

# Feasibility-Guided Planning over Multi-Specialized Locomotion Policies

Ying-Sheng Luo<sup>1\*</sup>, Lu-Ching Wang<sup>1\*</sup>, Hanjaya Mandala<sup>1\*</sup>, Yu-Lun Chou<sup>1\*</sup>, Guilherme Christmann<sup>1\*</sup>, Yu-Chung Chen<sup>2</sup>, Yung-Shun Chan<sup>2</sup>, Chun-Yi Lee<sup>2</sup>, and Wei-Chao Chen<sup>1</sup>

<sup>1</sup>Inventec Corporation <sup>2</sup>National Taiwan University

**Abstract**—Planning over unstructured terrain presents a significant challenge in the field of legged robotics. Although recent works in reinforcement learning have yielded various locomotion strategies, planning over multiple experts remains a complex issue. Existing approaches encounter several constraints: traditional planners are unable to integrate skill-specific policies, whereas hierarchical learning frameworks often lose interpretability and require retraining whenever new policies are added. In this paper, we propose a feasibility-guided planning framework that successfully incorporates multiple terrain-specific policies. Each policy is paired with a *Feasibility-Net*, which learned to predict feasibility tensors based on the local elevation maps and task vectors. This integration allows classical planning algorithms to derive optimal paths. Through both simulated and real-world experiments, we demonstrate that our method efficiently generates reliable plans across diverse and challenging terrains, while consistently aligning with the capabilities of the underlying policies.

## I. INTRODUCTION

Effective planning constitutes the foundation of robust robotic navigation on unstructured terrain and fundamentally depends on the underlying locomotion capabilities. Recent developments in both model-predictive control (MPC) [1] and model-free reinforcement learning (RL) [2] approaches have enabled modern robots to achieve exceptional performance in traversing challenging and unstructured terrain. Enhanced locomotion capabilities expand the accessible planning space through novel regions and enable robots to navigate more effectively across greater distances. However, with the increasing diversity of locomotion methods and skill specialization, the planning space becomes considerably complex. This complexity places substantial demands on planning methodologies, which must reason about the capabilities and limitations of available locomotion skills to generate valid and optimal trajectories. The complex relationship between locomotion diversity and planning complexity further highlights the critical importance of sophisticated planning frameworks that can effectively integrate and coordinate multiple locomotion modalities. These challenges thus motivate the need for innovative planning strategies that can effectively handle complex scenarios while maintaining path optimality across diverse locomotion capabilities.

Traditional planning approaches [3] attempt to address this complexity through classical robotics navigation methods that rely fundamentally on the elevation data transformed into occupancy grids and costmaps before being solved through

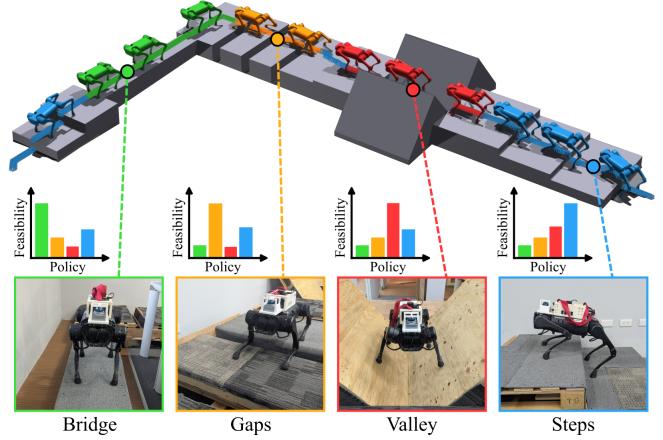


Fig. 1. Feasibility-guided planning enables optimal path selection & policy switching over mixed terrain via policy-specific feasibility representations.

graph search algorithms [3]. These methods require high-quality costmaps that demand extensive empirical tuning within the deployment environment. Classical approaches then establish thresholds to classify cells as obstacles or free space based on local features such as height and surface roughness. Nevertheless, this thresholding process becomes increasingly laborious and imprecise when multiple adaptable policies must be considered. Moreover, thresholding approaches fundamentally lack directional awareness, as traversal from a taller obstacle to a shorter one may be feasible while the reverse direction remains infeasible. To overcome these classical limitations, hierarchical planners trained on top of learned skills have emerged as a common strategy. These learning-based planners can reason about available skills through experience gained in simulated [4] or real-world rollouts [5]. Nevertheless, such enhanced reasoning capability comes at the expense of decision-making interpretability. The selection rationale for specific skills generally remains opaque to system designers and robot operators. Furthermore, the addition of new skills necessitates complete retraining of the entire system, which limits the scalability and adaptability of such approaches.

While classical methods suffer from interpretability and adaptability issues, RL-based planning approaches present their own challenges [6]. The core difficulty lies in addressing problem sets that require diverse and distinct locomotion capabilities, as training a single policy to handle comprehensive terrain variations often exceeds practical

\*Joint first authors.

tractability limits. To manage this complexity, skills are typically compartmentalized into distinct portions of the problem space and subsequently integrated through methods such as mixture-of-experts architectures [7] or hierarchical RL frameworks [8]. The independent skills can then be sequentially coordinated [9] through various coordination mechanisms, including timing switch identification [10], [11], continuous modulation techniques [12], or specialized gait switching methodologies [13], [14]. Despite these coordination strategies, RL-based approaches fundamentally require the planner to accommodate multiple adaptable policies while simultaneously evaluating their respective capabilities and limitations to achieve optimal skill selection and path generation. Such multi-policy coordination scenarios introduce significant overhead and can raise challenges in scalability as the number of locomotion skills increases.

