2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Learning Quadrupedal Locomotion with Impaired Joints Using Random Joint Masking

Mincheol Kim[1], Ukcheol Shin[2], Jung-Yup Kim[1]

***Abstract*— Quadrupedal robots have played a crucial role in various environments, from structured environments to complex harsh terrains, thanks to their agile locomotion ability. However, these robots can easily lose their locomotion functionality if damaged by external accidents or internal malfunctions. In this paper, we propose a novel deep reinforcement learning framework to enable a quadrupedal robot to walk with impaired joints. The proposed framework consists of three components: 1) a random joint masking strategy for simulating impaired joint scenarios, 2) a joint state estimator to predict an implicit status of current joint condition based on past observation history, and 3) progressive curriculum learning to allow a single network to conduct both normal gait and various joint-impaired gaits. We verify that our framework enables the Unitree's Go1 robot to walk under various impaired joint conditions in real-world indoor and outdoor environments.**

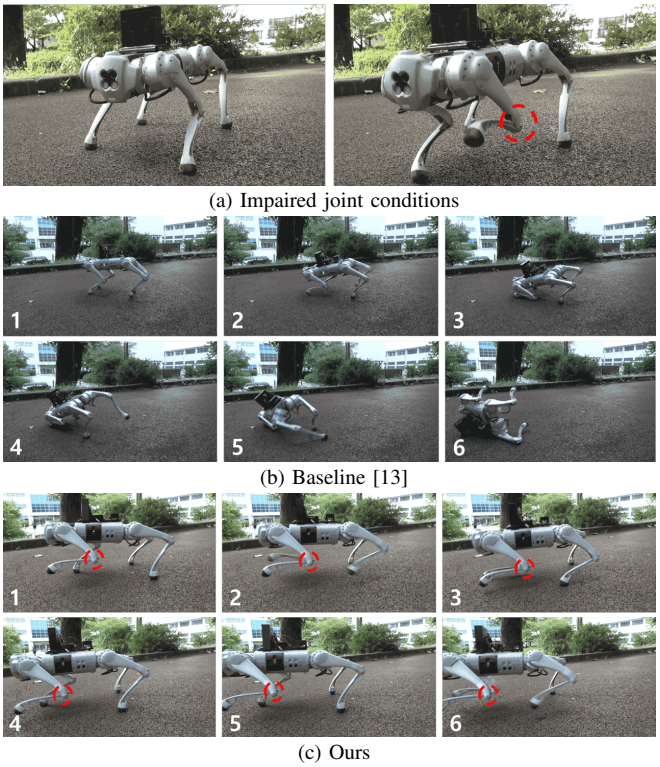
I. Introduction

Recently, quadrupedal robots have played a crucial role in various applications, such as human rescue, disaster response, and harsh terrain exploration [1], [2], [3]. To perform these applications in complex and dynamic environments, agile locomotion is a fundamental requirement for quadruped robots [4]. However, the locomotion ability can be impaired by external accidents, such as hit-by-obstacle and impact-by-objects, and internal hardware issues, such as joint malfunction and locking. These accidents and issues are more likely to occur in challenging and extreme environments.

Furthermore, the impaired locomotion ability can cause significant damage to both robots and humans, shortening the lifespan of the robot. As shown in Fig. 1-(a), if a joint is damaged by external or internal factors, it directly affects the locomotion ability and results in walking failure and robot body damage (*i.e.*, Fig. 1-(b)). Therefore, it is necessary to maintain robust locomotion ability even with impaired joints to ensure the safety of the robot itself and humans.

However, previous quadrupedal locomotion studies have been focused on agile quadrupedal locomotion in rough terrains [5], [6], [7] rather than fault-tolerant locomotion. Also, the existing fault-tolerant locomotion study just provides a naïve fall-recovery functionality [8], requires an accurate robot body model [9], [10], and not validated in real-world environments [11], [12]. Existing commercial robots also only provide basic fault detection or protective functions that shut down the system.

[1]M. Kim and J. Kim are with the Humanoid Robot Research Laboratory, Department of Mechanical Design and Robot Engineering, Seoul National University of Science and Technology, Seoul, 01811, Republic of Korea {kmc96, jyk76}@seoultech.ac.kr
[2] U. Shin is with Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15217, US ushin@andrew.cmu.edu



(a) Impaired joint conditions

(b) Baseline [13]

(c) Ours

Fig. 1: **Quadrupedal locomotion with impaired joints.** Our proposed framework aims to learn quadrupedal locomotion with impaired joints. In previous model [13], if a joint is damaged by external or internal factors (a), it directly affects the locomotion ability and results in walking failure and robot body damage (b). On the other hand, our proposed method (c) maintains stable locomotion capabilities even under impaired joint conditions (*i.e.*, red dotted circles).

