CS 5180 Reinforcement Learning

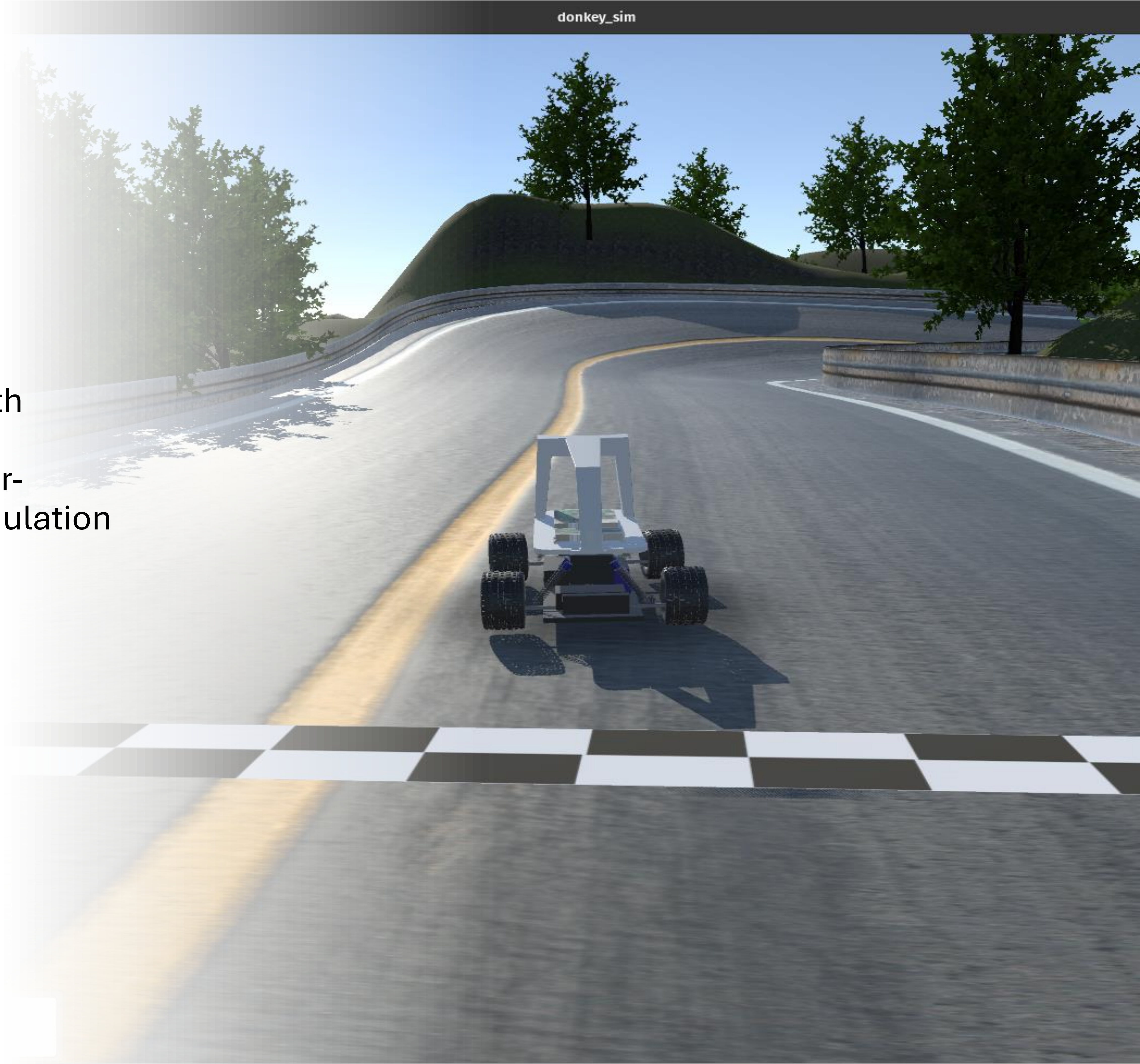
RL for Autonomous Car Racing

Girish Raut, Omkar Sargar

- **Project goal**: Develop an autonomous driving agent for racing circuits using reinforcement learning.
- **Motivation**: Address the need for real-time decision-making in dynamic environments and explore the challenges of high-performance racing.
- **Racing demands**: Optimal performance and precise maneuvers, unlike conventional autonomous driving focused on safety and efficiency.
- **Approach**: Utilize reinforcement learning to train the agent to master racing intricacies, integrating strategies and rapid decision-making.

Problem Formulation

Environment: Started with TORCS but switched to DonkeyCar Sim for a user-friendly and efficient simulation environment.



Observation space: Utilizes a front-facing camera providing RGB images (120x160x3 pixels) for track perception and decision-making.

Action space: Governed by throttle control for speed regulation and steering output for turning direction and sharpness, enabling precise maneuvers.

Reward Engineering: Experimented with various reward functions focusing on speed, collision penalties, and reverse-driving penalties to guide optimal performance.

- **Speed Reward**: Positive reward based on the car's velocity, encouraging faster lap times.
- **Collision Penalty**: Negative reward for collisions.
- **Reverse-driving Penalty**: Negative reward given for driving in the opposite direction as it is very undesirable.

