2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

SURF-Loco: Mastering Complex Industrial Terrains with 3D Surfel-based Reinforcement Learning for Legged Robots

Bailin He[1], Xiting Zhao[1,2,†], Qiao Sun[2], Xiaoyi Hu[2], Haojie Liu[2], Jiangwei Zhong[2], Wenqiang Zhang[1,*]

***Abstract*— Legged robots offer significant potential for navigating complex industrial terrains, but their capabilities are often constrained by perception systems struggling to interpret intricate 3D geometry. Conventional 2D/2.5D representations like depth or elevation maps fail to capture complex 3D geometry, leading to unsafe locomotion. This paper presents SURF-Loco, a novel framework that enables robust legged locomotion by learning directly from a 3D surfel-based model. Our approach uses surfels to create an omnidirectional representation that explicitly encodes the geometric properties necessary for stable locomotion. We integrate this structured 3D representation into an end-to-end Mixture-of-Experts (MoE) reinforcement learning policy. A variational autoencoder (VAE) distills the complex 3D surroundings into a compact latent context. This geometric context enables a gating network to dynamically select expert sub-policies for agile, context-aware actions. We validate our method on the Lenovo Daystar IS hexapod robot, achieving robust zero-shot sim-to-real transfer on a variety of challenging industrial obstacles.**

I. INTRODUCTION

The ongoing transformation toward Industry 4.0 demands robotic systems capable of operating autonomously in increasingly complex and unstructured industrial environments, where wheeled and tracked platforms are limited by challenges like perforated staircases, gaps and pipe networks. These environments, common in power substations, petrochemical plants, and construction sites, require a different approach to robotic locomotion.

Legged robots offer the potential to navigate such challenging terrains by leveraging discrete footholds and high-degree-of-freedom limbs [1] [2]. However, their performance is constrained by the perception-action loop. The critical challenge lies not in mechanical design or low-level control, but in how these robots perceive, represent, and reason about complex 3D geometry to make robust locomotion decisions.

Existing perceptive locomotion policies exhibit several limitations in cluttered industrial environments. Policies conditioned on forward-facing depth cameras [3], [4] or sparse LiDAR scans create large blind spots that make critical lateral and backward movements hazardous in confined spaces. While 2.5D elevation maps provide a more holistic view and are effective on many terrains [5], [6] , they are single-valued height functions, which fundamentally limits their ability to represent multi-level structures like overhangs or grated floors. Furthermore, their coarse resolution, often around 10 cm, makes them insensitive to ubiquitous, centimeter-scale hazards like thin cables or floor gaps. Figure 1 illustrates the differences between these traditional representations and our proposed surfel-based approach.

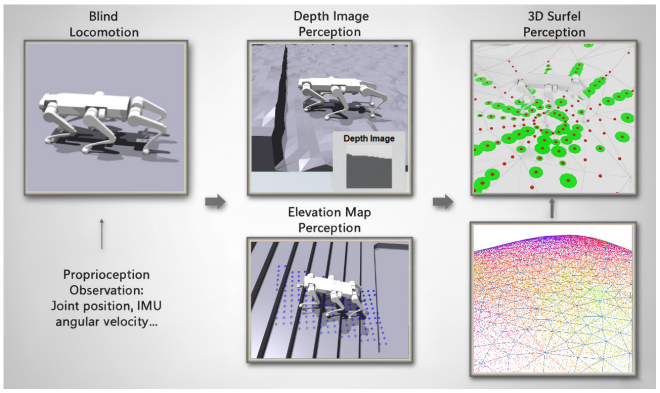

Fig. 1. Comparison of perception modalities for legged locomotion. From left to right: (1) Blind locomotion relies solely on proprioception, lacking foresight. (2) Depth image provides a limited, forward-facing view. (3) 2.5D elevation map fails to represent complex 3D geometry like the shown grated floor. (4) Our 3D surfel representation provides an omnidirectional and complete geometric model, enabling robust navigation on such challenging terrains.

To address these limitations, we introduce SURF-Loco: a 3D Surfel-based Reinforcement Learning framework for robust legged locomotion. Unlike search-and-rescue or exploratory scenarios, many industrial applications (e.g., routine inspection or maintenance in power plants and factories) feature largely static environments that can be pre-mapped. By using a pre-scanned map and a localization module, our locomotion policy is decoupled from a specific perception sensor, enhancing hardware flexibility. Our approach utilizes 3D surfels, which are oriented surface elements with position, normal, and radius, as the core geometric representation. Surfels provide several critical advantages: (1) Compared to unstructured point clouds, they possess inherent structure by encoding surface normals and radii, which is crucial for evaluating terrain stability. (2) Compared to voxel grids, they are more efficient at representing surfaces, avoiding the memory and resolution trade-offs associated with volumetric representations. (3) They naturally represent arbitrary 3D topologies including overhangs and hollow structures.

Our key innovation is coupling rich 3D representation with a Mixture-of-Experts (MoE) reinforcement learning architecture. A gating network uses geometric cues from the

[1]Shanghai Key Lab of Intelligent Information Processing, College of Computer Science and Artificial Intelligence, Fudan University, Shanghai, China. blhe25@m.fudan.edu.cn, wqzhang@fudan.edu.cn
[2]Robotics Lab, Lenovo CTO Organization, Shanghai, China. {zhaoxt9, sunqiao6, huxy31, liuhj31, zhongjw}@lenovo.com
*Corresponding author: Wenqiang Zhang wqzhang@fudan.edu.cn
†Corresponding author: Xiting Zhao zhaoxt9@lenovo.com

surfel map to dynamically select experts, producing stable and context-aware control. The rich geometric cues provided by surfels enable the gating network to reliably distinguish between scenarios requiring different expert strategies. By learning a direct perception-to-action policy in a single stage, our approach has agile and adaptive locomotion behaviors.

We validate our approach on the Lenovo Daystar IS hexapod robot. Unlike quadrupeds, hexapods can maintain at least three points of stable support throughout their gait cycle, providing the static stability required for precise maneuvering and stationary inspection tasks. Experiments demonstrate that our method successfully traverses challenging industrial obstacles and achieves robust zero-shot sim-to-real transfer.

The main contributions of this work are:

- We introduce SURF-Loco, the first 3D surfel-based perception and control framework for legged locomotion, which provides an omnidirectional, high-resolution 3D geometric representation to capture the full complexity of industrial environments.
- We develop a Mixture-of-Experts (MoE) reinforcement learning architecture that directly uses structured surfel geometry to learn specialized, context-aware locomotion skills within a single end-to-end policy.
- We validate our approach through extensive real-world deployments on hexapod robots, achieving zero-shot sim-to-real transfer for continuous traversal of challenging industrial terrains.