The main difficulty for fault-tolerant locomotion is that the locomotion strategies vary according to the damaged joint positions and status. For example, zero torque in the hip-roll position completely loses the whole Degree of Freedom (DoF) of one leg. However, motor locking up in the knee-pitch position may preserve most of the DoF of one leg. It is especially challenging in model-based control [14], [15], such as Model Predictive Control (MPC) and Whole Body Control (WBC), which need an accurate body model and time-consuming model parameter tuning. The model-based controls are highly limited in dealing with diverse impaired joint conditions. Therefore, we need to cover the diverse malfunction scenarios of joint positions and status to achieve comprehensive locomotion ability with impaired joints.

To this end, we propose a novel deep reinforcement learning framework to enable a quadrupedal robot to walk with impaired joints. The proposed framework contains three novel components: 1) a random joint masking strategy to provide diverse impaired joint scenarios, 2) a joint status estimator to judge each joint status and determine an appropriate locomotion strategy, and 3) progressive curriculum learning to make a single network conduct both normal gait and various joint-impaired gaits. To evaluate the effectiveness of the proposed method, we conduct thorough and various experiments, including real-world demonstrations. As shown in Fig. 1, our proposed framework enables the Unitree's Go1 robot to maintain robust locomotion capabilities even under various impaired joint conditions.

Our novelties can be summarized as follows:

1) We propose a random joint masking strategy that simulates diverse joint malfunction scenarios, such as zero torque or locking up, by randomly masking joint actions and torques.
2) We propose a joint status estimator that classifies whether each joint is normal or impaired based on previous observation history.
3) We propose a progressive curriculum learning to allow a single network to conduct both normal gait and various joint-impaired gaits.

II. Related Works

A. Deep Reinforcement Learning for Quadrupedal Robot

Deep reinforcement learning has been presented as an alternative direction for designing robot controllers without requiring explicit prior knowledge, such as dynamic models and inverse kinematics. The recent literature has been focused on reducing training time [13], agile quadrupedal locomotion [16], [17], and reducing sim-to-real gap [18], [19], [20]. Specifically, Rudin *et al.* [13] greatly shortens the entire training time by utilizing massive parallel GPU processing. A number of study makes the agent robustly or quickly walk in challenging terrains, such as high-speed running extensions [16] and fast locomotion on deformable terrains [17]. Another branch is reducing the domain gap between simulation and the real world by adapting environmental state [21], [22], using domain randomization [20], [23], [24], and leveraging teach-student training method [6], [21], [16].

However, all previous works assume the ideal condition that every joint is fully functional. In the real world, it is often violated by external accidents and internal hardware issues and leads to a complete loss of robot locomotion functionality. Therefore, this paper focuses on learning robust quadruped locomotion under an impaired joint condition, which has been less explored despite its importance.

B. Fault Tolerance Control for Quadrupedal Locomotion

Most studies about fault-tolerant quadrupedal locomotion have been proposed in model-based control systems [9], [25], [10]. They usually aimed to solve a single joint locking case with inverse kinematics [9], gait modeling [25], and whole-body control [10]. However, these methods often relied on a specific robot dynamic model, case-by-case models for various failure cases, and time-consuming model parameter tuning. A few fault tolerance algorithms utilizing deep reinforcement learning have been proposed [11], [12]. However, the proposed methods were only evaluated in a simulation environment and are not publicly available.

Unlike the previous method, our proposed method handles various joint failure cases, including zero joint torque and joint locking with varying joint positions. Furthermore, the proposed method is evaluated in both simulation and real-world indoor and outdoor environments to verify its effectiveness.

III. Proposed Learning Strategy

A. Overview of framework

Our proposed framework consists of three components, as shown in Fig. 2; random joint masking to simulate diverse impaired joint conditions (Sec. III-B), joint status estimator θ^S to judge the current joint status (Sec. III-C), and progressive curriculum learning (Sec. III-D) to allow a single policy network π to conduct diverse normal and impaired quadrupedal locomotion.

Architecture. Our framework contains teacher-student joint status estimators (*i.e.*, θ^T and θ^S) and shared policy network π. Our proposed framework utilizes teacher-student knowledge distillation [26] and privileged observation [21] to train a student joint status estimator. The policy network, which is shared between teacher and student networks, is trained through Proximal Policy Optimization (PPO) [27].

Training data acquisition. Given privileged observation[1] e_t, the teacher joint status estimator θ^T embeds the current environmental information as latent vector z_t. The latent z_t and observation[2] o_t fed into policy network π. After that, the policy network predicts the desired action[2] a_t for all joints. At this moment, the actions and joint torques are randomly masked (*i.e.*, disabled) based on the current curriculum scenario. The masked action forces the robot to move without utilizing the masked joint. This process is iterated to make training data within the current curriculum scenario. As a result, the policy network is trained to predict appropriate locomotion behavior according to the current joint status.

Reward. The reward is calculated after conducting the masked action. We utilized the same reward function used in the previous studies [13], [16], [28]. However, we found some reward terms that promote normal gait locomotion are ineffective for learning impaired quadrupedal locomotion. Therefore, we exclude the three reward terms (*i.e.*,Table I) for impaired joint scenarios. However, we still include the three terms for normal joint scenarios. Based on the training

[1]Our privileged observation includes ground friction, ground restitution, and joint mask value for all joint status (*i.e.*, 0 is normal, 1 to 12 indicates each impaired joint condition), defined as $e_t = [fr_t, gr_t, m_t]^T$, where $fr_t \in \mathbb{R}^1$, $gr_t \in \mathbb{R}^1$, and $m_t \in \mathbb{R}^1$

[2]Please refer to Appendix for observation and action details.