Algorithms

Proximal Policy Optimization (PPO):

- On-policy algorithm, ensuring stable and reliable training
- Potential for suboptimal exploration in complex environments
- Less sample-efficient compared to off-policy algorithms

Soft Actor-Critic (SAC):

- Off-policy algorithm, enabling efficient use of experience replay
- Encourages exploration through entropy maximization
- Capable of learning diverse strategies for different scenarios
- Potential for overestimation of value functions

Truncated Quantile Critics (TQC):

- Captures the distribution of the value function, accounting for uncertainty
- Robust to outliers and multimodal value distributions
- Increased computational complexity due to multiple quantile predictions
- Potential for instability during training if not tuned properly

Experiments

Safe Driving:

Objective:

Train the agent to navigate the track safely while maximizing progress.

Action Space:

- Throttle: [0, 1]
- Steer: [-0.5, 0.5]

Reward Function:

- Penalizes going off-track, collisions and driving in reverse.
- Encourages going fast while staying safe (close to the center).

Aggressive Racing:

Objective:

Train the agent to adopt more aggressive racing lines while maintaining control and avoiding penalties.

Action Space:

- Throttle: [0, 1]
- Steer: [-0.5, 0.5]

Reward Function:

- Penalizes collisions and driving in reverse.
- Encourages going fast (no condition for safety).

Aggressive Racing with braking:

Objective:

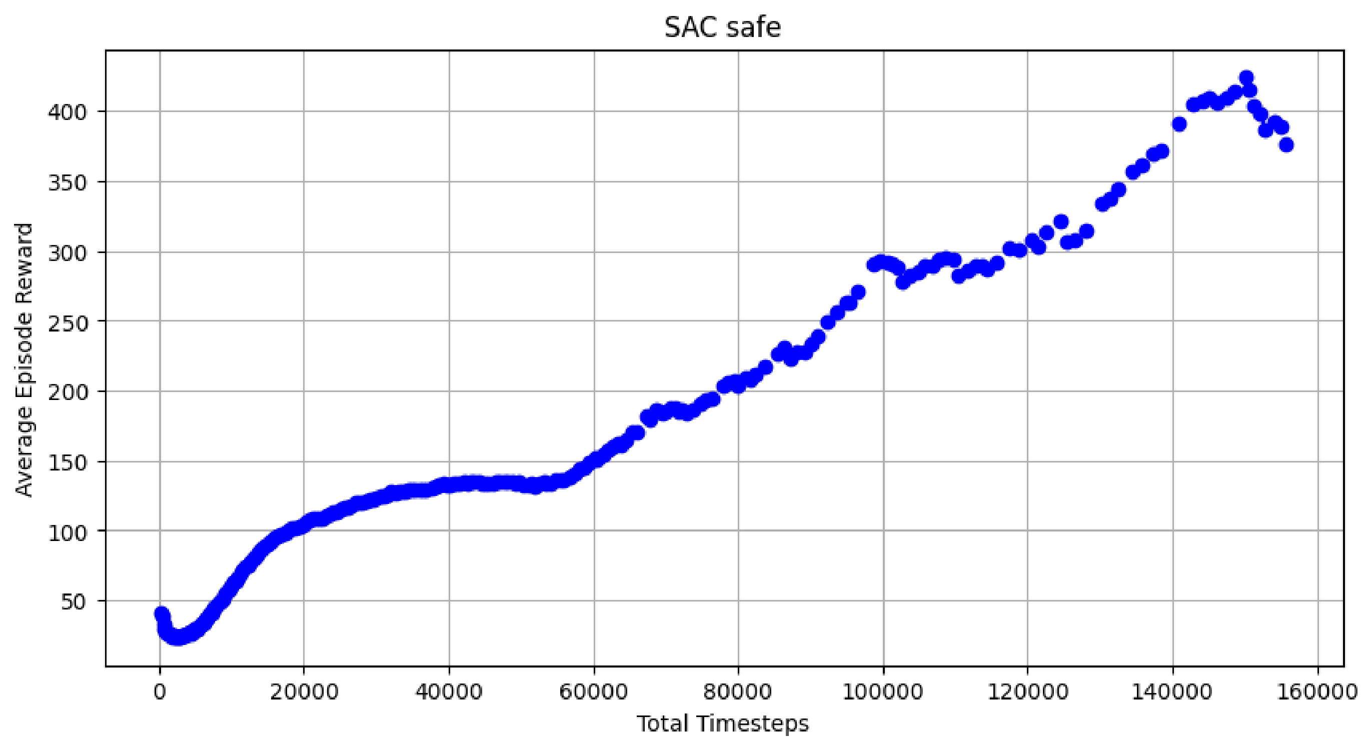
Train the agent to adopt more aggressive racing lines while maintaining control and avoiding penalties. Allow braking to potentially help the car to carry more speed into the corners and then brake while turning.

