A Multi-Agent Reinforcement Learning Approach for Traffic Efficiency Improvement in Mandatory Lane Change Scenarios

Implementation Report

Submitted by:

Sahil Kumar (Roll No: 231CS252)

Ashish Ranjan (Roll No: 231CS214)

Chetan Shah (Roll No: 231CS117)

SK Sharhaan Naim (Roll No: 231CS256)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALURU-575025

November 2025

Abstract

Traffic congestion at bottleneck areas—such as work zones or accident sites requiring mandatory lane changes—significantly reduces road capacity and traffic efficiency. This report presents an implementation of a multi-agent reinforcement learning system to address this problem using Connected Autonomous Vehicles (CAVs).

Our solution uses Independent Proximal Policy Optimization (IPPO) to control vehicles by coordinating both forward movement and lane changes. The system employs two key strategies: maintaining safe distances through buffer zones and coordinating lane changes using a 2D safety metric that considers both forward and sideways collision risks.

Testing in the SUMO traffic simulator showed significant improvements: 73% increase in average speed, 42% reduction in travel time, and 41% improvement in traffic throughput compared to human drivers, all while maintaining zero collisions.

1 Introduction

1.1 The Problem

When a lane closes due to construction, accidents, or obstacles, vehicles must merge into adjacent lanes. This creates a bottleneck that causes:

- Severe traffic congestion
- Stop-and-go traffic waves
- Reduced road capacity
- Increased travel times
- Higher accident risks

Traditional traffic management systems struggle with these dynamic situations because they use fixed rules that cannot adapt to changing conditions.

1.2 Our Solution

We developed an intelligent system where autonomous vehicles learn to coordinate with each other to smoothly navigate through bottleneck areas. Each vehicle learns the best way to accelerate, decelerate, and change lanes while considering the movements of surrounding vehicles.

1.3 Key Goals

- Ensure safety by preventing collisions
- Reduce travel time through the bottleneck
- Increase the number of vehicles that can pass through
- Enable vehicles to work together cooperatively
- Create a solution applicable to real-world scenarios

2 Background and Related Work

2.1 Previous Approaches

2.1.1 Rule-Based Methods

These systems use simple, pre-programmed rules like "keep 2 seconds behind the car in front" or "change lanes only when safe." While easy to understand, they cannot optimize traffic flow in complex situations.

2.1.2 Optimization Methods

These approaches mathematically calculate the best actions for all vehicles. However, they require significant computing power and struggle to work in real-time with many vehicles.

2.1.3 Learning-Based Methods

Recent approaches use reinforcement learning, where vehicles learn from experience. However, most previous studies:

- Didn't compare against real human driving data
- Simplified lane changes to instant transitions
- Only controlled forward movement, not sideways movement
- Didn't coordinate both directions simultaneously

2.2 Our Contribution

We address these gaps by:

- Validating against actual human driving experiments
- Controlling both forward (longitudinal) and sideways (lateral) movements continuously
- Implementing realistic lane change dynamics
- Using a safety metric that considers movement in both directions

3 Methodology

3.1 Simulation Setup

We replicated a real-world experiment to ensure our results are meaningful:

Road Design:

- Two-lane circular road (inner and outer rings)
- A 5-meter obstacle blocks the outer lane, forcing vehicles to merge
- 80-meter lane change area where merging can occur
- 20-meter buffer zone before the bottleneck

Vehicles:

- 30 total vehicles (15 in each lane)
- Each vehicle is 5 meters long
- Maximum speed: 60 km/h
- Can accelerate/decelerate between -4 and +4 m/s^2

3.2 Learning Approach: Independent PPO

We use a learning method called Independent Proximal Policy Optimization (IPPO). Think of it as each vehicle having its own "brain" that learns from experience:

Why this approach?