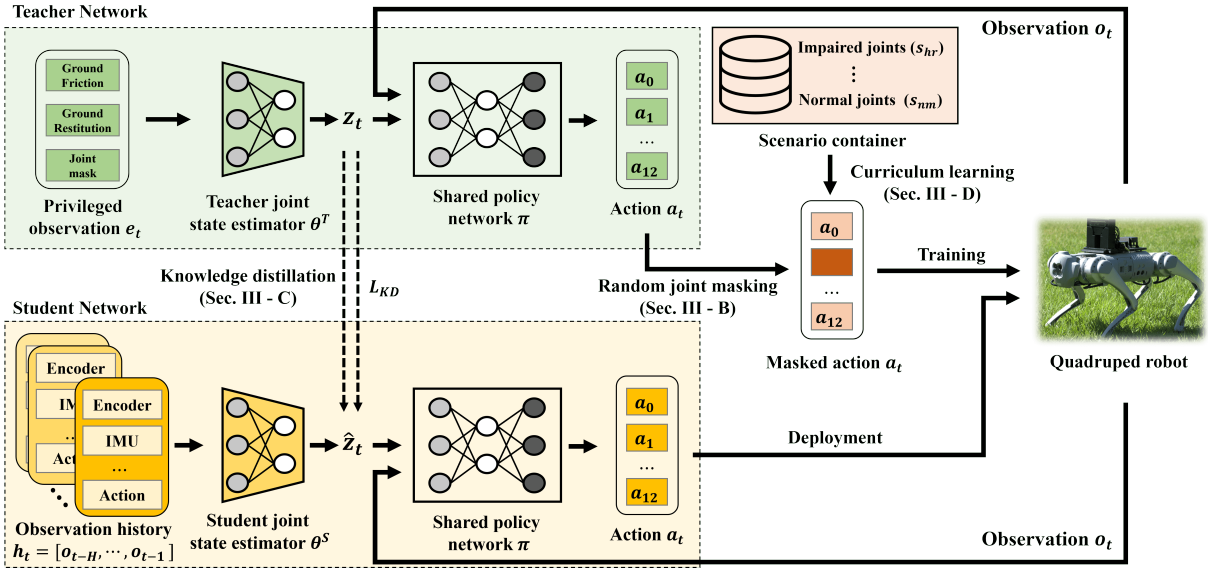


Fig. 2: **Overall pipeline of our proposed training framework.** Our proposed method aims to learn quadrupedal locomotion with impaired locomotion. For this purpose, we first enforce impaired joint scenarios to the robot by randomly masking joint action a_t (Sec. III-B). Also, to make the robot estimate the current joint status (*i.e.*, whether all joints are available or partially impaired) through observation history, we perform imitation learning between teacher and student latent vectors $\hat{z}_t$ and z_t (Sec. III-C). Additionally, we propose a progressive curriculum learning strategy to enable comprehensive locomotion ability for a normal joint and various impaired joint conditions (Sec. III-D). Based on the proposed components, the robot is able to walk in not only a normal joint condition but also various impaired joint conditions.

TABLE I: **Selectively utilized reward terms**. The three reward terms promote normal gait locomotion (*e.g.*, The values (C_{foot}^{cmd}, s_y^{cmd}) are generated from gait algorithm [28]). We exclude the three terms for impaired joint scenarios. Please refer to these works [13], [16], [28] for full reward table and descriptions.

Reward	Equation	weight
Swing phase tracking	$\sum[1 - C_{\text{foot}}^{\text{cmd}}(\theta^{\text{cmd}}, t)] \exp\left(-\lvert f^{\text{foot}}\rvert^2/0.25\right)$	4.0
Stance phase tracking	$\sum[C_{\text{foot}}^{\text{cmd}}(\theta^{\text{cmd}}, t)] \exp\left(-\lvert V_{xy}^{\text{foot}}\rvert^2/0.25\right)$	4.0
Foot swing tracking	$(p_{x,y,foot}^{f} - p_{x,y,foot}^{f,\text{cmd}}(s_y^{\text{cmd}}))^2$	-10.0

data acquisition process and reward terms, the policy network π is optimized using PPO algorithm [27] to maximize total expected reward Eq. (1).

$$J(\pi) = \mathbb{E}_{r\sim p(r|\pi)}\left[\sum_{t=0}^{T-1}\gamma^t r_t\right], \tag{1}$$

where γ is the discount factor, r_t is the weighted sum reward term at the time t, and T is the scenario length.