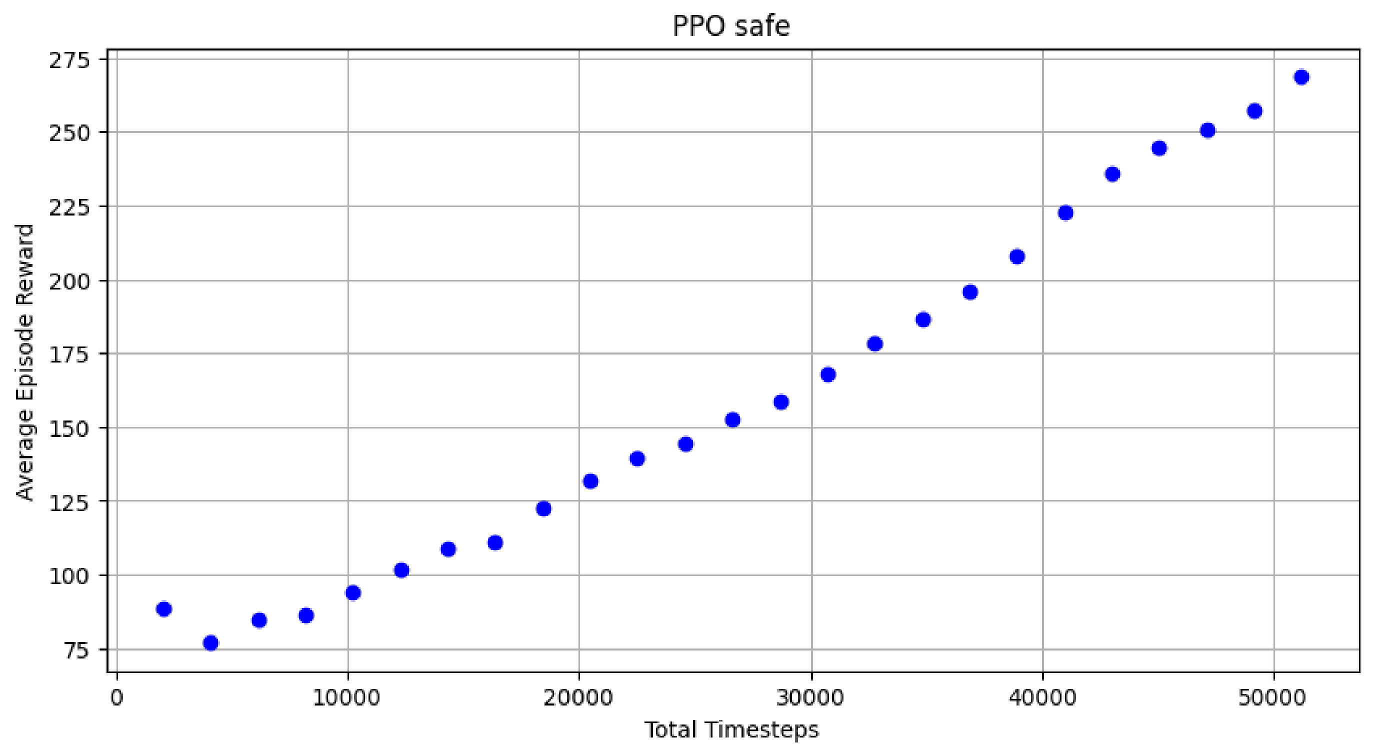
Action Space:

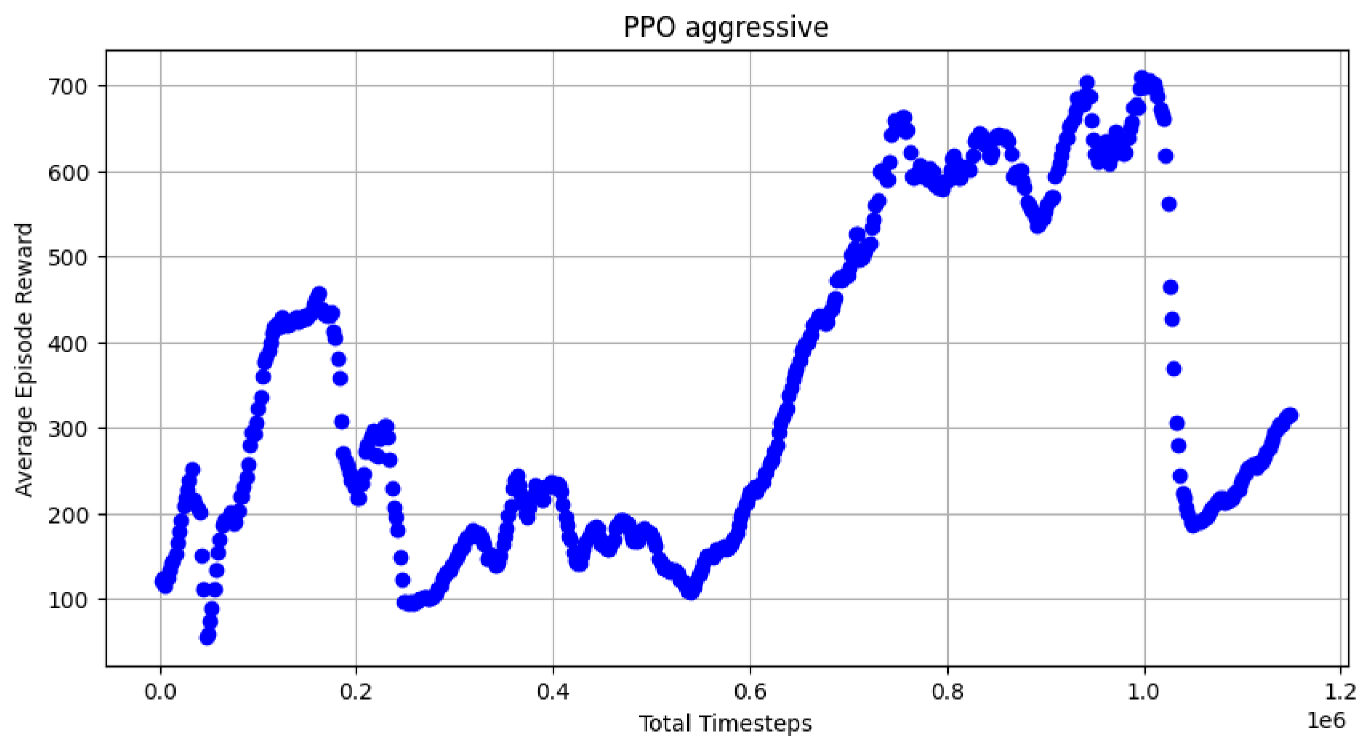
- Throttle: [-0.5, 1]
- Steer: [-0.5, 0.5]

