CMoE: Contrastive Mixture of Experts for Motion Control and Terrain Adaptation of Humanoid Robots

Shihao Ma[1,*], Hongjin Chen[1,*], Zijun Xu[1,*], Yi Zhao[1], Ke Wu[1], Ruichen Yang[1], Leyao Zou[1], Zhongxue Gan[1,†], and Wenchao Ding[1,†]

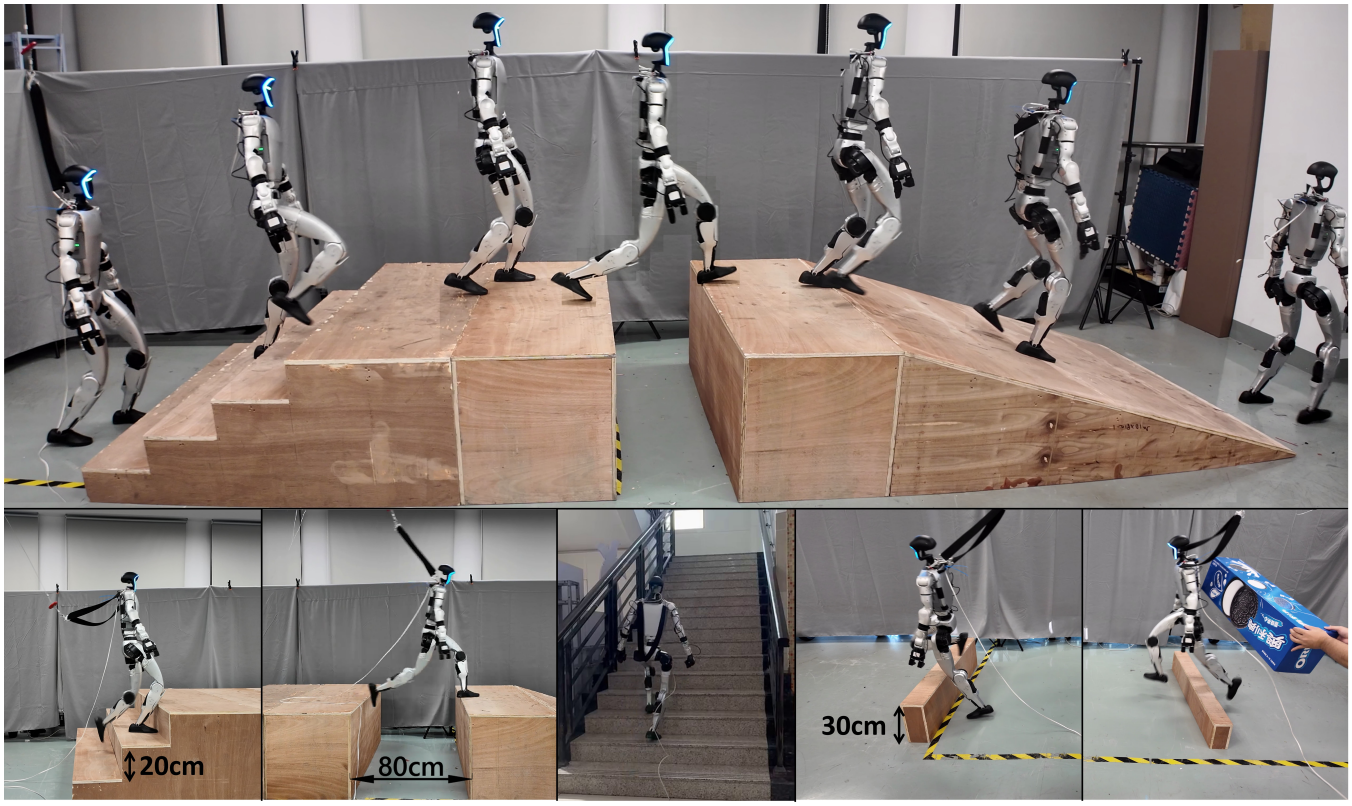

Fig. 1: We propose a mixture-of-experts model-based architecture that enables humanoid robots to simultaneously navigate a variety of challenging terrains. We validate our strategy on complex mixed terrains and non-training environments.

***Abstract*— For effective deployment in real-world environments, humanoid robots must autonomously navigate a diverse range of complex terrains with abrupt transitions. While the Vanilla mixture of experts (MoE) framework is theoretically capable of modeling diverse terrain features, in practice, the gating network exhibits nearly uniform expert activations across different terrains, weakening the expert specialization and limiting the model's expressive power. To address this limitation, we introduce CMoE, a novel single-stage reinforcement learning framework that integrates contrastive learning to refine expert activation distributions. By imposing contrastive constraints, CMoE maximizes the consistency of expert activations within the same terrain while minimizing their similarity across different terrains, thereby encouraging experts to specialize in distinct terrain types. We validated our approach on the Unitree G1 humanoid robot through a series of challenging experiments. Results demonstrate that CMoE enables the robot to traverse continuous steps up to 20 cm high and gaps up to 80 cm wide, while achieving robust and natural gait across diverse mixed terrains, surpassing the limits of existing methods. To support further research and foster community development, we release our code publicly.**

[1]College of Intelligent Robotics and Advanced Manufacturing, Fudan University, Shanghai, China, 200433

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62403142, in part by the Science and Technology Commission of Shanghai Municipality under Grant 24511103100, and in part by the Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX0103).

*Equal contribution. †Corresponding authors.

Project Page: `https://hoshi-no-ai.github.io/CMoE`

I. Introduction

Humanoid robots are expected to operate in abrupt transition real-world environments, where they frequently encounter dynamically changing terrains, such as gravel paths that transition into slopes or stepping stones interspersed with deformable surfaces[1][2]. Walking in such environments requires not only the ability to maintain stability on a single type of terrain but also the agility to rapidly adapt to heterogeneous terrains while preserving continuous motion.

