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

MAKP: Multi-mode Accurate Kicking Policy for Humanoid Robots

Zheng Zhang[1,2*], Kaiyang Xu[1*], Zhanxiang Cao[1,2], Yizhi Chen[2], Peng Wang[1], Haoyang Li[1,2], Yang Zhang[1], Shengcheng Fu[2], Xin Shen[1], Xiaokang Yang[1], Yue Gao[1,2†],

***Abstract*— Humanoid robot soccer players face fundamental challenges in achieving stable motion execution and ball trajectory control, particularly under balance constraints during single-leg support phases. In this paper, we introduce MAKP (Multi-mode Accurate Kicking Policy), a novel motion generation-based end-to-end kicking paradigm that enables humanoid robots to perform accurate ball kicking while executing diverse kicking motions. MAKP uniquely integrates a diffusion-based motion generator to produce various kicking trajectories and employs a three-stage learning strategy to address the inherent trade-off between motion similarity and kicking performance. Stage I focuses on stable motion tracking and single-leg balance maintenance, while Stage II optimizes ball kicking capabilities. In Stage III, we introduce a Multi-Critic mechanism combined with curriculum learning to further enhance the balance between kicking accuracy, motion similarity and robot stability. Real-world experiments on the Booster T1 platform validate the effectiveness of our approach.**

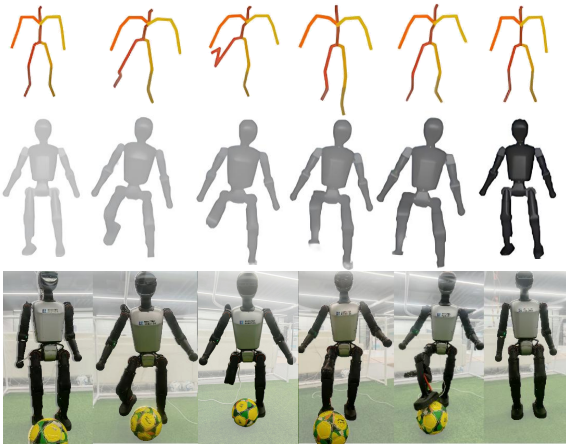

Fig. 1. **Snapshot of MAKP's multi-stage learning pipeline.** Top: Motion Diffusion Model generates skeletal trajectories. Middle: Trained policy with retargeted motion datasets deployed in MuJoCo. Bottom: Trained policy execution on physical robots.

I. Introduction

In recent years, humanoid robots have attracted increasing attention in sports applications, such as table tennis [1], soccer [2] [3], and other athletic domains. The emergence of international robotic soccer competitions like RoboCup [4] has further accelerated the development of robotic soccer technologies [5]. Among these capabilities, accurate kicking is one of the most fundamental skills, directly influencing both the competitive performance and practical utility of robotic soccer players.

Previous studies, particularly in RoboCup [6] [7] [8] , have explored ball-kicking using modularized methods with predesigned fixed motions. While these methods can achieve basic kicking functionality under specific conditions, they do not address the problem of shooting toward specific targets, limiting the robot's ability to actively control the ball's direction. Model-based policies [9] attempt to plan ball-kicking trajectories with precision but struggle to handle uncertainty factors in real-world environments, particularly the difficult-to-model contact dynamics between the foot and the deformable soccer ball, as well as uncertainties in ball-ground rolling friction. These challenges result in discrepancies between actual and desired positions.

Emerging reinforcement learning (RL) methods show promise in addressing these challenges by learning complex skills through environmental interaction without relying on explicit system models. Related work by UC Berkeley [10] demonstrated that hierarchical reinforcement learning (HRL) approaches can effectively achieve high-level kicking motion design for quadrupedal robot soccer players. However, quadrupedal robots benefit from superior static and dynamic stability due to their four-leg structure, making kicking motions easier to perform. In contrast to quadrupedal systems, bipedal robots face significantly greater challenges in soccer applications due to their inherently unstable two-legged locomotion and the complex balance requirements during dynamic kicking motions. Recent advances [11] [12] [13] have shown that deep reinforcement learning (DRL) can successfully tackle these stability challenges, demonstrating that bipedal robots can learn agile soccer skills including robust kicking, ball dribbling, and recovery behaviors through end-to-end RL training. Nevertheless, these promising results have yet to be validated on large-scale humanoid robots, where the increased complexity in dynamics, actuator characteristics, and physical constraints may pose additional challenges for RL-based skill acquisition.

In response to these challenges, we propose MAKP, an innovative framework for accurate ball kicking in humanoid robot soccer. This framework combines motion tracking with reinforcement learning, addressing complex ball interaction

This work was supported by the National Natural Science Foundation of China (Grant No. 62373242 and No. 92248303).

*These authors contributed equally.
[1]Shanghai Jiao Tong University, Shanghai, China
[2]Shanghai Innovation Institute, Shanghai, China
†Corresponding author. Email: yuegao@sjtu.edu.cn

problems through curriculum training strategies. We train the kicking policies in simulation and successfully transfer them to the physical humanoid robot Booster T1. Our contributions are summarized as follows:

1) We present a novel end-to-end whole-body control paradigm for humanoid robots that enables the execution of human-like kicking motions with accurate ball kicking capabilities.
2) We innovatively apply a diffusion-based motion generator and integrate it with the end-to-end learning framework, enhancing the diversity and naturalness of learned kicking behaviors
3) We achieve successful transfer in both sim-to-sim and sim-to-real scenarios. Extensive real-world experiments on the Booster T1 robot validate the effectiveness of our approach.

II. Related Works

A. Accurate Ball Kicking

Kicking accuracy is a fundamental challenge in robotic soccer, with recent research exploring end-to-end learning solutions. Pena et al. [14] proposed an omni-directional policy by optimizing parameters, but it targets small robots like NAO. Ji et al. [10] achieved accurate kicking for quadrupeds using hierarchical RL, yet without addressing bipedal single-leg balance. Currently, Kong et al. [15] introduced a progressive perception-action framework for humanoids, focusing on lightweight perception for generalization rather than explicitly balancing motion fidelity with kicking accuracy.