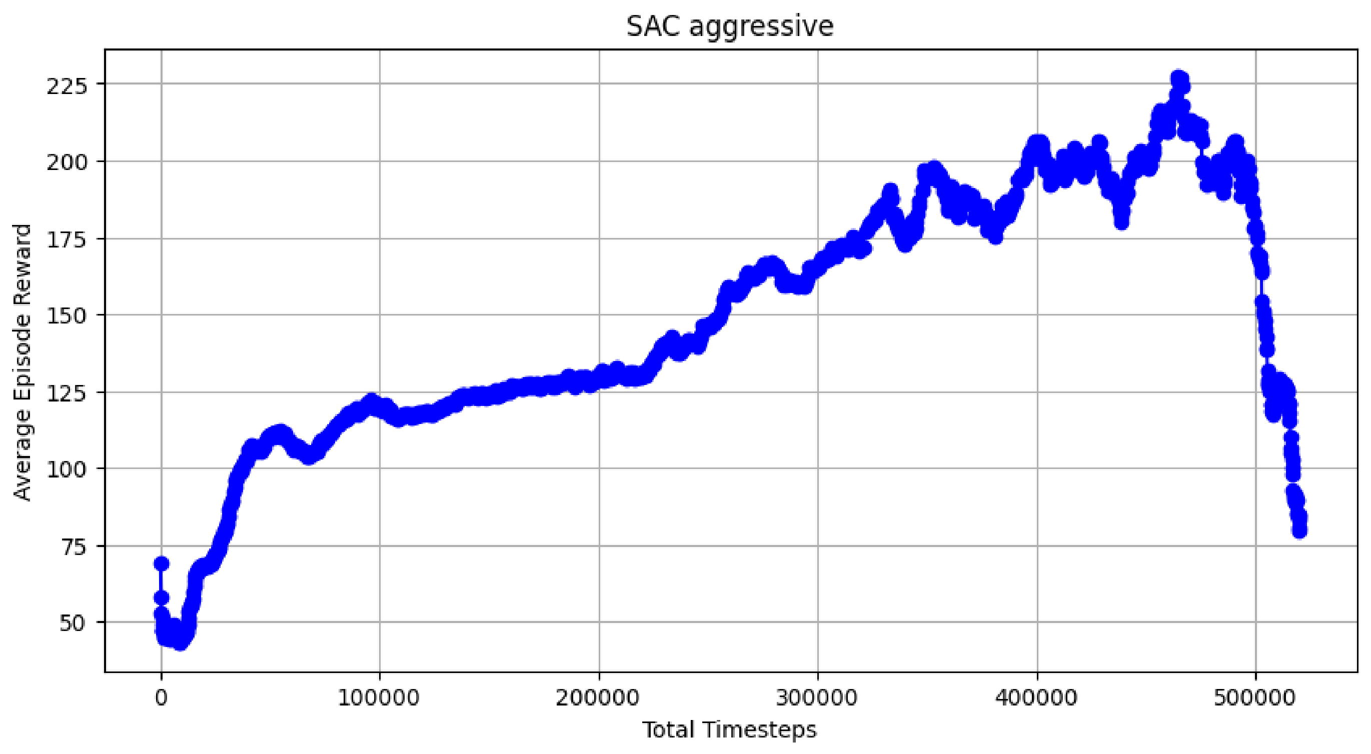
Reward Function:

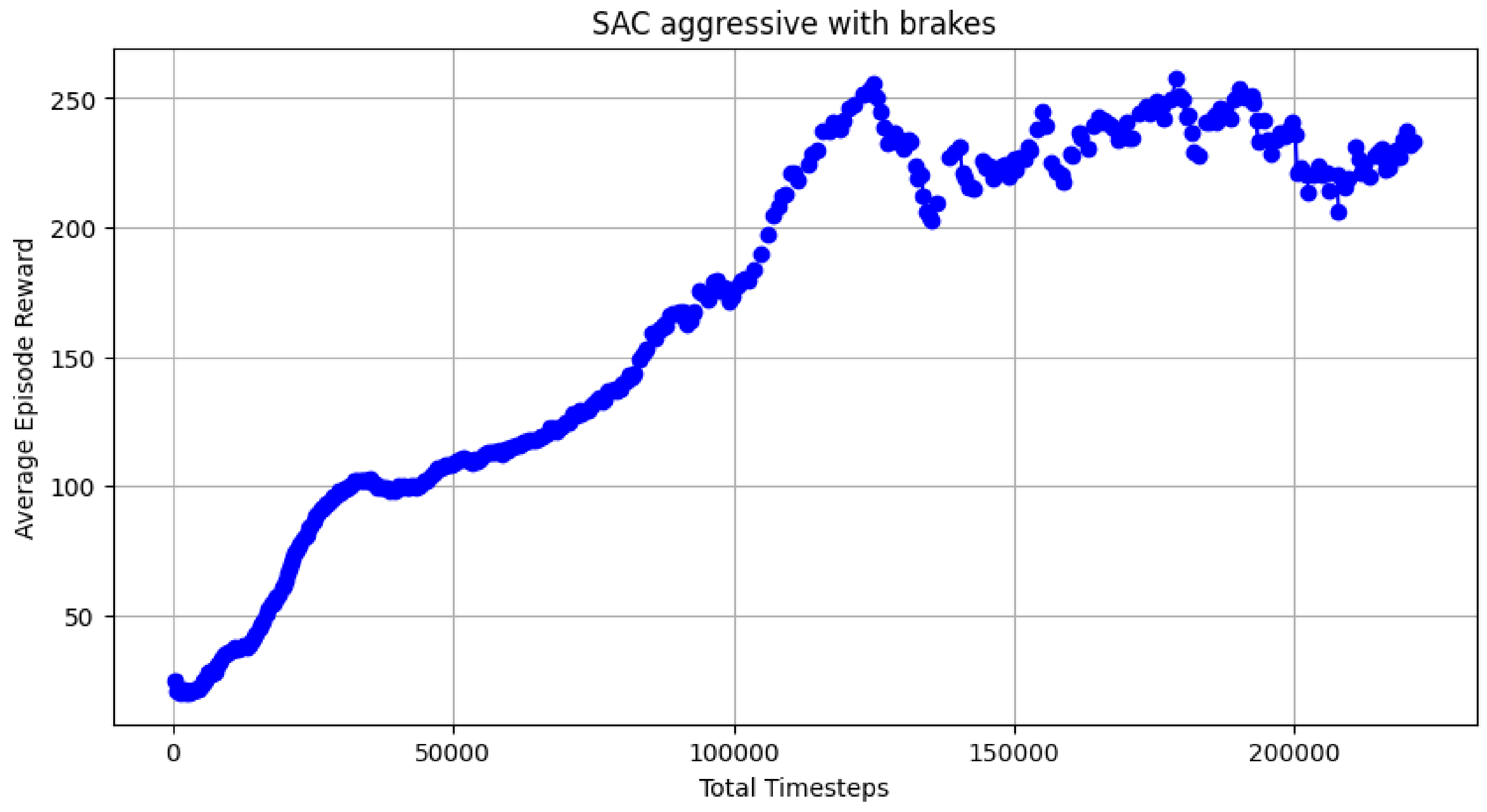
- Penalizes collisions and driving in reverse.
- Encourages going fast (no condition for safety).