To enable robots to adapt to various terrains, previous

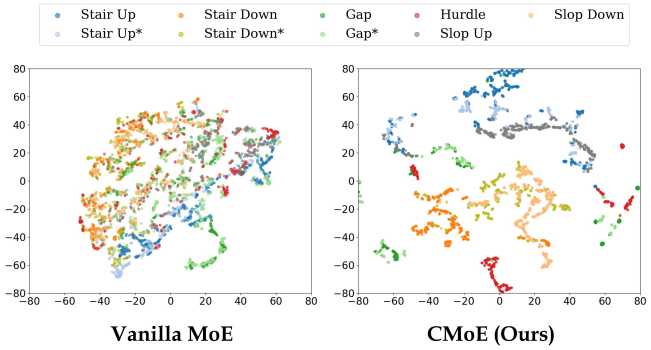

Fig. 2: t-SNE visualization of the experts activation of **Vanilla MoE** and **our method** across different terrains. ("*" indicates a simple version of this terrain)

works[3][4] have proposed a two-phase training approach, where the model is first trained on a single terrain and then distilled in a second phase. This method allows the model to focus on the details of each terrain while avoiding catastrophic forgetting that may occur when training on multiple terrains simultaneously. However, although this two-phase training approach enhances multi-terrain learning, it increases the training time and introduces the risk of overfitting.

To address these issues, some studies[5][6] have employed the MoE (Mixture of Experts) method, which can simultaneously consider features from multiple terrains. By dynamically activating experts, the MoE network adapts to various terrains and demonstrates good generalization capability. However, we found that it is not effective at activating skills based on environmental features. As shown in Fig. 2, the network using the Vanilla MoE strategy exhibits dispersed expert activation across both similar and different terrains, without forming any distinct clusters. This lack of clear clustering prevents the network from effectively adapting to specific terrain features, leading to less efficient skill activation and terrain handling.

To provide terrain adaptability for the robot, enabling it to effectively activate skills based on environmental features, we propose CMoE, a novel single-stage reinforcement learning framework that integrates the Mixture of Experts (MoE) architecture with contrastive learning. Unlike Vanilla MoE, CMoE maximizes the diversity of expert activation distributions within different terrains. Our method effectively activates skills based on environmental features and allocates experts with terrain-specific expertise when transitioning between terrains.

To achieve this, our framework first encodes the perceptual data using an autoencoder (AE), capturing the latent representation of the state of the environment. Next, contrastive learning is applied to improve the terrain perception gating network. This method maximizes the consistency of expert activation distributions within the same terrain while minimizing the similarity of expert activation distributions across different terrains. By minimizing the contrastive loss between expert activation levels and terrain information, the gating network is guided to output activations that reflect more terrain-specific differences, thereby encouraging the expert networks to specialize in certain types of terrain. Finally, a perceptual gating network dynamically adjusts the activation of experts based on environmental perception. The outputs of each expert are fused according to their activation levels, resulting in an action distribution.

To validate the effectiveness of CMoE, we conducted simulation and field experiments on a Unitree robot. Results show that even when trained simultaneously on eight different environments, CMoE achieves a higher learning ceiling on each terrain. Our robot can use a single policy to traverse obstacles up to 30 cm high, continuous steps up to 20 cm high, and gaps up to 80 cm wide, as shown in the Fig. 1. Surpassing the most challenging terrains studied in existing research. Furthermore, CMoE effectively activates specific skills based on environmental characteristics, more efficiently allocating expert activations during terrain transitions. This enables CMoE to excel on mixed terrains.

Summary of our contributions:

- We propose a single-stage, end-to-end framework that directly maps multimodal sensory inputs to robot actions, integrating estimators to process both historical information and terrain-specific data.
- We introduce a novel MoE actor-critic model, augmented with a contrastive learning objective to enhance the robots ability to adapt its terrain response strategy across complex, heterogeneous surfaces.
- Extensive experiments conducted on the Unitree-G1 robot demonstrate our approach's state-of-the-art performance, and we release our code to further contribute to the research community.

II. Related Work

A. Learning-based Humanoid Locomotion

With the development of legged robots, especially quadruped robots, motion control methods have matured[7][8][9]. Humanoid robots, in particular, have attracted considerable attention due to their ability to better adapt to human-designed environments[10][11]. However, walking on two legs requires precise center of gravity control and coordination, especially on complex terrain. To better enable robots to perceive their environment and their own state, some work[12] uses AE or VAE estimators to encode historical proprioceptive information and employs contrastive loss for state prediction, providing more information. However, due to the lack of environmental sensors, these methods have yet to fully realize the potential of robots.

In response to these limitations, models incorporating external perception have shown promise in enhancing mobility performance. Research has begun using depth cameras or lidar to provide terrain perception information for reinforcement learning models. However, depth camera-based perception strategies are limited by the camera's narrow field of view, resulting in robots being limited to forward