B. Learning-Based Whole-body Control

Traditional humanoid robot kicking methods often rely on pre-defined motion sequences [16] [17], which can achieve basic leg and foot movements but lack whole-body coordination. Recently, researchers have employed human motion data and learning-based methods to develop unified controllers for simulated characters capable of reproducing a wide variety of motions and performing diverse skills with human-like behavior [18] [19] [20]. Recent works [21] [22] [23] achieved stable, coordinated whole-body motion for humanoid robots through imitation learning from human motion data. However, these imitation learning methods focus solely on motion tracking and lack the ability to interact physically with dynamic objects, making it difficult to adjust kicking policies based on specific targets.

Unlike previous methods that concentrate on leg and foot movement sequences, our learning-based approach is specifically optimized for accurate interaction between humanoid robots and objects. By learning stable full-body coordination in Stage I and fine-tuning the kicking policy in Stages II and III, our controller ensures kicking accuracy while maintaining the naturalness of movements.

C. Motion Generation

The acquisition and retargeting of high-quality kicking motions pose significant challenges. Current motion acquisition approaches can be categorized into three main types: motion capture, video-based extraction, and generative model synthesis.

The AMASS (Archive of Motion Capture As Surface Shapes) dataset [24] contains 40 hours of motion capture data in SMPL [25] parameters, providing resources for kicking motions. Building upon AMASS, HumanML3D [26] incorporates language descriptions to create a large-scale dataset paired with text and motion data.

Video-based motion extraction significantly expands data sources, with VIBE [27] using temporal information and adversarial training to estimate temporally consistent SMPL parameters from video sequences. For soccer motions requiring large-scale displacement and global motion trajectories, GLAMR [28] estimates both local poses and world-coordinate trajectories. Luo et al. [18] propose a two-stage pipeline for human-to-humanoid motion retargeting, preserving key motion characteristics.

In the area of text-driven motion generation, the Motion Diffusion Model (MDM) [29] is the first to apply diffusion processes to human motion generation, producing high-quality motion sequences through iterative denoising. Our work employs the MDM to establish a flexible and scalable motion-generation pipeline, contrasting with motion capture by leveraging AMASS data for on-demand, diverse motion creation without dedicated capture sessions.

III. Method

The MAKP framework uses a three-stage training strategy to achieve accurate ball kicking for humanoid robot soccer players, as shown in Figure 2. First, the Motion Diffusion Model (MDM) generates diverse kicking motions based on task requirements, providing reference trajectories for handling tasks at various angles and distances. In Stage I, the robot tracks reference kicking motions. Stage II refines the robot's ability to kick the ball toward target positions. Stage III introduces a multi-critic architecture with MT-Critic and Kick-Critic modules, enabling adaptive control to balance motion similarity and accuracy. We will demonstrate why the multi-critic mechanism is necessary in section III-D and prove its effectiveness in the experimental section IV.

A. Problem Formulation

We formulate the humanoid robot soccer kicking task as a goal-conditioned reinforcement learning (RL) problem. Given the robot's proprioception p_t (including the last action $a_{t-1} \in \mathbb{R}^{23}$, root angular velocity $\omega_t^{root} \in \mathbb{R}^3$, joint positions $q_t \in \mathbb{R}^{23}$, joint velocity $\dot{q}_t \in \mathbb{R}^{23}$, and root projected gravity $g_t \in \mathbb{R}^3$), the ball's initial position in the robot base frame $b_0 \in \mathbb{R}^3$, and the target's initial position in the robot base frame $g_0 \in \mathbb{R}^3$, the objective is to learn a policy π that outputs actions $a_t \in \mathbb{R}^{23}$ at a frequency of 200 Hz. Additionally, the robot's observation includes a time phase variable $\phi \in [0, 1]$, where $\phi = 0$ represents the start of the motion and $\phi = 1$ represents the end [19]. The observation also includes a 4-step history of the robot's state p_t, defined as:

$$o_t \triangleq [p_{t-3:t}, b_0, g_0, \phi]. \tag{1}$$

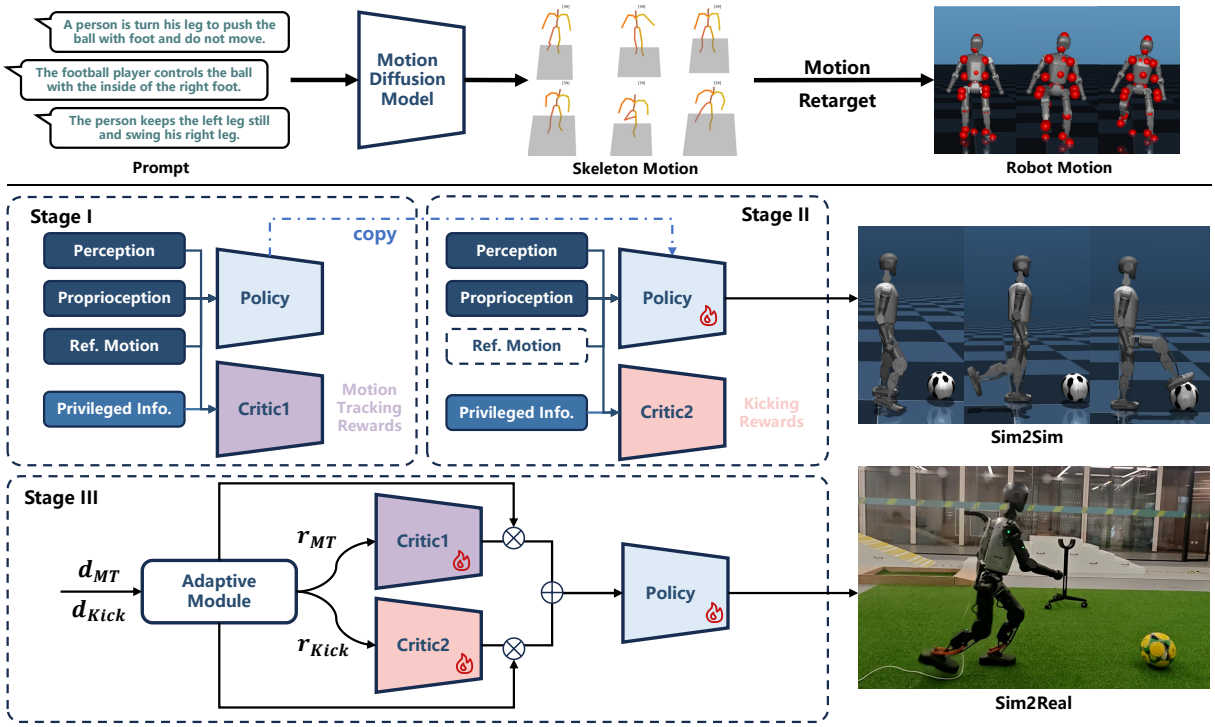

Fig. 2. **Overview of the MAKP framework.** Our approach consists of three main training stages: (1) **Stage I (Pre-train): Motion Tracking** – The robot learns to track reference kicking motions using proprioception, reference motion data, and imitation reward functions. (2) **Stage II (Fine-tune): Kicking Accuracy Optimization** – Building upon the motion tracking foundation, the robot learns to accurately kick the ball toward target positions with kicking task-specific reward functions, while preserving motion similarity through fine-tuning. (3) **Stage III (Post-train): Multi-Critic** – A multi-critic architecture with separate MT-Critic and Kick-Critic modules enables robust policy to learn through adaptive control, effectively balancing motion similarity and kicking accuracy. The framework supports both Sim2Sim validation and Sim2Real transfer to physical robots. Flame icons highlight the novel contribution within each block.

The action $a_t \in \mathbb{R}^{23}$ corresponds to the target joint positions and is passed to a PD controller that actuates the robot's degrees of freedom.

B. Motion Generation Using Motion Diffusion Model

We use the Motion Diffusion Model (MDM) [29] to generate reference motion sequences for human kicking, which are then rendered into 3D human meshes through the SMPL model [25] and retargeted [18] to the T1 robot. MDM learns the distribution of motion data to generate realistic human motions that satisfy physical constraints. For soccer applications, it generates kicking sequences based on task specifications such as ball and target positions and kicking style. Given conditional input, MDM generates motion sequences through a diffusion process [30]:

$$M = \{\theta_1, \theta_2, ..., \theta_T\} = MDM(C, noise), \tag{2}$$

where $\theta_t \in \mathbb{R}^{69}$ represents human joint angles at frame t, and C denotes the conditioning information guiding motion generation.

The joint angles $\{\theta_t\}_{t=1}^{T}$ generated by MDM are rendered into a 3D human mesh using the SMPL model:

$$(V_t, J_t) = SMPL(\theta_t, \beta), \quad t = 1, 2, \ldots, T, \tag{3}$$

where V_t and J_t denote the 3D mesh vertices and skeleton joints at time t, respectively. Finally, by mapping human joints to Booster T1, we optimize the robot's joint angle sequence using gradient descent:

$$\min L_{total} = L_{pos} + \lambda_1 L_{smo} + \lambda_2 L_{con} + \lambda_3 L_{vel}, \tag{4}$$

where L_{pos} is the position matching loss, L_{smo} is the smoothness loss, L_{con} is the constraint loss, and L_{vel} is the velocity constraint loss.

C. Two-stage Pre-train

To enable the robot to learn different motions for accurate kicking, we propose a progressive two-stage pre-training strategy. Stage I focuses on basic motion control, balance maintenance, and trajectory tracking, allowing the robot to track kicking motions while maintaining stability. Building upon Stage I, Stage II further optimizes kicking accuracy, where the robot learns to accurately kick the ball toward specific targets. This progressive approach reduces learning complexity while improving training efficiency and convergence stability.

1) Stage I Motion tracking: The goal of Stage I is to enable the robot to learn basic motion tracking and balance skills.

Termination Curriculum for Lower Body Tracking Tolerance: Training a policy to track kicking motions in simulation is challenging due to the trade-off between single-leg balance and leg swing. For instance, when imitating large-amplitude kicking motions, the policy often fails early

in training, opting for small-amplitude kicking to maintain balance and avoid falling penalties. To address this issue, we introduce a termination curriculum [31] defined by the following rule:

$$\tau_{\text{termination}}^{k+1} = \begin{cases} \tau_{\text{termination}}^{k} \cdot (1+\beta_1), & \text{if } \bar{L} < L_{\text{down}} \\ \tau_{\text{termination}}^{k} \cdot (1-\beta_1), & \text{if } \bar{L} > L_{\text{up}} \\ \tau_{\text{termination}}^{k}, & \text{otherwise,} \end{cases} \tag{5}$$

where $\bar{L}$ is the average episode length, and β_1 is the curriculum degree parameter. In early training, we set a generous termination threshold of 0.8 m, allowing the policy to prioritize learning basic single-leg stance balance without overly focusing on tracking similarity. As the robot improves its balance capabilities, we gradually tighten the threshold to 0.15 m, increasing the requirements for accurate kicking.