In light of these fundamental limitations across existing planning paradigms, we propose a feasibility-guided planning methodology that addresses the core challenge of coordinating multiple specialized locomotion policies for navigation across heterogeneous terrain, as shown in Fig. 1. Rather than relying on binary occupancy classifications or opaque learned planners, our approach transforms elevation maps into interpretable, policy-specific feasibility representations that preserve the transparency of classical methods while harnessing the capabilities of modern locomotion skills. The key insight underlying our methodology is that effective multi-policy planning requires explicit modeling of each policy's terrain-specific capabilities. We achieve this through *Feasibility-Net* models that are jointly trained with their corresponding locomotion policies to estimate both traversal feasibility and distributional characteristics of training terrain. These models enable us to generate terrain familiarity-weighted feasibility values across elevation maps using a sliding window approach. The resulting representations are inherently grounded in the actual capabilities of their paired policies, enabling traditional graph search algorithms to plan over fused feasibility tensors while maintaining full interpretability of planning decisions. Critically, this architecture supports seamless integration of new policies without retraining. Our contributions are summarized as follows:

- A novel per-policy feasibility estimation framework that jointly learns locomotion-specific traversal predictions and terrain distribution models, thereby enabling robust assessment across diverse and unfamiliar environments.
- A unified training paradigm that seamlessly integrates RL policy optimization with supervised feasibility learning, eliminating the need for separate data generation pipelines while maintaining the training efficiency.
- A directional feasibility tensor representation that transforms elevation maps into plannable cost functions, while supporting dynamic policy selection and interpretable planning without high-level controller training.

## II. RELATED WORK

**Learning-Based Traversability Prediction.** Learning-based traversability prediction has emerged as a critical

component for autonomous navigation in unstructured environments [15]–[17]. These approaches leverage multi-modal sensor data to model terrain navigability and demonstrate adaptive capabilities when encountering novel environmental conditions [18], [19]. The complexity of legged locomotion dynamics poses additional challenges for traversability assessment compared to wheeled systems. Contemporary methods address this complexity through the fusion of visual and proprioceptive information for terrain-aware locomotion [2], [20], [21], while risk-aware prediction frameworks aim to improve reliability [22]–[24]. Nevertheless, generalization across diverse terrain types and robustness to sensor noise remain persistent challenges [25], [26]. While these approaches provide valuable terrain assessment capabilities, they typically produce single traversability estimates that do not account for the varying capabilities of different locomotion policies. Our approach addresses this limitation by grounding feasibility predictions in the specific capabilities of individual locomotion policies, enabling explicit multi-policy planning across heterogeneous terrain configurations.

**Skill Chaining/Hierarchical RL.** Hierarchical RL and skill chaining methodologies enable autonomous agents to address long-horizon tasks through systematic decomposition into reusable behavioral components [27], [28]. Traditional hierarchical approaches employ high-level controllers to orchestrate the activation of specialized lower-level skills [29], [30]. These frameworks have demonstrated success in legged locomotion applications [31]–[34], particularly through the integration of navigation planners with learned locomotion policies [35]. However, hierarchical approaches introduce fundamental trade-offs between capability and interpretability. The learned high-level controllers typically operate as black-box systems, making it difficult to understand policy selection rationales or to integrate new skills without extensive retraining. Furthermore, these methods often require careful reward engineering and hyperparameter tuning for the hierarchical structure itself. Our methodology circumvents these limitations by employing interpretable graph search algorithms over learned feasibility representations, eliminating the need for high-level controller training while maintaining transparency in policy selection decisions.

## III. METHODOLOGY

Our approach addresses the multi-policy planning challenge through a unified framework that integrates locomotion policy learning with feasibility estimation, as illustrated in Fig. 2. The framework is motivated by the need to coordinate multiple specialized locomotion policies while maintaining interpretability and avoiding the limitations of hierarchical controllers. The system processes three key inputs from the simulation environment: *proprioceptive observations for policy control*, *local height maps for terrain assessment*, and *task vectors encoding velocity commands*. The core innovation lies in the joint training paradigm where locomotion policies and their corresponding *Feasibility-Net* models are optimized simultaneously through shared environment interactions. While the locomotion policy learns via standard

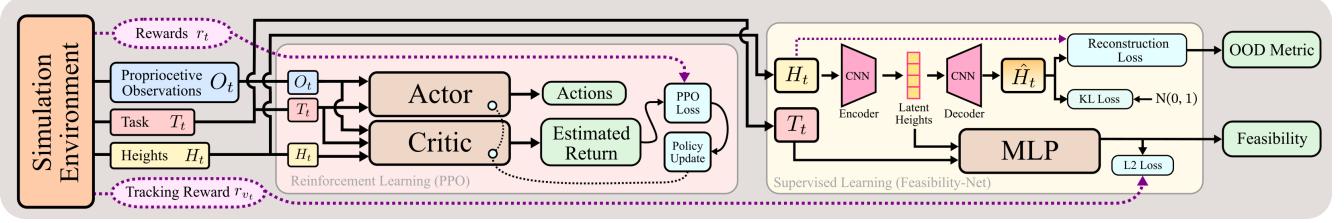


Fig. 2. Overview of the feasibility-aware planning framework. Locomotion policies and *Feasibility-Net* models are jointly trained using shared environment rollouts, where the feasibility models learn to predict velocity tracking performance and terrain distributions simultaneously with policy optimization.

