

RoboNaldo: Accurate, Stable and Powerful Humanoid Soccer Shooting via Motion-Guided Curriculum Reinforcement Learning

Anonymous Author(s)

Affiliation

Address

email



Figure 1: **RoboNaldo: accurate, stable and powerful whole-body soccer shooting on a real grassy soccer field.** Ghost-overlaid time sequence of a Unitree G1 humanoid executing a directed penalty kick.

Abstract: Elite humanoid soccer shooting requires whole-body stability, high-impulse whole-body interactions, and accuracy to targets. Motion tracking-driven reinforcement learning (RL) provides stability in whole-body movement coordination, but a fixed reference makes it hard to adapt to varied ball positions and strike timings; in contrast, task reward-driven RL struggles to explore and discover valid kicks from scratch. We therefore introduce **RoboNaldo**, a three-stage motion-guided curriculum RL framework for high-impulse humanoid interaction. A single human-kick reference is used as a scaffold and progressively shifts optimization towards shooting performance. The curriculum first learns stable whole-body kicking prior, then adapts the kick to free-kick settings where the ball is stationary at random positions, and finally extends it to moving-ball shooting through a locomotion-command and kick-trigger interface. A high-level heuristic planner controls this interface during training, while alternative high-level controllers can drive the same low-level policy at inference. In simulation, RoboNaldo demonstrates free-kick shot error 50% lower and shoot velocity $2.96\times$ than prior work baselines. In real world on a Unitree G1 with onboard perception, RoboNaldo attains 0.73 m and 0.86 m average target shooting error from 3 m away in free-kick and moving-ball cases, accordingly. And the post-contact ball velocity reaches 13.10 m/s, which is 59–71% of reported professional open-play shot speed.

Keywords: Athletic robots, humanoid-object interaction, sim-to-real transfer

1 Introduction

Soccer shooting provides a compact but demanding benchmark for athletic humanoid interaction. A successful shot requires the robot to direct the ball towards a target while generating a high-velocity

25 strike and maintaining whole-body stability. This couples several challenging control problems,
 26 including single-leg balance, millisecond-scale foot–ball contact, and generalization across diverse
 27 ball and target configurations. Unlike sustained-contact tasks such as lifting or pushing, shooting
 28 relies on a brief, high-impulse collision generated by an end-effector that simultaneously contributes
 29 to whole-body support and balance.

30 Recent humanoid soccer systems [1, 2, 3, 4] have made rapid progress in locomotion, ball chasing,
 31 and goal-directed kicking. However, their shooting evaluations remain limited in accuracy, power,
 32 and deployment scope. For example, the ball direction cosine metric in [4] exceeds 90% even
 33 when the shot-on-goal error is 2.4 m from a 5 m shooting distance. Among prior systems, only [3]
 34 considers moving-ball kicks, but it does not evaluate shooting accuracy or kick power, and the system
 35 is not demonstrated in outdoor deployment. These limitations point to a broader methodological
 36 challenge: existing ingredients for humanoid kicking provide either motion-level stability or task-level
 37 objectives, but they do not directly resolve the contact selection and timing decisions required for
 38 accurate shooting. Motion priors make the robot stable enough to kick, but a fixed reference cannot
 39 choose a new contact point, aim direction, or strike time [5, 4, 3]. Task rewards specify where the
 40 ball should go, but pure RL must discover balance, swing coordination, contact, and aim from sparse
 41 delayed feedback; AMP controllers [6, 2] encourage human-like behavior but still leave when-to-kick
 42 and where-to-aim hard to supervise. As a result, prior work has not jointly demonstrated *stability*,
 43 *accuracy*, *power*, and *generalizability* for humanoid soccer shooting.

44 We present **RoboNaldo**, a learning framework for accurate, stable and high-impulse humanoid
 45 interactions like soccer shooting. The key idea is to stage the problem according to what each learning
 46 signal can reliably provide. In *Stage 1*, motion tracking learns a stable kicking prior that captures
 47 balance and swing structure. In *Stage 2*, shooting rewards and diverse free-kick configurations teach
 48 the policy to adapt its strike for target-directed accuracy. In *Stage 3*, locomotion-command/kick-
 49 trigger interface converts moving-ball shooting into approach control plus contact-timing decisions.
 50 After training, the Stage 3 policy can be driven by a heuristic controller or by other high-level
 51 commands, such as human commands or a co-trained high-level neural policy.

52 We validate RoboNaldo in both simulation and real-world experiments. In simulation, Stage 2
 53 free-kicks achieve 0.899 m average error from 5 m, 65.5% of the shot within 1 m error, and 14.79 m/s
 54 ball speed, giving $0.5\times$ the error and $2.96\times$ the speed of previous work; Stage 3 preserves high-power
 55 moving-ball shooting with 63.3% shots within 1 m error. On hardware, RoboNaldo runs fully onboard
 56 on a Unitree G1 with LiDAR-camera perception and performs directed shots on a real grassy soccer
 57 field, as shown in Fig. 1. It achieves 0.73 m average low-target free-kick error from 3 m, ball speeds
 58 up to 13.10 m/s, and 74% contact in human-passed moving-ball trials. These results show that
 59 the proposed curriculum learns a shooting policy that is stable, accurate, and powerful across the
 60 evaluated free-kick and moving-ball regimes. Our contributions can be summarized as follows:

- 61 • We demonstrate powerful, and accurate real-world humanoid soccer shots with generaliza-
 62 tion and stability across ball velocities and positions.
- 63 • We propose a staged curriculum for high-impulse humanoid interactions. Starting from
 64 learning a motion prior, then learns contact and timing adaptations, converting general
 65 moving ball shooting into a learnable problem.
- 66 • RoboNaldo achieves accurate and powerful shooting, attaining 0.73 m and 0.86 m mean
 67 shot error in free-kick and moving-ball cases; the ball speeds up to 13.10 m/s (47.2 km/h),
 68 reaching $\sim 71\%$ of female and $\sim 59\%$ of professional male open-play shot speed at the UEFA
 69 European Championships [7].

70 2 Related Work

71 **Humanoid soccer and the shooting gap.** RL has produced sports robots for quadruped soccer [8],
 72 table tennis [9], badminton [10], and egocentric humanoid kicking and goalkeeping [1, 11, 3, 2, 12, 4,
 73 13]. A useful shot demands stability, point-level accuracy, high post-contact speed, and generalization

across ball configurations. Tracking controllers inherit stable coordination but constrain strike timing; AMP relaxes hard reference constraints but weakly supervises aim; pure task RL struggles with sparse rewards in a large humanoid control space, leading to credit-assignment issues which drops learning performances. The closest concurrent system, PAiD [4], shares staged training and egocentric sensing but targets goal-region entry, not point-level placement; its cosine metric cannot resolve sub-metre accuracy differences. RoboNaldo targets point-level accuracy, high speed, moving-ball shooting, onboard sensing, and outdoor deployment jointly (Table 1).

Humanoid whole-body object interaction. Contact-rich loco-manipulation systems [14, 15, 16, 17, 18] assume dwell-time contact and accumulate interaction rewards across multiple steps. Soccer shooting instead has a sub-10 ms impulse and delayed ball–target feedback; RoboNaldo addresses this with an Instant Interaction Reward and a Densified Shooting Reward that extrapolates post-contact ball state.

Curriculum reinforcement learning. Curriculum RL ramps task difficulty [19, 20] and is common in legged locomotion [21]. For soccer shooting, increasing ball-position diversity before the policy can kick yields no reward signal; RoboNaldo co-designs reward weights and task difficulty across three behavior-triggered stages.