Reward Terms: We design the motion tracking reward r_{MT} by combining four categories of terms: 1) penalties for constraint violations, 2) regularization for smooth control, 3) high-level tracking task-specific rewards, and 4) low-level kicking task-specific rewards. The specific terms and their associated weights are detailed in Table I. The reward r_{kick} consists of four key components:

- *Hit target*: A sparse reward of 200.0 is assigned when the ball successfully hits the target, serving as a strong incentive to complete the kicking task:

$$r_{\text{hit}} = \begin{cases} 200.0, & \text{if the ball hits the target} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

- *Ball target distance*: This term encourages the robot to kick the ball toward the target. If the distance between the ball and the target is within the target hitting threshold, the reward is 1.0; otherwise, it decays exponentially with the distance, formulated as:

$$r_{\text{dist}} = \begin{cases} 1.0, & \text{if } d < d_{\text{threshold}} \\ \exp(-\alpha \cdot d), & \text{otherwise,} \end{cases} \tag{7}$$

  where d is the distance between the ball and the target, $d_{\text{threshold}}$ is the target hitting threshold, and α is the decay parameter.
- *Balance*: A reward of 1.0 is given for maintaining balance, which is essential for controlled kicking motions:

$$r_{\text{balance}} = \exp\left(-\sum_{i=1}^{2} g_{t,i}^2\right), \tag{8}$$

  where $g_{t,i}$ is the i-th component of the root projected gravity vector g_t.
- *Ball contact*: A reward of 1.0 is provided when the change in contact force on the ball exceeds a threshold, indicating an effective kicking interaction:

$$r_{\text{contact}} = \begin{cases} 1.0, & \text{if } \Delta F > F_{\text{threshold}} \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

  where ΔF is the change in contact force and $F_{\text{threshold}}$ is the contact force threshold.

Domain Randomizations: To improve the robustness of the pre-trained policy, we randomize the parameters [32] [33] as detailed in Table II.

2) Stage II Kicking Accuracy Optimization: Building upon Stage I, Stage II introduces high-level kicking task-specific rewards to optimize kicking accuracy.

Curriculum of Target Hitting Threshold and Target Generation Range: To gradually enhance kicking accuracy, we introduce a dual-parameter curriculum incorporating both target hitting threshold and target generation range according to the following rule:

$$\tau_{\text{hit}}^{k+1} = \begin{cases} \tau_{\text{hit}}^{k} \cdot (1-\beta_2), & \text{if } \bar{S} > S_{\text{up}} \\ \tau_{\text{hit}}^{k} \cdot (1+\beta_2), & \text{if } \bar{S} < S_{\text{down}} \\ \tau_{\text{hit}}^{k}, & \text{otherwise,} \end{cases} \tag{10}$$

$$r_{x/y/z}^{k+1} = \begin{cases} r_{x/y/z}^{k} \cdot (1+\beta_3), & \text{if } \bar{S} > S_{\text{up}} \\ r_{x/y/z}^{k} \cdot (1-\beta_3), & \text{if } \bar{S} < S_{\text{down}} \\ r_{x/y/z}^{k}, & \text{otherwise,} \end{cases} \tag{11}$$

where $\bar{S}$ is the average success rate, and β_2, β_3 are curriculum degree parameters. In early training, we set a generous target hitting threshold of 0.3m, where a kick is considered successful when the ball-to-target distance falls below this threshold. Meanwhile, target positions are constrained to a narrow lateral range of ± 0.4m along the y-axis in the robot's reference frame. As training progresses and kicking accuracy improves, we gradually reduce the target hitting threshold to 0.05m while expanding the target generation range to ± 0.8m, requiring accurate target hitting across a broader area.

Reward Terms: We decrease the weights of tracking task-specific rewards, and increase the weights of kicking task-specific rewards to guide the agent to achieve accurate ball-kicking toward the target, while maintaining balance and foot-ball interaction.

Domain Randomizations: For accurate kicking, we removed external force application and center-of-mass offset terms, while adding ball initial position perturbation and ball observation noise.

A detailed summary of reward terms and domain randomizations for Stages I and II is provided in Table I and II.

D. Stage III: Multi-Critic Mechanism

A single critic network evaluating all rewards tends to neglect components with smaller weights, leading to unbalanced policy learning due to improper reward weighting. This results in policies that fail to perform natural motions or achieve accurate kicking. To address this issue, we introduce a Multi-Critic mechanism [34], where multiple critics perform distributed evaluation, allowing for more effective estimation of rewards within each group. Additionally, we present an adaptive style control module that dynamically adjusts the balance based on the current motion tracking error d_{MT} and kicking error d_{kick}.

Error Metrics Definition: We define two fundamental error metrics that capture the robot's performance in different aspects:

$$d_{MT} = ||q_t - q_t^{expert}||_2 + ||\dot{q}_t - \dot{q}_t^{expert}||_2 \tag{12}$$

TABLE I

REWARD TERMS FOR TWO-STAGE PRE-TRAIN

Term	Equation	Stage I	Stage II
	Penalty		
DoF pos. limits	$\mathbb{I}$(limit violation)	-10.0	-10.0
DoF vel. limits	$\mathbb{I}$(limit violation)	-5.0	-5.0
Torque limits	$\mathbb{I}$(limit violation)	-5.0	-5.0
Termination	$\mathbb{I}$(termination)	-200.0	-200.0
	Regularization		
Torques	$\|\boldsymbol{\tau}\|^2$	$-1e^{-6}$	$-1e^{-6}$
Action rate	$\|\mathbf{a}_t - \mathbf{a}_{t-1}\|^2$	-0.5	-0.5
Feet orientation	$\|\boldsymbol{\phi}_{\text{feet}}\|^2$	-2.0	-2.0
Feet heading	$\|\theta_{\text{feet}} - \theta_{\text{root}}\|$	-0.1	-0.1
Slippage	slip^2	-1.0	-1.0
	Task Reward		
Body pos.	$\exp\{-\|\mathbf{p} - \mathbf{p}_{\text{ref}}\|^2\}$	1.0	0.5
Body pos. (feet)	$\exp\{-\|\mathbf{p}_{\text{ft}} - \mathbf{p}_{\text{ref}}\|^2\}$	2.1	1.0
Body ang. vel.	$\exp\{-\|\boldsymbol{\omega} - \boldsymbol{\omega}_{\text{ref}}\|^2\}$	0.5	0.5
DoF pos.	$\exp\{-\|\mathbf{q} - \mathbf{q}_{\text{ref}}\|^2\}$	0.75	0.75
VR 3-point	$\exp\{-\|\mathbf{p}_{\text{3pt}} - \mathbf{p}_{\text{ref}}\|^2\}$	1.6	1.0
Body rot.	$\exp\{-\|\mathbf{R} - \mathbf{R}_{\text{ref}}\|_F^2\}$	0.5	0.5
Body vel.	$\exp\{-\|\mathbf{v} - \mathbf{v}_{\text{ref}}\|^2\}$	0.5	0.5
DoF vel.	$\exp\{-\|\dot{\mathbf{q}} - \dot{\mathbf{q}}_{\text{ref}}\|^2\}$	0.5	0.5
	Kicking Task-Specific Reward		
Hit target	See Equation 6.	200.0	200.0
Ball target dist	See Equation 7.	1.0	5.0
Balance	See Equation 8.	0.5	2.0
Ball contact	See Equation 9.	0.5	1.0