B. Random Joint Masking Strategy

Learning quadrupedal locomotion with impaired joints in the real world is a challenging problem due to various hardware and software issues, such as the difficult data acquisition process, potential severe damage to the robot body, non-observable information, and inaccurate robot state. Therefore, simulating impaired joint conditions is a safe and reliable solution to learning quadrupedal locomotion with impaired joints. To this end, we propose a random joint masking strategy to simulate diverse and plentiful joint malfunction scenarios, such as zero torque and locking up, at various joint positions. During the training process, the proposed strategy randomly sets joint torque as zero (*i.e.*, the joint is not functional anymore) and also sets the predicted action value of masked joint position as zero (*i.e.*, force not to use the deactivated joint). The detailed process is as follows.

1) Joint malfunction scenario assignment: Based on the current curriculum learning level, the current scenario assigns a normal joint condition or various joint malfunction scenarios (*e.g.*, malfunction in front-left-hip-pitch joint). If the current scenario is a normal joint condition, the predicted action a_t is directly delivered to the robot. If not, the predicted action and torque at the assigned joint position are masked out. Note that, during the training process, the multiple agents simultaneously conduct various normal and malfunction scenarios [13].

2) Joint action and torque masking: When the joint malfunction position is assigned, the torque τ_{des} and action value a_t of the assigned position are set to zero, as shown in Eq. (2) and Eq. (3).

$$\tau_{\text{des}}(p) \leftarrow \begin{cases} 0 & \text{if impaired joint} \\ \tau_{\text{des}}(p) & \text{otherwise} \end{cases}, \tag{2}$$

where p indicates joint position. Making the desired torque zero means the joint is not functional.

$$a_t(p) \leftarrow \begin{cases} 0 & \text{if impaired joint} \\ a_t(p) & \text{otherwise} \end{cases} \tag{3}$$

Also, masking the predicted action value as zero forces the agent not to use the deactivate joint. Therefore, the proposed strategy provides diverse and plentiful experiences in which various joints are not functional and won't move. As a result, the agent needs to learn a locomotion strategy that doesn't use the impaired joint.

C. Joint Status Estimator

Another important requirement for impaired quadruped locomotion is to identify the current joint status, whether the joints are fully functional or damaged, based on the observable information in the real world. This enables the agent to determine an appropriate locomotion strategy according to the current joint status. For this purpose, we utilize teacher-student knowledge distillation [26] and the idea of privileged observation [6], [21], [16]. The previous works [6], [21], [16] estimate embedded states of environmental variables that are challenging to know in the real world, such as ground friction and restitution, based on past observation history.

By extending this idea, we propose a joint status estimator θ^S that can estimate implicit representation of the current joint functionality and environmental variables based on the past observation history. The joint status estimator is trained with knowledge distillation from teacher estimator θ^T to student estimator θ^S.

1) Teacher joint status estimator: We fed the joint mask (ranging from 0 to 12) along with environmental variables to the teacher estimator as a privileged observation e_t. The joint mask indicates that all joints are fully functional or a certain joint is malfunctioning. Then, the teacher network implicitly represents the privileged observation as latent vector z_t.

2) Student joint status estimator: On the other hand, student estimators cannot leverage privileged observation in the real world. Therefore, the student estimator aims to estimate the same latent vector of the teacher network based on available observation information in the real world (*i.e.*, $[o_{t-H}, ..., o_{t-1}]$, where H is set to 30). To this end, the student estimator is trained to minimize the difference between the teacher latent vector z_t and student latent vector $\hat{z}_t$, as shown in Eq. (4). This knowledge distillation loss L_{KD} enables the student estimator to estimate the implicit representation for the current joint status and environmental variable information with a given past observation history.

$$L_{KD}(\hat{z}_t, z_t) = \|\hat{z}_t - z_t\|^2 \tag{4}$$

D. Progressive Curriculum Learning for Unified Policy

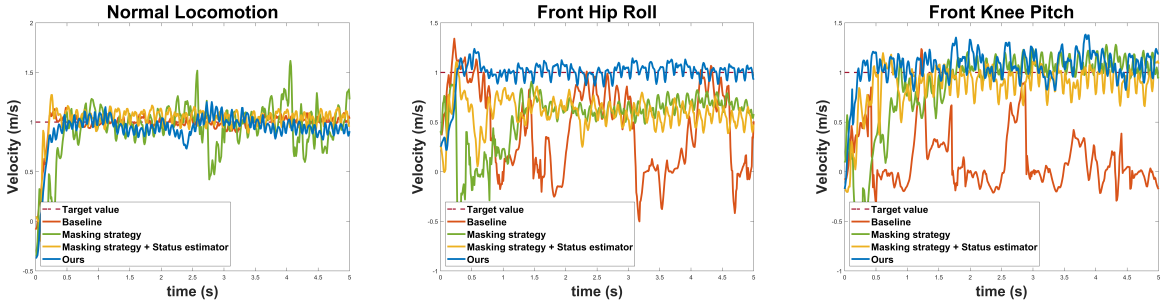
Locomotion strategies may vary depending on the impaired joint position and status. Training each network for each locomotion strategy greatly increases memory requirement and reduces the flexibility and scalability of the network. However, training a single network to handle various locomotion strategies is also a challenging problem because of the catastrophic forgetting [29], [30], [31]. In other words, the network forgets previously trained locomotion strategy while only preserving the newly learned strategy. Therefore, it leads to imbalanced performance for the various impaired joint scenarios.