Table 1: Comparison of different humanoid soccer systems. We compare target-point accuracy (Point-acc), reported ball speed (Speed), moving-ball shooting (Moving), egocentric sensing (Ego), and outdoor demonstration (Out). † denotes concurrent work.

Method	Shooting capability			Deployment	
	Point-acc	Speed	Moving	Ego	Out
STOFT [11]	×	×	×	×	×
Reactive [2]	×	×	×	✓	✓
Striker [12]	×	×	×	✓	×
HumanX [†] [3]	×	×	✓	×	×
PAiD [†] [4]	×	×	✓	✓	✓
RoboNaldo(Ours)	✓	✓	✓	✓	✓

3 Method

RoboNaldo trains a unified policy for locomotion and soccer shooting using a **three-stage task curriculum** (Fig. 2). We begin by formulating the task and specifying the observation space (Sec.3.1). We then introduce the three-stage curriculum that progressively develops shooting skills (Sec.3.2). Finally, we present the reward design that encourages accurate, powerful, and stable shots (Sec. 3.3).

3.1 Task Formulation

Soccer shooting task. The task is specified by the ball state and a static target point on the goal line. We express ball and target observations in the robot frame for policy input, and evaluate shot accuracy in the world frame. Free-kick shooting refers to the stationary-ball setting, where the ball starts from rest at a sampled location and the robot must strike it toward the target. Moving-ball shooting extends this setting by assigning the ball an incoming velocity and randomized arrival timing; the robot must therefore both reach a feasible contact configuration and trigger the kick at the appropriate time. Let $p_b(t)$ denote the ball position after contact and let p_t denote the target point. A trial is successful if $\min_{t \geq t_c} \|p_b(t) - p_t\|_2 \leq d_{\text{thresh}}$, where t_c is the first valid foot–ball contact. This post-contact definition ensures that accuracy is measured only after the robot has launched or redirected the ball.

Policy observation. The policy receives motion-reference cues, proprioceptive history, previous actions, and ball/target observations. In Stages 1 and 2, the reference-anchor cue specifies the motion-reference anchor used for tracking. In Stage 3, this anchor cue is replaced by the locomotion command used by the moving-ball planner; the kick-trigger signal determines when the reference switches from locomotion to kicking. The full per-dimension observation vector is given in Appendix.

3.2 Curriculum Design

RoboNaldo uses a three-stage curriculum to separate the main difficulties of humanoid shooting: acquiring a stable whole-body kick, adapting the strike to different stationary-ball configurations, and

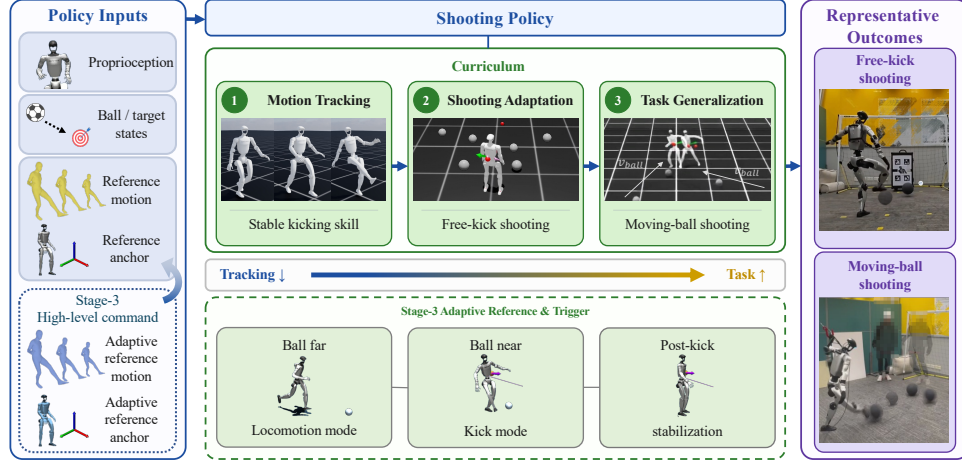


Figure 2: **RoboNaldo motion-guided curriculum.** A single low-level policy is trained sequentially across three stages. Stage 1 learns a stable kick by tracking a retargeted human motion. Stage 2 introduces shooting task rewards to adapt the kick for accurate stationary-ball free-kicks. Stage 3 replaces the reference motion and anchor cue with high-level locomotion and kick-trigger decision commands, enabling the policy to approach, trigger, strike, and stabilize when shooting moving balls.

118 timing the kick for moving balls. The stages are trained sequentially, with each stage initialized from
 119 the previous checkpoint.

120 **Stage 1: Shooting Motion Tracking.** The reference is a side-foot kick retargeted from human video
 121 via GVHMR [22] and GMR [23]. Following BeyondMimic [24], the policy imitates this reference
 122 without ball or task rewards, establishing coordination before aim/contact adaptation. Stages advance
 123 after task reward plateaus, action-noise standard deviation converges, and behavior emerges.

124 **Stage 2: Shooting Adaptation.** Stage 2 introduces the ball, the target, and the shooting rewards
 125 while randomizing the ball spawn position around the nominal contact location. The policy must
 126 therefore adapt the demonstrated kick rather than simply replay it: the approach, contact point, strike
 127 direction, and impact speed all depend on the sampled ball and target. The resulting checkpoint
 128 serves as the free-kick policy and provides the initialization for moving-ball training.

129 **Stage 3: Task Generalization.** Moving-ball shooting adds a temporal-alignment challenge. At
 130 realistic pass speeds, useful foot–ball contact occurs within only a few control steps, while the motion
 131 reference still imposes a nominal swing phase. Directly continuing Stage 2 under fast incoming-
 132 ball randomization provides little reliable contact signal. Stage 3 therefore makes timing explicit
 133 by introducing a high-level locomotion command and a kick trigger, dividing each episode into
 134 locomotion, kick, and post-kick stabilization modes.

135 **Heuristic planner.** Before contact, the policy locomotes toward the ball using discounted motion-
 136 tracking rewards, goal-directed velocity, and gait-quality terms. The planner steers the anchor toward
 137 the predicted ball position $\tilde{\mathbf{p}}b = \mathbf{p}b, \perp + \mathbf{v}b, \perp T_h$ while aligning the robot with the ball direction. It
 138 triggers the kick when the estimated closest body-frame approach distance $d_{\min}$ to the nominal
 139 foot-strike offset satisfies $(d_{\min} < r_{\text{thr}}) \wedge (z_b < h_{\max}) \wedge (\phi < \phi_{\text{kick}})$, meaning that the ball is
 140 reachable, kickable, and the kick phase has not started. The motion clock is then snapped to ϕ_{kick} ,
 141 with small training-time phase jitter for robustness, and the policy switches to full-weight kick
 142 tracking. After contact, a post-kick stabilization reference is held for T_{stab} steps before resetting the
 143 clock, discouraging unstable impacts that score high shot rewards but fail immediately after the kick.

144 **High-level control interface.** The locomotion command and kick-trigger signal form a modular
 145 interface between the high-level planner and the low-level shooting policy. The rule-based planner
 146 used during training can therefore be replaced at deployment by another high-level controller without
 147 retraining the low-level policy.

3.3 Reward Design

The per-step reward is $r = r_{\text{motion}}w_{\text{motion}} + r_{\text{task}}w_{\text{task}} + r_{\text{reg}}w_{\text{reg}}$, the complete formulas of each reward and stage-wise weights are shown in Appendix.

Motion tracking reward. r_{motion} follows BeyondMimic [24]: exponential-kernel tracking on 14 body parts, with extra foot-velocity weight for contact sensitivity.

Regularization reward. r_{reg} penalizes action rate, smoothness, foot slip, airborne time, joint velocities/torques, and joint-limit violations (Table B.2 and Section B).