II. RELATED WORKS

Research in legged robot locomotion has made significant strides, broadly categorized by the perceptual information used to inform control. One major branch focuses on proprioceptive or 'blind' locomotion, which relies solely on the robot's internal state and has achieved remarkable agility on simpler terrains using techniques like privileged learning [7], [8]. Another branch incorporates exteroceptive sensing, primarily using 2D depth images or 2.5D elevation maps to provide foresight, enabling navigation of more complex obstacles [4], [5]. More recently, a third category has emerged that leverages true 3D representations like point clouds or voxels to tackle geometrically challenging environments [9]. In this section, we survey these categories, analyze their respective strengths and limitations, and select representative state-of-the-art methods from each for subsequent experimental comparison with our proposed framework.

A. Proprioceptive (Blind) Locomotion

Policies conditioned solely on proprioceptive feedback, such as joint states and IMU data, have demonstrated remarkable robustness and agility on flat or moderately challenging terrain [7], [8]. To enhance their adaptability, some approaches learn to implicitly infer environmental properties from the robot's dynamic response to contact [10] or even develop a form of "terrain imagination" that operates without direct exteroceptive input [11]. Concurrently, architectural innovations, such as Mixture of Experts (MoE) frameworks, have enabled single policies to master diverse repertoires of skills without the typical gradient conflicts of multitask learning [12]. However, they lack the foresight required for proactive planning and struggle with complex terrains involving large gaps, high obstacles, or cluttered spaces where precise foot placement is paramount [9].

B. Vision-based Locomotion with 2D/2.5D Perception

To enable foresight, many approaches incorporate exteroceptive sensing, primarily using depth cameras to provide a 2D or 2.5D representation of the upcoming terrain. Early successes in this area demonstrated highly dynamic behaviors, such as parkour, by feeding low-resolution depth images directly into the control policy [4], [13], [14]. However, these methods are sensitive to visual occlusions, sensor noise, and limited fields of view.

More advanced techniques build richer representations from partial visual input to address these shortcomings. These include learning world models to predict future perceptual states [15], using geometric priors to "imagine" occluded regions [16], or explicitly learning an online terrain reconstruction module to provide a denoised, local heightmap to the policy [17], [18]. Despite significant performance improvements, these methods are fundamentally limited by their 2.5D world understanding, failing to capture complex 3D geometry like overhangs or vertical structures.

C. Locomotion with Structured 3D Representations

Another line of work processes raw sensor data, typically from LiDAR, into more structured 3D representations. A common approach is to generate local elevation maps, which provide a stable and robust input for the policy. These maps have been used in various forms, from sampling discrete points [6], [19] and guiding foothold selection [20], to more advanced multi-layer structures designed to represent complex terrains with overhangs [21]. These methods benefit from LiDAR odometry [22] and elevation mapping [5] pipelines that filter out sensor noise for sim2real deployment.

To tackle environments where a 2.5D heightmap is insufficient, a recent trend is moving towards true 3D representations. Researchers have explored using 3D occupancy voxels for navigating confined spaces [9], fusing multi-modal data into semantic Bird's Eye View (BEV) maps for holistic scene understanding [23], and processing raw 3D point clouds directly with specialized network architectures [24]. Others have used VAE-based world models for perception denoising [25] or integrated high-frequency LiDAR-IMU fusion to enhance agility [26]. These methods confirm the value of 3D information, but often rely on dense, computationally intensive representations like voxels or require complex networks to interpret unstructured point clouds.

D. Surfels for 3D Mapping and Navigation

Surfels have proven highly effective in 3D reconstruction and SLAM systems, from early dense fusion methods to modern, efficient LiDAR-inertial odometry frameworks [27]–[29], offering an efficient and accurate representation of surfaces. Each surfel encodes position, normal, and radius,

providing rich geometric information in a compact form. Surfel-based maps have also been used to generate risk-aware traversability maps for quadruped path planning [30]. However, the direct use of surfels as a primary perceptual input for a low-level, learning-based locomotion policy remains unexplored.

Our work bridges this gap by proposing surfels as the geometric representation for an end-to-end locomotion controller. The properties of surfels, such as surface normals indicating contact feasibility, radius informing foothold stability, and a sparse surface-only encoding, are exceptionally well-aligned with the demands of agile legged locomotion. We demonstrate this by learning a policy that directly consumes a surfel-based model of the environment, achieving robust performance on a hexapod robot.

III. METHODOLOGY

Our proposed framework, SURF-Loco, empowers legged robots with robust locomotion capabilities in complex 3D environments through a novel surfel-based perception pipeline combined with a Mixture-of-Experts (MoE) reinforcement learning policy. This section details the core components of our approach: first, how the robot perceives the world using a rich 3D surfel representation; second, the architecture of our control policy that translates perception into action; and finally, the reward function designed to foster stable and effective locomotion. The structure, depicted in Fig. 2, processes proprioceptive state history and two distinct forms of exteroceptive information derived from surfel-based perception pipeline.

A. Surfel-based Perception Pipeline

As illustrated in Fig. 2, instead of traditional 2.5D elevation maps, our policy uses a more expressive surfel-based representation of the 3D surroundings. The perception pipeline first generates a local surfel map and then encodes it into a compact latent feature vector.

1) Local Surfel Map Generation: Our perception process begins with the robot's accurate pose (position and orientation) within an existing mesh of the environment. In simulation, this pose is directly obtained from the physics engine. We then generate a local, ego-centric surfel map by performing raycasting from the robot's base into the surrounding mesh. To ensure this perception pipeline is fast enough, we use the NVIDIA Warp high-performance simulation framework to implement the raycasting on the GPU [31]. Building upon the approach of [32], which uses Warp for fast LiDAR simulation, we modify the ray-tracing kernel to additionally extract the surface normal and estimate the local surface radius at each point of intersection. This allows efficient, parallel generation of our rich surfel map.

To balance comprehensive awareness with computational efficiency, we employ a non-uniform spherical grid for raycasting. Informed by prior works that utilize spherical grids [9] and prioritize near-ground perception [33], we configured a 17×17 array of rays. This resolution was experimentally chosen as a trade-off: coarser grids resulted in sparse maps that caused the robot to step into blind spots, while denser grids incurred high computational costs. The 17×17 array captures sufficient detail for stable locomotion. The grid's structure enables dense sampling of the ground and efficient, sparser omnidirectional awareness of larger obstacles:

- **Downward Ground Scan (17×11):** A dense grid is cast downwards to capture detailed ground geometry, crucial for precise and stable locomotion.
- **Surrounding Scan (17×6):** Sparser grids are cast horizontally and upwards to perceive obstacles, walls, and overhanging structures, ensuring omnidirectional awareness.