Algorithm 1: Progressive Curriculum Learning

```
Initialize policy network π, status estimators θ^T, θ^S;
initialize curriculum scenario container C = [s_nm];
Empty buffers D_1, D_2;
for 0 ≤ itr ≤ N_itr do
    for 0 ≤ i ≤ N_env do
        s_i ← AssignSenario(C); # Sec. III-B
        for 0 ≤ t ≤ T do
            z_t ← θ^T(e_t);, ẑ_t ← θ^S(h_t);
            a_t ← π(o_t, z_t);
            a_t ← Masking(a_t, s_i); # Sec. III-B
            o_{t+1}, e_{t+1}, r_t ← envs[i].step(a_t);
            Store (o_t, e_t, a_t, r_t), (ẑ_t, z_t) in D_1, D_2;
        end
    end
    Update π^T and θ^T using PPO [27];
    Update θ^S with L_KD; # Sec. III-C
    Empty buffers D_1, D_2;
    # Sec. III-D, Progressive curriculum learning
    R_avg ← Σ_{i=0}^{N_env} J(π)/N_env;
    if R_avg > th_level1 then
        C = [s_nm, s_kp];
        if R_avg > th_level2 then
            C = [s_nm, s_kp, s_hp];
            if R_avg > th_level3 then
                C = [s_nm, s_kp, s_hp, s_hr];
            end
        end
    end
end
```

To resolve this issue and make a single network encompass both normal and various joint-impaired locomotion capabilities, we propose a progressive curriculum learning method. The proposed method progressively includes more difficult joint-impaired conditions while preserving overall performance over the current curriculum learning level. As shown in algorithm 1, the initial curriculum level only includes a normal joint condition s_{nm}, the easiest and basic locomotion. After that, whenever the reward mean value R_{avg}, averaged over the current curriculum level, satisfies certain criteria, the curriculum learning level gradually includes more difficult impaired joint conditions in the order of knee-pitch, hip-pitch, and hip-roll. If the curriculum level reaches the final stage, all joint conditions (*i.e.*, normal and abnormal joints) are equally assigned to the scenarios.

Note that the order is decided by the disturbance level based on empirical findings. When one of the legs is impaired because of a joint malfunction, the impaired leg can be considered a disturbance factor to the other three legs. In this case, we found the most significant disturbance occurs when the torque of the Hip-Roll joint, which is closest to the robot's body, becomes zero. The tendency of disturbance decreases as the impaired joint is far from the robot's body (*i.e.*, knee-pitch is the least disturbance). Therefore, we proceeded with the curriculum learning level in the order of knee-pitch, hip-pitch, and hip-roll.



(a) Normal joint condition (b) Zero-torque (Hip-roll) (c) Zeor-torque (Knee-pitch)

Fig. 3: **Response graph comparison for body velocity command in Isaac Gym simulation [32].** We compare baseline [13] and variants of the proposed method (*i.e.*, an ablation study of each component). Given the forward body velocity command (*i.e.*, 1.0 m/s), all methods show similar performance in a normal joint condition (a). However, when a joint has malfunctioned (*i.e.*, zero-torque) at a randomly joint (b,c), our proposed framework immediately responds and shows stable tracking ability.

TABLE II: **Tracking error comparison in Issac Gym simulation**. Given the forward velocity command, we compare the tracking error of each method with the Root-Mean-Square-Error (RMSE) metric. The best and runner-up performances in each block are highlighted in **bold** and underline.

Joint status	Impaired Joint position		RMSE ↓			
			Baseline [13]	+Masking	+Masking +Status Est	Ours
Normal	None		**0.1469**	0.2079	0.3009	0.1699
Zero Torque	Front Left	HR	0.7089	0.5290	0.4364	0.1507
		HP	0.9747	0.4899	0.2980	0.2668
		KP	0.9252	0.4214	0.2580	0.1920
	Avg		0.8696	0.4801	0.3308	**0.2031**
	Rear Right	HR	0.8928	0.2300	0.2475	0.1986
		HP	1.0320	0.2584	0.1513	0.1959
		KP	0.8081	0.3604	0.2003	0.1492
	Avg		0.9109	0.2829	0.1997	**0.1812**
Locked Up	Front Left	HR	0.2503	0.4803	0.2789	0.2003
		HP	0.2456	0.5075	0.2896	0.2333
		KP	0.2831	0.6684	0.2158	0.2456
	Avg		0.2596	0.5520	0.2614	**0.2264**
	Rear Right	HR	0.6966	0.2704	0.1484	0.2001
		HP	0.7086	0.3935	0.1405	0.1542
		KP	0.7139	0.2614	0.2265	0.2298
	Avg		0.7063	0.3084	**0.1718**	0.1947
Total Avg			0.5787	0.3662	0.2529	**0.1951**

IV. Experimental Results

A. Quadrupedal Locomotion in Isaac Gym Simulation [32]

We evaluate our proposed method with the baseline algorithm [13] and the variants of the proposed method. Unfortunately, fault-tolerant reinforcement learning methods [11], [12] are publicly unavailable, so we were not able to compare these methods. The experimental results are shown in Fig. 3 and Tab. II. 'Masking strategy' indicates we have added the proposed joint masking strategy (*i.e.*, Sec III-B) to the Baseline [13] method. 'Masking strategy + status estimator' means we further have added joint status estimator (*i.e.*, Sec III-C). 'Ours' is the final proposed framework. For the evaluation, we gave the forward body velocity command of 1.0m/s and measured the tracking accuracy of each method using the Root-Mean-Square-Error (RMSE) metric.