Soccer shooting task reward. r_{task} covers foot/CoM approach, ball speed, contact orientation, the Instant Interaction Reward (Eq. (1)), and a densified shot-placement reward. For shot placement, we combine a sparse hit signal $r_{\text{sparse goal}} = \mathbf{1}[\hat{d} \leq d_{\text{threshold}}]$ with a dense distance signal $r_{\text{dense goal}} = \exp(-\hat{d}^2/\sigma_{\text{goal}}^2)$, where $\hat{d} = \min_{t' \leq t} \|\mathbf{p}_{\text{ball}} - \mathbf{p}_{\text{target}}\|$, with $d_{\text{threshold}} = 0.5$ m in Stage 2 and 0.35 m in Stage 3. To provide earlier aiming feedback than the episode-minimum distance alone, we also ballistically extrapolate the post-contact ball state to its goal-line crossing and reward the predicted target error at each post-contact step, thereby densifying the long-horizon placement objective without changing the evaluation metric. Stage 3 additionally activates velocity-tracking and feet air-time locomotion terms.

Instant interaction reward. HDMI [14] uses a position–force product, which is effective for sustained contact but collapses for soccer shooting: meaningful foot–ball contact lasts only 3–5 physics steps at 200 Hz. RoboNaldo instead rewards the full contact lifecycle—approach, impact, and post-contact ball outcome:

$$r_{\text{interact}} = (R_{\text{contact}} + R_{\text{goal}})(R_{\text{vel}} + R_{\text{force}})/4, \quad (1)$$

where R_{contact} shapes approach, R_{force} and R_{vel} demand high-impulse impact, and R_{goal} preserves result-level aiming gradients. The additive groups prevent reward collapse around the brief contact window while still penalizing weak, slow, or misdirected strikes.

4 Simulation Experiments

Setup and metrics. We train in Isaac Lab [25] with 4096 parallel environments on an NVIDIA RTX 4090 and evaluate on 16,384 held-out episodes. Targets are sampled on an $8 \text{ m} \times 2 \text{ m}$ goal plane 5 m ahead. In *free-kick*, a stationary ball is sampled in a $1 \text{ m} \times 1 \text{ m}$ frontal square; in *moving-ball shooting*, a ball is launched at 0–5 m/s toward a random point in a $2 \text{ m} \times 2 \text{ m}$ frontal square. Reported metrics are alive rate, shot error (mean per-episode nearest ball-target distance), success rates (error $< 0.5/1.0$ m), peak ball speed, and ball-contact rates.

Experimental results. We evaluate RoboNaldo against naive and existing methods in free-kick and moving-ball regimes. Pure PPO/AMP rarely make stable directed contact; pure motion tracking remains a replayed kick. Stage 2 learns accurate, powerful free-kicks but does not zero-shot to moving balls, while Stage 3 generalizes to both regimes. Degradation concentrates at extreme lateral and high targets, as visualized in the target-conditioned heatmaps in Fig. E.1.

Ablations. We ablate the staged curriculum, Stage 3 mechanism, and interaction reward in Table 2. Each stage addresses a distinct failure mode: without Stage 1, task RL lacks a whole-body scaffold and struggles to survive; without Stage 2, the policy lacks a reliable free-kick interaction primitive and converges to a poor local optimum which is only tracking the motions; without the Stage 3 planner, the kick timing fails to adapt; all leading to drops in performances. Removing adaptive sampling or stabilization phase severely disrupts the robustness and stability completing the shooting task, whose alive rate drops from 98.8% to 32.8% and 24.4% accordingly, though while maintaining

Table 2: **Simulation results and ablations.** Shot err. is nearest ball–target distance; suc. denotes success rates with shot error less than 0.5/1.0 m; v_{ball}^{max} is peak post-contact ball speed. Free-kick uses a stationary ball; moving-ball shooting redirects a ground ball launched at 0–5 m/s. Contact denotes the ratio of envs where the robot contacts with ball. Ablated variants warm-start from the corresponding RoboNaldo stage checkpoint. Nominal policies and metrics are bold.

Entry	Regime	0.5m suc. (%) $\uparrow$	1.0m suc. (%) $\uparrow$	Shot err. (m) $\downarrow$	v_{ball}^{max} (m/s) $\uparrow$	Contact (%) $\uparrow$	Alive (%) $\uparrow$
<i>Main comparison: free-kick</i>							
PPO [26]	Free-kick	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0	4.721 $\pm$ 0.007	0.780 $\pm$ 0.003	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0
AMP [6]	Free-kick	0.9 $\pm$ 0.1	3.3 $\pm$ 0.1	3.733 $\pm$ 0.004	1.771 $\pm$ 0.051	38.6 $\pm$ 2.8	41.6 $\pm$ 2.3
PAiD [4]	Free-kick	8.2 $\pm$ 0.7	19.5 $\pm$ 0.5	1.850 $\pm$ 0.026	4.986 $\pm$ 0.017	67.7 $\pm$ 0.3	67.3 $\pm$ 0.5
Stage 1 motion tracking	Free-kick	2.6 $\pm$ 0.1	9.2 $\pm$ 0.3	2.835 $\pm$ 0.008	1.353 $\pm$ 0.006	27.9 $\pm$ 0.6	99.9 $\pm$ 0.1
RoboNaldo stage 2	Free-kick	28.8 $\pm$ 1.3	65.5 $\pm$ 0.6	0.899 $\pm$ 0.012	14.792 $\pm$ 0.011	82.1 $\pm$ 0.3	100.0 $\pm$ 0.0
RoboNaldo stage 3	Free-kick	36.3 $\pm$ 0.1	66.5 $\pm$ 0.2	1.055 $\pm$ 0.012	14.178 $\pm$ 0.034	81.1 $\pm$ 0.1	99.7 $\pm$ 0.1
<i>Main comparison: moving-ball shooting</i>							
PPO [26]	Moving	0.6 $\pm$ 0.1	2.3 $\pm$ 0.1	3.962 $\pm$ 0.004	4.251 $\pm$ 0.013	0.5 $\pm$ 0.1	0.0 $\pm$ 0.0
AMP [6]	Moving	0.9 $\pm$ 0.1	3.5 $\pm$ 0.1	3.575 $\pm$ 0.015	4.334 $\pm$ 0.012	12.5 $\pm$ 0.2	15.7 $\pm$ 0.3
PAiD [4]	Moving	3.5 $\pm$ 0.4	9.3 $\pm$ 0.6	2.149 $\pm$ 0.034	5.030 $\pm$ 0.005	50.6 $\pm$ 0.6	48.1 $\pm$ 0.5
Stage 2 zero-shot	Moving	1.9 $\pm$ 0.1	7.2 $\pm$ 0.1	3.159 $\pm$ 0.003	4.250 $\pm$ 0.009	15.3 $\pm$ 0.4	37.4 $\pm$ 0.2
RoboNaldo stage 3	Moving	32.4 $\pm$ 0.6	63.3 $\pm$ 0.6	1.131 $\pm$ 0.004	13.875 $\pm$ 0.004	79.2 $\pm$ 0.1	98.8 $\pm$ 0.1
<i>Ablations: curriculum</i>							
Stage 0–2	Free-kick	0.0 $\pm$ 0.0	0.3 $\pm$ 0.1	4.302 $\pm$ 0.016	4.586 $\pm$ 0.210	18.8 $\pm$ 0.5	1.2 $\pm$ 0.1
Stage 1–3	Moving	1.7 $\pm$ 0.2	6.4 $\pm$ 0.2	3.453 $\pm$ 0.033	4.196 $\pm$ 0.005	13.9 $\pm$ 0.8	65.3 $\pm$ 0.5
Stage 2–3 w/o planner	Moving	0.9 $\pm$ 0.3	3.9 $\pm$ 0.3	3.283 $\pm$ 0.024	4.710 $\pm$ 0.006	19.1 $\pm$ 1.1	18.9 $\pm$ 1.0
<i>Ablations: mechanism</i>							
W/o adaptive sampling	Moving	4.5 $\pm$ 0.2	12.6 $\pm$ 0.3	3.398 $\pm$ 0.011	12.084 $\pm$ 0.064	28.9 $\pm$ 0.2	32.8 $\pm$ 2.0
W/o stabilization	Moving	15.3 $\pm$ 0.2	25.6 $\pm$ 0.3	3.107 $\pm$ 0.036	12.967 $\pm$ 0.023	34.1 $\pm$ 1.0	24.4 $\pm$ 1.2
<i>Ablations: interaction reward</i>							
HDMI-style reward	Free-kick	2.5 $\pm$ 0.1	8.9 $\pm$ 0.2	2.865 $\pm$ 0.001	2.715 $\pm$ 0.012	45.0 $\pm$ 0.4	98.7 $\pm$ 0.2