TABLE II

DOMAIN RANDOMIZATION PARAMETERS FOR TWO-STAGE PRE-TRAIN

Parameter	Range	Stage I	Stage II
	Ball State Perturbation		
Initial Position	xy:$\mathcal{U}(-0.02, 0.02)$	✗	✓
Detection	$\sigma = 0.10$ m;	✗	✓
	Robot Dynamics Perturbation		
Torque RFI	$\mathcal{U}(0.5, 1.5)$	✓	✓
	Robot Physical Parameters		
Link Mass	$\mathcal{U}(0.8, 1.2)\times$ default	✓	✓
Base Mass	–	✗	✗
	Control System Parameters		
Proportional Gain	$\mathcal{U}(0.75, 1.25) \times K_p$	✓	✓
Derivative Gain	$\mathcal{U}(0.75, 1.25) \times K_d$	✓	✓
Control Delay	$[0, 2]$steps(Max 60 ms delay)	✓	✓
	Environment Parameters		
Friction Coefficient	$\mathcal{U}(0.5, 1.25)$	✓	✓
	Disabled Parameters (for Kicking Accuracy		
Push Robots	–	✓	✗
Base COM Offset	–	✓	✗

The motion tracking error d_{MT} quantifies the deviation between the robot's current joint configuration and the expert demonstration. Here, $q_t \in \mathbb{R}^{23}$ represents the robot's current joint positions, $q_t^{expert} \in \mathbb{R}^{23}$ denotes the corresponding expert joint positions from the reference trajectory, and $\dot{q}_t$, $\dot{q}_t^{expert} \in \mathbb{R}^{23}$ are the respective joint velocities. This metric captures both positional and velocity tracking accuracy, ensuring smooth motion execution.

$$d_{kick} = ||p_{target}^{ball} - p_{actual}^{ball}||_2. \tag{13}$$

The kicking error d_{kick} measures the Euclidean distance between the ball's target position $p_{target}^{ball} \in \mathbb{R}^3$ and its actual position $p_{actual}^{ball} \in \mathbb{R}^3$ after the kicking action. This metric directly reflects the precision of the kicking task.

Adaptive Weight Computation: To dynamically balance between motion similarity and kicking accuracy, we employ a softmax-based weighting scheme that adaptively adjusts the influence of each critic based on current performance:

$$w_t^{MT} = \frac{\exp(-\lambda_1 d_{MT})}{\exp(-\lambda_1 d_{MT}) + \exp(-\lambda_2 d_{kick})} \tag{14}$$

$$w_t^{kick} = \frac{\exp(-\lambda_2 d_{kick})}{\exp(-\lambda_1 d_{MT}) + \exp(-\lambda_2 d_{kick})}. \tag{15}$$

These equations implement a competitive softmax normalization scheme where the weights w_t^{MT} and w_t^{kick} represent the relative importance of motion tracking and kicking accuracy, respectively. The hyperparameters λ_1 and λ_2 control the sensitivity of each weight to its corresponding error metric.

Multi-Critic Value Function: The final value function combines the outputs of two specialized critic networks through the adaptive weights:

$$V_\phi(s_t) = w_t^{MT} \cdot V_{\phi_1}^{MT}(s_t) + w_t^{kick} \cdot V_{\phi_2}^{kick}(s_t). \tag{16}$$

Here, $V_{\phi_1}^{MT}(s_t)$ is a critic network specialized in evaluating motion tracking rewards r_{MT}, while $V_{\phi_2}^{kick}(s_t)$ focuses on kicking task-specific rewards r_{kick}. Each critic develops expertise in its respective domain, avoiding the interference and bias that occurs when a single critic attempts to balance conflicting objectives.

Training Strategy: During training, both critics are updated using temporal difference learning with their respective reward signals:

$$\mathcal{L}_{\phi_1} = \mathbb{E}[(V_{\phi_1}^{MT}(s_t) - r_{MT} - \gamma V_{\phi_1}^{MT}(s_{t+1}))^2] \tag{17}$$

$$\mathcal{L}_{\phi_2} = \mathbb{E}[(V_{\phi_2}^{kick}(s_t) - r_{kick} - \gamma V_{\phi_2}^{kick}(s_{t+1}))^2]. \tag{18}$$

This Multi-Critic architecture enables the policy to automatically prioritize motion tracking when kicking accuracy is satisfactory, and vice versa, leading to more balanced and robust learning outcomes. The adaptive weighting mechanism ensures that neither objective is completely neglected during training, addressing the challenge of multi-objective optimization [35].

IV. EXPERIMENTS

A. Experimental Setup

To evaluate the diversity and accuracy of the proposed method, we utilize MDM to generate three distinct kicking motions, **Kick-0 (a left-foot inside kick)**, **Kick-1 (a right-foot instep kick)**, **Kick-2 (a right-foot inside kick)**. These motions serve as expert demonstrations for our three-stage training framework, providing diverse kicking policies adaptable to different ball and target positions. We train our policy using the PPO algorithm [36] in IsaacGym, validate it in MuJoCo for sim-to-sim transfer, and finally deploy it on real-world Booster T1 for sim-to-real evaluation.