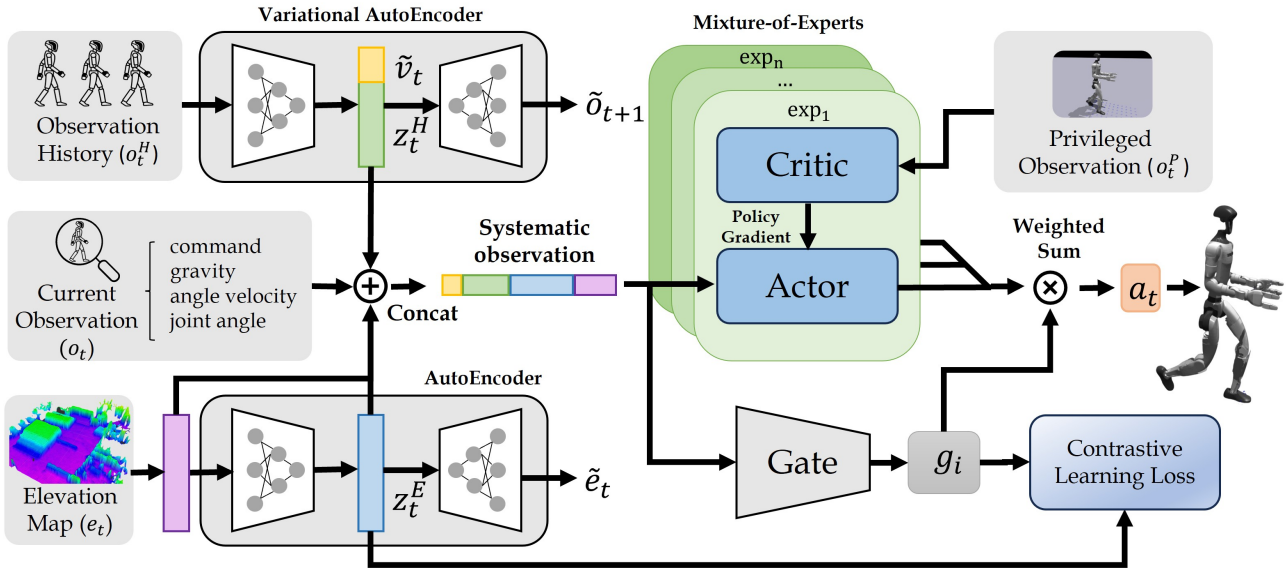

Fig. 3: Overview of our framework. We encode historical information and the elevation map into explicit and implicit representations, which are fused with the current observation to form the system observation. These are then fed into the MoE structure, consisting of multiple experts and a gating network, which outputs expert activations and performs contrastive learning with the encoded environmental data.

movement[13][14][15][16]. LiDAR is widely used due to its wide field of view and higher accuracy. Compared to blind strategies, robots using elevation maps as external perception can better acquire environmental information and make more adaptive actions[17][18].

B. Mixture of Experts in Locomotion

Since the introduction of MoE[19][20], it has been regarded as an effective mechanism for addressing gradient conflicts and task interference in multi-task reinforcement learning[21] due to its modularity and interpretability[22][23]. It has been widely applied in the fields of autonomous driving[24], natural language processing[25], and computer vision[26].

In recent years, MoE strategies have also been applied in the robotics field, with the introduction of a mixed expert model for quadruped reinforcement learning[27]. The MoE-Loco method[5] uses MoE to alleviate gradient conflicts in quadruped and bipedal locomotion tasks, but it lacks an environmental perception system to regulate the MoE. MoRE[6] focuses on training a variety of humanoid gaits through MoE, improving the gait learning ability. However, the gait selection is essentially a behavior under human intervention, which does not reflect the robot's autonomy. Moreover, since the expert activation is almost unrelated to the terrain, this does not enhance the robot's terrain traversal capability. In addition, the current MoE still suffers from the lazy gating problem[28][29][30].

III. Method

A. System Overview

As shown in Fig. 3, CMoE aims to improve the multi-terrain adaptability of humanoid robots by combining contrastive learning and MoE. Specifically, we introduce the CMoE framework, which synergistically integrates three key components: a VAE to infer the robot's true state from incomplete sensory data, an unsupervised contrastive learning mechanism that learns a latent representation to effectively distinguish between terrains, and a MoE architecture that decomposes the complex control task for specialized action generation. By unifying state estimation, terrain perception, and action selection, our framework enables the development of a highly generalized policy, thereby markedly improving the robot's capabilities in multi-terrain locomotion.

B. Problem Description

We define multi-terrain locomotion as a Markov Decision Process (MDP) with a tuple $\langle S, A, T, R, \gamma \rangle$, where S is the state space, A is the action space, and $s_t \subseteq S$ represents the robots state. The system's dynamics are governed by the transition probability T, and the agent receives a reward from $R(s, a)$. The discount factor $\gamma \in [0, 1]$ balances immediate and future rewards. We use Proximal Policy Optimization (PPO) to learn the optimal policy π^* that maximizes the expected discounted return:

$$\pi^* = \arg\max \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right]. \tag{1}$$

C. Information Encoding

The proprioception observation o_t from sensors is:

$$\mathbf{o_t} = [\omega_t, g_t, c_v^t, \theta_t, \dot{\theta}_t, a_{t-1}], \tag{2}$$

including robot angular velocity ω_t, gravity direction g_t, velocity command c_v^t, joint angle θ_t, velocity $\dot{\theta}_t$, and the last action a_{t-1}.

We propose a context-state distillation model to capture the robots proprioception. The model contains two separate estimators: the first is a VAE-based estimator that predicts

the robots body state, and the second is used for extracting features from the elevation map.

The encoder maps the input observation ${\mathbf{o_t}}^H$ to robot body velocity $\mathbf{v}t$ and latent representation z_t^H, which is then decoded to $\tilde{\mathbf{o}}t+1$. Specifically, we employ a β-variational autoencoder (β-VAE) for the autoencoder setup. According to [31], estimating body velocity is crucial for terrain traversal. The context-state distillation model involves a hybrid loss function:

$$\mathcal{L}_{\text{CS}} = \text{MSE}(\tilde{\mathbf{v}}_t, \mathbf{v}_t) + \mathcal{L}_{\text{VAE}}, \tag{3}$$

where $\text{MSE}(\tilde{\mathbf{v}}_t, \mathbf{v}_t)$ and $\mathcal{L}_{\text{VAE}}$ represent body velocity loss and VAE reconstruction loss, respectively. $\mathbf{v}_t$ is the ground truth given by simulator. The VAE reconstruction loss, $\mathcal{L}_{\text{VAE}}$ is formulated as:

$$\mathcal{L}_{\text{VAE}} = \text{MSE}(\tilde{\mathbf{o}}_{t+1}, \mathbf{o}_{t+1}) + \beta D_{\text{KL}}(q(\mathbf{z}_t^H \mid \mathbf{o}_t^H) \parallel p(\mathbf{z}_t^H)), \tag{4}$$

here, $\tilde{\mathbf{o}}_{t+1}$ is the reconstructed observation at the next time step, $q(\mathbf{z}_t^H \mid \mathbf{o}_t^H)$ is the posterior distribution of $\mathbf{z}_t^H$ given $\mathbf{o}_t^H$, and $p(\mathbf{z}_t^H)$ is the prior, assumed to be a standard Gaussian distribution.

The other estimator uses an autoencoder to extract features from the elevation map [32] [33] for self-prediction, with the loss function given by:

$$\mathcal{L}_{\text{AE}} = \text{MSE}(\tilde{\mathbf{e}}_t, \mathbf{e}_t), \tag{5}$$

where $\mathbf{e}_t$ is the ground truth elevation map from the simulator.

D. Mixture of Experts Policy

Gradient conflicts typically arise in multitask reinforcement learning [34]. To tackle these challenges, we incorporate a MoE architecture into both the actor and critic networks within the PPO framework.

