HIERARCHICAL REINFORCEMENT LEARNING FOR AUTONOMOUS PARKING

Combining A* Path Planning with PPO Control

MSML 642 ROBOTICS

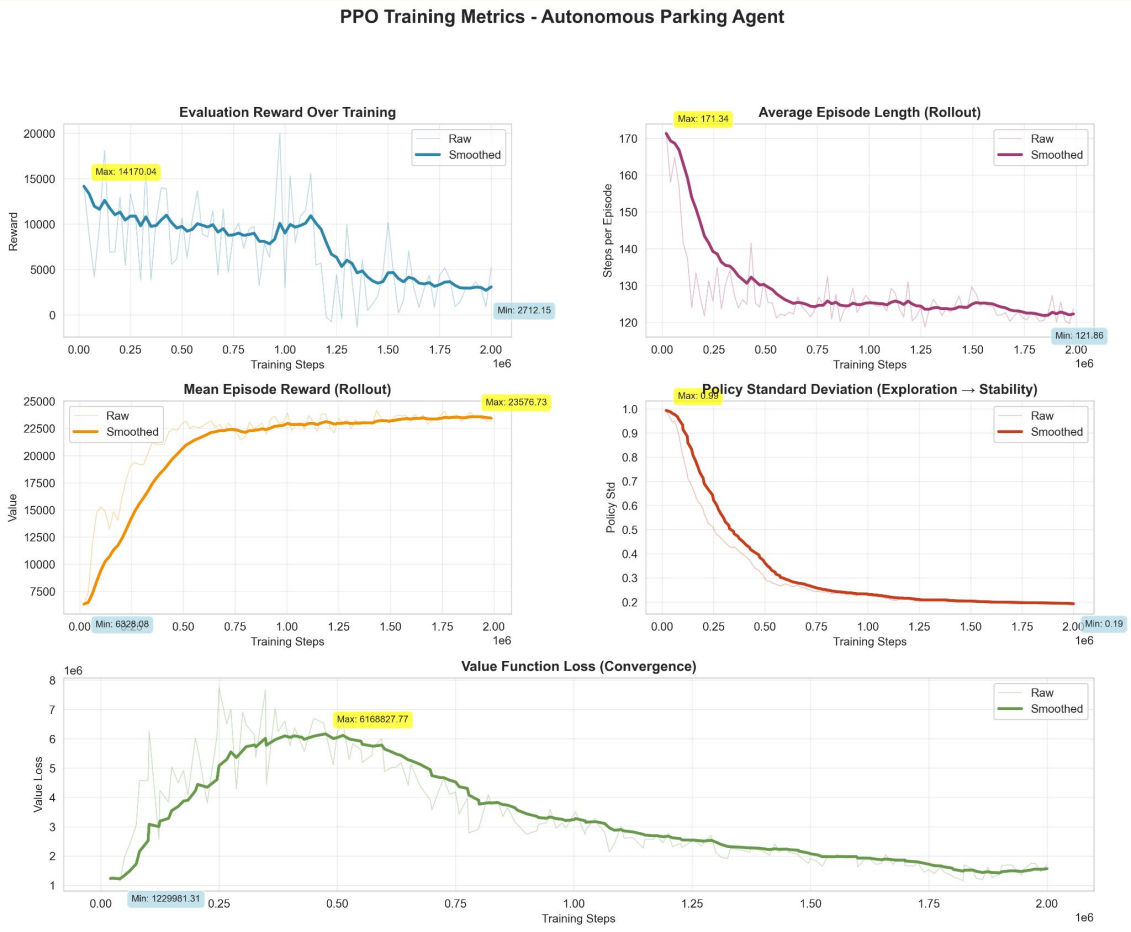
Vishwaksena Dingari
121113764

CURRENT RESULTS

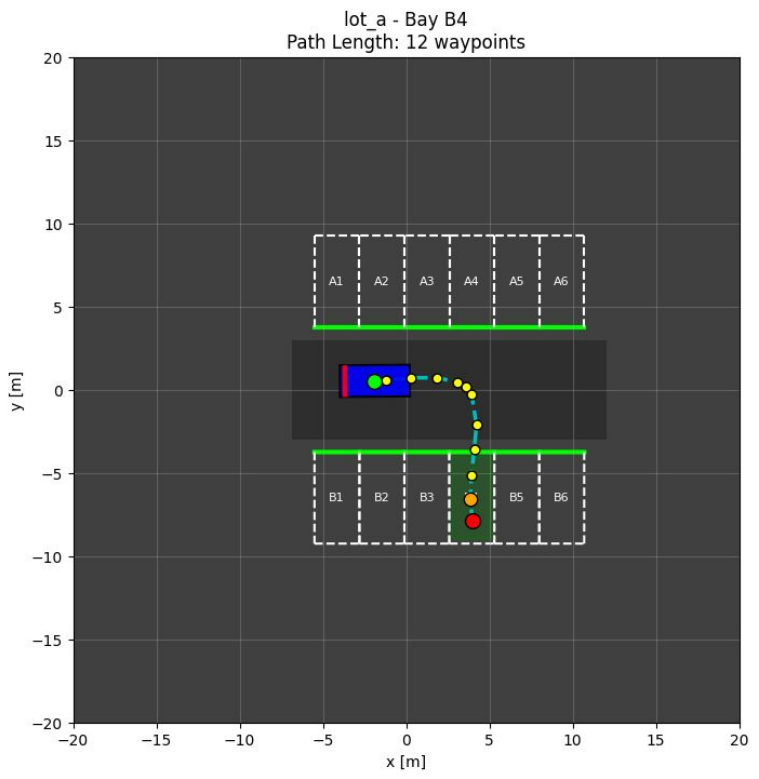
Training Summary (2M steps, ~12 hours)

