li, Silvio Savarese, Hyowon Gweon, C. Karen Liu, Jiajun Wu, and Li Fei-Fei. Behavior-1k: A human-centered, embodied ai benchmark with 1, 000 everyday activities and realistic simulation. *CoRL*, 2024. 3

[32] Quanyi Li, Zhenghao Peng, Lan Feng, Qihang Zhang, Zhenghai Xue, and Bolei Zhou. Metadrive: Composing diverse driving scenarios for generalizable reinforcement learning. *TPAMI*, 2022. 2, 3

[33] Zhongyu Li, Xue Bin Peng, Pieter Abbeel, Sergey Levine, Glen Berseth, and Koushil Sreenath. Robust and versatile bipedal jumping control through reinforcement learning. In *RSS*, 2023. 3

[34] Minghuan Liu, Zixuan Chen, Xuxin Cheng, Yandong Ji, Ri-Zhao Qiu, Ruihan Yang, and Xiaolong Wang. Visual whole-body control for legged loco-manipulation. In *CoRL*, 2024. 3

[35] Sohani Liyanage, Hussein Dia, Rusul Abduljabbar, and Saeed Asadi Bagloee. Flexible mobility on-demand: An environmental scan. *Sustainability*, 2019. 3

[36] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu based physics simulation for robot learning. In *NeurIPS Datasets and Benchmarks*. 2, 3, 5

[37] Mark Martinez, Chawin Sitawarin, Kevin Finch, Lennart Meincke, Alex Yablonski, and Alain Kornhauser. Beyond grand theft auto v for training, testing and enhancing deep learning in self driving cars. *arXiv preprint arXiv:1712.01397*, 2017. 3

[38] Mahmoud Masoud, Mohammed Elhenawy, Mohammed H Almannaa, Shi Qiang Liu, Sebastien Glaser, and Andry Rakotonirainy. Heuristic approaches to solve e-scooter assignment problem. *IEEE access*, 2019. 3

[39] ITF Safe Micromobility. Report by the international transport forum oecd/itf. In *International Transport Forum: Paris, France*, 2020. 1

[40] P Midgley. Shared smart bicycle schemes in european cities. *Global Transport Knowledge Partnership*, 2009. 3

[41] Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science robotics*, 2022. 3

[42] Dimitris Milakis, Laura Gedhardt, Daniel Ehebrecht, and Barbara Lenz. Is micro-mobility sustainable? an overview of implications for accessibility, air pollution, safety, physical activity and subjective wellbeing. *Handbook of sustainable transport*, 2020. 1

[43] Mayank Mittal, Calvin Yu, Qinxi Yu, Jingzhou Liu, Nikita Rudin, David Hoeller, Jia Lin Yuan, Ritvik Singh, Yunrong Guo, Hammad Mazhar, et al. Orbit: A unified simulation framework for interactive robot learning environments. *RAL*, 2023. 2, 3, 5, 8

[44] Nvidia Corp. Isaac sim. https://developer.nvidia.com/isaac/sim, 2024. Accessed: 2024-11. 3

[45] Nvidia Corp. Nvidia omniverse. https://developer.nvidia.com/omniverse, 2024. Accessed: 2024-11. 2, 4

[46] Nvidia Corp. Physx. https://developer.nvidia.com/physx-sdk, 2024. Accessed: 2024-11. 2

[47] Giulia Oeschger, Páraic Carroll, and Brian Caulfield. Micromobility and public transport integration: The current state of knowledge. *Transportation Research Part D: Transport and Environment*, 2020. 1, 3

[48] Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. Virtualhome: Simulating household activities via programs. In *CVPR*, 2018. 3

[49] Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander William Clegg, Michal Hlavac, So Yeon Min, Vladimir Vondrus, Théophile Gervet, Vincent-Pierre Berges, John M. Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Jitendra Malik, Devendra Singh Chaplot, Unnat Jain, Dhruv Batra, Akshara Rai, and Roozbeh Mottaghi. Habitat 3.0: A co-habitat for humans, avatars, and robots. In *ICLR*, 2023. 3

[50] Martin L Puterman. Markov decision processes. *Handbooks in operations research and management science*, 1990. 7

[51] Alexander Rutherford, Benjamin Ellis, Matteo Gallici, Jonathan Cook, Andrei Lupu, Gardar Ingvarsson, Timon Willi, Akbir Khan, Christian Schroeder de Witt, Alexandra Souly, et al. Jaxmarl: Multi-agent rl environments in jax. *arXiv preprint arXiv:2311.10090*, 2023. 5

[52] Manolis Savva, Jitendra Malik, Devi Parikh, Dhruv Batra, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, and Vladlen Koltun. Habitat: A platform for embodied AI research. In *ICCV*, 2019. 3