PPO actor-critic architecture [36], the *Feasibility-Net* employs supervised learning to predict velocity tracking rewards based on the same environmental data. The *Feasibility-Net* architecture incorporates a variational autoencoder that models terrain height distributions alongside a multilayer perceptron (MLP) head that generates feasibility predictions, enabling robust performance assessment under distribution shift conditions through out-of-distribution (OOD) detection. This joint training enables each policy to develop terrain specialization while its paired feasibility model learns terrain-specific performance predictions that are inherently grounded in the policy’s actual capabilities. During deployment, the framework transforms elevation maps into directional feasibility tensors that encode predicted policy performance across different movement directions, which are subsequently fused to create unified cost functions compatible with standard graph search algorithms. This design enables optimal path discovery with full transparency in policy selection decisions while supporting seamless integration of new policies without the need for retraining.

#### A. Learning Policy-Specific Feasibility Models

We next detail the architecture and training methodology of *Feasibility-Net*, which serves as the core mechanism that enables each locomotion policy to develop terrain-specific performance predictions while maintaining robust generalization capabilities. *Feasibility-Net* takes two primary inputs: (1) a task vector encoding velocity commands, and (2) a local heightmap sampled from the robot’s immediate vicinity. The network architecture employs an MLP head that outputs normalized feasibility scores representing the expected velocity tracking performance. We define feasibility as the predicted velocity tracking reward, which is learned through supervised regression with  $L_2$  loss minimization. The feasibility scores are bounded between 0.0 and 1.0, where 0.0 indicates complete tracking failure and 1.0 represents perfect velocity following. This formulation integrates seamlessly into standard RL pipelines by utilizing identical rollout data already collected during policy training. The velocity tracking reward function is formally defined as:

$$r_v = \exp(-\|v - \hat{v}\|_2/\sigma), \quad (1)$$

where  $v$  represents the agent’s root velocity vector,  $\hat{v}$  denotes the velocity command from the task vector, and  $\sigma = 0.25$  serves as a scaling parameter for controlling sensitivity.

A critical challenge in deploying specialized locomotion policies across diverse environments is the inevitable presence of distribution shift between training and deployment terrains. Feasibility predictions become unreliable when the model encounters terrain features that differ significantly from the training distribution. To address this fundamental limitation, we augment the *Feasibility-Net* with a variational autoencoder branch that explicitly models the distribution of training heightmaps. The VAE reconstruction error serves as a metric for OOD detection, where heightmaps exhibiting high reconstruction loss indicate terrain configurations that deviate substantially from the training distribution. This mechanism enables the system to modulate feasibility confidence based on terrain familiarity, thereby improving robustness during deployment. The complete *Feasibility-Net* optimization combines both feasibility prediction and distribution modeling through the following joint loss function:

$$\mathcal{L}_{\text{feas}} = \left\| \hat{f} - r_v \right\|_2^2 + \alpha \mathcal{L}_{\text{VAE}}, \quad (2)$$

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}}(\mathbf{H}_{x,y}) + \beta \mathcal{L}_{\text{KL}}(\mathbf{H}_{x,y}), \quad (3)$$

where  $\hat{f}$  denotes the *Feasibility-Net* output,  $\mathbf{H}_{x,y}$  represents the heightmap sampled around the robot’s position, and the weighting parameters are set to  $\alpha = 1.0$  and  $\beta = 5 \times 10^{-4}$ . The VAE loss components, denoted as  $\mathcal{L}_{\text{VAE}}$ , follow standard formulations [37], incorporating both reconstruction loss and Kullback-Leibler divergence regularization with respect to a Gaussian prior distribution.

#### B. Elevation Map to Feasibility Tensor Transformation

We now detail the deployment-time transformation process that converts elevation maps into actionable feasibility representations. This serves as the critical bridge between the learned feasibility models and the subsequent planning operations. An elevation map is formally represented as a 2D grid  $\mathcal{M} \in \mathbb{R}^{W \times H}$ , where each cell encodes the terrain height at spatial coordinates  $(x, y)$ . The feasibility computation across the complete map  $\mathcal{M}$  employs a sliding window approach, as in Fig. 3, that ensures spatial consistency with the training methodology established in the previous section. For each spatial location  $(x, y)$ , we extract a local heightmap patch  $\mathbf{H}_{x,y} \in \mathbb{R}^{w \times h}$  where the dimensions correspond precisely to the patch resolution used during training. Each patch is then paired with a synthetic velocity command vector that represents the intended direction from the current position.

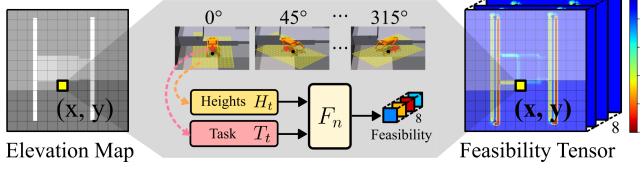


Fig. 3. Sliding window methodology for elevation map to feasibility tensor transformation. Local heightmap patches are extracted at each spatial location and processed through multiple directions to generate policy-specific feasibility predictions.