- Episode Reward increased from 6K → 23K
- Episode Length stable at ~125 steps (efficient navigation)
- Policy Std decreased 0.99 → 0.19 (exploration → control)
- Value Loss converged from 6M → 1.7M

Interpretation:
Agent is consistently improving, stabilizing, and following A* waypoint plans with high precision.



AUTONOMOUS PARKING IN LOTS



- INPUT:
  - Robot state (x, y, yaw, v)
  - Goal bay location, yaw
  - 64-ray Lidar scan (360°, 20m)
- OUTPUT:
  - Steering ∈ [-1, 1]
  - Throttle ∈ [-1, 1]
- CHALLENGES
  - Long Horizon: 10-25 meter navigation, 500-2000 timesteps/episode
  - Multi-Orientation: Bays face different directions, Rotation-invariant parking, Progressive curriculum
  - Precision Parking: Position: <0.5m error, Orientation: <10° error, Must stop completely

HIERARCHICAL ARCHITECTURE

HIGH-LEVEL PLANNER (A*)

- 0.25 m grid-based search
- Collision + corridor checks
  - Prevent cutting through parking bays
  - Forces path to stay inside driving lanes
- B-Spline smoothing
  - Removes sharp A* corners
  - Produces drivable curvature (car-like motion)

LOW-LEVEL CONTROLLER (PPO)

- Observation (~80 dims):
  - 64-ray lidar
  - Vehicle state (x, y, yaw, v)
  - Bay geometry (cx, cy, yaw error)
  - Waypoint tracking (distance + heading)
  - Corridor offset + progress signals
- Actions (continuous):
  - Steering $\in [-1, 1]$
  - Throttle $\in [-1, 1]$

REWARD FUNCTION DESIGN

NAVIGATION PHASE (r_nav) (dist > 3.0m)	PARKING PHASE (r_park) (dist < 3.0m)
• Waypoint Progress: +0.1 per meter (only if moving forward)	• Lateral Alignment: +20.0 (1 - |cy|/0.8)
• Path Velocity: +0.5 × v_path (reward approaching goal)	• Yaw Alignment: +15.0 (1 - |Δθ|/1.5)
• Backward Penalty: -2.0 × |v| (strong discouragement)	• Depth Progress: +10.0 × Δcx (drive deeper into bay)
• Anti-freeze: -0.01 (keep moving)	• Success Bonus: +500.0 (dominates episode return)
• Distance Penalty: -0.002 × dist	
• Corridor Drift Penalty: −corridor_weight × |lateral_offset|	

<u>UNIVERSAL PENALTIES</u>

- Collision: -100
- Timeout: -50

Phase Blending:r_total = α(d) × r_nav + (1-α(d)) × r_park
where α(d) = d / 3.0 (weight decreases as approach)

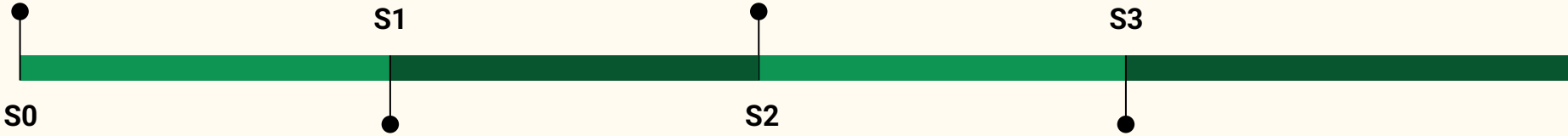
CURRICULUM LEARNING

Stage 0 - Aligned Spawn

- Aligned Start,
- Lot A
- 6 Bays
- Learn Basics and Park

Stage 2 - Lot A

- Stage 1 + Other side
- Spawns Along the road
- Random spawn
- Navigate the space and Park



Stage 1 - Lot A - Side A

- Add Cars in Other Bays
- Spawns along the road
- Random Spawn 22m max
- Navigate the space and Park

Stage 2 - Lot A + Lot B

- Stage 2 + Lot B
- Added Lot B, a t-shaped lot
- Random spawn along the roads
- Navigate the space and Park

FUTURE WORK & BROADER IMPACT

IMMEDIATE NEXT STEPS

- Complete curriculum, Expected: 10-30% success rate
- Sim-to-real transfer with TurtleBot3
- Multi-agent coordination

KEY LESSONS

- Hierarchical RL works (20x speedup)
- Curriculum essential (300x improvement)
- Reward engineering critical

REFERENCES

- https://arxiv.org/html/2407.01333v1
- https://github.com/Harish-Balaji-B/Ros2_humble_Autonomous_Parking_System_using_Turtlebot3_Waffle/
- https://github.com/aatmaj28/Autonomous-Car-Parking-RL
- https://arxiv.org/abs/1707.06347
- https://arxiv.org/abs/1805.08296