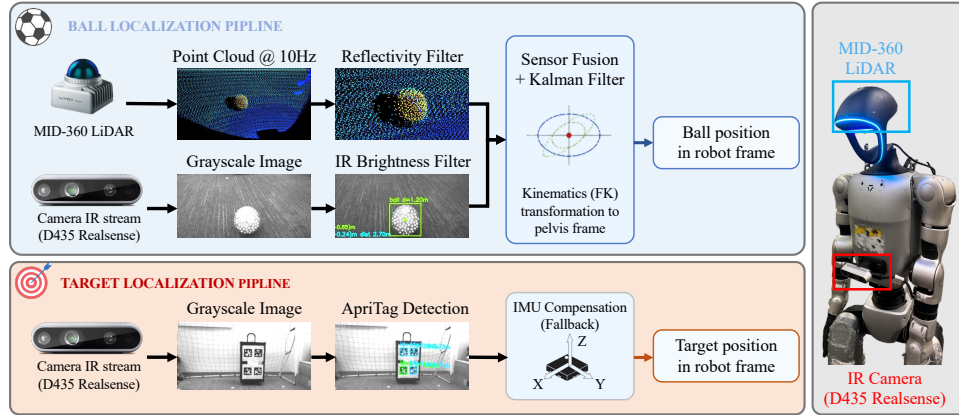


Figure 3: **Real-world perception stack.** A head-mounted MID-360 LiDAR tracks the retro-reflective ball at close range, while a chest-mounted IR camera extends ball detection beyond the LiDAR near-field. Both estimates are fused by a constant-velocity Kalman filter and transformed into the pelvis frame for policy input. The target board is localized by AprilTags from the same chest camera.

190 high shooting power. Replacing our instant interaction reward with HDMI-style sparser ones drops
191 contact accuracy performance, with free-kick success from 28.8% to 2.5%.

192 5 Real-World Experiments

193 5.1 Hardware Setup and Perception

194 Hardware experiments use a Unitree G1 [27] (29 DOF, 35 kg) and a size-5 soccer ball on indoor floors
195 and outdoor football fields. The ONNX policy runs onboard at 50 Hz over DDS, with no motion-
196 capture or offboard state estimation. A head-mounted Livox MID-360 and chest-mounted RealSense
197 D435 localize the ball and AprilTag target (Fig. 3), so the trials exercise the full perception–control
198 loop rather than a privileged-state replay. The same deployed controller is used for indoor calibration,
199 point-target free-kicks, human-passed moving balls, and outdoor football-field demonstrations.

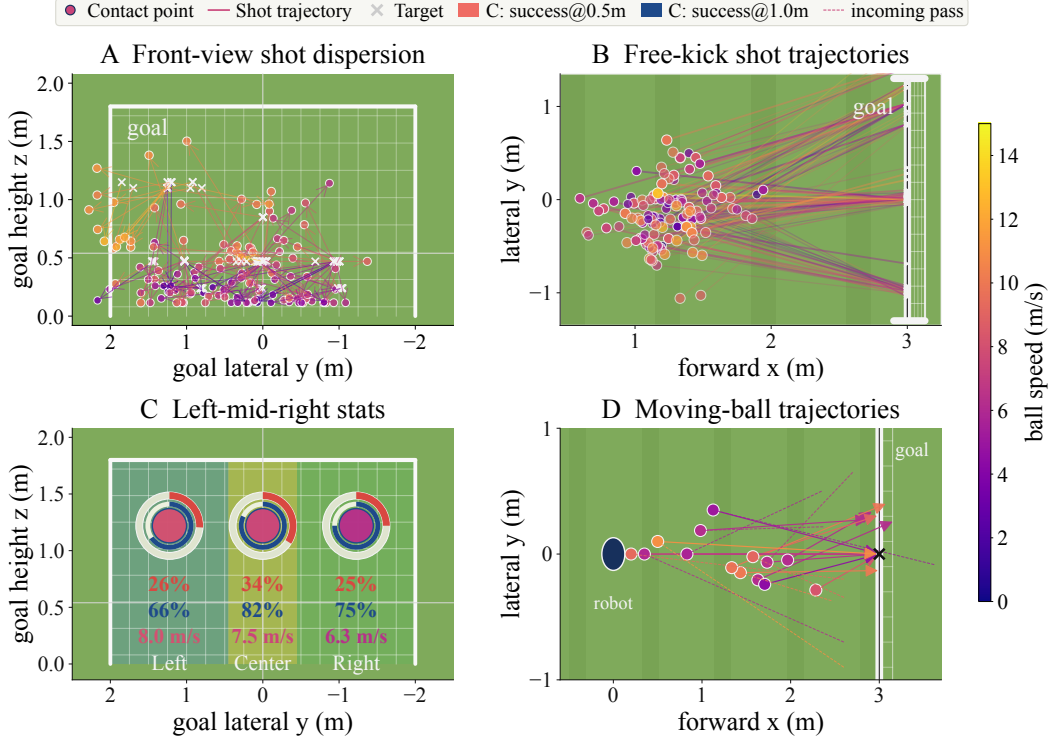


Figure 4: **Real-robot shooting overview.** A: front-goal view shot dispersion; B: free-kick shot trajectories; C: per-region accuracy success rates and speed; D: moving-ball shot trajectories.

200 5.2 Experimental Results

201 We report two trial counts in Table 3: n_{total} counts all recorded attempts, while n_{valid} counts attempts
 202 that successfully launch the ball and therefore have a measurable trajectory. Accuracy metrics,
 203 including the 0.5 m and 1.0 m success rates, shot error, and ball speed, are computed over n_{valid}
 204 launches. Contact% and Alive% are reported over n_{total} attempts to separately characterize end-to-end
 205 reliability.

206 RoboNaldo transfers to hardware with reliable free-kick performance across target locations. Out of
 207 136 recorded free-kick attempts, the robot launches 124 valid shots, corresponding to 91.2% contact,
 208 and remains alive in all trials. Conditioned on valid launches, 31.5% and 80.6% of shots land within
 209 0.5 m and 1.0 m of the target, respectively, with a mean radial error of 0.73 m and a mean ball speed of
 210 7.42 m/s. The left, center, and right target breakdown shows that the policy has a natural right-footed
 211 bias: shots towards left-side has larger speed; center targets are the most accurate, while off-center
 212 targets remain close in 1.0 m success, despite the increased task difficulty.