A fundamental challenge in feasibility assessment lies in accounting for the directional nature of locomotion capabilities. Since robot movement involves both position and orientation, and terrain traversability often exhibits directional dependencies, we must evaluate feasibility across multiple orientational configurations. To address this requirement, we systematically rotate each heightmap patch  $\mathbf{H}_{x,y}$  across  $d$  discrete directional configurations. The *Feasibility-Net* processes each rotated patch  $\mathbf{H}_{x,y}$  alongside its corresponding task vector to generate a directional feasibility estimate  $\mathbf{f}_{x,y} \in \mathbb{R}^d$ . Our implementation employs  $d = 8$  directional samples with uniform  $45^\circ$  angular spacing, resulting in an 8-channel feasibility tensor  $\mathcal{F} \in \mathbb{R}^{W \times H \times 8}$ . Each tensor channel encodes the predicted feasibility for executing movement commands from position  $(x, y)$  along the corresponding direction. While the directional feasibility estimates provide comprehensive spatial coverage, deployment across diverse environments necessitates explicit handling of distribution shift. Recall from the previous section that our *Feasibility-Net* incorporates VAE-based distribution modeling specifically to address this challenge. During the transformation process, we leverage this capability to modulate feasibility confidence based on terrain familiarity. The out-of-distribution adjustment weight  $w_{\text{ood}}$  is defined as follows:

$$w_{\text{ood}}(x, y) = \text{clip}\left(1 - \frac{\mathcal{L}_{\text{recon}}(\mathbf{H}_{x,y}) - \tau_{\text{low}}}{\tau_{\text{high}} - \tau_{\text{low}}}, 0, 1\right), \quad (4)$$

where  $\mathcal{L}_{\text{recon}}(\mathbf{H}_{x,y})$  is the reconstruction loss for  $\mathbf{H}_{x,y}$ , and the threshold parameters  $\tau_{\text{low}} = 0.1$  and  $\tau_{\text{high}} = 2.0$  define the acceptable reconstruction error range. This weighting mechanism approaches unity for terrain patches consistent with the training distribution and diminishes toward zero as reconstruction error increases, thereby reducing confidence in feasibility predictions for unfamiliar terrain configurations. The final weighted feasibility representation is computed as:

$$\hat{\mathcal{F}}(x, y) = w_{\text{ood}}(x, y) \cdot \mathbf{f}_{x,y}. \quad (5)$$

The transformation process produces policy-specific feasibility tensors that encode both locomotion capabilities and terrain familiarity across directional configurations. These representations form the input to our proposed multi-policy planning framework, which we detail in the following section, where multiple feasibility tensors are fused to enable coordinated navigation over specialized locomotion policies.

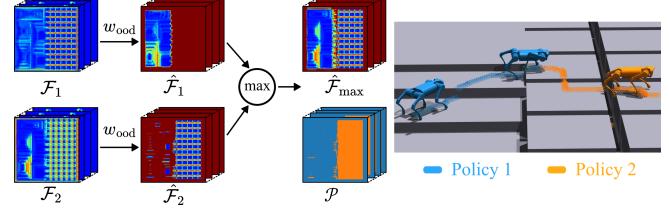


Fig. 4. Multi-policy feasibility tensor fusion and planning framework. Individual policy-specific feasibility representations are combined through maximum fusion to create unified cost functions, enabling graph search algorithms to discover optimal paths with transparent policy selection.

### C. Feasibility Tensor Fusion and Graph-Based Planning

We next present the multi-policy coordination framework that enables unified planning across specialized locomotion capabilities. Our approach employs a collection of velocity-tracking locomotion policies  $\{\pi_1, \pi_2, \dots, \pi_n\}$ , where each policy specializes in distinct terrain configurations, alongside their corresponding *Feasibility-Net* models  $\{F_1, F_2, \dots, F_n\}$ . These policy-model pairs undergo joint optimization through the alternating training paradigm established in previous sections. Given an  $\mathcal{M}$ , the framework generates multiple policy-specific feasibility tensors  $\{\hat{\mathcal{F}}_1, \hat{\mathcal{F}}_2, \dots, \hat{\mathcal{F}}_n\}$  through the transformation methodology detailed previously. The central challenge lies in coordinating these specialized representations to enable coherent navigation decisions that leverage the strengths of each policy while avoiding their limitations. We address this through a fusion strategy that creates a unified feasibility representation by selecting the maximum feasibility score across all policies:

$$\hat{\mathcal{F}}_{\max}(x, y, d) = \max_i \hat{\mathcal{F}}_i(x, y, d), \quad (6)$$

where  $\hat{\mathcal{F}}_i(x, y, d)$  represents the OOD-weighted feasibility for policy  $i$  at position  $(x, y)$  and direction  $d$ . This fusion mechanism ensures that the unified representation captures the optimal locomotion capability available at each location.

$\hat{\mathcal{F}}_{\max}$  serves as the foundation for graph-based path planning using Dijkstra's algorithm, which is employed rather than heuristic methods like A\* as the direction-dependent and policy-dependent cost structure violates admissibility requirements for optimal A\* performance. The edge cost between neighboring positions is defined as the inverse of  $\hat{\mathcal{F}}_{\max}$ . Upon completion, the framework determines the appropriate locomotion policy for each trajectory segment. The policy selection mechanism identifies which specialized policy provided the maximum feasibility score during fusion:

$$\mathcal{P}(x, y, d) = \arg \max_i \hat{\mathcal{F}}_i(x, y, d). \quad (7)$$

This strategy ensures optimal policy activation while maintaining interpretability of decision rationale. Fig. 4 depicts the pipeline from individual feasibility representations, fusion, to final policy assignment along the planned trajectory.