Specifically, each expert module comprises a dedicated actor-critic pair, where each critic evaluates only its corresponding actor using privileged observation. Every expert receives the estimated body velocity $\mathbf{v}_t^p$, along with implicit contextual state variables $\mathbf{z}_t^E$ and $\mathbf{z}_t^H$, current observation $\mathbf{o}_t^c$ and the elevation map $\mathbf{e}_t$, and produces either an action or a value estimate. To ensure consistency between policy evaluation and action generation, the same gating network is shared across both the actor and critic MoE components. The final output is obtained as a weighted sum of the expert responses after softmax normalization:

$$\mu_{\text{weighted}} = \sum_{i=1}^{N} \text{softmax}(g_i) \cdot \mu_i, \tag{6}$$

where μ_i means the i-th expert outputs and g_i is the expert activation.

E. Terrains Contrastive Learning

We introduce a novel method, terrains contrastive learning, in humanoid locomotion. Through this approach, terrains are encoded into latent features and enhance the relation with MoE gate network. Specially, we first adopt two MLP to transform the gate output and elevation map into the same dimension g_t^z and e_t^z, respectively. Then, in the contrastive learning process, if a pair of $\langle g^z, e^z \rangle$ belong to the same trajectory, they are considered the positive samples. Otherwise, they are negative. The optimization is inspired by SwAV[35]. To predict the cluster assignment probability $\mathbf{p}_t^{\text{g}}$ and $\mathbf{p}_t^{\text{e}}$ from g_t^z and e_t^z. we apply a $\mathcal{L}_2$-normalization on the prototype to obtain normalized matrix $\mathbf{E} = \{\bar{\mathbf{e}}_1, ..., \bar{\mathbf{e}}_K\}$, and then take a softmax over the dot products of source vectors or target vectors with all the prototypes:

$$\mathbf{p}_t^{\text{g}} = \frac{\exp(\frac{1}{\tau} {g_t^z}^\top \mathbf{e}_k)}{\sum_{k'} \exp(\frac{1}{\tau} {g_t^z}^\top \mathbf{e}_{k'})}, \quad \mathbf{p}_t^{\text{e}} = \frac{\exp(\frac{1}{\tau} {e_t^z}^\top \mathbf{e}_k)}{\sum_{k'} \exp(\frac{1}{\tau} {e_t^z}^\top \mathbf{e}_{k'})}, \tag{7}$$

where $\mathbf{p}_t^{\text{g}}$ and $\mathbf{p}_t^{\text{e}}$ are the predicted probability that terrain map to individual cluster with index k, while τ is a temperature parameter.

To obtain $(\mathbf{q}_1^{\text{g}}, ..., \mathbf{q}_K^{\text{g}})$ and $(\mathbf{q}_1^{\text{e}}, ..., \mathbf{q}_K^{\text{e}})$ for the aforementioned predicted probabilities, while avoiding trivial solutions, the Sinkhorn-Knopp algorithm[36] is applied to both

TABLE I: Terrain description

Terrain	Description	Range
Slope	Slope angle	0-20°
Stairs	Step height	0.05-0.23m
Gap	Ditch width	0.1-0.8m
Hurdle	Hurdle height	0.2-0.4m
	Hurdle width	0.1-0.3m
Discrete	Irregular protrusion height	0.1-0.2m
Mix1	Mixed terrain of gaps and steps,gaps width	0.1-0.8m
	Mixed terrain of gaps and steps,steps height	0.1-0.15m
Mix2	Mixed terrain of single-log bridge and steps, bridge width	0.5-1.0m
	Mixed terrain of single-log bridge and steps, step height on the bridge	0.1-0.25m

TABLE II: Reward Functions

Term	Equation	Weight
velocity tracking	$\exp\left\{-\lVert\mathbf{v}_{xy} - \mathbf{v}_{xy}^c\rVert_2^2/\sigma\right\}$	2.0
yaw tracking	$R_{\text{yaw}} = \exp\left(-\lvert\psi_{\text{cmd}} - \psi\rvert\right)$	2.0
z velocity	$\mathbf{v}_z^2$	-1.0
roll-pitch velocity	$\lVert\boldsymbol{\omega}_{xy}\rVert_2^2$	-0.05
orientation	$\lVert\mathbf{g}_x\rVert_2^2 + \lVert\mathbf{g}_y\rVert_2^2$	-2.0
base height	$(h - h^{\text{target}})^2$	-15.0
feet stumble	$\bigvee_{i\in\text{feet}}\{\lVert\mathbf{F}_{i,xy}\rVert_2 > 3\cdot\lvert F_{i,z}\rvert\}$	-1.0
collision	$\sum_{i\in\mathcal{I}_{\text{penalty}}}\mathbb{I}(\lVert\mathbf{F}_i\rVert_2 > 0.1)$	-15.0
feet distance	$\min(\lvert p_{y,0} - p_{y,1}\rvert - d_{min}, d_{max} - d_{min})$	0.8
feet air time	$\sum_{i=1}^2\left(t_{\text{air},i} - t_{\text{air}}^{\text{target}}\right)\cdot\mathbb{F}_i$	1.0
feet ground parallel	$\sum_{i=1}^2\text{Var}(\mathbf{p}_{z,i})$	-0.02
hip dof error	$\sum_{i\in\text{hip joints}}\left\lvert\theta_i - \theta_i^{\text{default}}\right\rvert^2$	-0.5
dof acc	$\sum_{i\in\text{all joints}}\ddot{\theta}_i^2$	$-2.5e-7$
dof vel	$\sum_{i\in\text{all joints}}\dot{\theta}_i^2$	$-5.0e-4$
torques	$\sum_{i\in\text{all joints}}\tau_i^2$	$-1.0e-5$
action rate	$\lVert\mathbf{a}_t - \mathbf{a}_{t-1}\rVert_2^2$	-0.3
dof pos limits	$\text{ReLU}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\max}) + \text{ReLU}(\boldsymbol{\theta}_{\min} - \boldsymbol{\theta})$	-2.0
dof vel limits	$\text{ReLU}(\lvert\dot{\boldsymbol{\theta}}\rvert - \lvert\dot{\boldsymbol{\theta}}_{\max}\rvert)$	-1.0
torque limits	$\sum_{i\in\text{all joints}}\text{RELU}\left(\hat{\tau}_i - \hat{\tau}_i^{\max}\right)$	-1.0
feet edge	$\mathbf{1}_{\text{foot at edge of the terrain}}$	-1.0

TABLE III: Quantitative Comparison in Simulation. Metrics Include Success Rate and Average Travel Distance.