213 Moving-ball trials expose a different bottleneck. Human passes introduce timing variation and
 214 off-reference approach states, reducing contact reliability, but the kick remains accurate once a clean
 215 launch is produced. Among 27 moving-ball attempts, the robot launches 20 valid shots; conditioned
 216 on these launches, 30.0% and 70.0% land within 0.5 m and 1.0 m of the target, with 0.86 m mean shot
 217 error at 7.10 m/s. We observe in real-world experiment that the main degradation in the moving-ball
 218 setting comes from the sim-to-real gap of kick-timing trigger and imperfect incoming passes from
 219 human.

220 Compared to professional soccer players, RoboNaldo reaches 13.10 m/s peak ball speed, 59–71%
 221 of reported professional open-play shot velocities [7], indicating high-impulse contact rather than
 222 conservative low-velocity strikes.

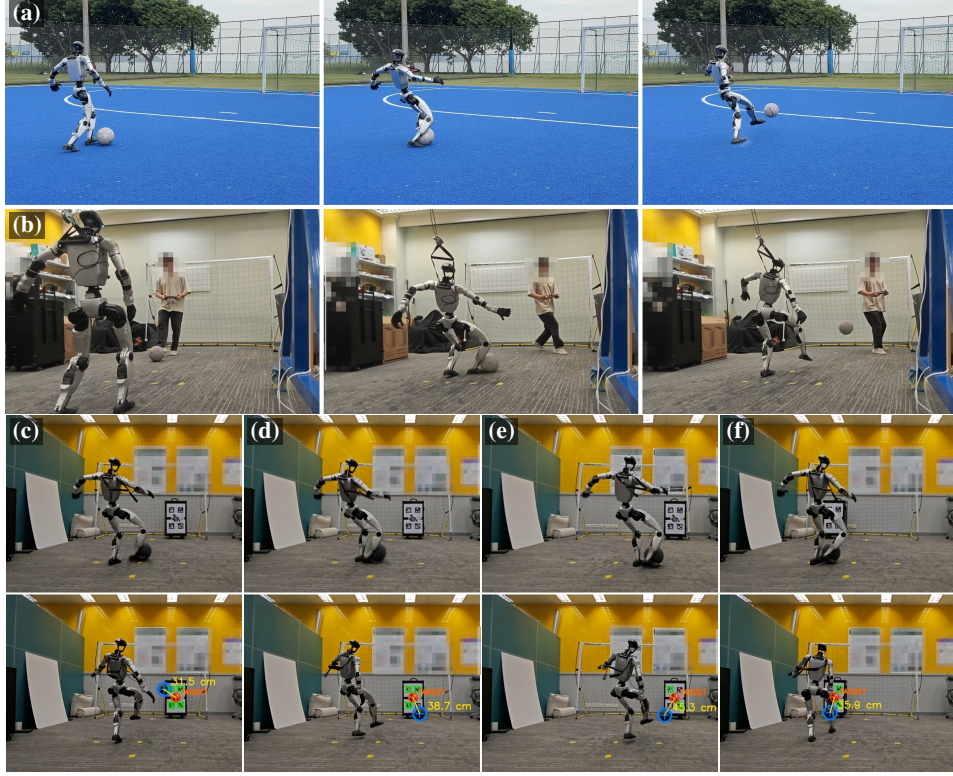


Figure 5: **Real-robot shooting showcase.** (a) outdoor free-kick and (b) moving-ball one-touch shooting are time-lapses read left-to-right. (c)–(f) are four free-kick trials read top-to-bottom (contact frame above, landing outcome below), with landing errors of 31.5, 38.7, 45.3, and 35.9 cm.

Table 3: **Real-world shooting results.** Metrics follow Table 2; moving-ball targets are random.

	n_{total}	n_{valid}	0.5m suc. (%) $\uparrow$	1.0m suc. (%) $\uparrow$	Shot err. (m) $\downarrow$	v_{ball} (m/s) $\uparrow$	Contact (%) $\uparrow$	Alive (%) $\uparrow$
<i>Free-kick</i>								
Overall	136	124	31.5	80.6	0.73 ± 0.39	7.42 ± 2.62	91.2	100.0
Left	58	49	30.6	77.6	0.74 ± 0.33	7.97 ± 2.98	84.5	100.0
Center	50	48	35.4	85.4	0.65 ± 0.34	7.47 ± 2.30	96.0	100.0
Right	28	27	25.9	77.8	0.85 ± 0.52	6.35 ± 2.06	96.4	100.0
<i>Moving-ball</i>	27	20	30.0	70.0	0.86 ± 0.47	7.10 ± 2.40	74.1	88.9

6 Discussion and Conclusion

We presented RoboNaldo, a motion-guided curriculum RL framework for stable, accurate, and high-impulse humanoid interactions. RoboNaldo learns and demonstrates perceptive, powerful and accurate free-kicks and interactive moving-ball shots on a physical Unitree G1 with generalizability and stability. In real-world experiments, RoboNaldo achieves 0.73 m average target error from 3 m, ball speeds up to 13.10 m/s, and 74% contact in human-passed moving-ball trials. These results suggest that motion-guided curriculum learning is effective for high-impulse humanoid interaction tasks requiring balance, aim, power, and contact timing.

Limitations and future work. RoboNaldo currently relies on a single reference kick motion and a handcrafted high-level trigger policy for stage transitions and kick timing. This limits its ability to handle richer soccer scenarios that require multiple skills and autonomous decision making. Future work will investigate multi-skill motion priors and train high-level policies for skill selection, transition and timing in more complex cases. The perception module is also tailored to a retro-reflective ball; extending it to active egocentric vision to perceive natural ball is an important next step.

References

- [1] 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*, 9(89):eadi8022, 2024.
- [2] Y. Wang, C. Luo, P. Chen, J. Liu, W. Sun, T. Guo, K. Yang, B. Hu, Y. Zhang, and M. Zhao. Learning vision-driven reactive soccer skills for humanoid robots. *arXiv preprint arXiv:2511.03996*, 2025.
- [3] Y. Wang, Q. Zhao, Y. F. Lau, R. Yu, H. W. Tsui, Q. Chen, J. Wang, J. Pang, and P. Tan. HumanX: Toward agile and generalizable humanoid interaction skills from human videos. *arXiv preprint arXiv:2602.02473*, 2026.
- [4] 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. *arXiv preprint arXiv:2602.05310*, 2026.
- [5] 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)*, 37(4):1–14, 2018.
- [6] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (TOG)*, 40(4):1–20, 2021.
- [7] B. De-la Cruz-Torres, A. Ruiz-de Alarcón-Quintero, and M. Navarro-Castro. The role of shot velocity in advanced post-shot metrics: Evidence from the UEFA european football championships. *Data*, 11(2):39, 2026. doi:10.3390/data11020039.
- [8] Z. Su, Y. Gao, E. Lukas, Y. Li, J. Cai, F. Tulbah, F. Gao, C. Yu, Z. Li, Y. Wu, and K. Sreenath. Toward real-world cooperative and competitive soccer with quadrupedal robot teams. In *Conference on Robot Learning (CoRL)*, 2025. arXiv:2505.13834.
- [9] 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.
- [10] Y. Ma, A. Cramariuc, F. Farshidian, and M. Hutter. Learning coordinated badminton skills for legged manipulators. *Science Robotics*, 10:eadu3922, 2025.
- [11] W. Li, J. Ma, M. Lu, and P. Lu. Like playing a video game: Spatial-temporal optimization of foot trajectories for controlled football kicking in bipedal robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3565–3572, 2025. arXiv:2510.01843.
- [12] Z. Xu, M. Seo, D. Lee, H. Fu, J. Hu, J. Cui, Y. Jiang, Z. Wang, A. Brund, J. Biswas, and P. Stone. Learning agile striker skills for humanoid soccer robots from noisy sensory input. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2026. arXiv:2512.06571.
- [13] J. Ren, J. Long, T. Huang, H. Wang, Z. Wang, F. Jia, W. Zhang, J. Wang, P. Luo, and J. Pang. Humanoid goalkeeper: Learning from position conditioned task-motion constraints. *arXiv preprint arXiv:2510.18002*, 2025.
- [14] H. Weng, Y. Li, N. Sobanbabu, Z. Wang, Z. Luo, T. He, D. Ramanan, and G. Shi. HDMI: Learning interactive humanoid whole-body control from human videos. *arXiv preprint arXiv:2509.16757*, 2025.
- [15] S. Zhao, Y. Ze, Y. Wang, C. K. Liu, P. Abbeel, G. Shi, and R. Duan. Resmimic: From general motion tracking to humanoid whole-body loco-manipulation via residual learning. *arXiv preprint arXiv:2510.05070*, 2025.