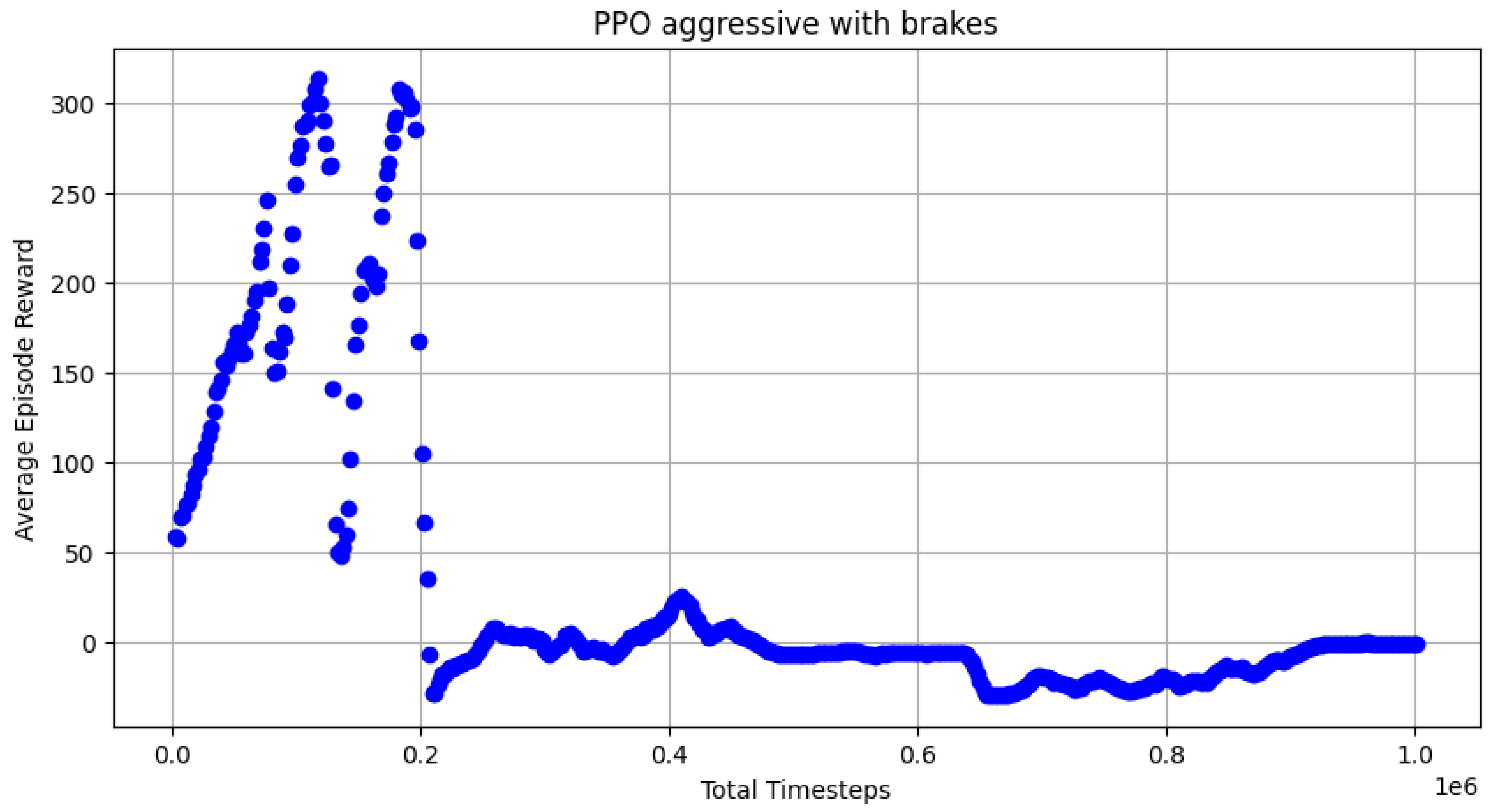

SAC safe
Average Episode Reward
Total Timesteps
50
100
150
200
250
300
350
400
0
20000
40000
60000
80000
100000
120000
140000
160000