For each ray that intersects the mesh, a surfel is generated. Each surfel is a 5-dimensional vector encoding key geometric properties: (1) normalized distance to the intersection point; (2) estimated local surface radius, indicating patch size; and (3-5) Euler angles (z, y, x) representing the surface normal. This process yields a structured spherical surfel map $\mathbf{S}_t \in \mathbb{R}^{17\times17\times5}$ at each timestep t.

2) Perception Encoding via a Variational Autoencoder: The raw $17 \times 17 \times 5$ surfel map is too high-dimensional to serve directly as a policy input. To distill this rich geometric information into a compact and robust representation, we employ a Multilayer Perceptron (MLP) based Variational Autoencoder (VAE), which we term the *Surfel Encoder*.

The VAE consists of a probabilistic encoder $q_\phi(\mathbf{z}_t|\mathbf{S}_t)$ and a decoder $p_\theta(\hat{\mathbf{S}}_t|\mathbf{z}_t)$. The full surfel map $\mathbf{S}_t$ is first flattened into a 1445-dimensional vector and passed through the MLP-based encoder. The encoder outputs the parameters of a diagonal Gaussian distribution: a mean vector $\boldsymbol{\mu}_z$ and a log-variance vector $\log \boldsymbol{\sigma}_z^2$. A latent vector $\mathbf{z}_t \in \mathbb{R}^L$ (dimension $= 36$) is then sampled using the reparameterization trick:

$$\mathbf{z}_t = \boldsymbol{\mu}_z + \boldsymbol{\sigma}_z \odot \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \tag{1}$$

This latent vector $\mathbf{z}_t$ serves as a holistic, compressed representation of the robot's full 3D surroundings. While implicit, it effectively encodes geometric information such as obstacle occupancy and terrain elevation, providing global context to the policy.

The VAE is optimized as part of the overall learning process to minimize a composite loss function that balances reconstruction quality with latent space regularization:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + D_{\text{KL}}(q_\phi(\mathbf{z}_t|\mathbf{S}_t)||\mathcal{N}(0, \mathbf{I})) \tag{2}$$

The first term, $\mathcal{L}_{\text{recon}}$, is the mean squared error (MSE) between the original surfel map $\mathbf{S}_t$ and its reconstruction $\hat{\mathbf{S}}_t$ from the decoder, ensuring geometric fidelity:

$$\mathcal{L}_{\text{recon}} = ||\mathbf{S}_t - \hat{\mathbf{S}}_t||_2^2 \tag{3}$$

The second term is the Kullback-Leibler (KL) divergence between the encoder's output distribution and a standard normal prior $\mathcal{N}(0, \mathbf{I})$. This term acts as a regularizer on the latent space. We use its analytical form for diagonal Gaussians:

$$D_{\text{KL}}(q_\phi(\mathbf{z}_t|\mathbf{S}_t)||\mathcal{N}(0, \mathbf{I})) = 0.5 \sum_{j=1}^{L} (\mu_{z,j}^2 + \sigma_{z,j}^2 - \log(\sigma_{z,j}^2) - 1) \tag{4}$$

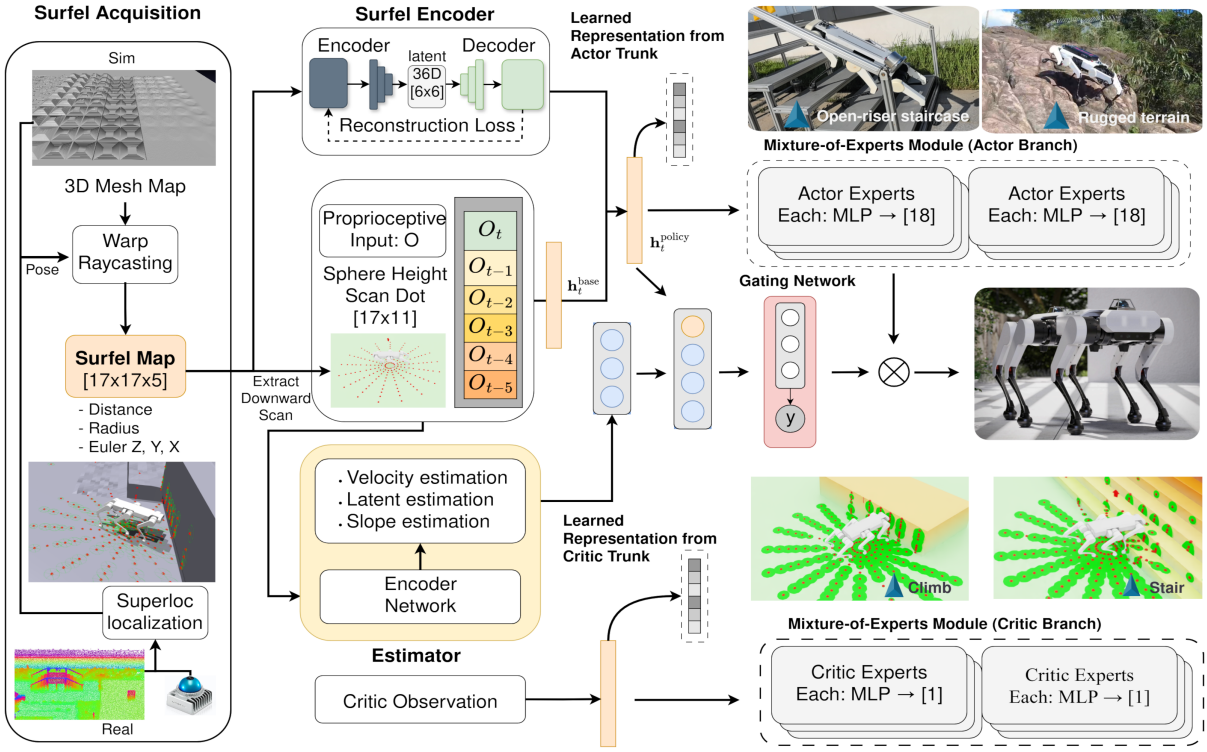

Fig. 2. The overall framework of SURF-Loco. The policy processes proprioceptive history and two streams of geometric information from the surfel map: a downward ground scan and a compressed global context from the VAE. This fused representation is fed into a Mixture-of-Experts (MoE) actor, where a gating network dynamically selects expert sub-policies to generate the final action.