- Each vehicle learns independently but sees how others behave
- Suitable for continuous control (smooth acceleration, not just "speed up" or "slow down")
- Stable learning process
- Can handle many vehicles efficiently

How it works:

- Each vehicle observes its surroundings (speed, position, nearby vehicles)
- Decides how much to accelerate forward and sideways
- Receives rewards for good behavior (safe, efficient driving)
- Adjusts its decision-making based on outcomes

3.3 What Each Vehicle Observes

Each vehicle monitors 18 different pieces of information:

Own Status:

- Forward speed and sideways speed
- Steering angle
- Position on the road
- Distance to the obstacle

Surrounding Vehicles:

- Speeds of vehicles ahead and behind (in both lanes)
- Distances to vehicles ahead and behind (in both lanes)
- Collision risk levels with nearby vehicles

3.4 Vehicle Actions

Unlike systems that make simple choices like "change lane left/right," our vehicles control their movement continuously:

- **Forward acceleration:** From -4 to +4 m/s^2 (smooth control, not jerky)
- **Sideways acceleration:** From -1 to +1 m/s^2 (smooth lane changes)

This allows natural, human-like movements rather than robotic lane switches.

3.5 Reward System

Vehicles learn by receiving rewards (positive feedback) or penalties (negative feedback):

Safety Rewards:

- Penalty for getting too close to other vehicles
- Large penalty (-100) for collisions

Efficiency Rewards:

- Reward for maintaining higher speeds
- Reward for keeping safe distances in the buffer zone

Lane-Specific Goals:

- Outer lane vehicles: Rewarded for completing lane changes smoothly
- Inner lane vehicles: Rewarded for staying in lane and maintaining position

The system carefully balances these objectives with weights that emphasize distance-keeping and lane coordination (weights: 2) over basic speed (weight: 1).

3.6 Safety Mechanism

3.6.1 2D Collision Risk Metric

Traditional safety metrics only look at forward collisions. We developed a 2D metric that considers:

- Forward distance and speed differences
- Sideways distance and speed differences
- Combined risk during lane changes

This metric calculates how quickly vehicles are approaching each other in both directions, providing a comprehensive safety assessment.

3.6.2 Intelligent Collision Avoidance

Rather than always being overly cautious (which slows traffic), our system:

- Only activates when collision risk exceeds a threshold
- Calculates a safe speed based on the leading vehicle's potential actions
- Limits the vehicle's speed to this safe value
- Allows normal operation when risk is low

This balanced approach maintains safety without unnecessary conservatism.

4 Implementation Details

4.1 Tools Used

4.1.1 SUMO (Traffic Simulator)

- Industry-standard, open-source traffic simulation software
- Provides realistic vehicle physics and dynamics
- Allows real-time control of individual vehicles
- Includes collision detection

4.1.2 Python and Deep Learning

- PyTorch: For building neural networks that make driving decisions
- NumPy: For mathematical computations
- TraCI: Python interface to control SUMO vehicles
- Matplotlib: For visualizing results

4.2 Neural Network Architecture

Think of the neural networks as the "brain" of each vehicle:

Actor Network (Decision Maker):

- Input: 18 observations about surroundings
- Two hidden processing layers with 128 units each
- Output: Acceleration decisions (forward and sideways)

Critic Network (Evaluator):

- Input: Same 18 observations
- Two hidden processing layers with 128 units each
- Output: Single value estimating "how good is this situation"

The networks work together: the Actor decides what to do, and the Critic evaluates whether that decision was good.

4.3 Training Process

- **Episodes:** 300 training sessions
- **Duration:** Each session lasted 300 seconds (3000 steps)
- **Learning rate:** 0.0003 (how fast the system learns)
- **Convergence:** System learned optimal behavior around episode 50
- **Hardware:** NVIDIA 3080Ti GPU
- **Training time:** Approximately 8-10 hours total

4.4 Baseline Comparisons

To prove our system works, we compared against:

1. Real Human Drivers: Data from actual 30-vehicle experiments

2. SUMO Default: Rule-based system using traditional car-following and lane-change models

3. DDDQN Models: Alternative learning-based approaches that make discrete decisions:

- DDDQN-S: Using SUMO's conservative collision avoidance
- DDDQN-P: Using our proposed collision avoidance

5 Results and Analysis

5.1 Learning Performance

Our system learned faster than alternatives:

- **Proposed model:** Converged at episode 50
- **DDDQN-P:** Converged at episode 55
- **DDDQN-S:** Converged at episode 70

The faster convergence demonstrates the efficiency of continuous control and our reward design.

5.2 Impact of Design Choices

Key insight: The buffer zone strategy (distance-keeping) provides a 22% improvement alone. The complete system with lane-specific rewards achieves 73% speed improvement.