To assess the contributions of each component to kicking accuracy in our method, we performed an ablation study,

comparing our full three-stage method with the following baselines:

- **Stage I Motion Tracking Only (MT)**: Only tracks the kicking motions generated by MDM, without further fine-tuning using high-level kicking task-specific rewards.
- **Stage II Two-Stage Training w/o Multi-Critic (MT+Kick)**: A basic two-stage framework that adopts a traditional Single-Critic architecture.
- **Stage III Full Multi-Critic Method (Multic-Critic)**: The complete framework incorporating both two-stage pre-train and Multi-Critic optimization.

For fairness, all baseline comparisons used the same training hyper-parameters. All methods were trained using the same expert demonstration data generated by MDM for three kicking motions.

B. Metrics

We evaluated the methods from three key dimensions:

- **Kicking Accuracy (Acc.)**: A trial is considered successful if the position of the ball falls within a 0.2m radius of the specified target.
- **Tracking Error (Err.)**: The time-averaged sum of absolute differences between all joint angles and their reference trajectories:

$$\text{Err.} = \left(\sum_{t=1}^{T} \sum_{i=1}^{N} ||q_i(t) - q_i^{\text{ref}}(t)||_2 \right) / T, \qquad (19)$$

  where $q_i(t)$ is the actual angle of joint i at time t, $q_i^{\text{ref}}(t)$ is the reference angle from the expert demonstration, N is the total number of joints, and T is the total number of time steps.
- **Balance Stability (Fall Rate)**: The percentage of trials in which the robot loses balance and falls either during or immediately after the kicking process.

For each method, 42 kicking trials were conducted using the Kick-0, Kick-1 and Kick-2 motions.

C. Results And Analyses

Table III presents the quantitative evaluation results of different methods.

TABLE III
DIFFERENT METHODS ANALYSIS RESULTS

Motion	Methods	Acc.(%)	Err.	Fall(%)
Kick-0	MT	57.1	**1.13**	19.0
	MT+Kick	61.9	1.37	23.8
	Multi-Critic	**76.2**	**1.15**	**16.7**
Kick-1	MT	52.4	**1.19**	26.2
	MT+Kick	61.9	1.47	33.3
	Multi-Critic	**66.7**	**1.28**	**19.0**
Kick-2	MT	42.9	**0.89**	16.7
	MT+Kick	54.8	1.43	26.2
	Multi-Critic	**64.3**	**1.13**	**14.3**

In terms of Kicking Accuracy, the baseline MT method shows the poorest performance with 57.1% / 52.4% / 42.9%, but maintains smallest Tracking Error and relatively low Fall Rate across Kick-0 / Kick-1 / Kick-2 respectively, indicating it focuses primarily on motion imitation rather than kicking accuracy. After Stage II, MT+Kick demonstrates a significant improvement in Kicking Accuracy (61.9% / 61.9% / 54.8%), but with the largest Tracking Error (1.37 / 1.47 / 1.43) and the highest Fall Rate, which indicates that our task-specific rewards greatly encourage the robot to fine-tune its kicking direction while ignoring the similarity and balance. Therefore, simply increasing the weight of the kicking reward leads to overemphasizing task completion while neglecting motion quality. It's noteworthy that the Multi-Critic approach effectively compromises to achieve the highest Kicking Accuracy while maintaining relatively low Tracking Error and the lowest Fall Rate (16.7% / 19.0% / 14.3%). This balanced performance demonstrates that the multi-critic architecture effectively learns to coordinate between competing objectives without requiring explicit reward weight tuning.

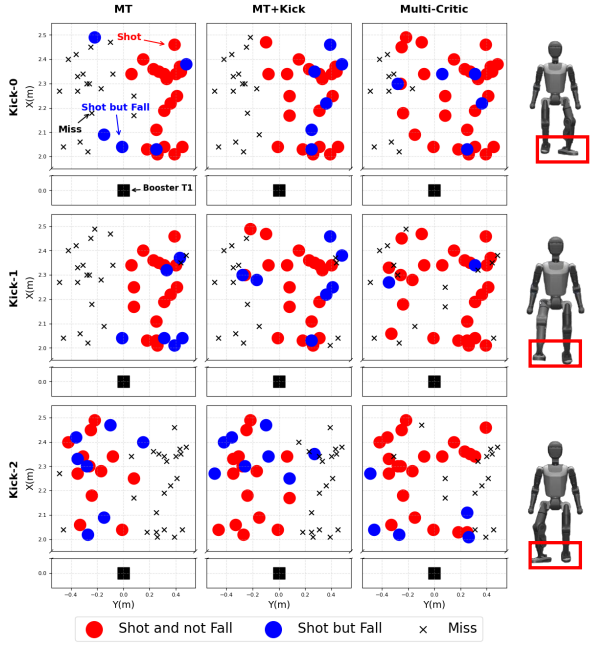

Fig. 3. **Kicking Results of Three Kicking Motions Under Different Methods.** Red dots represent successful kicks that hit the target; blue dots indicate that the kick hits the target, but the robot falls; gray crosses (×) denote kicks that miss the target, and black squares show the robot's initial position. Each subplot shows 42 kicking trials in the target area. MT shows conservative behavior with low accuracy but stable performance. MT+Kick improves accuracy (61.9% / 61.9% / 54.8%) but increases instability. Multi-Critic achieves highest accuracy while maintaining lowest fall rates (16.7% / 19.0% / 14.3%).

1) Kicking Accuracy: Figure 3 visualizes the kicking results for all three methods across different kicking motions. As shown, MT can only kick the ball in the direction of its natural leg movement. However, after Stages II and III, both MT+Kick and Multi-Critic methods can kick the ball almost anywhere within the target area, demonstrating much better control and flexibility.