This training objective encourages the VAE to learn a latent space that not only accurately represents the input geometry but is also well-structured and beneficial for the downstream control task.

B. SURF-Loco Control Policy

The SURF-Loco policy is a Mixture-of-Experts (MoE) architecture that learns specialized locomotion skills. Distinct from other MoE approaches [12], [21], [34], the proposed framework integrates a Variational Autoencoder (VAE) for perception processing with an MoE policy, all optimized within a single-stage, end-to-end training process.

1) State Representation and Multi-stage Encoding: As shown in the left and middle portions of Fig. 2, the policy's understanding of the world is formed from three sources:

1) **Proprioceptive History ($\mathbf{O}_t$):** A history of the last 6 timesteps of the robot's internal state $\mathbf{o}_t$. Each $\mathbf{o}_t \in \mathbb{R}^{69}$ contains joint information, IMU readings, and gait commands, as detailed in Table I. This history is denoted $\mathbf{O}_t \in \mathbb{R}^{6\times 69}$.
2) **Downward Ground Scan ($\mathbf{e}_t$):** A direct spherical scan of the ground. This input is extracted from the downward-facing 17×11 height of the full surfel map, flattened to a vector $\mathbf{e}_t \in \mathbb{R}^{187}$. It provides the policy with explicit, fine-grained information for local terrain.
3) **Global Geometric Context ($\mathbf{z}_t$):** The latent vector $\mathbf{z}_t \in \mathbb{R}^{36}$ from the VAE encoder, representing the entire omnidirectional surfel map. This provides broader contextual awareness of surrounding obstacles and terrain structure.

Using these two perception inputs enables the policy to learn fine-grained local geometry for immediate foothold decisions while also leveraging a broader global context for strategic maneuvers and obstacle avoidance.

TABLE I

ACTOR OBSERVATION COMPONENTS AT EACH TIMESTEP ($d_o = 69$).

Variable	Description	Dimension
$\mathbf{c}$	command (e.g., velocity)	$\mathbb{R}^3$
$\mathbf{g}$	projected gravity	$\mathbb{R}^3$
$\boldsymbol{\omega}$	angular velocity	$\mathbb{R}^3$
$\mathbf{q}$	joint positions	$\mathbb{R}^{18}$
$\dot{\mathbf{q}}$	joint velocities	$\mathbb{R}^{18}$
$\mathbf{a}_{t-1}$	previous action	$\mathbb{R}^{18}$
$\sin(2\pi\,\tilde{\phi}_t)$	gait phase (sine)	$\mathbb{R}^1$
$\cos(2\pi\,\tilde{\phi}_t)$	gait phase (cosine)	$\mathbb{R}^1$
u_t^{stand}	stand command flag	$\{0,1\}$
P_t	gait period	$\mathbb{R}^1_{>0}$
d_t	duty factor	$(0,1)$
u_t^{swap}	phase swap flag	$\{0,1\}$
	Total	$\mathbb{R}^{69}$

Note. Normalized phase update: $\phi_{t+1} = (\phi_t + \Delta t/P_t) \bmod 1$; swapped phase: $\tilde{\phi}_t = \left(\phi_t + \frac{1}{2}\,u_t^{\text{swap}}\right) \bmod 1$. Stand/swap flags $u_t^{\text{stand}},\, u_t^{\text{swap}} \in \{0,1\}$.

The encoding process fuses these information streams as follows: First, the proprioceptive history and the local ground scan are concatenated and processed by a *base trunk encoder* f_{base} to create a base latent representation:

$$\mathbf{x}_t = \left[\,\text{vec}(\mathbf{O}_t);\ \mathbf{e}_t\,\right] \in \mathbb{R}^{414+187=601} \tag{5}$$

$$\mathbf{h}_t^{\text{base}} = f_{\text{base}}(\mathbf{x}_t) \in \mathbb{R}^{256} \tag{6}$$

Next, this base representation is concatenated with the global geometric context from the VAE. This combined vector is passed through a second *fusion trunk encoder* f_{fusion} to produce the final policy latent state $\mathbf{h}_t^{\text{policy}}$:

$$\mathbf{h}_t^{\text{fused}} = [\mathbf{h}_t^{\text{base}};\ \mathbf{z}_t] \in \mathbb{R}^{256+36=292} \tag{7}$$

$$\mathbf{h}_t^{\text{policy}} = f_{\text{fusion}}(\mathbf{h}_t^{\text{fused}}) \in \mathbb{R}^{256} \tag{8}$$

This final state $\mathbf{h}_t^{\text{policy}}$ encapsulates all available information and serves as the input to the actor and gating networks.

2) Mixture-of-Experts Actor with Gaussian Mixture Policy: Given the rich latent state $\mathbf{h}_t^{\text{policy}}$, the actor generates actions using an MoE architecture. It consists of K expert networks and a gating network. Each expert k outputs the parameters of a diagonal Gaussian distribution:

$$\pi_k(\mathbf{a}_t|\mathbf{h}_t^{\text{policy}}) = \mathcal{N}\big(\boldsymbol{\mu}_k(\mathbf{h}_t^{\text{policy}}),\ \text{diag}(\boldsymbol{\sigma}_k^2(\mathbf{h}_t^{\text{policy}}))\big)$$

The gating network g_ψ computes a set of weights $\mathbf{w}_t$ that determine the contribution of each expert:

$$\mathbf{w}_t = g_\psi(\mathbf{h}_t^{\text{policy}}) \in \Delta^{K-1}, \qquad \sum_{k=1}^{K} w_{t,k} = 1$$

The final action distribution is a weighted sum of the expert distributions:

$$\pi(\mathbf{a}_t|\mathbf{h}_t^{\text{policy}}) = \sum_{k=1}^{K} w_{t,k}\, \pi_k(\mathbf{a}_t|\mathbf{h}_t^{\text{policy}})$$

During training, actions are sampled from this mixture. During deployment, the deterministic mean action is used for stability:

$$\hat{\mathbf{a}}_t = \sum_{k=1}^{K} w_{t,k}\, \boldsymbol{\mu}_k(\mathbf{h}_t^{\text{policy}})$$

To prevent expert collapse, the diverse terrain curriculum enforces specialization. In rare cases where experts collapse on hard terrains, initializing on simpler terrains and resuming training resolves it. Furthermore, high-frequency expert switching is regularized by action rate and smoothness penalties.