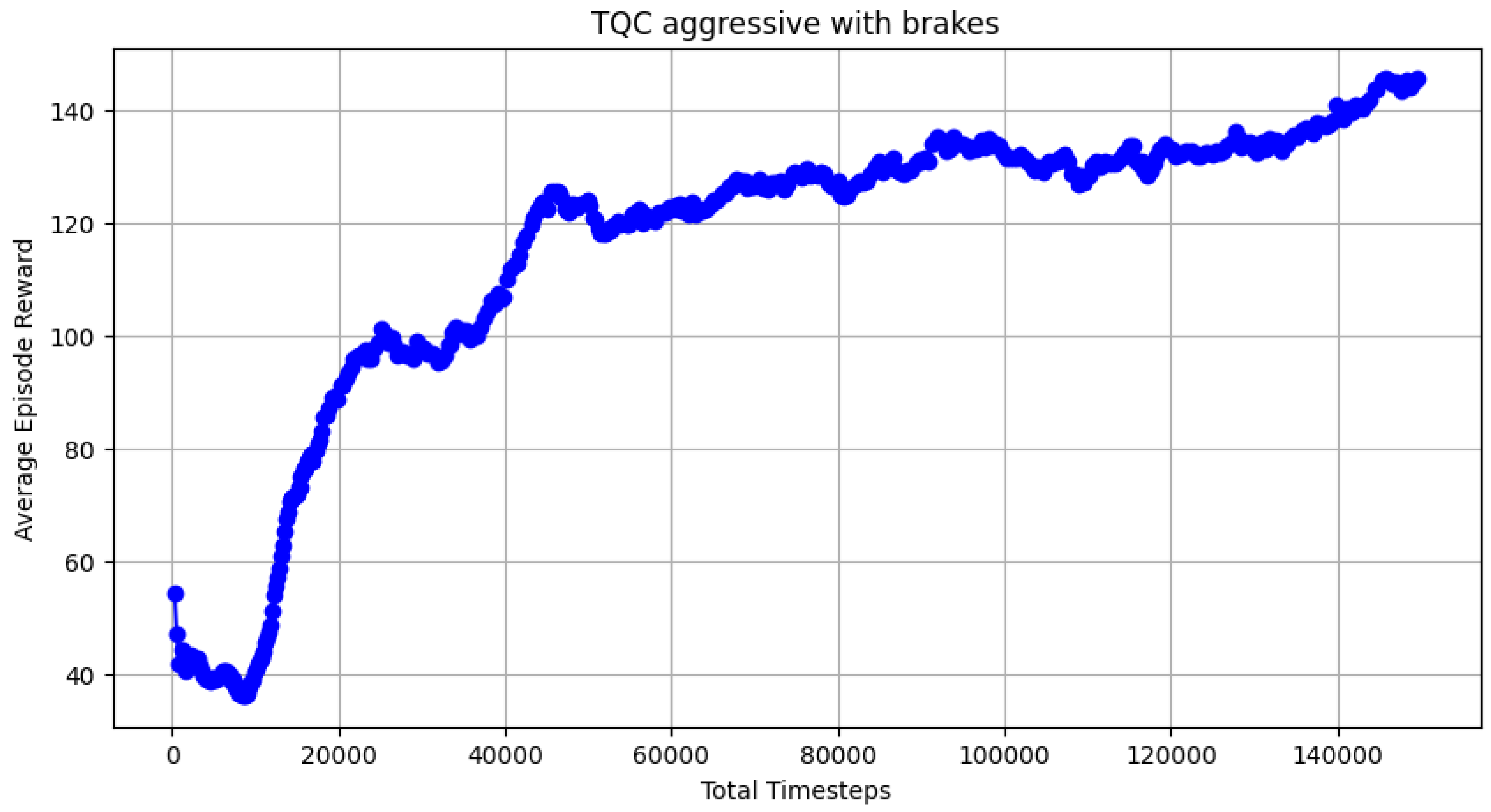

PPO safe
Average Episode Reward
Total Timesteps
75
100
125
150
175
200
225
250
275
0
10000
20000
30000
40000
50000


PPO aggressive
Average Episode Reward
Total Timesteps
1e6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
100
200
300
400
500
600
700

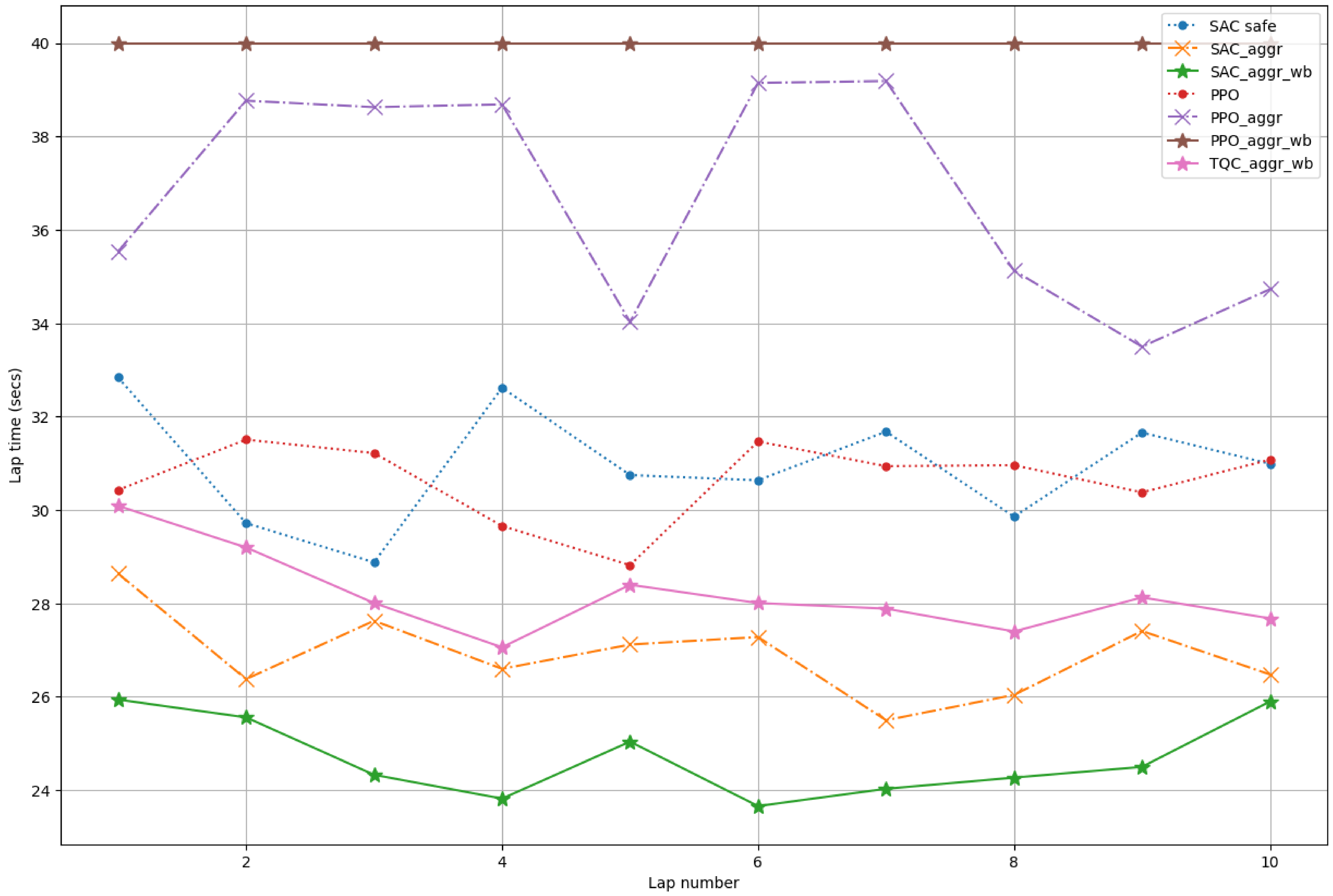

SAC aggressive
Average Episode Reward
Total Timesteps
0
100000
200000
300000
400000
500000
50
75
100
125
150
175
200
225