## IV. SIMULATION EVALUATION

We evaluate our feasibility-guided planning framework through comprehensive simulation studies using a

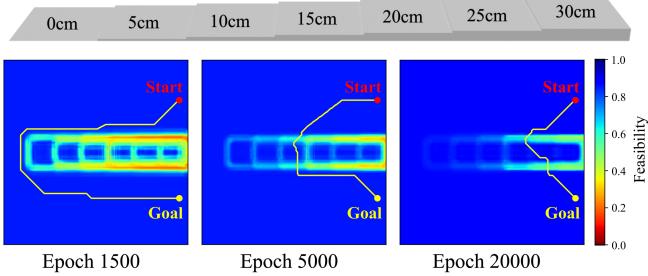


Fig. 5. Feasibility-guided planning adaptation across training progression on stepped terrain. As locomotion capabilities develop, the planner progressively selects shorter and more direct routes, demonstrating dynamic adaptation to policy skill evolution.

TABLE I

PLANNING STATISTICS FOR STEP POLICIES ACROSS DIFFERENT TRAINING STAGES ON STAIRS-LIKE TERRAIN

	Epoch 1.5K	Epoch 5K	Epoch 20K
Path Length ↓	26.75m	14.84m	9.73m
Success Rate ↑	100.00%	100.00%	100.00%
SPL ↑	0.19	0.34	0.51

quadrupedal robot platform. The experiments follows established protocols from DreamWaQ [38], implementing a modified PPO training loop [36] that alternates between locomotion policy optimization and supervised *Feasibility-Net* training. Each locomotion policy specializes in a specific terrain configuration, with its corresponding *Feasibility-Net* learning to predict terrain-specific traversal performance via the joint training methodology described in previous sections.

#### A. Feasibility-Guided Planning Adaptation

We first examine how the planning system adapts its route selection as locomotion capabilities develop during training. This analysis demonstrates the dynamic relationship between policy capability and planning behavior, and validates that feasibility predictions can accurately reflect the evolving skills of their associated locomotion policies. The experimental protocol analyzes planning behavior at three representative training checkpoints: early-stage (limited capability), mid-stage (developing capability), and late-stage (mature capability) on a stepped terrain environment with varying obstacle heights. At each checkpoint, we compute plans between fixed start and goal positions and generate feasibility maps that visualize how the planner utilizes feasibility estimates for route selection, as shown in Fig. 5. The results demonstrate clear adaptation to evolving policy capabilities. During early training stages, when the locomotion policy lacks step-climbing ability, the planner generates conservative routes that follow extended detours through flat terrain regions. As training progresses and feasibility estimates improve, the planner begins routing through medium-height obstacles and produces shorter path lengths. At the final training stage, when the policy achieves robust step-climbing performance, the planner selects direct trajectories through previously avoided regions. Table I presents the quan-

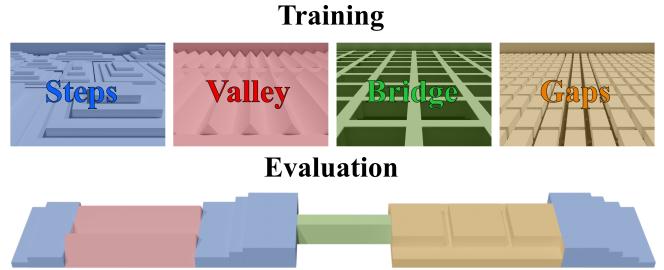


Fig. 6. Training and evaluation terrain configurations. Individual policies are trained on specialized terrain types (steps, gaps, bridge, valley) and evaluated on both individual terrains and a composite mixed environment combining all terrain features.

TABLE II

CONTROLLER SUCCESS RATES ACROSS TERRAIN TYPES COMPARING SPECIALIZED POLICIES, OUR APPROACH, AND A GENERAL POLICY

Terrain	Policy					
	Steps	Gaps	Bridge	Valley	Ours	General
Steps	<b>99.99</b>	0.00	0.00	61.30	99.96	98.75
Gaps	0.00	99.45	0.00	0.00	<b>99.49</b>	0.00
Bridge	0.00	0.00	98.07	0.00	<b>100.00</b>	0.00
Valley	0.00	0.00	0.00	99.60	<b>99.69</b>	58.94
Mixed	0.00	0.00	0.00	0.00	<b>98.60</b>	0.00

titative analysis. The planning maintains 100% success rates across all training stages and demonstrates that feasibility predictions consistently align with actual policy capabilities. The Success weighted by Path Length (SPL) metric [39] increases from 0.19 to 0.51 and reflects the planner’s ability to exploit improving locomotion skills for path optimization. These reveal that feasibility-guided planning ensures navigation reliability and dynamical adaptation to leverage policy capabilities for path generation.

#### B. Multi-Policy Planning Across Heterogeneous Terrains

We evaluate the multi-policy coordination framework across diverse terrain configurations. This experiment aims to demonstrate the framework’s ability to leverage specialized locomotion policies for navigation across heterogeneous environments that exceed the capabilities of individual policies. We train four distinct locomotion policies alongside their corresponding *Feasibility-Net* models, each specializing in a specific terrain type: (1) stepped obstacles, (2) gap traversal, (3) narrow bridge navigation, and (4) valley terrain. To assess both individual policy performance and multi-policy coordination capabilities, we construct a composite evaluation environment that integrates all four terrain features into a unified map, as illustrated in Fig. 6. This design enables direct comparison between specialized policies, generalist approaches, and our feasibility-guided coordination framework. The evaluation protocol employs 40 unique navigation tasks with randomized start and goal positions within each single-terrain environment. Each task consists of 250 independent trials. This produces 10,000 trials per terrain type for statistical significance. The mixed terrain evaluation employs fixed start and goal positions due to the singular entry and