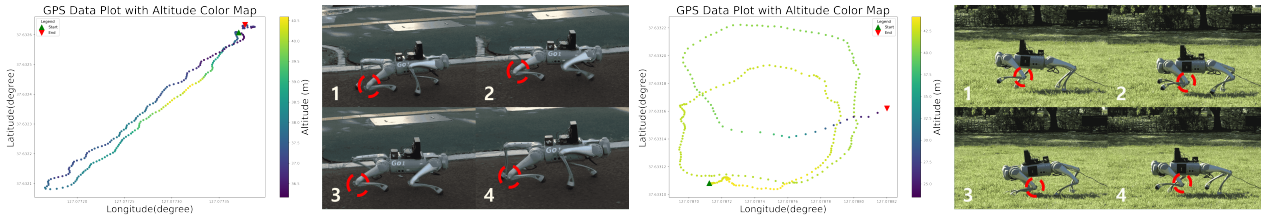
All methods show similar performance in a normal joint condition (Fig. 3-(a)). However, when a certain joint has malfunctioned (*i.e.*, zero-torque) randomly, baseline [13] loses the locomotion capability and doesn't follow the command at all. On the other hand, our proposed framework immediately responds and shows stable and reliable tracking ability (Fig. 3-(b,c)). As shown in Tab. II, our proposed method achieves the best and second-best performance over a normal joint and various impaired joint conditions.

The table also shows the effectiveness of each proposed component of our proposed framework. Applying the random joint masking strategy to the baseline greatly decreased the tracking errors in impaired joint conditions. However, the performance tendency was quite imbalanced according to each impaired joint scenario. This is because the agent didn't know about the current joint conditions, whether it is normal or impaired. The joint status estimator resolved the issue and let the model know which joint was damaged or fully functional. However, still, the performance tendency was imbalanced and the normal gait performance was decreased. Progressive curriculum learning resolved the imbalanced performance issue and boosted the overall locomotion ability over various joint conditions. Based on the three proposed components, we finally built a single neural network to possess diverse quadrupedal locomotion capabilities in normal and various impaired joint conditions.

B. Empirical Analysis of Impaired Quadrupedal Locomotion in Real-world Laboratory Environments

In this section, we empirically analyze the performance of impaired quadrupedal locomotion in a real-world indoor environment by using a motion capture system. We applied our learned model to the Unitree's Go1 robot without any fine-tuning stage. For the experiments, we give joint failure scenarios to the robot by forcing the joint at a fixed state (*i.e.*, locked up) or setting the joint torque as zero (*i.e.*, zero-torque). Also, we assign forward-moving and rotation-in-place tasks in normal joint and joint malfunction conditions. We measure the robot's body and yaw velocity using six OptiTrack motion capture cameras to evaluate the performance of forward-moving and in-place rotation ability. The body and yaw velocities are averaged over a 5m distance and 10 seconds.

The experimental results are shown in Table III. The Unitree's Go1 robot has reported 1.35m/s body velocity and 3.46 rad/s yaw velocity in a normal joint condition. Impaired



(a) Course 1 (b) Snapshot in Course 1 (c) Course 2 (d) Snapshot in Course 2

Fig. 4: **Impaired quadrupedal locomotion in real-world outdoor environments.** The trajectory was recorded using an RTK-GPS attached to the robot's body. Course 1 consists of gentle incline, decline, and level areas with clean and hard ground. Course 2 is deformable ground covered with soft grass and small pebbles.

TABLE III: **Performance comparison of impaired quadrupedal locomotion using motion capture data**.

Joint status	Impaired Joint position		Motion capture data	
			Body velocity(m/s)	Body Yaw(rad/s)
Normal	None		**1.3536**	**3.4643**
Zero-Torque	Front Left	HR	0.9964	2.6009
		HP	1.1506	2.6628
		KP	1.2687	2.7093
	Avg		**1.1386**	**2.6577**
	Rear Right	HR	1.044	3.3191
		HP	1.3121	3.532
		KP	1.3171	3.0147
	Avg		**1.2244**	**3.2886**
Locked up	Front Left	HR	0.9702	3.2931
		HP	1.0050	3.7662
		KP	1.0580	3.9492
	Avg		**1.0111**	**3.6695**
	Rear Right	HR	1.0682	3.4407
		HP	1.0416	3.6306
		KP	1.1703	3.6719
	Avg		**1.0934**	**3.5811**