2) Tracking Accuracy: To evaluate the tracking accuracy of different methods and motions over time steps, we performed an ablation study using the same baselines, with the

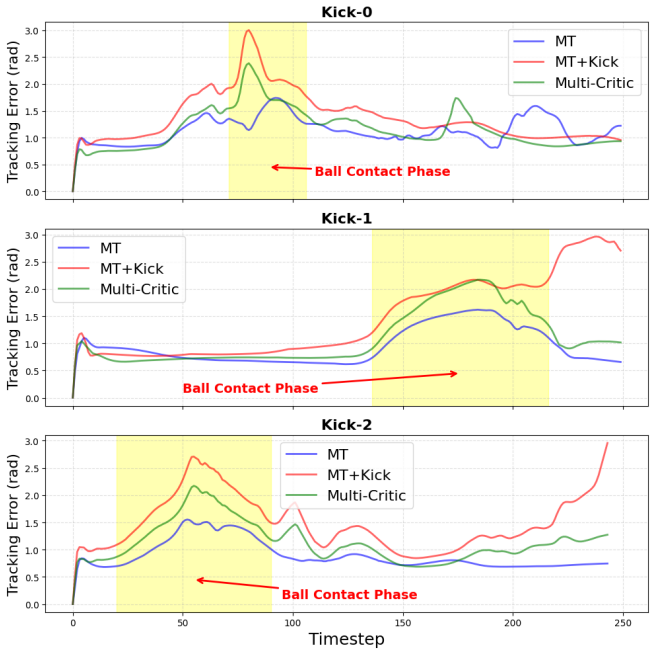

Fig. 4. **Temporal Analysis of Motion Tracking Performance During Soccer Kicking.** Comparison of tracking errors for MT (blue), MT+Kick (red), and Multi-Critic (green) methods across Kick-0, Kick-1, and Kick-2 motions, showing error peaks during the ball-contact phase and post-contact recovery performance.

evaluation metric being the only difference:

- **Tracking Error**: The sum of absolute differences between all joint angles and their reference trajectories over time steps:

$$e(t) = \sum_{i=1}^{N} ||q_i(t) - q_i^{\text{ref}}(t)||_2. \tag{20}$$

For each method, 5 kicking trials were conducted using the Kick-0, Kick-1 and Kick-2 motions with the same target position and the same initial ball position.

Figure 4 shows the average Tracking Error over time steps. As shown, all methods experience a peak during the ball contact phase when the robot deviates from the reference trajectory for accurate ball kicking. While Multi-Critic method (green) shows slightly higher tracking errors over time steps compared to MT method (blue), it demonstrates better overall stability and recovery performance compared to MT+Kick method (red), which is particularly evident in the post-contact stabilization phase, making it more suitable for robust and accurate kicking applications.

Figure 5 analyzes the motion differences among three different methods during the ball-contact phase. While the standard MT method can effectively track reference trajectories, the low weight assigned to kicking task-specific rewards results in potential failure to make contact with the ball during the kicking process. The MT+Kick method, by increasing the kicking task-specific rewards, leads to severe motion distortions. Although it significantly improves kicking accuracy, it may cause larger tracking errors relative

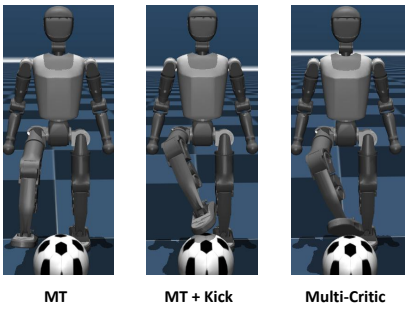

Fig. 5. **Comparison of Motion Control Strategies in Soccer-Kicking Tasks.** Left: MT method maintains motion similarity but fails ball contact. Middle: MT+Kick achieves ball contact but with motion distortion. Right: Multi-Critic balances motion similarity and kicking performance.

to the standard reference trajectory during the post-kick recovery phase, making the robot prone to falling. The multi-critic method effectively balances the tracking and the kicking task, achieving accurate ball kicking while maintaining minimal tracking errors.

D. Deployment

We deploy our MAKP framework on the physical Booster T1 humanoid robot platform. The robot stands 1.18 meters tall and features 23 degrees of freedom distributed across the upper body (10 DOF), waist (1 DOF), and lower body (12 DOF). The robot is equipped with a comprehensive sensor suite including an IMU and joint encoders for real-time state estimation, as well as a RealSense D455 Depth Camera for visual perception [37]. All onboard computations are performed using an NVIDIA AGX embedded computing platform, providing sufficient processing power for real-time policy execution at 200Hz control frequency.

V. Conclusions and Analysis

We present MAKP, a novel end-to-end framework for accurate kicking control in humanoid robots. By combining Motion Diffusion Model with a three-stage reinforcement learning strategy, MAKP achieves precise ball trajectory control while maintaining motion diversity.

Despite the significant progress achieved by MAKP, several limitations remain to be addressed in future work. First, the system currently focuses on accurate control of ground-rolling balls, while the control capabilities for aerial ball trajectories have not been fully explored, limiting the trajectory control scope. Second, the onboard camera only uses first-frame data for decision-making, which limits the system's dynamic adaptation capabilities to a certain extent, restricting visual decision-making data utilization. Finally, current kicking angle constraints (e.g., the left foot can only kick rightward) restrict the diversity of the system's kicking motions, constraining the overall motion flexibility.

Future research directions include extending control to spatial kicking with complex actions like chip shots and volleys, and extend our framework to a unified controller between multiple kicking motions.

REFERENCES

[1] Z. Su, B. Zhang, N. Rahmanian, Y. Gao, Q. Liao, C. Regan, K. Sreenath, and S. S. Sastry, "Hitter: A humanoid table tennis robot via hierarchical planning and learning," *arXiv preprint arXiv:2508.21043*, 2025.

[2] C. Lin, D. Affinita, M. E. Zimmatore, D. Nardi, D. D. Bloisi, and V. Suriani, "Self-supervised feature extraction for enhanced ball detection on soccer robots," *arXiv preprint arXiv:2506.16821*, 2025.