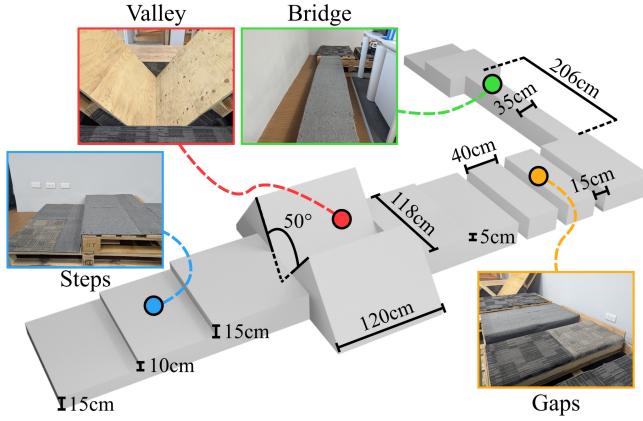


Fig. 7. Real-world planning on mixed terrains (steps, gaps, bridge, valley).

exit configuration of the composite environment. A trial is considered a failure if (1) the planner fails to generate a valid path due to insufficient feasibility across available routes, or (2) the planner successfully produces a path but the agent cannot reach the designated goal in the prescribed time limit.

Table II presents the comparative performance analysis across individual specialized policies, our feasibility-guided planning framework, and a baseline general policy trained jointly across all terrain types. The specialized policies achieve near-optimal performance within their respective domains, with success rates of 99.99% on stepped terrain and 98.07% on bridge navigation. However, these policies exhibit complete failure when deployed outside their specialized domains. The general policy demonstrates moderate performance on certain terrain types, achieving 98.75% success on steps and 58.94% on valley terrain, but suffers complete failure on others. This performance degradation illustrates the fundamental challenge of single-policy approaches across diverse terrain configurations, where conflicting optimization objectives and the need for terrain-specific reward engineering limit overall generalization capability. Our feasibility-guided planning framework addresses these limitations through dynamic policy selection based on local terrain characteristics. The framework consistently achieves performance equivalent to the best specialized policy for each terrain type, with success rates of 99.96% on steps, 99.49% on gaps, and 100% on bridge terrain. Most significantly, the framework maintains robust performance of 98.60% on the composite mixed terrain environment. This performance demonstrates that dynamic policy switching based on feasibility assessment enables the retention of specialized locomotion capabilities while providing the flexibility necessary for navigation across heterogeneous environmental conditions.

## V. REAL-WORLD VALIDATION

We validate the proposed methodology by constructing a heterogeneous test environment comprising four distinct terrain sections that correspond to our trained policy specializations. The environment includes stepped obstacles with heights ranging from 5 to 15 cm, gap traversal sections

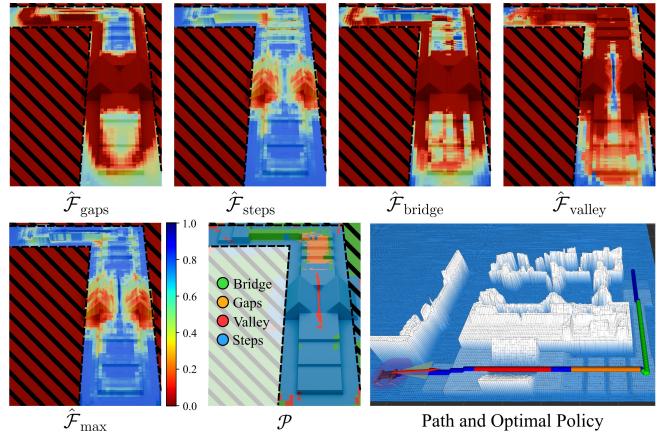


Fig. 8. Real-world feasibility tensor generation and policy selection. An elevation map is transformed into policy-specific feasibility representations, which enable the planner to select feasible paths and optimal locomotion policies based on local terrain characteristics.

TABLE III  
POLICY SWITCHING SUCCESS RATE ON DIFFERENT TERRAINS

From Terrain	Steps	Valley	Steps	Gaps	Bridge
To Terrain	Valley	Steps	Gaps	Bridge	Steps
Success Rate	95%	100%	100%	85%	100%

featuring platforms separated by 15 cm spacing, a narrow bridge segment with 35 cm width, and a valley terrain characterized by V-shaped geometry with 50° inclination angles. Fig. 7 presents the detailed geometric parameters alongside a comprehensive view of the mixed terrain configuration.

The environment is scanned using a Livox MID-360 LiDAR sensor, and the resulting point cloud undergoes denoising and downsampling to 5 cm resolution through voxel-based filtering. The processed point cloud is converted into an elevation map, which we transform into feasibility representations  $\hat{\mathcal{F}}_{\text{steps}}$ ,  $\hat{\mathcal{F}}_{\text{gaps}}$ ,  $\hat{\mathcal{F}}_{\text{bridge}}$ , and  $\hat{\mathcal{F}}_{\text{valley}}$  following the methodology described in Section III-B, as illustrated in Figure 8. The feasibility maps assign low scores to regions where the associated locomotion policy lacks specialization. During execution, the policy index map  $\mathcal{P}$  directs the controller to activate the appropriate policy at each location. We employ a Unitree A1 quadruped equipped with the same LiDAR sensor for real-time localization. Note that the supplementary video demonstrates the complete sequence.

