

ESE 6510: Physical Intelligence Drone Racing Strategy Design

Kartik Virmani, Dhyey Shah

Abstract

In this report we present our complete drone racing strategy for the Isaac-Quadcopter-Race-v0 environment, integrating a custom refined reward structure, observation design and reset distribution. We implemented a gate-aware progress reward, velocity-based shaping, stability penalties, and a linear lap-time objective that strongly encourages aggressive but reliable traversal of the race course. A multi-frame observation space—combining world, body, and gate-relative coordinates—provided the policy with rich geometric context. We further applied domain randomization over thrust-to-weight ratio, aerodynamic coefficients, and PID gains to improve sim-to-real stability. Together, these components yielded a fast, stable racing policy capable of consistent multi-lap behavior under significant dynamics variation under a 5k iteration training and resulting in a race time of 20.5 secs for 3 laps in our observation.

1 Reward Function Design

Our approach enables the quadcopter to complete multiple laps around a fixed sequence of racing gates with high speed and robustness. We tried multiple strategy parameters but our final technical contributions included a strengthened gate-passing detector, progress-based reward shaping, and a straight-line speed bonus that encourages aggressive but stable flight behavior.

2 Reward Function Design

The reward function combines multiple components to guide the quadcopter toward optimal racing behavior:

2.1 Gate Passing Reward

Gate detection uses the gate-local coordinate frame. The drone's position relative to gate i is:

$$\begin{aligned}\mathbf{p}_t^g &= (x_t^g, y_t^g, z_t^g)^\top \\ &= \text{subtract_frame_transforms}(\mathbf{w}_{i_t}, \mathbf{q}_{i_t}^{\text{gate}}, \mathbf{p}_t)\end{aligned}$$

A successful gate pass occurs when:

$$\begin{aligned}x_t^{g,\text{prev}} &> 0 \quad (\text{was behind gate}) \\ x_t^g &< 0.18 \quad (\text{crossed gate plane}) \\ |y_t^g| &< 0.60, |z_t^g| < 0.60 \quad (\text{within aperture})\end{aligned}$$

The gate reward is:

$$r_{\text{gate}} = 10 \cdot \mathbf{1}(\text{gate_passed}_t)$$

2.2 Progress and Velocity Rewards

Planar distance to current gate:

$$d_t = \|\mathbf{p}_{t,xy}^* - \mathbf{p}_{t,xy}\|_2$$

Progress reward based on distance improvement:

$$\begin{aligned}\Delta d_t &= d_{t-1} - d_t \\ r_{\text{prog}} &= \text{clip}(\Delta d_t, -1, 1)\end{aligned}$$

Velocity toward gate reward:

$$\begin{aligned}\mathbf{u}_t &= \frac{\mathbf{p}_t^* - \mathbf{p}_t}{\|\mathbf{p}_t^* - \mathbf{p}_t\|_2 + \epsilon} \\ v_t^{\parallel} &= \mathbf{v}_t^\top \mathbf{u}_t \\ r_{\text{vel}} &= \text{clip}(v_t^{\parallel}, -1, 20) \\ r_{\text{back}} &= -\text{clip}(-v_t^{\parallel}, 0, 2)\end{aligned}$$

2.3 Stability and Penalty Terms

Heading alignment with gate direction:

$$\begin{aligned}a_t &= (\mathbf{f}_t^{\text{world}})^\top \mathbf{u}_t \\ r_{\text{head}} &= \text{clip}(a_t, -1.5, 1)\end{aligned}$$

Tilt penalty using roll ϕ_t and pitch θ_t :

$$\begin{aligned}T_t &= |\phi_t| + |\theta_t| \\ p_{\text{tilt}} &= \text{clip}(T_t - 0.8, 0, 2) \\ r_{\text{tilt}} &= -p_{\text{tilt}}\end{aligned}$$

Angular velocity penalty:

$$r_{\text{ang}} = -0.1 \|\boldsymbol{\omega}_t^b\|$$

Crash penalty (after 100-step grace period):

$$r_{\text{crash}} = -\mathbf{1}(\|\mathbf{F}_t^{\text{contact}}\|_2 > 10^{-8})$$

Lap-time reward based on completing a lap:

$$r_{\text{lap}} = (t_{\text{target}} - t_{\text{lap}}) \cdot \mathbf{1}(\text{lap_completed})$$

2.4 Complete Reward Function

$$\begin{aligned}r(t) &= w_p r_{\text{prog}} + w_v r_{\text{vel}} + w_g r_{\text{gate}} + w_h r_{\text{head}} \\ &\quad + w_t r_{\text{tilt}} + w_{\omega} r_{\text{ang}} + w_c r_{\text{crash}} \\ &\quad + w_b r_{\text{back}} + w_{\ell} r_{\text{lap}}\end{aligned}$$

3 Observation Space Design

The observation vector (31 dimensions) provides comprehensive state information:

3.1 Drone State (13D)

- World position: (x, y, z)
- Body-frame linear velocity: (v_x^b, v_y^b, v_z^b)
- Body-frame angular velocity: $(\omega_x^b, \omega_y^b, \omega_z^b)$
- Euler angles: (ϕ, θ, ψ)
- Quaternion scalar: q_w

3.2 Gate Information (13D)

- Current gate body position: (g_x^b, g_y^b, g_z^b)
- Gate direction (body frame): \hat{d}^b (3D)
- Gate distance: d_g
- Drone in gate frame: (x^g, y^g, z^g)
- Next gate body position: (ng_x^b, ng_y^b, ng_z^b)

3.3 Progress and History (5D)

- Normalized gates passed
- Previous action (4 motor commands)

This multi-frame representation enables both local control and global navigation.