3) Critic with Mixture-of-Experts Value Function: The critic network, which estimates the state-value function, also utilizes an MoE architecture. It receives a more comprehensive set of information, including privileged environment data available only in simulation, to provide a better training signal. The critic observation, detailed in Table II, includes proprioceptive state, base velocity, external disturbances, the local surfel scan, and projected terrain normals.

TABLE II

CRITIC OBSERVATION COMPONENTS AT EACH TIMESTEP($d_{\text{CRITIC}} = 265$).

Variable	Description	Dimension
$\mathbf{o}_t^{\text{current}}$	Current proprioceptive state (as in Table I)	$\mathbb{R}^{69}$
$\mathbf{v}_t^{\text{base}}$	Scaled base linear velocity (privileged)	$\mathbb{R}^3$
$\mathbf{d}_t$	External disturbance (privileged)	$\mathbb{R}^3$
$\mathbf{e}_t$	Downward Ground Scan (17x11 grid)	$\mathbb{R}^{187}$
$\mathbf{n}_t^{\text{terrain}}$	Projected terrain normal (privileged)	$\mathbb{R}^3$
	Total	$\mathbb{R}^{265}$

TABLE III

REWARD FUNCTIONS. POSITIVE = REWARD; NEGATIVE = PENALTY.

Type	Item	Formula	Weight
Tracking			
	Tracking lin vel	$r_{\text{lin}} = \begin{cases}\exp\big(-\sigma_{\text{lin}}\|\mathbf{c}_{xy}-\mathbf{v}_{xy}\|_2^2\big), & \text{move}\\ \exp\big(-2\sigma_{\text{lin}}\|\mathbf{c}_{xy}-\mathbf{v}_{xy}\|_1\big), & \text{stand}\end{cases}$	3.8
	Tracking ang vel	$r_{\text{ang}} = \begin{cases}\exp\big(-\sigma_{\text{ang}}(c_\psi-\omega_z)^2\big), & \text{move}\\ \exp\big(-2\sigma_{\text{ang}}\lvert c_\psi-\omega_z\rvert\big), & \text{stand}\end{cases}$	1.7
	Contact timing match	$r_{\text{ct}} = \sum_{i=1}^{N_f}\mathbf{1}\big[(F_{i,z}>1) \Leftrightarrow (\text{phase}_i \bmod 1 < 0.56)\big]\cdot(1-\mathbf{1}_{\text{climb/gap}})$	0.185
Motion / Effort			
	Torques (L2)	$r_\tau = \sum_j \tau_j^2$	-1.0×10^{-4}
	DOF acc	$r_{\ddot{q}} = \sum_j \left(\frac{\dot{q}_{t-1}^{(j)}-\dot{q}_t^{(j)}}{\Delta t}\right)^2$	-2.5×10^{-7}
	Action rate	$r_{\Delta a} = \sum_{j\in\mathcal{A}}\big(a_t^{(j)}-a_{t-1}^{(j)}\big)^2$	-0.1
	Smoothness (2nd)	$r_{\Delta^2 a} = \sum_{j\in\mathcal{A}}\big(a_t^{(j)}-2a_{t-1}^{(j)}+a_{t-2}^{(j)}\big)^2$	-0.03
	DOF vel (L2)	$r_{\dot{q}} = \sum_j(\dot{q}^{(j)})^2$	-1.0×10^{-4}
	DOF vel limits	$r_{\text{vel-lim}} = \sum_j \max\big(0,\ \lvert\dot{q}^{(j)}\rvert - \alpha\,\dot{q}_{\max}^{(j)}\big)$	-0.1
	Torque limits	$r_{\text{tor-lim}} = \sum_j \max\big(0,\ \lvert\tau^{(j)}\rvert - \beta\,\tau_{\max}^{(j)}\big)$	-0.05
	Joint power	$r_{\text{pwr}} = \sum_j \lvert\dot{q}^{(j)}\rvert\,\lvert\tau^{(j)}\rvert$	-1.0×10^{-4}
	Feet slip	$r_{\text{slip}} = \sum_{i=1}^{N_f}\mathbf{1}[F_{i,z}>1]\,\|\mathbf{v}_{i,xy}\|_2^2$	-0.12
	Collision	$r_{\text{coll}} = \sum_{b\in\mathcal{B}}\mathbf{1}\big[\|\mathbf{F}_b\|>0.1\big]$	-1.0
Posture & Stability			
	Base height error	$r_h = (h-h^\star)^2$	-0.0
	Lin vel (z)	$r_{v_z} = v_z^2$	-1.0
	Ang vel (x–y)	$r_{\omega_{xy}} = \|\boldsymbol{\omega}_{xy}\|_2^2$	-0.05
	Orientation (tilt)	$r_{\text{tilt}} = (n_x^2+n_y^2)\cdot(1.25)^{\mathbf{1}[c_x<0]}\cdot(1-\mathbf{1}_{\text{climb/gap}})$	-12.0
	Feet contact forces	$r_{F\text{foot}} = \sum_{i=1}^{N_f}\max\big(0,\ \|\mathbf{F}_i\|-F_{\max}\big)$	-5.0×10^{-4}
	Feet stumble	$r_{\text{stum}} = \mathbf{1}\big[\exists i:\ \|\mathbf{F}_{i,xy}\|>5\,\lvert F_{i,z}\rvert\big]$	-0.4
	Feet air-time (reward)	$r_{\text{air}} = \sum_{i=1}^{N_f}\min(T_i,0.5)\cdot\mathbf{1}[\text{first\_contact}(i)]$	0.5

C. Reward Function Design

The design of the reward function is critical for shaping the desired locomotion behavior. Our goal is to train the hexapod robot to maintain a stable, alternating tripod gait while accurately tracking velocity commands across diverse and challenging terrains. Unlike policies for highly dynamic parkour [35], we prioritize safe industrial locomotion by incorporating safety-oriented penalties, for instance on joint torques, to limit aggressive, high-impact motions such as jumping. The reward function is a weighted sum of multiple terms categorized into tracking, motion efficiency, stability, and gait regulation. A comprehensive list of all reward components and their corresponding weights is provided in Table III.

IV. EXPERIMENTS

We conduct extensive experiments in both simulation and the real world to validate the effectiveness of SURF-Loco. Our evaluation is designed to answer three key questions: (1) How crucial is the Mixture-of-Experts (MoE) architecture for learning diverse locomotion skills? (2) How does SURF-Loco perform against state-of-the-art methods on challenging industrial terrains that are intractable for 2.5D representations? (3) Can the learned policy achieve robust zero-shot transfer to a physical hexapod robot?