- [16] D. Kalaria, S. S. Harithas, P. Katara, S. Kwak, S. Bhagat, S. Sastry, S. Sridhar, S. Vemprala, A. Kapoor, and J. C.-K. Huang. DreamControl: Human-inspired whole-body humanoid control for scene interaction via guided diffusion. *arXiv preprint arXiv:2509.14353*, 2025.
- [17] Y. Zhang, Y. Yuan, P. Gurunath, I. Gupta, S. Omidshafiei, A.-a. Agha-mohammadi, M. Vazquez-Chanlatte, L. Pedersen, T. He, and G. Shi. FALCON: Learning force-adaptive humanoid loco-manipulation. In *Learning for Dynamics and Control Conference (LADC)*, 2026. arXiv:2505.06776.
- [18] S. Xu, H. Y. Ling, Y.-X. Wang, and L.-Y. Gui. InterMimic: Towards universal whole-body control for physics-based human-object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. arXiv:2502.20390, Highlight Paper.
- [19] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 41–48, 2009.
- [20] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181):1–50, 2020.
- [21] N. Rudin, D. Hoeller, P. Reist, and M. Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In *Conference on Robot Learning*, pages 91–100, 2022.
- [22] Z. Shen, H. Pi, Y. Xia, Z. Cen, S. Peng, Z. Hu, H. Bao, R. Hu, and X. Zhou. World-grounded human motion recovery via gravity-view coordinates. In *SIGGRAPH Asia 2024 Conference Papers*, 2024. arXiv:2409.06662.
- [23] J. P. Araújo, Y. Ze, P. Xu, J. Wu, and C. K. Liu. Retargeting matters: General motion retargeting for humanoid motion tracking. *arXiv preprint arXiv:2510.02252*, 2025.
- [24] 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.
- [25] M. Mittal, C. Yu, Q. Yu, J. Liu, N. Rudin, D. Hoeller, J. L. Yuan, R. Singh, Y. Guo, H. Mazhar, et al. Orbit: A unified simulation framework for interactive robot learning environments. In *IEEE Robotics and Automation Letters*, volume 8, pages 3740–3747, 2023.
- [26] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. In *arXiv preprint arXiv:1707.06347*, 2017.
- [27] Unitree Robotics. Unitree g1 humanoid robot. <https://www.unitree.com/g1>, 2024.
- [28] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. In *Advances in Neural Information Processing Systems*, 2021.
- [29] 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 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, 2017.
- [30] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3803–3810, 2018.
- [31] G. Jocher and J. Qiu. Ultralytics YOLO11. <https://docs.ultralytics.com/models/yolo11/>, 2024.
- [32] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5026–5033, 2012.

328 A Observation Space: Per-Dimension Breakdown

Table A.1: Policy observation vector (547 dims). [†]In Stage 3 the anchor reference is replaced by a 9-dim heuristic locomotion command (target body velocity/heading), keeping the total at 547. The critic additionally receives privileged, noise-free ground-truth body positions/orientations/velocities.

Group	Component	Dim.
Motion reference	reference joint positions, $\mathbf{q}_{\text{ref}}$ (29 DOF)	29
	reference joint velocities, $\dot{\mathbf{q}}_{\text{ref}}$ (29 DOF)	29
Anchor reference [†]	anchor (torso) position in body frame, $\mathbf{p}_{\text{anc}}$	3
	anchor orientation (6D rotation), $\mathbf{R}_{\text{anc}}$	6
Proprioception ($\times 5$ history)	base angular velocity	15
	joint positions (29 DOF)	145
	joint velocities (29 DOF)	145
	previous actions (29 DOF)	145
Exteroception ($\times 5$ history)	ball position in robot frame	15
	target position in robot frame	15
Total		547

329 B Reward Terms and Weights

330 Table B.1 lists every reward term. Motion-tracking terms use an exponential kernel $\exp(-\|\mathbf{e}\|^2/\sigma^2)$
 331 on the indicated error $\mathbf{e}$ with the listed base standard deviation σ_0 ; in Stages 2–3 the effective standard
 332 deviation is σ_0/α with $\alpha=2$, and in Stage 3 each motion-tracking term is additionally multiplied by
 333 the proximity-based relaxation factor $g_i(d_{\text{ball}})$ (Table B.3). Task terms are scaled by the goal weight
 334 w_g (0 in Stage 1, 1 thereafter); regularization terms are scaled by w_r (0.2 in Stages 1–2, 0.25 in
 335 Stage 3). The 14 tracked bodies are the pelvis, the bilateral hip-roll, knee, and ankle-roll links, the
 336 torso, and the bilateral shoulder-roll, elbow, and wrist-yaw links.

337 **Instant Interaction Reward factors.** The compact reward in Eq. (1) uses

$$\begin{aligned}
 R_{\text{contact}} &= \exp\left(-\frac{[\|\mathbf{p}_{\text{foot}} - \mathbf{p}_{\text{ball}}\| - r_b]_+}{\sigma_f}\right), & R_{\text{goal}} &= \exp\left(-\frac{[\hat{d} - \frac{\sigma_g}{2}]_+^2}{(2\sigma_g)^2}\right), \\
 R_{\text{vel}} &= 1 - \exp\left(-\frac{[\|\mathbf{v}_{\text{ball}}\| - v_{\text{min}}]_+^2}{10}\right), & R_{\text{force}} &= \min\left(\frac{f}{f_{\text{thresh}}}, 1\right),
 \end{aligned} \tag{2}$$

338 where $\hat{d} = \min_{t' \leq t} d_{\text{ball-target}}(t')$ and $[\cdot]_+ = \max(\cdot, 0)$.

339 **Densified Shooting Reward.** Shot placement is fully determined at the moment of impact, yet the
 340 episode-minimum ball–target distance reward yields signal only when the ball eventually reaches the
 341 goal line. To close this credit-assignment gap, at every post-contact step we ballistically extrapolate
 342 the ball’s current velocity under gravity to its goal-line crossing:

$$\begin{aligned}
 t^* &= \frac{p_{t,y} - p_{b,y}}{v_{b,y} + \epsilon}, & \hat{\mathbf{p}} &= \mathbf{p}_b + \mathbf{v}_b t^* - \frac{1}{2}g(t^*)^2\hat{\mathbf{z}}, \\
 r_{\text{densified}} &= \begin{cases} \exp(-\|\hat{\mathbf{p}} - \mathbf{p}_t\|^2/\sigma^2) & t^* > 0 \\ 0 & \text{else,} \end{cases}
 \end{aligned} \tag{3}$$

343 with $\hat{p}_z$ clamped to $\geq$ the ball radius. The condition $t^* > 0$ activates the reward only when the ball
 344 moves toward the goal; once the ball crosses the goal line the reward switches to the actual landing
 345 position.