4 Reset Strategy

4.1 Training Reset Distribution

1. Start behind gate 0 with random offset:

$$\begin{aligned} x_l &\sim \mathcal{U}(-3, -1) \\ y_l &\sim \mathcal{U}(-1, 1) \\ z_l &\sim \mathcal{U}(-0.3, 0.3) \end{aligned}$$

2. Convert to world frame using gate orientation

3. Set initial yaw to face gate with noise:

$$\psi_0 = \text{atan2}(y_g - y_0, x_g - x_0) + \mathcal{U}(-0.3, 0.3)$$

4. Add small roll/pitch noise: $\mathcal{U}(-0.1, 0.1)$

5. Initialize forward velocity: $s \sim \mathcal{U}(0, 0.5)$ toward gate

5 Reward Scales

Our final policy uses a tuned set of reward scales that balance aggressive racing behavior with stability and safety:

• Progress and Navigation:

- Progress toward gate: $w_p = 2.0$
- Gate pass bonus: $w_g = 10.0$
- Forward velocity shaping: $w_v = 3.0$
- Straightaway speed bonus: $w_{\text{speed}} = 1.5$

• Stability and Control:

- Tilt penalty: $w_t = 0.1$
- Forward-tilt shaping: $w_{ft} = 1.0$
- Angular velocity penalty: $w_\omega = 0.04$

• Safety and Termination:

- Crash penalty: $w_c = 6.0$
- Episode termination penalty: $w_{\text{death}} = -50.0$

• Lap-Time Objective:

- Lap-time reward scale: $w_\ell = 5.0$
- Time penalty scale: $w_{\text{time}} = 4.0$

6 Domain Randomization

To enhance robustness, we randomize dynamics parameters per environment:

- Thrust-to-weight ratio: $[0.95, 1.05] \times$ nominal
- Aerodynamic coefficients: $[0.5, 2.0] \times$ nominal
- PID gains: $[0.85, 1.15] \times$ nominal values
- Motor time constants and drag coefficients randomized

This randomization improves policy robustness against model inaccuracies and physical variations.

7 Results & Conclusion

Overall we experimented using strategies like per time-step penalties, curriculum based rewards, gate-lookahead observations, heading rewards etc., but our final implemented strategy demonstrated several key capabilities including stable yet competitive gate navigation with our local coordinate approach, high speed stability with the combination of tilt penalties, angular velocity penalties, and velocity shaping along with the straightaway speed bonus encouraging high-speed traversal between long gate segments, and lap time targets for competitive lap time completions. Our Lap time for 3 laps under our observation was 20.5 seconds, and the policy was robust for randomized resets and stable for long runs without any crashes. Our final policy results can be found attached below in Fig.1. along with our final run time [video link](#).

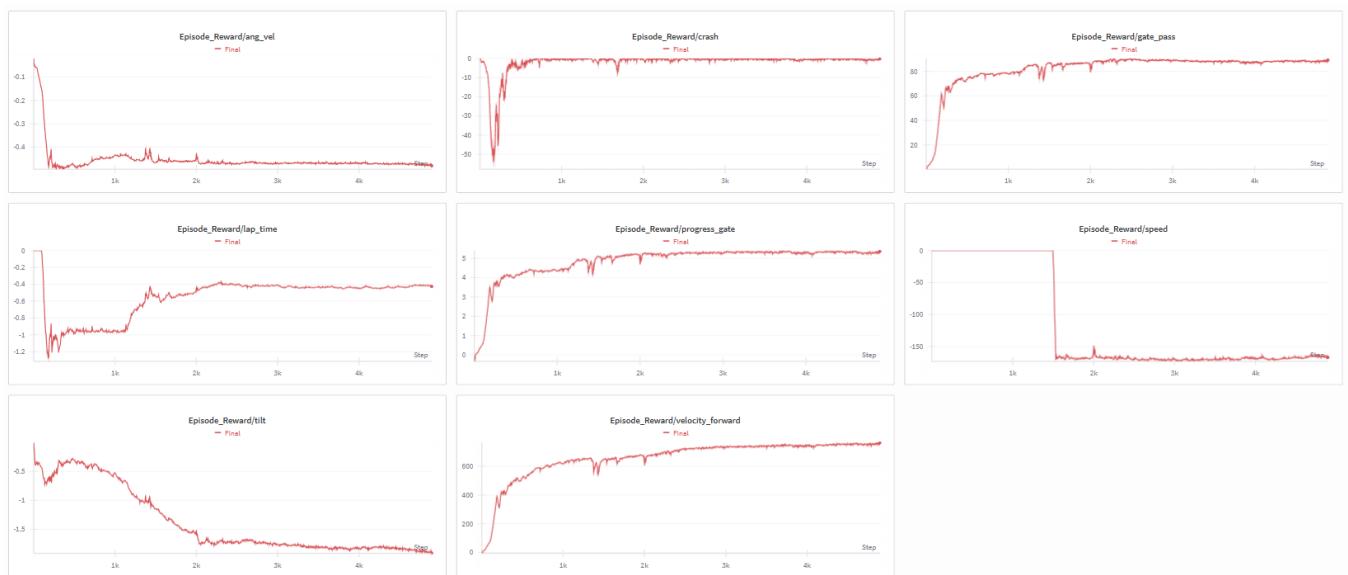


Figure 1: Drone racing policy results