A. Experimental Setup

Robot Platforms: We validate our approach on the Lenovo Daystar IS hexapod robot, shown in Fig. 3. The robot was equipped with a LiDAR for exteroceptive sensing, an IMU for state estimation, and joint encoders for proprioceptive feedback.

Simulation Environment and Terrains: All policies are trained in Isaac Gym. The simulation runs with a physics

Fig. 3. Lenovo Daystar IS hexapod robot used for real-world validation.

step of 500 Hz and a policy control frequency of 50 Hz. To create a diverse curriculum of industrial-style terrains, we build upon the widely used legged_gym framework [36]. However, its standard procedural terrain generation, which creates a uniform grid of vertices from a horizontal resolution (10cm by default), presents a significant challenge for our application. Simulating the intricate geometries found in industrial environments requires a fine resolution (e.g. 2 cm), which leads to an explosion in the number of triangles. This results in prohibitive VRAM consumption and drastically slows down simulation, rendering large-scale parallel training impractical.

TABLE IV

TERRAIN DESIGN AND DIFFICULTY LEVELS

Terrain Type	Specification
Flat ground	1 set, smooth flat surface (baseline)
Cobblestone	2 sets, surface roughness amplitude: 0.05 – 0.10 m
Slope	3 sets, slope incline: 0.10 – 0.46 rad
Gap	4 sets, gap width: 0.15 – 0.30 m
Climb	5 sets, climb height: 0.20 – 0.60 m
Stairs	6 sets, step height: 0.05 – 0.20 m

To overcome this bottleneck, we developed an automated mesh simplification algorithm. It selectively reduces mesh complexity to 12% of the original face count by compacting locally flat regions while preserving high-detail geometry using the mesh triangle normals. This method enables efficient, large-scale training on a rich set of terrains without sacrificing the fidelity of essential features. Our final curriculum includes flat ground, cobblestone, slopes, gaps, steps, and stairs, detailed in Table IV.

Baselines for Comparison: We compare SURF-Loco against three recent SOTA baseline methods:

- **HIM [10]**: A proprioceptive-only policy based on the Hierarchical Imagination Model (HIM) architecture. It represents a strong blind locomotion baseline.
- **PIM [6]**: A Perception-Informed Model that augments the Blind-HIM policy with a 2.5D height scan (elevation map) as input. As the official source code was unavailable, we performed our own implementation based on the public HIM codebase and integrated the 2.5D height scan input module.
- **WMP [15]**: A vision-based World Model Policy that uses a forward-facing depth camera stream. This approach learns to predict future states from image sequences but is constrained by a limited field of view. We adapted the official implementation to our hexapod robot model.

B. Ablation Study on MoE Architecture

We validated the MoE architecture, which used K=3 experts, by comparing it against a standard Multi-Layer Perceptron (MLP) baseline. Both were trained with consistent reward and terrain designs. As shown in Figure 4, the MoE policy demonstrated clear advantages over the MLP baseline, achieving higher task success rates and faster convergence across terrain difficulty levels. Moreover, it maintained higher average forward velocity at the target speed and yielded greater cumulative rewards. These results highlight the effectiveness of the MoE framework in accelerating learning and enhancing locomotion robustness.

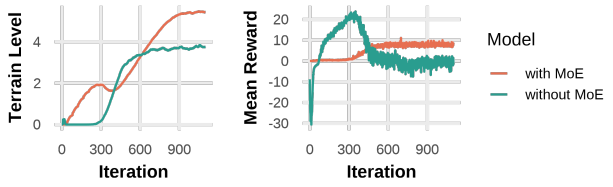

Fig. 4. Ablation of the MoE versus a standard MLP on the mixed-terrain curriculum. **Left:** terrain level (higher is harder) versus iteration. **Right:** mean episode reward versus iteration. Under identical reward and terrain settings, the MoE policy progresses to harder terrains earlier and attains higher rewards, indicating faster convergence and improved locomotion robustness.

C. Analysis of Expert Specialization

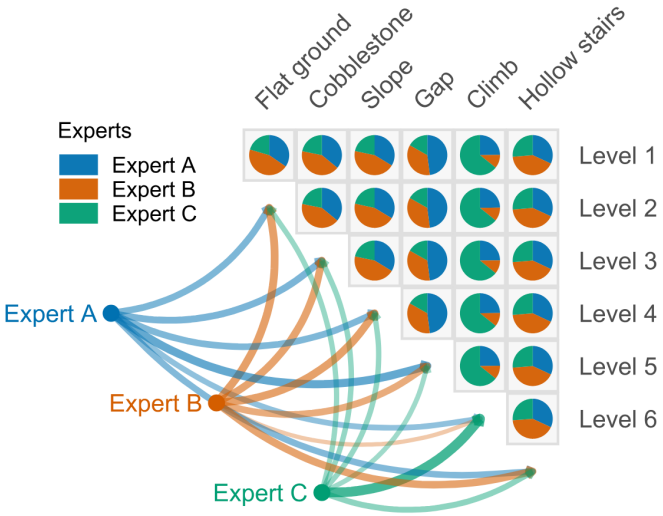

Fig. 5. Expert contribution patterns across terrains and difficulty levels. Each pie chart represents the relative probability distribution of three experts (A, B, and C) in solving locomotion tasks under different terrain types (columns) and difficulty levels (rows). The lower-left panel illustrates the global contribution tendencies of each expert, with arcs linking experts to their associated terrains and levels.

Figure 5 illustrates the distribution of expert contributions across terrain types (columns) and difficulty levels (rows), highlighting how expertise is shared, specialized, or complementary depending on environmental conditions. Each cell contains a pie chart representing the probability distribution of the three experts under the corresponding condition. While flat ground and slope exhibit relatively balanced contributions, gap and stair terrains rely more heavily on

TABLE V

PERFORMANCE COMPARISON ON CHALLENGING TERRAINS. 'MAX' INDICATES THE OBSTACLE WITH $\geq 30\%$ SUCCESS RATE. N = 10 TRIALS.