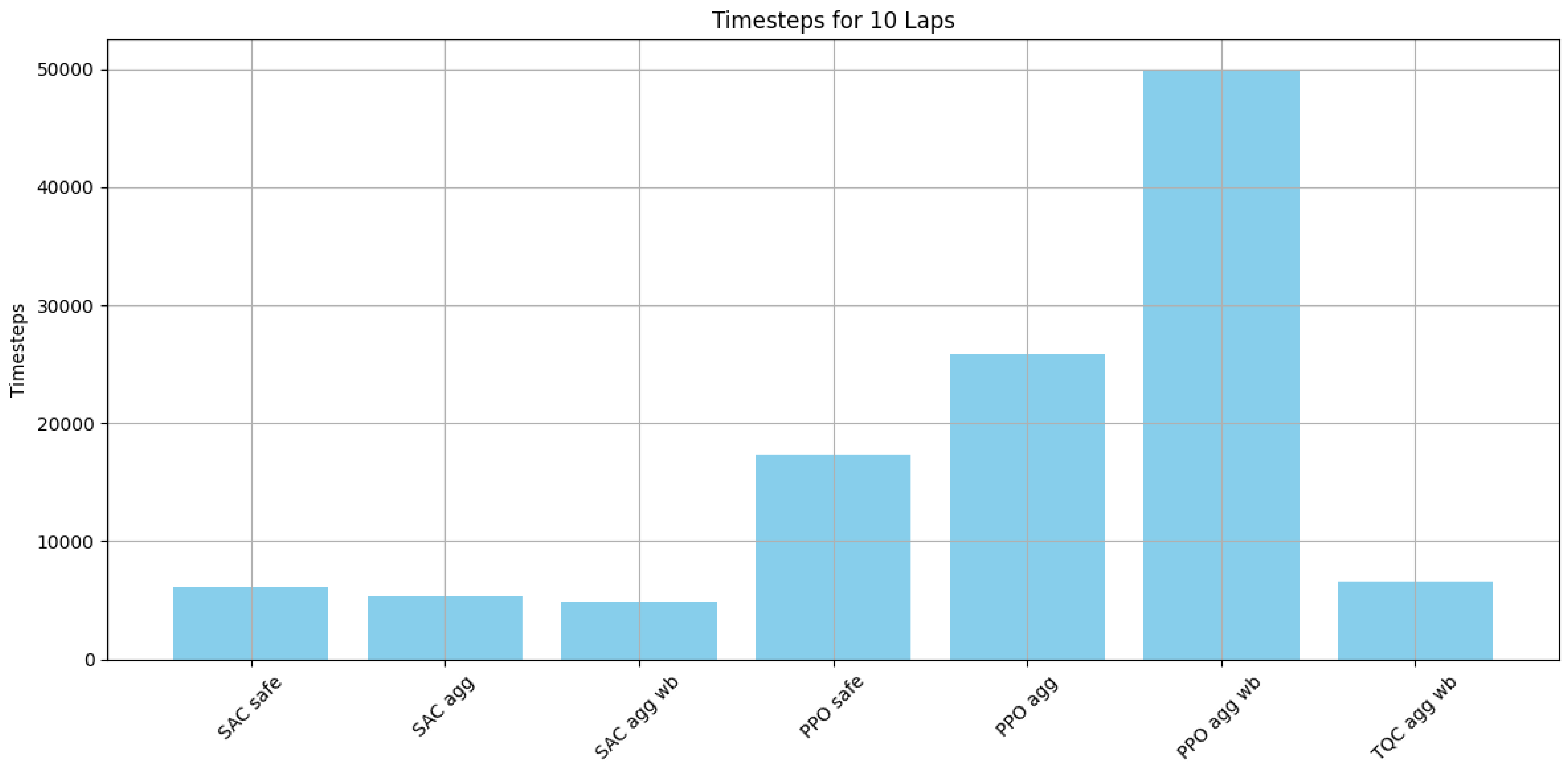

SAC aggressive with brakes
Average Episode Reward
Total Timesteps
250
200
150
100
50
0
50000
100000
150000
200000


PPO aggressive with brakes
Average Episode Reward
Total Timesteps
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
1e6


TQC aggressive with brakes
Average Episode Reward
Total Timesteps
140
120
100
80
60
40
0
20000
40000
60000
80000
100000
120000
140000

Performance comparison for 10 Laps




Timesteps for 10 Laps
Timesteps
0
10000
20000
30000
40000
50000
SAC safe
SAC agg
SAC agg wb
PPO safe
PPO agg
PPO agg wb
TQC agg wb