locomotion generally shows slightly worse but comparable results compared to normal quadrupedal locomotion. Compared to the baseline model [13] that was tumbled down under impaired joint conditions (*i.e.*, Fig. 1-(b)), we believe our proposed framework significantly increases the robot's stability and reliability regarding locomotion capability. Also, our proposed framework sustains the forward-moving and rotation tasks against diverse impaired joint positions (*e.g.*, hip-roll, hip-pitch, and knee-pitch). Still, there is a lot of room to improve in that the performance of impaired locomotion decreases according to the impaired joint positions and front-biased center of gravity. For example, malfunction at the hip-pitch position much degenerates locomotion capability than the keen-pitch joint, due to the disturbance level that is proportional to the distance to the robot body. We will investigate this problem in future work.

C. Impaired Quadrupedal Locomotion in Real-world Outdoor Environments

We further evaluated that our proposed framework can walk with impaired joints in real-world outdoor environments. To evaluate performance in outdoor environments, we selected two different courses. Course 1 consists of gentle incline slopes, decline slopes, and level areas with clean and hard grounds. Course 2 is almost a flat region but the terrain is deformable ground covered with soft grass and small pebbles. We measured the robot's trajectory using Real-Time Kinematic (RTK) GPS attached to the robot's body. The outdoor evaluation results are shown in Fig. 4. We can observe that the agent robustly walks in various terrains even under impaired joint conditions. The agent traverses two different terrains with a total distance of 0.5 km based on learned impaired quadrupedal locomotion strategies (*i.e.*, Fig. 4-(b,d)). Further results can be found in the supplementary video. In future work, we plan to study walking in challenging terrains such as mountains and caves.

V. Conclusion

In this paper, we proposed a novel deep reinforcement learning framework to enable a quadrupedal robot to walk with impaired joints. The proposed framework contains three novel components: 1) a random joint masking strategy to provide diverse impaired joint scenarios, 2) a joint status estimator to judge each joint status and determine an appropriate locomotion strategy, and 3) progressive curriculum learning to make a single network conduct both normal gait and various joint-impaired gaits. The proposed framework is thoroughly verified in simulation environments, real-world indoor environment using motion capture, and real-world outdoor environments. As a result, our proposed framework enables the Unitree's Go1 robot to maintain stable and reliable locomotion capability even under various impaired joint conditions.

ACKNOWLEDGMENT

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE, Korea) under the Industrial Technology Innovation Program. Grant No. 20026194, "Development of Human-Life Detection and Fire-Suppression Solutions based on Quadruped Robots for Firefighting and Demonstration of Firefighting Robots and Sensors".

Appendix

Observation. Observation $o_t \in \mathbb{R}^{62}$ consists of joint angles $q_t \in \mathbb{R}^{12}$, joint angular velocities $\dot{q}_t \in \mathbb{R}^{12}$, gravity vector $g_t \in \mathbb{R}^3$, foot touchdown location $f_t \in \mathbb{R}^4$, x and y velocity, yaw angular velocity, body height, foot swing frequency, body pitch and body roll $c_t \in \mathbb{R}^7$, current action $a_t \in \mathbb{R}^{12}$, and previous action $a_{t-1} \in \mathbb{R}^{12}$.

Action. Action a_t are defined as the angles of each joint, $a_t \in \mathbb{R}^{12}$ equal to the size of the robot's degrees of freedom (DOF), and the torques applied to each joint for joint position control are calculated using PD control.

REFERENCES

[1] C Dario Bellicoso, Marko Bjelonic, Lorenz Wellhausen, Kai Holtmann, Fabian Günther, Marco Tranzatto, Peter Fankhauser, and Marco Hutter. Advances in real-world applications for legged robots. *Journal of Field Robotics*, 35(8):1311–1326, 2018.

[2] Joshua Hooks, Min Sung Ahn, Jeffrey Yu, Xiaoguang Zhang, Taoyuanmin Zhu, Hosik Chae, and Dennis Hong. Alphred: A multi-modal operations quadruped robot for package delivery applications. *IEEE Robotics and Automation Letters*, 5(4):5409–5416, 2020.

[3] Zhiming Chen, Tingxiang Fan, Xuan Zhao, Jing Liang, Cong Shen, Hua Chen, Dinesh Manocha, Jia Pan, and Wei Zhang. Autonomous social distancing in urban environments using a quadruped robot. *IEEE Access*, 9:8392–8403, 2021.

[4] Marco Tranzatto, Takahiro Miki, Mihir Dharmadhikari, Lukas Bernreiter, Mihir Kulkarni, Frank Mascarich, Olov Andersson, Shehryar Khattak, Marco Hutter, Roland Siegwart, et al. Cerberus in the darpa subterranean challenge. *Science Robotics*, 7(66):eabp9742, 2022.

[5] Péter Fankhauser, Marko Bjelonic, C Dario Bellicoso, Takahiro Miki, and Marco Hutter. Robust rough-terrain locomotion with a quadrupedal robot. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5761–5768. IEEE, 2018.