Method	Perception Modality	Max Width of Gap (m)	0.4m Gap Success Rate	Max Stair Height (m)				0.2m Hollow Stairs Success Rate	Max Height of Climb (m)	0.6m Climb Success Rate
				↑	↓	←	→			
HIM	Proprioception	0.4	30%	0.23	0.23	0.23	0.23	40%	0.4	-
WMP	Depth Image	0.4	**90%**	0.23	-	-	-	10%	0.8	80%
PIM	Height Scan	0.4	60%	0.23	0.23	0.23	0.23	50%	0.8	70%
SURF-Loco (Ours)	3D Surfel	**0.6**	**90%**	0.23	0.23	0.23	0.23	**90%**	**0.82**	**90%**

specific experts. The arcs in the lower-left panel summarize global tendencies by linking each expert to the terrains and levels where they dominate for intuitive comparison. Overall, the figure highlights both local specialization and broader contribution patterns, offering insight into how the model decomposes complex locomotion tasks into expert-specific competencies. Notably, Expert B is generally engaged across most terrains, particularly in standard locomotion tasks, whereas Expert C contributes more prominently to climbing terrains and Expert A specializes in gap terrains. Unlike the approach in [12], our method emphasizes skill-oriented learning, such as expert coordination during suspension phases in climbing scenarios. In the hollow-stair terrain as well as when traversing flat ground, cobblestone, and slope terrains, Expert B exhibited superior proficiency in locomotion.

D. Comparative Evaluation on Challenging Terrains

Our comparative evaluation, summarized in Table V, reveals the crucial role of the underlying environmental representation in determining locomotion capability. Since all policies were trained on a similar curriculum that included normal stairs, it was expected that their performance on such terrain would be comparable. As expected, the proprioceptive-only HIM policy showed limitation on complex obstacles, confirming the general need for perception. The perception-enabled baselines performed better but revealed challenges specific to their sensory inputs. The vision-based WMP, for example, was effective on forward-facing tasks but struggled with omnidirectional movements and geometrically ambiguous terrains like hollow stairs, likely due to its constrained field of view and resolution. Similarly, the PIM policy, relying on a 2.5D height scan, demonstrated competent locomotion on many terrains but faced inherent difficulties with hollow stairs, as its representation leads to reduced capability.

In comparison, SURF-Loco demonstrated a consistent and high level of performance, particularly on the terrains that proved most challenging for the baselines. Its success on the hollow stairs suggests that the 3D surfel representation provides a more complete geometric understanding, allowing the policy to better identify valid footholds where 2.5D maps fall short. Moreover, the approach also showed enhanced capability in traversing high obstacles, indicating that its richer environmental model translates to more effective and robust locomotion strategies overall.

E. Real-World Deployment and Sim-to-Real Transfer

We validated our approach through both sim-to-sim and zero-shot sim-to-real transfers. The policy, trained in Isaac Gym, generalized effectively when deployed without modification in Isaac Lab and Webots, demonstrating robustness to different physics engines.

For hardware deployment, the policy was transferred zero-shot to the physical Lenovo Daystar IS hexapod robot. The control policy runs on an onboard Rockchip RK3588, while a dedicated Mini PC with an Intel i7-12700H CPU handles localization algorithm. This system uses *SuperOdom* [37] to fuse data from a Livox Mid-360 LiDAR and an IMU, localizing the robot against a handheld-scanned prebuilt map. To bridge the reality gap, our training regimen incorporates extensive domain randomization to address critical physical mismatches, including terrain friction, robot base and link masses, center of mass positions, and motor dynamics. We also model sensor latencies (100–300 ms) and inject noise into the raycasting measurements to mimic real-world sensor imperfections and localization drift.

We conducted experiments on complex terrains, including a 70cm high platform and hollow stairs. This successful zero-shot performance validates our surfel-based representation as a powerful sim-to-real interface, enabling the policy to ground its actions in stable 3D geometry and overcome the challenges of real-world dynamics and sensor noise.

V. CONCLUSIONS

We presented SURF-Loco, a framework that learns robust locomotion by directly leveraging 3D surfel representations. By integrating this rich geometric input with a Mixture-of-Experts policy, our approach enables specialized skills for navigating complex industrial terrains. Extensive experiments on a hexapod, including successful zero-shot sim-to-real transfer, demonstrate the effectiveness of our method over conventional 2.5D representations, especially on geometrically challenging obstacles like hollow stairs.

Despite these promising results, our approach has several limitations. Currently, it relies on a pre-scanned static map, making it vulnerable to incomplete or noisy map data. Furthermore, this restricts its use in dynamic settings where obstacles may shift or surfaces deform. Its real-world performance heavily depends on a reliable external localization system. Additionally, the human-engineered reward function makes training sensitive to hyperparameters. Its static components do not adapt to specific contexts or experts. Exploring simpler or adaptive reward mechanisms remains a key area for future improvement.

Looking ahead, we plan to extend the framework to handle dynamic environments by integrating online perception. Concurrently, we will explore methods to simplify the reward design while preserving performance. Another promising

direction is the fusion of semantic information with the geometric surfel representation, which could unlock more intelligent behaviors. Furthermore, extending SURF-Loco to platforms with lower static stability, such as quadrupeds and bipeds, is a promising direction to validate the 3D representation's generality. Finally, we intend to utilize the surfel-based model as a common geometric foundation for both locomotion and navigation planning, enabling the robot to tackle more complex, long-horizon tasks.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No.62576109, 62072112). We also thank Lenovo (Robotics Lab, CTO Organization) for providing the robot platform and engineering support.

REFERENCES

[1] M. Hutter, C. Gehring, D. Jud, A. Lauber, C. D. Bellicoso, V. Tsounis, J. Hwangbo, K. Bodie, P. Fankhauser, M. Bloesch *et al.*, "Anymal-a highly mobile and dynamic quadrupedal robot," in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016, pp. 38–44.

[2] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science robotics*, vol. 5, no. 47, p. eabc5986, 2020.

[3] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, "Legged locomotion in challenging terrains using egocentric vision," in *Conference on robot learning*. PMLR, 2023, pp. 403–415.

[4] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, "Extreme parkour with legged robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 11 443–11 450.

[5] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, and M. Hutter, "Elevation mapping for locomotion and navigation using gpu," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2273–2280.

[6] J. Long, J. Ren, M. Shi, Z. Wang, T. Huang, P. Luo, and J. Pang, "Learning humanoid locomotion with perceptive internal model," *arXiv preprint arXiv:2411.14386*, 2024.

[7] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on Robot Learning*, 2022, pp. 91–100.

[8] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," in *Robotics: Science and Systems*, 2021.

[9] T. Miki, J. Lee, L. Wellhausen, and M. Hutter, "Learning to walk in confined spaces using 3d representation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 8649–8656.

[10] J. Long, Z. Wang, Q. Li, L. Cao, J. Gao, and J. Pang, "Hybrid internal model: Learning agile legged locomotion with simulated robot response," in *The Twelfth International Conference on Learning Representations*, 2024.