Method	Success Rate							
	slope	**stair up**	**stair down**	**discrete**	**gap**	**hurdle**	**mix1**	**mix2**
CMoE	**0.991**	**0.886**	0.905	0.991	**0.974**	**0.987**	**0.767**	**0.747**
Vanilla MoE	0.957	0.798	**0.908**	0.987	0.818	0.970	0.605	0.662
Base	0.966	0.481	0.483	**1.000**	0.221	0.779	0.276	0.388
	Average Travel Distance (m)							
CMoE	**14.870**	**10.802**	**10.824**	13.440	**14.876**	13.470	**12.055**	**9.750**
Vanilla MoE	11.675	8.898	9.250	14.870	11.980	**14.780**	9.960	8.703
Base	12.917	8.210	8.726	**15.280**	7.385	10.124	8.209	8.848

encoders. Now that we have the cluster assignment gate outputs and elevations, the contrastive learning objective is simply to maximize the match accuracy:

$$\mathcal{J}^{\text{SwAV}} = -\frac{1}{2H}\sum_{t=1}^{H}(\mathbf{q}_t^{\text{g}}\log\mathbf{p}_t^{\text{e}} + \mathbf{q}_t^{\text{e}}\log\mathbf{p}_t^{\text{g}}). \tag{8}$$

IV. Implementation Details

A. Training Parameters

We trained 4096 environments in parallel on IsaacGym using an NVIDIA RTX 4090 computer for 20,000 epochs. We selected 5 experts and used an elevation map covering a 0.7m x 1.1m rectangle around the robot. The parameters used for contrastive learning were as follows: num prototype = 32 and temperature = 0.2, which were chosen to optimize model performance. The terrains used for training in the simulation included eight types of terrain: stairs, gaps, and flat ground, as shown in the table. To reduce the learning difficulty, we employed a curriculum learning mechanism [37]. Furthermore, since all terrains were trained in the same phase, to balance the difficulty of each terrain, we divided them into simple and complex categories. We then performed velocity command curriculum learning on complex terrains, gradually increasing the magnitude and direction of the velocity commands.

B. Domain Randomization

To enhance the robot's real-world motion capabilities, we pre-randomized the following parameters during simulation: joint mass and moment of inertia, friction, restitution, motor strength, and motor k_p and k_d. We also applied a perturbation of up to 30 N to the robot every 16 seconds. For the elevation map, we introduced both delay noise and Gaussian noise, and pre-randomized the offset and rotation of the elevation map. Specifically, we designed nonlinear salt-and-pepper noise to address the unstable extremes that may occur in the elevation map. The update formula for the height points $h(i)$ is as follows.

$$h(i) = \begin{cases} \mathcal{U}(M, 2M-m) & \text{with probability } p, \\ \mathcal{U}(2m-M, m) & \text{with probability } p, \\ h(i) & \text{with probability } 1-2p, \end{cases} \tag{9}$$

where M and m are the maximum and minimum values at the corresponding height points, $\mathcal{U}(a, b)$ represents drawing from a uniform distribution with the range from a to b. In more complex terrain, the salt-and-pepper noise intensity is higher. Furthermore, real-world elevation data is often limited by sensor resolution, exhibiting smooth curved edges rather than sharp corner transitions. To address this issue, we used another domain randomization technique to chamfer the right-angled edges in the elevation map used in the simulator, thereby more realistically simulating the real world. Real world experiments have verified that this method is effective in addressing noise issues in elevation maps.

C. Reward Function

In addition to referencing existing work and making some minor adjustments, we redefined the foot distance function and implemented penalties for distances that were either too small. We also designed a discontinuous reward that is enabled only under specific terrain conditions. For example, the foot edge reward is activated only when the robot is trained in a hurdle or gap environment. This is because touching the edge of a step or irregular terrain should not be penalized. For a detailed description of the reward function, see TABLE II.

V. Experiments

A. Motion Control Performance

We compare the proposed framework CMoE with the following baselines: (1) Base strategy: The actor-critic network without the MoE structure, but with the same number of parameters as our strategy. (2) Vanilla MoE strategy: Uses the most basic Vanilla MoE structure, but do not include the terrain encoder and contrastive learning.

We established a benchmark for the robot's motion performance across different tasks. Our benchmark consists of a 3 m $\times$ 18 m runway with different terrains. The terrain details are shown in Table I. Experimental environment: Each robot was instructed to walk at a speed of 0.8 m/s on a continuous terrain with obstacles for 20 seconds. If the termination conditions are met, the trial is recorded as a failure, and the environment is reset. Termination conditions include: (1) collision with parts other than the feet, and (2) torso roll or pitch deviation exceeds 1 degree. If the termination conditions are not met and the maximum running time is reached, the trial is recorded as a success.

Record the following indicators: (1) Success Rate: the proportion of robots that complete the entire journey to

the total number of robots in the experiment. (2) Average Travel Distance: the movement distance in the direction of movement under the travel time limit.

The experimental results shown in Table III indicate that CMoE outperforms other methods in both success rate and average travel distance, demonstrating strong terrain adaptability. It excels in complex terrains, such as stairs and gaps, achieving higher success rates and longer travel distances. This highlights CMoE's advantage in overcoming gradient conflicts and optimizing motion control. By integrating MoE and contrastive learning, CMoE dynamically adjusts its expert network, enhancing performance in diverse environments, particularly in mixed terrains like mix1 and mix2.

In contrast, Vanilla MoE performs well in some terrains but underperforms in complex ones like gap and mix2, due to the lack of contrastive learning, limiting its adaptability and performance in challenging environments.

B. *Effect of Contrastive Learning*

To explore how MoE allocates experts according to terrain and to control for variables, we compared our method (CMoE) with the Vanilla MoE model that does not use contrastive learning, as shown in the t-SNE plot in Fig. 2.