Table B.1: Reward terms. Motion-tracking terms use an exp. kernel $\exp(-\|\mathbf{e}\|^2/\sigma^2)$; σ_0 is the base std, with $\sigma=\sigma_0/\alpha$ ($\alpha=2$) in Stages 2–3 and a proximity-based relaxation factor $g_i(d_{\text{ball}})$ in Stage 3 (Table B.3). w_g : goal weight (0/1); w_r : regularization weight (0.2/0.25); per-stage scale shown as St. 1–2/St. 3.

Term	Quantity / form	Weight
<i>Motion tracking</i>		
anchor position ($\sigma_0=0.3$ m)	torso position vs. ref.	0.1
anchor orientation ($\sigma_0=0.4$ rad)	torso orientation vs. ref.	0.5
body positions ($\sigma_0=0.3$ m)	14 relative body positions	1.0
body orientations ($\sigma_0=0.4$ rad)	14 body orientations	1.0
body linear vel. ($\sigma_0=1.0$ m/s)	14 body linear velocities	1.0
body angular vel. ($\sigma_0=\pi$ rad/s)	14 body angular velocities	1.0
feet linear vel. ($\sigma_0=1.0$ m/s)	2 feet linear velocities	1.0
<i>Task ($\times w_g$)</i>		
error_ball_to_target	exp; $\min_t \ \mathbf{p}_b - \mathbf{p}_t\ $ ($\sigma=0.5/\alpha$); = 1 within thr.	$5w_g$
predicted_error_ball_to_target	exp; $\ \hat{\mathbf{p}} - \mathbf{p}_t\ $ (Eq. 3, $\sigma=1/\alpha$)	$10w_g$
goal_reward_burst	constant for 10 steps on success (delay 100)	$300w_g$
robot_ball_contact	ankle- $\mathbf{p}_{\text{exec}}$ proximity + contact force, buffered	$2w_g$
contact orientation	ankle velocity vs. ball→target	$1w_g$
ankle-ball / torso-ball distance	d_{ball} / torso-ball dist.	$1w_g$ / $1.5w_g$
ball_velocity	sat. Lorentzian on $\ \mathbf{v}_b\ $ post-contact	$0.5w_g$
alive	constant while upright	$0.5w_g$
<i>Regularization ($\times w_r$)</i>		
action rate	$-\ \Delta \mathbf{a}\ ^2$	-0.1 (1.0/1.5)
action smoothness	$-\ \Delta^2 \mathbf{a}\ ^2$	-10^{-5} (1.0/2.0)
feet slip	$-\ \mathbf{v}_{\text{foot}}\ ^2$ in stance	-10^{-2}
no-fly	$\neg [\text{both feet airborne}]$	-0.2 (1.0/1.25)
joint velocity	$-\ \dot{\mathbf{q}}\ ^2$	-10^{-5}
joint torque	$-\ \boldsymbol{\tau}\ ^2$	-2×10^{-7}
joint limit (always active)	$\neg [\text{violation}]$	-10 (1.0/1.5)
undesired contacts	contact off ankles/wrists	-0.1
feet contact time	smooth foot-contact duration	-0.05 (1.0/1.2)

Table B.2: **Three-stage curriculum.** Ball spawns relative to (1.0, −0.2, 0.115) m; [†]kick/loco mode. Full observation vector in Table A.1.

	S1	S2	S3
	Tracking	Adapt	Generalize
w_{task}	0.0	0.8	1.0
w_{reg}	0.0	0.2	0.25
w_{motion}	1.0	1.0	1.0/0.1 [†]
Ball spawn XY (m)	—	± 0.75	± 1.0
Ball vel. XY (m/s)	0	0.1 m/s	5 m/s
EEF Δp_z			
term. (m) [24]	0.25	0.5	0.5

346 **Proximity-based tracking relaxation.** In Stage 3, each motion-tracking term is multiplied by a
347 distance-conditioned relaxation factor:

$$g_i(d) = \mu_i + (1 - \mu_i) \cdot \text{clamp}\left(\frac{d - d_{\text{near}}}{d_{\text{far}} - d_{\text{near}}}, 0, 1\right), \quad d_{\text{near}} = 0.35 \text{ m}, \quad d_{\text{far}} = 1.0 \text{ m}, \quad (4)$$

348 with per-term near-ball scales μ_i given in Table B.3. Feet linear velocity is suppressed to $\mu=0.05$,
349 giving the policy near-total freedom in ankle placement during contact; body orientation is only
350 reduced to $\mu=0.6$, preserving balance during the lean. Unlike globally zeroing a tracking weight,
351 this design retains full reference pressure during approach and relaxes only the terms where the task
352 objective genuinely conflicts with motion tracking.

Table B.3: **Proximity-based tracking relaxation scales** μ_i (Stage 3, $d_{\text{ball}} \leq d_{\text{near}} = 0.35$ m).

Term	anchor pos.	anchor ori.	body pos.	body ori.	body lin. vel.	body ang. vel.	feet lin. vel.
μ_i	0.10	0.60	0.20	0.60	0.20	0.30	0.05

C Hyperparameters and Domain Randomization

PPO (RSL-RL). We optimize the policy with PPO [26] using the massively parallel RSL-RL implementation [21]. Clip ratio 0.2; entropy coefficient 0.001; value-loss coefficient 1.0 with clipped value loss; 5 learning epochs and 4 minibatches per update; 24 environment steps per rollout; adaptive learning rate (initial 2×10^{-5}) with KL target 0.01; discount $\gamma=0.99$; GAE $\lambda=0.95$; gradient-norm clip 1.0; up to 10^5 iterations per stage with checkpoints every 10^3 . Actor and critic are 3-layer MLPs (512→256→128, ELU) with empirical observation normalization; initial action noise std 1.0. The critic uses privileged, noise-free observations; the actor uses the noised observation of Appendix A.

Heuristic planner. Prediction horizon $T_h=0.4$ s; proximity radius $r_{\text{thr}}=0.25$ m; maximum kickable height $h_{\text{max}}=0.75$ m; locomotion-kick boundary frame $\tau_{\text{kick}}=240$; adaptive sampling window ± 10 frames; foot-strike offset $\mathbf{p}_{\text{offset}}=(0.15, -0.1)$ m. The desired anchor command is $\mathbf{p}_{\text{des}} = \mathbf{p}_{\text{anc},\perp} + \frac{1}{2} \min(\|\tilde{\mathbf{p}}_b - \mathbf{p}_{\text{anc},\perp}\|, 2) \hat{\mathbf{d}}_b$ and $\mathbf{q}_{\text{des}} = \text{norm}([1 + \mathbf{e}_x \cdot \hat{\mathbf{d}}_b, \mathbf{e}_x \times \hat{\mathbf{d}}_b])$, where $\tilde{\mathbf{p}}_b = \mathbf{p}_{b,\perp} + \mathbf{v}_{b,\perp} T_h$. The predicted closest-approach distance is $d_{\text{min}} = \|\Delta \mathbf{p} + t^* \mathbf{v}_{b,\perp}^{\text{body}}\|$, with $\Delta \mathbf{p} = \mathbf{p}_{b,\perp}^{\text{body}} - \mathbf{p}_{\text{offset}}$ and $t^* = \text{clamp}(-\Delta \mathbf{p} \cdot \mathbf{v}_{b,\perp}^{\text{body}} / \|\mathbf{v}_{b,\perp}^{\text{body}}\|^2, 0, T_h)$.