[6] Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning quadrupedal locomotion over challenging terrain. *Science robotics*, 5(47):eabc5986, 2020.

[7] Fumiya Iida, Rolf Pfeifer, et al. Cheap rapid locomotion of a quadruped robot: Self-stabilization of bounding gait. In *Intelligent autonomous systems*, volume 8, pages 642–649. IOS Press Amsterdam, 2004.

[8] Joonho Lee, Jemin Hwangbo, and Marco Hutter. Robust recovery controller for a quadrupedal robot using deep reinforcement learning. *arXiv preprint arXiv:1901.07517*, 2019.

[9] Jung-Min Yang. Kinematic constraints on fault-tolerant gaits for a locked joint failure. *Journal of Intelligent and Robotic Systems*, 45:323–342, 2006.

[10] Junwen Cui, Zhan Li, Jing Qiu, and Tianxiao Li. Fault-tolerant motion planning and generation of quadruped robots synthesised by posture optimization and whole body control. *Complex & Intelligent Systems*, 8(4):2991–3003, 2022.

[11] Timothée Anne, Jack Wilkinson, and Zhibin Li. Meta-learning for fast adaptive locomotion with uncertainties in environments and robot dynamics. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4568–4575. IEEE, 2021.

[12] Wataru Okamoto, Hiroshi Kera, and Kazuhiko Kawamoto. Reinforcement learning with adaptive curriculum dynamics randomization for fault-tolerant robot control. *arXiv preprint arXiv:2111.10005*, 2021.

[13] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on Robot Learning*, pages 91–100. PMLR, 2022.

[14] Jared Di Carlo, Patrick M Wensing, Benjamin Katz, Gerardo Bledt, and Sangbae Kim. Dynamic locomotion in the mit cheetah 3 through convex model-predictive control. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 1–9. IEEE, 2018.

[15] Christian Gehring, Stelian Coros, Marco Hutter, Carmine Dario Bellicoso, Huub Heijnen, Remo Diethelm, Michael Bloesch, Péter Fankhauser, Jemin Hwangbo, Mark Hoepflinger, et al. Practice makes perfect: An optimization-based approach to controlling agile motions for a quadruped robot. *IEEE Robotics & Automation Magazine*, 23(1):34–43, 2016.

[16] Gabriel B Margolis, Ge Yang, Kartik Paigwar, Tao Chen, and Pulkit Agrawal. Rapid locomotion via reinforcement learning. *arXiv preprint arXiv:2205.02824*, 2022.

[17] Suyoung Choi, Gwanghyeon Ji, Jeongsoo Park, Hyeongjun Kim, Juhyeok Mun, Jeong Hyun Lee, and Jemin Hwangbo. Learning quadrupedal locomotion on deformable terrain. *Science Robotics*, 8(74):eade2256, 2023.

[18] Sylvain Koos, Jean-Baptiste Mouret, and Stéphane Doncieux. Crossing the reality gap in evolutionary robotics by promoting transferable controllers. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 119–126, 2010.

[19] Adrian Boeing and Thomas Bräunl. Leveraging multiple simulators for crossing the reality gap. In *2012 12th international conference on control automation robotics & vision (ICARCV)*, pages 1113–1119. IEEE, 2012.

[20] Jie Tan, Tingnan Zhang, Erwin Coumans, Atil Iscen, Yunfei Bai, Danijar Hafner, Steven Bohez, and Vincent Vanhoucke. Sim-to-real: Learning agile locomotion for quadruped robots. *arXiv preprint arXiv:1804.10332*, 2018.

[21] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for legged robots. *arXiv preprint arXiv:2107.04034*, 2021.

[22] Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Vassilios Tsounis, Vladlen Koltun, and Marco Hutter. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26):eaau5872, 2019.

[23] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017.

[24] Bhairav Mehta, Manfred Diaz, Florian Golemo, Christopher J Pal, and Liam Paull. Active domain randomization. In *Conference on Robot Learning*, pages 1162–1176. PMLR, 2020.

[25] CF Pana, IC Resceanu, and DM Patrascu. Fault-tolerant gaits of quadruped robot on a constant-slope terrain. In *2008 IEEE International Conference on Automation, Quality and Testing, Robotics*, volume 1, pages 222–226. IEEE, 2008.

[26] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

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

[28] Gabriel B Margolis and Pulkit Agrawal. Walk these ways: Tuning robot control for generalization with multiplicity of behavior. In *Conference on Robot Learning*, pages 22–31. PMLR, 2023.

[29] Craig Atkinson, Brendan McCane, Lech Szymanski, and Anthony Robins. Pseudo-rehearsal: Achieving deep reinforcement learning without catastrophic forgetting. *Neurocomputing*, 428:291–307, 2021.

[30] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[31] Prakhar Kaushik, Alex Gain, Adam Kortylewski, and Alan Yuille. Understanding catastrophic forgetting and remembering in continual learning with optimal relevance mapping. *arXiv preprint arXiv:2102.11343*, 2021.

[32] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.