[53] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 7

[54] Dhruv Shah and Sergey Levine. Viking: Vision-based kilometer-scale navigation with geographic hints. In *RSS*, 2022. 3

[55] Susan Shaheen, Elliot Martin, and Adam Cohen. Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in north america. 2013. 1, 3

[56] Susan A Shaheen, Stacey Guzman, and Hua Zhang. Bikesharing in europe, the americas, and asia: past, present, and future. *Transportation research record*, 2010. 3

[57] Bokui Shen, Fei Xia, Chengshu Li, Roberto Martín-Martín, Linxi Fan, Guanzhi Wang, Claudia Pérez-D'Arpino, Shyamal Buch, Sanjana Srivastava, Lyne Tchapmi, Tchapmi Micael, Vainio Kent, Wong Josiah,

Fei-Fei Li, and Savarese Silvio. igibson 1.0: a simulation environment for interactive tasks in large realistic scenes. In *IROS*, 2021. 3

[58] Tanya Short and Tarn Adams. *Procedural generation in game design*. CRC Press, 2017. 4

[59] Laura M. Smith, J. Chase Kew, Tianyu Li, Linda Luu, Xue Bin Peng, Sehoon Ha, Jie Tan, and Sergey Levine. Learning and adapting agile locomotion skills by transferring experience. In *RSS*, 2023. 3

[60] Maks Sorokin, Jie Tan, C Karen Liu, and Sehoon Ha. Learning to navigate sidewalks in outdoor environments. *RAL*, 2022. 2, 3

[61] Andrew Szot, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John M. Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, Oleksandr Maksymets, Aaron Gokaslan, Vladimir Vondrus, Sameer Dharur, Franziska Meier, Wojciech Galuba, Angel X. Chang, Zsolt Kira, Vladlen Koltun, Jitendra Malik, Manolis Savva, and Dhruv Batra. Habitat 2.0: Training home assistants to rearrange their habitat. In *NeuIPS*, 2021. 3

[62] A Tiwari. Micro-mobility: the next wave of urban transportation in india. *YS Journal, January*, 2019. 1

[63] Tien Dung Tran, Nicolas Ovtracht, and Bruno Faivre d'Arcier. Modeling bike sharing system using built environment factors. *Procedia Cirp*, 2015. 3

[64] Nathan Tsoi, Alec Xiang, Peter Yu, Samuel S Sohn, Greg Schwartz, Subashri Ramesh, Mohamed Hussein, Anjali W Gupta, Mubbasir Kapadia, and Marynel Vázquez. Sean 2.0: Formalizing and generating social situations for robot navigation. *RAL*, 2022. 3

[65] U.S. Department of Transportation. Fatality analysis reporting system (fars). https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars. 2

[66] U.S. Department of Transportation. Transportation reports and publications. https://www.transportation.gov/, 2024. Accessed: 2024-11. 6

[67] U.S. Federal Highway Administration. National household travel survey (nhts). https://nhts.ornl.gov/, 2017. Accessed: 2024-11. 6

[68] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *ISRR*, 2011. 5

[69] Hanqing Wang, Jiahe Chen, Wensi Huang, Qingwei Ben, Tai Wang, Boyu Mi, Tao Huang, Siheng Zhao, Yilun Chen, Sizhe Yang, et al. Grutopia: Dream general robots in a city at scale. *arXiv preprint arXiv:2407.10943*, 2024. 3

[70] Gary White and Siobhan Clarke. Urban intelligence with deep edges. *IEEE Access*, 2020. 3

[71] Wayne Wu, Honglin He, Jack He, Yiran Wang, Chenda Duan, Zhizheng Liu, Quanyi Li, and Bolei Zhou. Metaurban: An embodied ai simulation platform for urban micromobility. *arXiv preprint arXiv:2407.08725*, 2024. 2

[72] Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, and Ziwei Liu. Citydreamer: Compositional generative model of unbounded 3d cities. In *CVPR*, 2024. 3

[73] Hong Yang, Qingyu Ma, Zhenyu Wang, Qing Cai, Kun Xie, and Di Yang. Safety of micro-mobility: Analysis of e-scooter crashes by mining news reports. *Accident Analysis & Prevention*, 2020. 3

[74] Naoki Yokoyama, Ram Ramrakhya, Abhishek Das, Dhruv Batra, and Sehoon Ha. Hm3d-ovon: A dataset and benchmark for open-vocabulary object goal navigation. *arXiv preprint arXiv:2409.14296*, 2024. 3

[75] Shougao Zhang, Mengqi Zhou, Yuxi Wang, Chuanchen Luo, Rongyu Wang, Yiwei Li, Xucheng Yin, Zhaoxiang Zhang, and Junran Peng. Cityx: Controllable procedural content generation for unbounded 3d cities. *arXiv preprint arXiv:2407.17572*, 2024. 3