The t-SNE plot of our method shows that similar terrains are clearly clustered based on their degree of similarity, while dissimilar terrains exhibit distinct boundaries. Similar terrains can be grouped and stratified according to their similarity. For example, ascending steps, simple steps, and uphill terrain are clustered together, but also stratified based on terrain difficulty. Furthermore, we observe that in the t-SNE plot of our method, ascending steps are significantly farther away from descending steps. This suggests that the model does not simply categorize steps as one terrain type, but instead differentiates between ascending steps and descending steps, which aligns with human cognition of walking.

In contrast, the expert weights in the Vanilla MoE model do not show significant changes with different terrains. This indicates that the Vanilla MoE is unable to understand the specialization differences between terrains, limiting the experts ability to specialize in different environments. In contrast, CMoE effectively allocates experts according to terrain, as demonstrated by the distinct clustering of expert weights in the t-SNE plot.

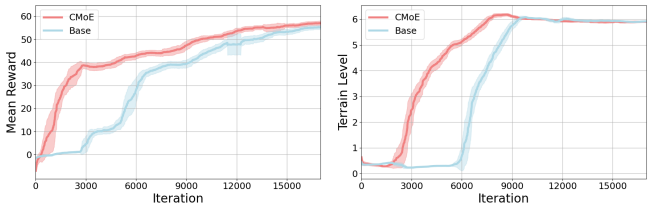

Fig. 4: The training curve during the multi-terrain training phase, depicting the reward curve and terrain level change with training iterations.

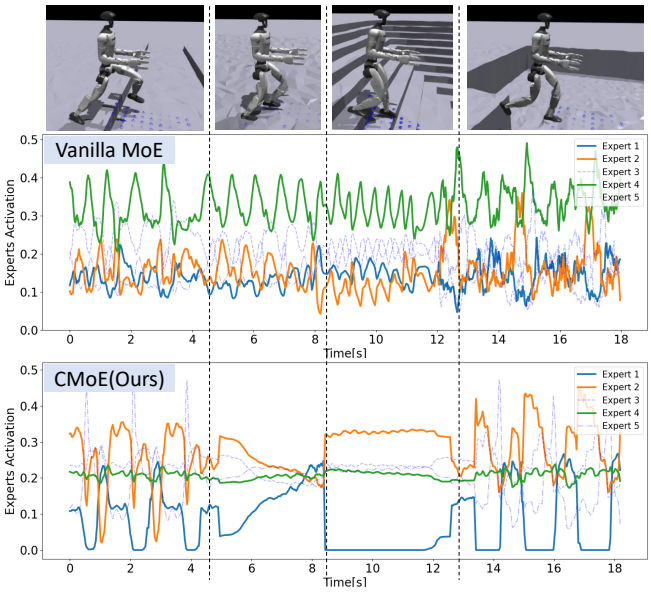

Fig. 5: The robot passes through four types of terrain: hurdles, uphill, downstairs, and gaps. The Vanilla MoE method and our method show the changes in the expert activation level over time.

C. *Expert Behavior Analysis*

To visually demonstrate the contribution of MoE to the expert activation levels across different terrains, as shown in Fig. 5, we designed four different terrains, including three consecutive 30cm hurdles, a 15-degree uphill slope, 10 steps downhill, and three 60cm gaps. We then had robots using the Vanilla MoE and CMoE methods walk across this mixed terrain, recording the changes in expert activation levels as the terrain varied.

It can be observed that in Vanilla MoE, the activation level of each expert remains within a small fixed range with minimal fluctuations, and there is no noticeable change as the terrain changes. In contrast, in CMoE, each expert's activation level exhibits jumps and a wider range of variations, with clear changes occurring as the terrain switches.

Additionally, in the CMoE plot, we can see that Expert 1 only appears in locations where the terrain ascends, such as uphill slopes and hurdles, which require foot lifting. On the other hand, Expert 2 performs oppositely, with a higher weight in areas where the terrain descends.

We hypothesize that Expert 1 specializes in terrains such as ascending steps. To verify this hypothesis, we removed Expert 1's output from the evaluation environment, while keeping the outputs of other experts unchanged. The experimental results showed that the robot attempted to lift its leg before stepping up the stairs but failed, causing it to stumble and fall. However, in the downhill scenario, the robot successfully navigated the stairs. This indicates that while the robot failed to lift its legs properly, its other movements were unaffected, allowing it to navigate terrains that do not require leg lifting, such as downhill slopes and descending stairs. This indicates that Expert 1's specialization is indeed in ascending step terrains, making it the expert for handling such terrains.

D. Training Performance

To test our method's improvements in training speed and efficiency, we trained using both the Base method and our own method with curriculum learning, recording changes in terrain level and reward. As shown in Fig. 4, the CMoE methods terrain level increased first and reached its maximum value, indicating that it learned to navigate the most difficult terrain first. The CMoE method outperformed the Base method in both reward growth rate and maximum value, suggesting that the gating network's terrain classification reduces learning difficulty and improves efficiency.

E. Real-world Experiment

To validate the effectiveness of our approach, we deployed the trained CMoE policy on a Unitree G1 humanoid robot. We used radar to collect point cloud information and combined it with the robot's positioning system to obtain elevation map observations. This data was then packaged and sent to our network, allowing the policy to be directly transferred from the simulated environment to the robot.

We then conducted experiments on various terrains to test the robot's real-world performance. For gap terrain, we tested the robot's maximum traversable width of 80 cm, which, to our knowledge, is the largest among existing methods. For step terrain, we prepared three different step heights: 12, 15, and 20 cm. Our strategy enabled the robot to successfully traverse the most challenging steps, surpassing existing methods in the literature [38] (which only support a maximum height of 15 cm). For hurdle terrain, the robot was able to cross thresholds of 30 cm in height. Finally, the robot can easily go up and down a 17-degree slope.

In addition to these basic terrains, we also tested the robots performance on a mixed terrain, which consisted of steps, gaps, slopes, and hurdles. This more complex terrain significantly challenged the robot's ability to adapt to rapidly changing environments. As shown in Fig. 6 and Fig. 1, the Unitree G1 robot successfully completed a series of terrain crossings, including combinations of the aforementioned single terrains, demonstrating the effectiveness of our method in perceiving various terrains and executing extreme parkour maneuvers.