Simulation. 4096 parallel environments on a 10×10 m grid in GPU-accelerated simulation [28]; PhysX GPU solver with 8 position and 4 velocity iterations; physics step 5 ms (200 Hz) with $4 \times$ decimation (50 Hz control); episodes of 10 s (500 steps). The ball is a rigid sphere (radius 0.115 m, mass ≈ 0.41 kg, restitution 0.95) with contact sensors; the active ankle-roll link has a contact sensor for kick detection.

Domain randomization. Following standard sim-to-real randomization practice [29, 30], per episode: robot collision material: static friction $\mathcal{U}[0.3, 1.6]$, dynamic friction $\mathcal{U}[0.3, 1.2]$, restitution $\mathcal{U}[0, 0.5]$; joint default-position offsets ± 0.01 rad; torso center-of-mass offset ± 0.025 m (fore-aft) and ± 0.05 m (lateral/vertical); actuator command delay of a few control steps. At 1–3 s intervals a random base push is applied: linear velocity ± 0.5 m/s (horizontal), ± 0.2 m/s (vertical); angular velocity ± 0.52 rad/s (roll/pitch), ± 0.78 rad/s (yaw). Ball spawn position and (in Stage 3) initial velocity are randomized as in Table B.2.

D Deployment Details

Table D.1: **Perception modules.** *Fast ball*: robust to 15 m/s motion blur. *Onboard*: no external infra.

Module	Range	Lat. (ms)	Fast ball	Onboard
MoCap (Vicon)	room	<5	✓	×
RGB (D435 color)	6 m	30–60	×	✓
Depth (D435 depth)	3 m	30–60	–	✓
LiDAR (MID-360)	40 m	10–20	✓	✓

Perception. *Ball (LiDAR, near range).* A Livox MID-360 is rigidly mounted on the head with a downward pitch so that its dense returns cover the floor from a few decimetres to about two metres in front of the robot. Each ~ 10 Hz point cloud is filtered by (i) reflectivity (the retro-reflective soccer ball returns are far brighter than surroundings), (ii) a distance gate, and (iii) a region of interest; a sphere of known radius is fit (least-squares with outlier rejection) to the surviving points, the center is smoothed by a constant-velocity Kalman filter, transformed to the pelvis frame, and published

on `rt/ball_state`. *Ball (IR+grayscale camera, far range)*. Beyond the LiDAR coverage we localize the ball from the chest-mounted RealSense D435 *infrared* (grayscale) stream rather than the RGB image: the retro-reflective ball is a bright blob in the IR image, segmented by intensity thresholding plus a known-radius circularity check, then back-projected and fused into the same Kalman filter as the LiDAR estimate. Keying on retro-reflective brightness avoids the failure modes of the appearance-based detectors we also tried on the D435 RGB stream—a YOLO11m [31] TensorRT detector and an HSV+MOG2 background-subtraction detector—namely HSV color segmentation being easily disturbed by lighting and background, and the YOLO detector losing the ball under fast-motion blur. Robust long-range tracking of a fast moving ball remains part of the planned extension to ball-stopping and interception. *Target (camera)*. An AprilTag detector on the RealSense D435 localizes the goal board; alternatively a fixed coordinate is used. Detections older than 300 ms are flagged invalid.

Joint-index remapping. Isaac Lab enumerates the 29 DOF in breadth-first order over the kinematic tree, whereas the Unitree SDK and the MuJoCo [32] model use depth-first order. A fixed permutation (and its inverse) is applied when packing the observation and when mapping policy actions to motor commands. Getting this permutation wrong silently produces a plausible-looking but non-functional policy, so we list it explicitly in the released code.

Initialization and control loop. On entry to the policy state, the five-step observation history buffers are zero-filled and a 10-step linear warmup blends the robot’s current joint configuration $\mathbf{q}_0$ into the first policy output $\mathbf{a}_1$: $\mathbf{q}_{\text{target}}^{(t)} = (1 - t/10) \mathbf{q}_0 + (t/10) \mathbf{a}_1$ for $t = 1, \dots, 10$. The control loop runs at 50 Hz: it reads `rt/lowstate` (joint state and IMU, published at 500 Hz by the firmware) and `rt/ball_state/rt/target_state`, runs ONNX inference, and publishes `rt/lowcmd` (per-joint position target with PD gains, CRC-checked). The RoboNaldo policy is one state of an on-robot finite-state machine; an immediate damping (passive) mode is bound to a dedicated button for safety.

E Additional Simulation Results

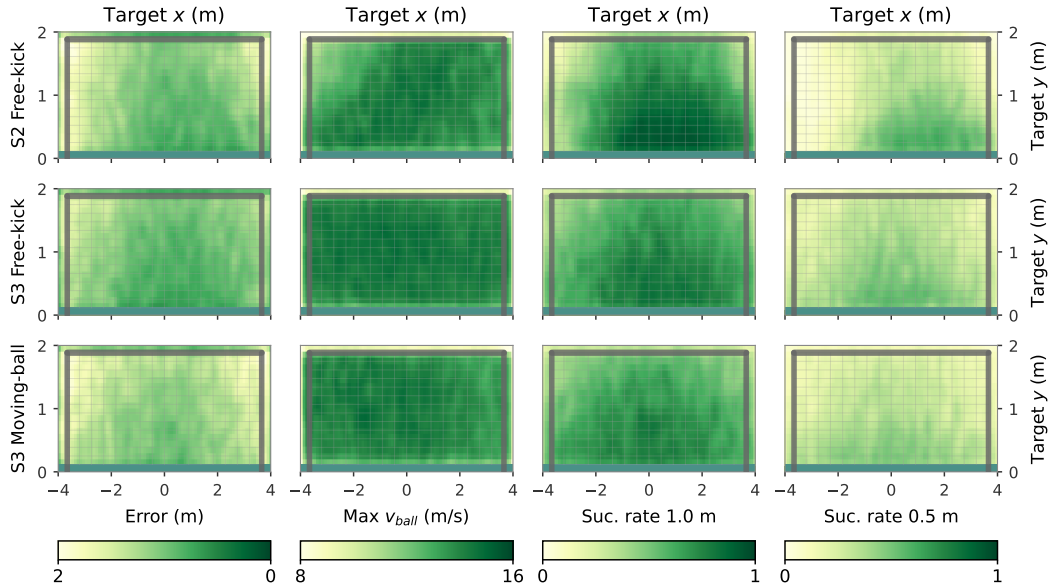


Figure E.1: **Target-conditioned shooting heatmaps.** Rows correspond to Stage 2 free-kick, Stage 3 free-kick, and Stage 3 moving-ball shooting. Each heatmap visualizes one shooting metric over goal-plane target locations: shot error, peak ball speed, 1.0 m success rate, or 0.5 m success rate.

F Additional Results

We plan the following diagnostics to further characterize the system beyond the main simulation and hardware evaluations:

- **LiDAR ball-localization accuracy.** Position error vs. a motion-capture or tape-measure ground truth across the working range, and its effect on kick accuracy.
- **Proximity relaxation scale sensitivity.** Performance vs. a uniform scaling of all near-ball scales μ_i and vs. the $(d_{\text{near}}, d_{\text{far}})$ window, to show robustness to these choices.
- **Whole-body coordination.** Joint-angle trajectories over a canonical kick cycle, showing the swing-leg hip/knee/ankle chain coupled with arm counter-rotation; comparison against the motion reference.
- **Extended hardware analysis.** Per-target failure-mode breakdown (perception failure / contact miss / balance loss) with Wilson confidence intervals, and additional kick-sequence montages including failure cases.
- **Weight-only ablation.** Stage-3 training with the ball-spawn range held fixed and only w_g varied, to isolate the goal weight from the exploration-difficulty schedule.