Table III presents the policy switching success rates across terrain transitions under 20 attempts. Bridge and valley terrain segments exhibit increased difficulty in real-world deployment due to their requirement for precise specialized locomotion expertise. The reduced success rates for entering these terrain types stem primarily from localization imprecision, where premature policy switching causes the robot to become immobilized within the current terrain segment. The zero switching cost assumption between policies introduces additional challenges during locomotion transitions. Individual policy success rates within their respective specialization

domains reach 90% for steps, 100% for gaps, 85% for bridge traversal, and 100% for valley navigation. The experimental results indicate that *Feasibility-Net* successfully selects appropriate policies at optimal locations while maintaining feasible trajectory generation, achieving an overall success rate of 70% for complete mixed terrain traversal. In comparison, the general policy achieves 95% success on steps, 15% on bridge and 45% on valley terrain but fails completely on gaps segments, resulting in 0% success for the identical navigation task.

## VI. CONCLUSION

We present a feasibility-guided planning framework that enables coordinated navigation across heterogeneous terrain through multiple specialized locomotion policies. The approach addresses fundamental limitations in classical planning methods, namely the reliance on binary occupancy representations and hierarchical reinforcement learning approaches, which suffer from interpretability challenges and retraining requirements. Our core contribution focuses on the joint training paradigm that simultaneously optimizes locomotion policies with corresponding *Feasibility-Net* models, learning policy-specific traversal predictions grounded in actual locomotion capabilities while incorporating out-of-distribution detection for robust deployment. Our experimental validation demonstrates the effectiveness of the framework in multiple dimensions. The planning system successfully adapts its route selection to the locomotion capabilities evolved during training, maintaining the success rate of 100% at different training stages while achieving improvements in path efficiency with SPL increasing from 0.19 to 0.51. Multi-policy coordination experiments show that our approach consistently matches specialized policy performance within respective domains while attaining 98.60% success rate on mixed terrain environments, compared to complete failure of baseline general policies. Real-world deployment on a Unitree A1 quadruped achieves an overall success rate of 70% across four physical terrain configurations while maintaining interpretable policy selection and supporting seamless integration of new locomotion skills without system retraining.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support from the National Science and Technology Council (NSTC) in Taiwan under grant numbers NSTC 114-2221-E-002-069-MY3, NSTC 113-2221-E-002-212-MY3, and NSTC 114-2218-E-A49-026. The authors would also like to express their appreciation for the supports from NVIDIA Corporation and NVIDIA AI Technology Center (NVAITC). Furthermore, the authors extend their gratitude to the National Center for High-Performance Computing for providing the necessary computational and storage resources.

## REFERENCES

- [1] J. Di Carlo, P. M. Wensing, B. Katz, *et al.*, “Dynamic Locomotion in the MIT Cheetah 3 Through Convex Model-Predictive Control,” in *2018 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Oct. 2018, pp. 1–9.
- [2] J. Lee, J. Hwangbo, L. Wellhausen, *et al.*, “Learning quadrupedal locomotion over challenging terrain,” *Science Robotics*, vol. 5, no. 47, eabc5986, 21, 2020.
- [3] S. Macenski, F. Martín, R. White, and J. G. Clavero, “The Marathon 2: A Navigation System,” in *2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Oct. 2020, pp. 2718–2725.
- [4] D. Hoeller, N. Rudin, D. Sako, and M. Hutter, “ANYmal parkour: Learning agile navigation for quadrupedal robots,” *Science Robotics*, vol. 9, no. 88, eadi7566, 13, 2024.
- [5] K. Weerakoon, A. J. Sathyamoorthy, M. Elnoor, and D. Manocha, “VAPOR: Legged Robot Navigation in Unstructured Outdoor Environments using Offline Reinforcement Learning,” in *2024 IEEE Int. Conf. Robot. Autom. ICRA*, May 2024, pp. 10344–10350.
- [6] P. Henderson, R. Islam, P. Bachman, *et al.*, “Deep Reinforcement Learning That Matters,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 29, 2018.
- [7] J. Won, D. Gopinath, and J. Hodgins, “A scalable approach to control diverse behaviors for physically simulated characters,” *ACM Trans. Graph.*, vol. 39, no. 4, 33:33:1–33:33:12, 12, 2020.
- [8] X. B. Peng, M. Chang, G. Zhang, *et al.*, “MCP: Learning Composable Hierarchical Control with Multiplicative Compositional Policies,” in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.
- [9] G. Konidaris and A. Barto, “Skill Discovery in Continuous Reinforcement Learning Domains using Skill Chaining,” in *Advances in Neural Information Processing Systems*, vol. 22, Curran Associates, Inc., 2009.
- [10] J. H. Soeseno, Y.-S. Luo, T. P.-C. Chen, and W.-C. Chen, “Transition Motion Tensor: A Data-Driven Approach for Versatile and Controllable Agents in Physically Simulated Environments,” in *SIGGRAPH Asia 2021 Technical Communications*, Dec. 14, 2021, pp. 1–4.
- [11] G. Christmann, Y.-S. Luo, J. H. Soeseno, and W.-C. Chen, “Expanding Versatility of Agile Locomotion through Policy Transitions Using Latent State Representation,” in *2023 IEEE Int. Conf. Robot. Autom. ICRA*, May 2023, pp. 5134–5140.
- [12] G. Christmann, Y.-S. Luo, and W.-C. Chen, “Expert Composer Policy: Scalable Skill Repertoire for Quadruped Robots,” in *2024 IEEE Int. Conf. Robot. Autom. ICRA*, May 2024, pp. 9727–9734.
- [13] G. Bledt, M. J. Powell, B. Katz, *et al.*, “MIT Cheetah 3: Design and Control of a Robust, Dynamic