Finally, we conducted robustness experiments to further evaluate the robots performance, as shown in Fig. 7(a). The step structure in this environment differs in width and height from the stairs in the training environment. Some steps even have protruding edges, which impose higher demands on the robots stair-climbing gait. Despite these challenges, the robot still demonstrated excellent stability during long-distance walking and stair navigation. Additionally, we tested the robustness of our method against disturbances, such as rope dragging and object collisions. As shown in Fig. 7(b), when the robot was crossing a 30 cm obstacle and was struck by an object during interference, it remained stable and successfully passed through. These results showcase the strong robustness of our method.

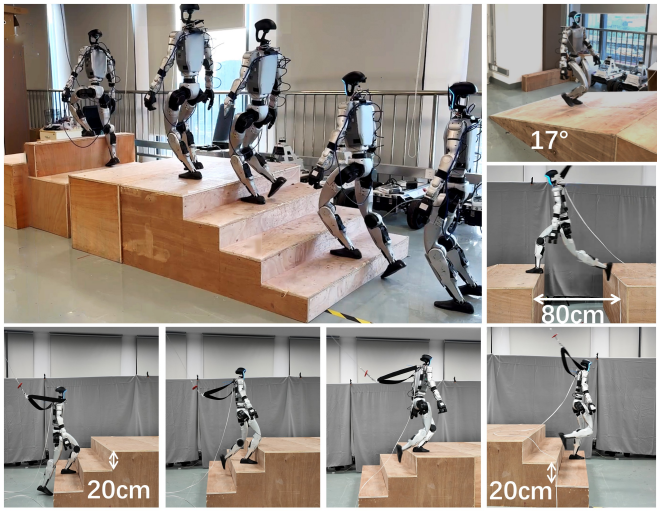

Fig. 6: In mixed terrain, the robot climbed 15 cm steps, a 60 cm gap, a 30 cm hurdle, and an uphill slope. In single terrain testing, the robot was able to ascend three steps of 20 cm in height with a steady pace, as well as descend stairs.

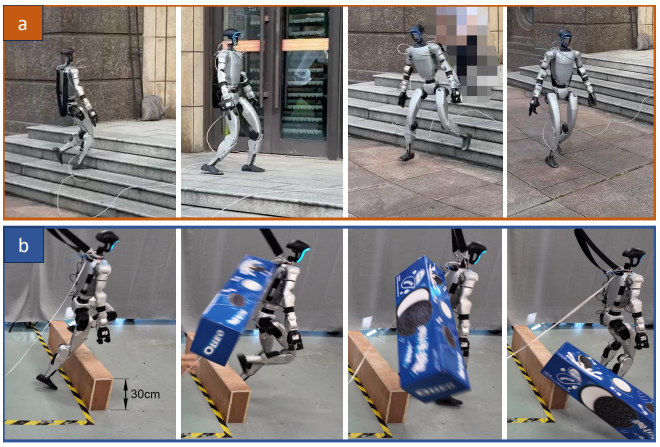

Fig. 7: (a) In an outdoor environment, the robot can navigate four continuous sections of untrained terrain with step edges while maintaining continuous movement. (b) Even when encountering human intervention while crossing a 30 cm high obstacle, it can maintain a stable posture and successfully pass the obstacle.

VI. Conclusions

We propose the CMoE, which integrates MoE and contrastive learning to enhance the robots adaptability across diverse terrains. Through a series of experiments, we have demonstrated the superior performance of CMoE in various terrains, including improvements in success rate, travel distance, and robustness in complex environments. Compared to the traditional Vanilla MoE method, CMoE more effectively allocates the expert network, particularly in handling complex terrains, and shows greater stability and generalization ability. The experimental results highlight CMoE's significant advantages in enhancing the robot's control and adaptability, enabling it to handle rapidly changing environments and perform challenging tasks. Future work will extend this approach to whole body control, enabling coordinated full-body parkour.

REFERENCES

[1] Hang Lai, Jiahang Cao, Jiafeng Xu, Hongtao Wu, Yunfeng Lin, Tao Kong, Yong Yu, and Weinan Zhang. World model-based perception for visual legged locomotion. *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11531–11537, 2024.

[2] Alberto Romay, Stefan Kohlbrecher, David C Conner, Alexander Stumpf, and Oskar von Stryk. Template-based manipulation in unstructured environments for supervised semi-autonomous humanoid robots. In *2014 IEEE-RAS International Conference on Humanoid Robots*, pages 979–986. IEEE, 2014.

[3] Ziwen Zhuang, Zipeng Fu, Jianren Wang, Christopher G Atkeson, Soeren Schwertfeger, Chelsea Finn, and Hang Zhao. Robot parkour learning. In *Conference on Robot Learning*, 2023.

[4] Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. Humanoid parkour learning. *ArXiv*, abs/2406.10759, 2024.

[5] Runhan Huang, Shaoting Zhu, Yilun Du, and Hang Zhao. Moe-loco: Mixture of experts for multitask locomotion. *ArXiv*, abs/2503.08564, 2025.

[6] Dewei Wang, Xinmiao Wang, Xinzhe Liu, Jiyuan Shi, Yingnan Zhao, Chenjia Bai, and Xuelong Li. More: Mixture of residual experts for humanoid lifelike gaits learning on complex terrains. *ArXiv*, abs/2506.08840, 2025.

[7] Jinhan Li, Yifeng Zhu, Yuqi Xie, Zhenyu Jiang, Mingyo Seo, Georgios Pavlakos, and Yuke Zhu. Okami: Teaching humanoid robots manipulation skills through single video imitation. *ArXiv*, abs/2410.11792, 2024.

[8] Bart Jaap van Marum, Aayam Shrestha, Helei Duan, Pranay Dugar, Jeremy Dao, and Alan Fern. Revisiting reward design and evaluation for robust humanoid standing and walking. *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11256–11263, 2024.

[9] Yueqi Zhang, Quancheng Qian, Tai-Wei Hou, Peng Zhai, Xiaoyi Wei, Kangmai Hu, Jiafu Yi, and Lihua Zhang. Renet: Fault-tolerant motion control for quadruped robots via redundant estimator networks under visual collapse. *IEEE Robotics and Automation Letters*, 2025.