[3] F. Vahl, J. Griepenburg, J. Gutsche, J. Güldenstein, and J. Zhang, "Soccerdiffusion: Toward learning end-to-end humanoid robot soccer from gameplay recordings," *arXiv preprint arXiv:2504.20808*, 2025.

[4] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "Robocup: The robot world cup initiative," in *Proceedings of the first international conference on Autonomous agents*, 1997, pp. 340–347.

[5] P. Stone, *Intelligent autonomous robotics: A robot soccer case study*. Springer Nature, 2022.

[6] S. Behnke and J. Stückler, "Hierarchical reactive control for humanoid soccer robots," *International Journal of Humanoid Robotics*, vol. 5, no. 03, pp. 375–396, 2008.

[7] A. Carlos, E. Rajesh, C. Zhou *et al.*, "A modular architecture for humanoid soccer robots with distributed behavior control," *International Journal of Humanoid Robotics*, vol. 5, no. 3, pp. 397–416, 2008.

[8] M. Friedmann, J. Kiener, S. Petters, D. Thomas, O. Von Stryk, and H. Sakamoto, "Versatile, high-quality motions and behavior control of a humanoid soccer robot," *International Journal of Humanoid Robotics*, vol. 5, no. 03, pp. 417–436, 2008.

[9] A. C. S. Mota and M. R. Máximo, "Minimum time footstep planning for simulated robot soccer kicks using model predictive control," in *2024 Brazilian Symposium on Robotics (SBR), and 2024 Workshop on Robotics in Education (WRE)*. IEEE, 2024, pp. 97–102.

[10] Y. Ji, Z. Li, Y. Sun, X. B. Peng, S. Levine, G. Berseth, and K. Sreenath, "Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1479–1486.

[11] T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, J. Humplik, M. Wulfmeier, S. Tunyasuvunakool, N. Y. Siegel, R. Hafner *et al.*, "Learning agile soccer skills for a bipedal robot with deep reinforcement learning," *Science Robotics*, vol. 9, no. 89, p. eadi8022, 2024.

[12] B. Yan, H. Zhu, X. Li, and X. Li, "Reinforcement learning based soccer kicking for humanoid robots," in *International Conference on the Frontiers of Robotics and Software Engineering*. Springer, 2024, pp. 203–212.

[13] D. Tirumala, M. Wulfmeier, B. Moran, S. Huang, J. Humplik, G. Lever, T. Haarnoja, L. Hasenclever, A. Byravan, N. Batchelor *et al.*, "Learning robot soccer from egocentric vision with deep reinforcement learning," *arXiv preprint arXiv:2405.02425*, 2024.

[14] P. Pena, J. Masterjohn, and U. Visser, "An omni-directional kick engine for humanoid robots with parameter optimization," in *Robot World Cup*. Springer, 2017, pp. 385–397.

[15] J. Kong, X. Liu, Y. Lin, J. Han, S. Schwertfeger, C. Bai, and X. Li, "Learning soccer skills for humanoid robots: A progressive perception-action framework," 2026. [Online]. Available: https://arxiv.org/abs/2602.05310

[16] S. Feng, E. Whitman, X. Xinjilefu, and C. G. Atkeson, "Optimization-based full body control for the darpa robotics challenge," *Journal of field robotics*, vol. 32, no. 2, pp. 293–312, 2015.

[17] B. Hu, Z. Liang, X. Jiang, and Z. Liu, "Kicking motion design of humanoid robots using gradual accumulation learning method," in *2025 37th Chinese Control and Decision Conference (CCDC)*. IEEE, 2025, pp. 660–665.

[18] Z. Luo, J. Cao, K. Kitani, W. Xu *et al.*, "Perpetual humanoid control for real-time simulated avatars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 895–10 904.

[19] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions On Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[20] C. Tessler, Y. Guo, O. Nabati, G. Chechik, and X. B. Peng, "Maskedmimic: Unified physics-based character control through masked motion inpainting," *ACM Transactions on Graphics (TOG)*, vol. 43, no. 6, pp. 1–21, 2024.

[21] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan *et al.*, "Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills," *arXiv preprint arXiv:2502.01143*, 2025.

[22] Q. Liao, T. E. Truong, X. Huang, Y. Gao, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," *arXiv preprint arXiv:2508.08241*, 2025.

[23] Z. Chen, M. Ji, X. Cheng, X. Peng, X. B. Peng, and X. Wang, "Gmt: General motion tracking for humanoid whole-body control," *arXiv preprint arXiv:2506.14770*, 2025.

[24] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "Amass: Archive of motion capture as surface shapes," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5442–5451.

[25] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 851–866.

[26] L. Bensabath, M. Petrovich, and G. Varol, "A cross-dataset study for text-based 3d human motion retrieval," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1932–1940.

[27] M. Kocabas, N. Athanasiou, and M. J. Black, "Vibe: Video inference for human body pose and shape estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5253–5263.

[28] Y. Yuan, U. Iqbal, P. Molchanov, K. Kitani, and J. Kautz, "Glamr: Global occlusion-aware human mesh recovery with dynamic cameras," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 038–11 049.

[29] G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, and A. H. Bermano, "Human motion diffusion model," *arXiv preprint arXiv:2209.14916*, 2022.

[30] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[31] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[32] F. Muratore, C. Eilers, M. Gienger, and J. Peters, "Data-efficient domain randomization with bayesian optimization," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 911–918, 2021.

[33] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. 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)*. IEEE, 2017, pp. 23–30.

[34] J.-P. Sleiman, M. Mittal, and M. Hutter, "Guided reinforcement learning for robust multi-contact loco-manipulation," in *8th Annual Conference on Robot Learning (CoRL 2024)*, 2024.

[35] C. Liu, X. Xu, and D. Hu, "Multiobjective reinforcement learning: A comprehensive overview," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 3, pp. 385–398, 2014.

[36] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[37] R. Varghese and M. Sambath, "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International conference on advances in data engineering and intelligent computing systems (ADICS)*. IEEE, 2024, pp. 1–6.