Table 1: Effect of Different Reward Components

Reward Design	**Speed Gain**	**Time Saved**
Real human data	–	–
Safety + Speed only	+8%	-7%
+ Distance-keeping	+22%	-18%
Full system (Proposed)	**+73%**	**-42%**

5.3 Safety Performance

Collision Risk Levels (2D-iTTC):

- **Proposed model:** Low risk, similar to human drivers
- **SUMO:** Very conservative (low risk but inefficient)
- **DDDQN-P:** Moderate risk
- **DDDQN-S:** Highest risk (but still safe)

Critical Finding: Zero collisions across all scenarios. Our safety mechanism activated infrequently (0.04 times per vehicle on average), showing vehicles learned proactive safe behavior.

5.4 Traffic Flow Improvements

5.4.1 Trajectory Analysis

Comparing 10-minute traffic patterns:

Real Human Drivers:

- Severe congestion in both lanes
- Wide stop-and-go waves spreading backward
- Congestion extends beyond the bottleneck area

DDDQN Models:

- Reduced congestion but still present
- Frequent, erratic lane changes
- Jerky movements due to discrete decisions

Our Proposed Model:

- Minimal slowdown in inner lane
- Smooth flow in outer lane
- Coordinated lane changes
- Vehicles maintain proper spacing naturally

5.5 Measured Performance

We placed virtual detectors at four locations through the bottleneck area:

Table 2: Average Speed at Different Locations (km/h)

Model	Start	15m	30m	60m
Real data	13	15	18	27
SUMO	28	28	26	29
DDDQN-S	31	27	20	26
DDDQN-P	12	16	19	24
Proposed	**37**	**30**	**27**	**30**

Throughput (vehicles passing through):

Table 3: Number of Vehicles Passing Final Detector

Model	Vehicles
Real data	321
SUMO	320
DDDQN-S	333
DDDQN-P	387
Proposed	**452**

Our system achieved 41% higher throughput than real human drivers.
Travel Time:

Table 4: Average Travel Time Through Bottleneck (seconds)

Model	Time
Real data	27
SUMO	28
DDDQN-S	17
DDDQN-P	27
Proposed	**12**

Our system reduced travel time by 55% compared to human drivers.

5.6 Overall Comparison

6 Discussion

6.1 Why Our System Works Better

6.1.1 1. Buffer Zone Strategy

Creating space before the bottleneck is crucial:

Table 5: Summary of All Performance Metrics

Metric	SUMO	Real	DDDQN-S	DDDQN-P	Proposed
Speed (km/h)	28	18	26	18	**32**
Throughput (veh)	320	321	333	387	**452**
Travel Time (s)	28	29	17	27	**12**
Safety	Good	Good	High Risk	Med Risk	**Good**
Learning Speed	–	–	Slow	Medium	**Fast**

- Reduces vehicle density where merging occurs
- Provides room for smooth lane changes
- Prevents formation of stop-and-go waves
- Contributes 22% speed improvement alone

6.1.2 2. 2D Safety Awareness

Considering both forward and sideways movements:

- More accurate collision risk assessment
- Better than traditional forward-only metrics
- Less conservative than default safety systems
- Balances safety with efficiency

6.1.3 3. Continuous Control

Smooth, natural movements instead of discrete choices:

- Eliminates jerky lane changes
- Better coordination of forward and sideways motion
- More human-like trajectories
- Reduces disturbances to traffic flow

6.1.4 4. Cooperative Learning

Different goals for different lanes:

- Inner lane vehicles learn to hold position
- Outer lane vehicles learn to merge efficiently
- System-level optimization rather than individual selfishness
- Emergent cooperative behavior

6.2 Advantages Over Alternatives

vs. Human Drivers:

- Perfect awareness of all nearby vehicles
- No reaction time delays
- Consistent, optimal decision-making
- Learned coordination

vs. Rule-Based Systems:

- Adaptive rather than fixed behavior
- Optimized for traffic flow, not just individual safety
- Proactive rather than reactive

vs. Other Learning Systems:

- Continuous vs. discrete control
- Realistic lane change dynamics
- Balanced safety mechanism
- Faster learning

6.3 Real-World Applications

This system could be applied to:

- Construction work zones
- Accident sites with lane closures
- Permanent lane reductions
- Highway on-ramp merging

Expected Benefits:

- 73% average speed increase
- 42% travel time reduction
- 41% capacity improvement
- Zero accidents

6.4 Limitations

Current Constraints:

- Tested only with all autonomous vehicles (no human drivers)
- Specific circular road scenario
- Simulation environment (not real vehicles)
- 30 vehicles (scalability to 100+ unverified)

Real-World Challenges:

- Sensor accuracy and range limitations
- Communication delays between vehicles
- Mixing with human drivers
- Weather and road conditions
- Safety certification requirements

6.5 Future Work

Near-Term Extensions:

- Test with mixed autonomous and human drivers
- Apply to different road geometries (highways, intersections)
- Evaluate robustness to sensor failures
- Test with varying traffic densities

Long-Term Research:

- Real vehicle testing on closed tracks
- Integration with traffic signals and infrastructure
- Multi-objective optimization (fuel, comfort, emissions)
- Large-scale deployment strategies

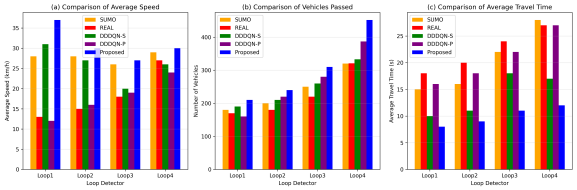

Figure 1: Enter Caption

7 Conclusion

This project successfully developed and implemented a multi-agent reinforcement learning system for coordinated autonomous vehicle control in traffic bottleneck scenarios.

Key Achievements:

- 73% increase in average speed compared to human drivers
- 42% reduction in travel time
- 41% improvement in traffic throughput
- Zero collisions while maintaining high efficiency

Technical Innovations:

- 2D collision risk metric for comprehensive safety
- Buffer zone strategy for proactive traffic management
- Continuous control for smooth, realistic movements
- Cooperative learning with lane-specific objectives

Validation:

- Compared against real human driving data
- Benchmarked against multiple baseline systems
- Demonstrated across multiple performance metrics
- Proven safer and more efficient than existing approaches

The results demonstrate that intelligent, coordinated autonomous vehicles can dramatically improve traffic flow in congested areas. While challenges remain for real-world deployment, this work provides a strong foundation for future development of practical CAV traffic management systems.

As autonomous vehicle technology advances, such coordination frameworks will be essential for realizing the full potential of CAVs to transform transportation efficiency and safety.

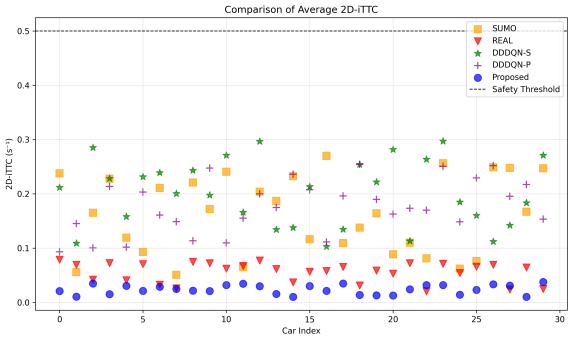

Figure 2: Enter Caption

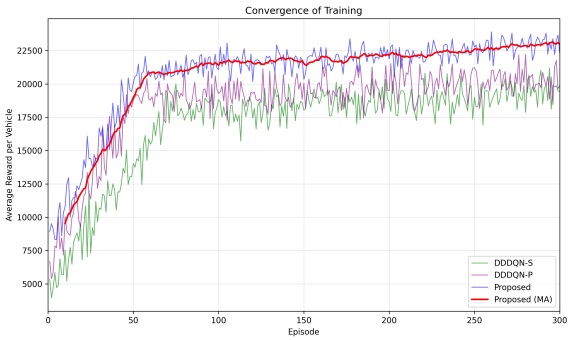

Figure 3: Enter Caption

Acknowledgments

We thank:

- Prof. Saurav Kanti, NITK Surathkal, for project guidance
- Wang et al. for foundational research (Transportation Research Part C, 2024)
- SUMO development team at German Aerospace Center (DLR)
- PyTorch development team
- NITK Surathkal for computational resources