[10] Zhaoyuan Gu, Junheng Li, Wenlan Shen, Wenhao Yu, Zhaoming Xie, Stephen McCrory, Xianyi Cheng, Abdulaziz Shamsah, Robert J. Griffin, C. Karen Liu, Abderrahmane Kheddar, Xue Bin Peng, Yuke Zhu, Guanya Shi, Quan Nguyen, Gordon Cheng, Huijun Gao, and Ye Zhao. Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning. *ArXiv*, abs/2501.02116, 2025.

[11] Xinyang Gu, Yen-Jen Wang, and Jianyu Chen. Humanoid-gym: Reinforcement learning for humanoid robot with zero-shot sim2real transfer. *ArXiv*, abs/2404.05695, 2024.

[12] Miguel 'Angel de Miguel, Jorge Beltr'an, Juan S. Cely, Francisco Mart'in, Juan Carlos Manzanares, and Alberto Garc'ia. I move therefore i learn: Experience-based traversability in outdoor robotics. *ArXiv*, abs/2507.00882, 2025.

[13] Ruihan Yang, Ge Yang, and Xiaolong Wang. Neural volumetric memory for visual locomotion control. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1430–1440, 2023.

[14] Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Legged locomotion in challenging terrains using egocentric vision. In *Conference on Robot Learning*, 2022.

[15] Ruiqi Yu, Qianshi Wang, Yizhen Wang, Zhicheng Wang, Jun Wu, and Qiuguo Zhu. Walking with terrain reconstruction: Learning to traverse risky sparse footholds. *ArXiv*, abs/2409.15692, 2024.

[16] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11443–11450, 2023.

[17] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model. *ArXiv*, abs/2411.14386, 2024.

[18] Huayi Wang, Zirui Wang, Junli Ren, Qingwei Ben, Tao Huang, Weinan Zhang, and Jiangmiao Pang. Beamdojo: Learning agile humanoid locomotion on sparse footholds. *ArXiv*, abs/2502.10363, 2025.

[19] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, page 7987, Feb 1991.

[20] M. I. Jordan. Hierarchical mixtures of experts and the em algorithm. In *IEE Colloquium on Advances in Neural Networks for Control and Systems*, 1994.

[21] Qing Wang, Xue Han, Jiahui Wang, Lehao Xing, Qian Hu, Lianlian Zhang, Chao Deng, and Junlan Feng. Multipl-moe: Multi-programming-lingual extension of large language models through hybrid mixture-of-experts. 2025.

[22] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, QuocV. Le, GeoffreyE. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv: Learning*, Jan 2017.

[23] Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. Gshard: Scaling giant models with conditional computation and automatic sharding. *Cornell University - arXiv*, Jun 2020.

[24] Lu Xu, Jiaqian Yu, Xiongfeng Peng, Yiwei Chen, Weiming Li, Jaewook Yoo, Sunghyun Chunag, Dongwook Lee, Daehyun Ji, and Chao Zhang. Mose: Skill-by-skill mixture-of-experts learning for embodied autonomous machines. 2025.

[25] Nan Du, Yanping Huang, Andrew M. Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, Barret Zoph, Liam Fedus, Maarten Bosma, Zongwei Zhou, Tao Wang, Yu Emma Wang, Kellie Webster, Marie Pellat, Kevin Robinson, Kathleen S. Meier-Hellstern, Toju Duke, Lucas Dixon, Kun Zhang, Quoc V. Le, Yonghui Wu, Z. Chen, and Claire Cui. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, 2021.

[26] Suning Huang, Zheyu Zhang, Tianhai Liang, Yihan Xu, Zhehao Kou, Chenhao Lu, Guowei Xu, Zhengrong Xue, and Huazhe Xu. Mentor: Mixture-of-experts network with task-oriented perturbation for visual reinforcement learning. *ArXiv*, abs/2410.14972, 2024.

[27] Wenxuan Song, Han Zhao, Pengxiang Ding, Can Cui, Shangke Lyu, Yaning Fan, and Donglin Wang. Germ: A generalist robotic model with mixture-of-experts for quadruped robot. *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11879–11886, 2024.

[28] Liangwei Nathan Zheng, Wei Emma Zhang, Mingyu Guo, Miao Xu, Olaf Maennel, and Weitong Chen. Rethinking gating mechanism in sparse moe: Handling arbitrary modality inputs with confidence-guided gate. *ArXiv*, abs/2505.19525, 2025.

[29] Jiamin Li, Qiang Su, Yitao Yang, Yimin Jiang, Cong Wang, and Hong-Yu Xu. Adaptive gating in mixture-of-experts based language models. In *Conference on Empirical Methods in Natural Language Processing*, 2023.

[30] R. Liu, Young Jin Kim, Alexandre Muzio, Barzan Mozafari, and Hany Hassan Awadalla. Gating dropout: Communication-efficient regularization for sparsely activated transformers. In *International Conference on Machine Learning*, 2022.

[31] Gwanghyeon Ji, Juhyeok Mun, Hyeongjun Kim, and Jemin Hwangbo. Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion. *IEEE Robotics and Automation Letters*, 7(2):46304637, April 2022.

[32] Péter Fankhauser, Michael Bloesch, and Marco Hutter. Probabilistic terrain mapping for mobile robots with uncertain localization. *IEEE Robotics and Automation Letters (RA-L)*, 3(4):3019–3026, 2018.

[33] Péter Fankhauser, Michael Bloesch, Christian Gehring, Marco Hutter, and Roland Siegwart. Robot-centric elevation mapping with uncertainty estimates. In *International Conference on Climbing and Walking Robots (CLAWAR)*, 2014.

[34] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Multi-task reinforcement learning without interference. In *Proc. Optim. Found. Reinforcement Learn. Workshop NeurIPS*, 2019.

[35] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments, 2021.

[36] Philip A Knight. The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275, 2008.

[37] N. Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. *ArXiv*, abs/2109.11978, 2021.

[38] Junfeng Long, Zirui Wang, Quanyi Li, Jiawei Gao, Liu Cao, and Jiangmiao Pang. Hybrid internal model: Learning agile legged locomotion with simulated robot response. *ArXiv*, abs/2312.11460, 2023.