- Quadruped Robot," in *2018 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Oct. 2018, pp. 2245–2252.
- [14] J. Humphreys, J. Li, Y. Wan, *et al.*, "Bio-Inspired Gait Transitions for Quadruped Locomotion," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6131–6138, 2023.
- [15] Q. Zhu, Z. Sun, S. Xia, *et al.*, "Learning-Based Traversability Costmap for Autonomous Off-Road Navigation," in *Intelligent Robotics*, Z. Hou, X. Fu, Q. Hu, *et al.*, Eds., Singapore: Springer Nature, 2025, pp. 301–312.
- [16] S. Gigante, "A Machine Learning-Based Approach for Traversability Analysis and Path Planning in Martian Rovers," laurea, Politecnico di Torino, Apr. 21, 2023, 97 pp.
- [17] M. V. Gasparino, A. N. Sivakumar, and G. Chowdhary, "WayFASTER: A Self-Supervised Traversability Prediction for Increased Navigation Awareness," in *2024 IEEE Int. Conf. Robot. Autom. ICRA*, May 2024, pp. 8486–8492.
- [18] M. V. Gasparino, A. N. Sivakumar, Y. Liu, *et al.*, "WayFAST: Navigation With Predictive Traversability in the Field," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10651–10658, 2022.
- [19] C. Sevastopoulos and S. Konstantopoulos, "A Survey of Traversability Estimation for Mobile Robots," *IEEE Access*, vol. 10, pp. 96331–96347, 2022.
- [20] H. Zhang, Z. Li, X. Zeng, *et al.*, "Traversability-Aware Legged Navigation by Learning from Real-World Visual Data." arXiv: 2410.10621. (Nov. 2024).
- [21] Y. Cheng, H. Liu, G. Pan, *et al.*, "Quadruped robot traversing 3D complex environments with limited perception," in *2024 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Oct. 2024, pp. 9074–9081.
- [22] L. Gan, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "Energy-Based Legged Robots Terrain Traversability Modeling via Deep Inverse Reinforcement Learning," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8807–8814, 2022.
- [23] M. Oh, B. Yu, I. M. A. Nahrendra, *et al.*, "TRIP: Terrain Traversability Mapping With Risk-Aware Prediction for Enhanced Online Quadrupedal Robot Navigation." arXiv: 2411.17134. (Nov. 2024).
- [24] L. Wellhausen and M. Hutter, "Rough Terrain Navigation for Legged Robots using Reachability Planning and Template Learning," in *2021 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Sep. 2021, pp. 6914–6921.
- [25] S. Ha, J. Lee, M. van de Panne, *et al.*, "Learning-based legged locomotion: State of the art and future perspectives," *The International Journal of Robotics Research*, vol. 44, no. 8, pp. 1396–1427, 1, 2025.
- [26] G. Haddeler, M. Y. Chuah, Y. You, *et al.*, "Traversability analysis with vision and terrain probing for safe legged robot navigation," *Front. Robot. AI*, vol. 9, 22, 2022.
- [27] Z. Huang, "Discovering and Exploiting Skills in Hierarchical Reinforcement Learning," *IEEE Access*, vol. 12, pp. 163042–163055, 2024.
- [28] L.-J. Lin, "Hierarchical learning of robot skills by reinforcement," in *IEEE International Conference on Neural Networks*, Mar. 1993, 181–186 vol.1.
- [29] S. Pateria, B. Subagdja, A.-h. Tan, and C. Quek, "Hierarchical Reinforcement Learning: A Comprehensive Survey," *ACM Comput. Surv.*, vol. 54, no. 5, 109:1–109:35, 5, 2021.
- [30] J. B. Evans and Ö. Şimşek, "Creating Multi-Level Skill Hierarchies in Reinforcement Learning," *Advances in Neural Information Processing Systems*, vol. 36, pp. 48472–48484, 15, 2023.
- [31] X. B. Peng, G. Berseth, K. Yin, and M. Van De Panne, "DeepLoco: Dynamic locomotion skills using hierarchical deep reinforcement learning," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–13, 31, 2017.
- [32] D. Jain, A. Iscen, and K. Caluwaerts, "Hierarchical Reinforcement Learning for Quadruped Locomotion," in *2019 IEEE/RSJ Int. Conf. Intell. Robots Syst. IROS*, Nov. 2019, pp. 7551–7557.
- [33] T. Li, N. Lambert, R. Calandra, *et al.*, "Learning Generalizable Locomotion Skills with Hierarchical Reinforcement Learning," in *2020 IEEE Int. Conf. Robot. Autom. ICRA*, May 2020, pp. 413–419.
- [34] J. Lee, M. Bjelonic, A. Reske, *et al.*, "Learning robust autonomous navigation and locomotion for wheeled-legged robots," *Science Robotics*, vol. 9, no. 89, eadi9641, 24, 2024.
- [35] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning," in *Proceedings of the 5th Conference on Robot Learning*, PMLR, Jan. 11, 2022, pp. 91–100.
- [36] J. Schulman, F. Wolski, P. Dhariwal, *et al.*, "Proximal Policy Optimization Algorithms." arXiv: 1707.06347. (Aug. 2017).
- [37] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes." arXiv: 1312.6114. (Dec. 2022).
- [38] I. M. Aswin Nahrendra, B. Yu, and H. Myung, "DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning," in *2023 IEEE Int. Conf. Robot. Autom. ICRA*, May 2023, pp. 5078–5084.
- [39] P. Anderson, A. Chang, D. S. Chaplot, *et al.*, "On Evaluation of Embodied Navigation Agents." arXiv: 1807.06757. (Jul. 2018).