[11] I. Nahrendra, B. Yu, and H. Myung, "Dreamwaq: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning," *arXiv preprint arXiv:2301.10602*, 2023.

[12] R. Huang, S. Zhu, Y. Du, and H. Zhao, "Moe-loco: Mixture of experts for multitask locomotion," *arXiv preprint arXiv:2503.08564*, 2025.

[13] T. Qian, H. Zhang, Z. Zhou, H. Wang, M. Cai, and Z. Kan, "Leeps: Learning end-to-end legged perceptive parkour skills on challenging terrains," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 12 904–12 909.

[14] E. Chane-Sane, J. Amigo, T. Flayols, L. Righetti, and N. Mansard, "Soloparkour: Constrained reinforcement learning for visual locomotion from privileged experience," *arXiv preprint arXiv:2409.13678*, 2024.

[15] H. Lai, J. Cao, J. Xu, H. Wu, Y. Lin, T. Kong, Y. Yu, and W. Zhang, "World model-based perception for visual legged locomotion," *arXiv preprint arXiv:2409.16784*, 2024.

[16] S. Li, S. Luo, J. Wu, and Q. Zhu, "Move: Multi-skill omnidirectional legged locomotion with limited view in 3d environments," *arXiv preprint arXiv:2412.03353*, 2024.

[17] R. Yu, Q. Wang, Y. Wang, Z. Wang, J. Wu, and Q. Zhu, "Walking with terrain reconstruction: Learning to traverse risky sparse footholds," *arXiv preprint arXiv:2409.15692*, 2024.

[18] S. Luo, S. Li, R. Yu, Z. Wang, J. Wu, and Q. Zhu, "Pie: Parkour with implicit-explicit learning framework for legged robots," *IEEE Robotics and Automation Letters*, 2024.

[19] Y. Zhao, T. Wu, Y. Zhu, X. Lu, J. Wang, H. Bou-Ammar, X. Zhang, and P. Du, "Zsl-rppo: Zero-shot learning for quadrupedal locomotion in challenging terrains using recurrent proximal policy optimization," *arXiv preprint arXiv:2403.01928*, 2024.

[20] H. Wang, Z. Wang, J. Ren, Q. Ben, T. Huang, W. Zhang, and J. Pang, "Beamdojo: Learning agile humanoid locomotion on sparse footholds," *arXiv preprint arXiv:2502.10363*, 2025.

[21] Y. Chen, J. Ma, Z. Luo, Y. Han, Y. Dong, B. Xu, and P. Lu, "Learning autonomous and safe quadruped traversal of complex terrains using multi-layer elevation maps," *IEEE Robotics and Automation Letters*, vol. 10, no. 10, pp. 9606–9613, 2025.

[22] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[23] Q. Zhang, Z. Zhang, W. Cui, J. Sun, J. Cao, Y. Guo, G. Han, W. Zhao, J. Wang, C. Sun *et al.*, "Humanoidpano: Hybrid spherical panoramic-lidar cross-modal perception for humanoid robots," *arXiv preprint arXiv:2503.09010*, 2025.

[24] I. Nahrendra, B. Yu, M. Oh, D. Lee, S. Lee, H. Lee, H. Lim, and H. Myung, "Obstacle-aware quadrupedal locomotion with resilient multi-modal reinforcement learning," *arXiv preprint arXiv:2409.19709*, 2024.

[25] W. Sun, B. Cao, L. Chen, Y. Su, Y. Liu, Z. Xie, and H. Liu, "Learning perceptive humanoid locomotion over challenging terrain," *arXiv preprint arXiv:2503.00692*, 2025.

[26] Z. Wang, Y. Li, L. Xu, H. Shi, Z. Ma, Z. Chu, C. Li, F. Gao, K. Yang, and K. Wang, "Sf-tim: A simple framework for enhancing quadrupedal robot jumping agility by combining terrain imagination and measurement," *arXiv preprint arXiv:2408.00486*, 2024.

[27] T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison, "Elasticfusion: Dense slam without a pose graph." in *Robotics: science and systems*, vol. 11, no. 3. Rome, 2015.

[28] X. Chen, A. Milioto, E. Palazzolo, P. Giguere, J. Behley, and C. Stachniss, "Suma++: Efficient lidar-based semantic slam," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4530–4537.

[29] T.-M. Nguyen, D. Duberg, P. Jensfelt, S. Yuan, and L. Xie, "Slict: Multi-input multi-scale surfel-based lidar-inertial continuous-time odometry and mapping," *IEEE Robotics and Automation Letters*, vol. 8, no. 4, pp. 2102–2109, 2023.

[30] M. Oh, B. Yu, I. Nahrendra, S. Jang, H. Lee, D. Lee, S. Lee, Y. Kim, M. K. Christiansen, H. Lim *et al.*, "Trip: Terrain traversability mapping with risk-aware prediction for enhanced online quadrupedal robot navigation," *arXiv preprint arXiv:2411.17134*, 2024.

[31] M. Macklin, "Warp: A high-performance python framework for gpu simulation and graphics," https://github.com/nvidia/warp, March 2022, nVIDIA GPU Technology Conference (GTC).

[32] Z. Wang, T. Ma, Y. Jia, X. Yang, J. Zhou, W. Ouyang, Q. Zhang, and J. Liang, "Omni-perception: Omnidirectional collision avoidance for legged locomotion in dynamic environments," *arXiv preprint arXiv:2505.19214*, 2025.

[33] B. Ma, N. Xu, C. Qi, X. Liu, Y. Mo, J. Wang, and C. Lu, "Ppl: Point cloud supervised proprioceptive locomotion reinforcement learning for legged robots in crawl spaces," *arXiv preprint arXiv:2508.09950*, 2025.

[34] D. Wang, X. Wang, X. Liu, J. Shi, Y. Zhao, C. Bai, and X. Li, "More: Mixture of residual experts for humanoid lifelike gaits learning on complex terrains," *arXiv preprint arXiv:2506.08840*, 2025.

[35] Z. Zhuang, Z. Fu, J. Wang, C. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, "Robot parkour learning," in *Conference on Robot Learning (CoRL)*, 2023.

[36] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on robot learning*. PMLR, 2022, pp. 91–100.

[37] S. Zhao, H. Zhu, Y. Gao, B. Kim, Y. Qiu, A. M. Johnson, and S. Scherer, "Superloc: The key to robust lidar-inertial localization lies in predicting alignment risks," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025.