Joint Pedestrian and Vehicle Traffic Optimization in Urban Environments using Reinforcement Learning

Bibek Poudel[1], Xuan Wang[2], Weizi Li[1], Lei Zhu[3], and Kevin Heaslip[4]

***Abstract*— Reinforcement learning (RL) holds significant promise for adaptive traffic signal control. While existing RL-based methods demonstrate effectiveness in reducing vehicular congestion, their predominant focus on vehicle-centric optimization leaves pedestrian mobility needs and safety challenges unaddressed. In this paper, we present a deep RL framework for adaptive control of eight traffic signals along a real-world urban corridor, jointly optimizing both pedestrian and vehicular efficiency. Our single-agent policy is trained using real-world pedestrian and vehicle demand data derived from Wi-Fi logs and video analysis. The results demonstrate significant performance improvements over traditional fixed-time signals, reducing average wait times per pedestrian and per vehicle by up to 67% and 52% respectively, while simultaneously decreasing total wait times for both groups by up to 67% and 53%. Additionally, our results demonstrate generalization capabilities across varying traffic demands, including conditions entirely unseen during training, validating RL's potential for developing transportation systems that serve all road users.**

I. Introduction

Traffic congestion has become a silent tax on modern civilization. Each year, drivers in U.S. cities waste an average of 54 hours stuck in traffic [1], costing over 160 billion in lost productivity [2]. With urban cores in metropolitan areas experiencing an increase of traffic inflow, up to 25% year-over-year [3], congestion is set to worsen as urbanization continues. To address this challenge, traffic control systems have evolved from traditional handcrafted rules and actuated systems to Adaptive Traffic Signal Control (ATSC). Driven by cost-effectiveness [4], [5], increasing availability of traffic data [6], and advances in optimization techniques [7], ATSC has become a central focus of intelligent transportation systems research [8]. While ATSC has reduced congestion and improved vehicular flow, its evolution has largely overlooked a critical stakeholder: pedestrians.

Currently, pedestrian fatalities in the U.S. have reached their highest level in 41 years, averaging about 21 deaths per day [9], [10]. Urban areas are particularly affected, with 84% of these fatalities occurring in cities and 76% taking place at non-intersection locations such as mid-block crossings and commercial strips [11]. A key factor is the prevalence of unsignalized mid-block crosswalks, where pedestrians lack clear guidance and drivers may not yield, increasing accident risks [12]. Implementing signalized crosswalks at these locations, with dedicated pedestrian phases and clear right-of-way signals, has proven effective in reducing vehicle-pedestrian conflicts and enhancing rule compliance [13]. The inclusion of pedestrian-friendly features in ATSC frameworks ensures that traffic control systems not only alleviate congestion but also proactively protect vulnerable road users.

As urban traffic control grows increasingly complex, ATSC strategies are leveraging real-time sensor data and computational methods. Reinforcement learning (RL) has emerged as a leading approach due to its ability to learn adaptive policies directly from data without relying on traffic models or handcrafted rules. Recent RL-based methods have primarily focused on improving vehicular throughput, neglecting pedestrian efficiency or modeling pedestrian behavior in overly simplified scenarios [14]–[16]. This vehicle-centric approach fails to capture the complex interactions between vehicles and pedestrians, especially in urban environments characterized by high pedestrian volumes, frequent mid-block crossings, and diverse mobility demands. Developing control policies that simultaneously optimize pedestrian and vehicular efficiency in realistic urban settings remains an open challenge.

We introduce an RL framework for corridor-level control of eight traffic signals, jointly optimizing for vehicular and pedestrian efficiency. Our contributions are

- we develop a single policy that effectively manages high traffic volumes (up to 6,000 pedestrians/hr and 558 vehicles/hr) based on real-world demand data derived from Wi-Fi logs and video analysis;
- our results demonstrate substantial performance improvements over traditional fixed-time signals, reducing average wait time per vehicle by up to 52% and average wait time per pedestrian by up to 67%;
- we provide insights into the behaviors learned by our policy, including its ability to coordinate multiple independent signals to create a "green wave" effect and its responsiveness to real-time traffic, as demonstrated by its adaptive phase switching.

The code, data, and videos are available in our GitHub: github.com/poudel-bibek/Urban-Control.

II. Related Work

Conventional adaptive traffic control systems have relied on model-based [17] or rule-based [18] approaches. De-

[1]Bibek Poudel and Weizi Li are with Min H. Kao Department of Electrical Engineering and Computer Science at University of Tennessee, Knoxville, TN, USA bpoudel3@vols.utk.edu, weizili@utk.edu

[2]Xuan Wang is with Department of Electrical and Computer Engineering at George Mason University, Fairfax, VA, USA xwang64@gmu.edu

[3]Lei Zhu is with Department of Industrial and Systems Engineering at University of North Carolina at Charlotte, Charlotte, NC, USA lei.zhu@charlotte.edu

[4]Kevin Heaslip is with Department of Civil and Environmental Engineering at University of Tennessee, Knoxville, TN, USA kheaslip@utk.edu

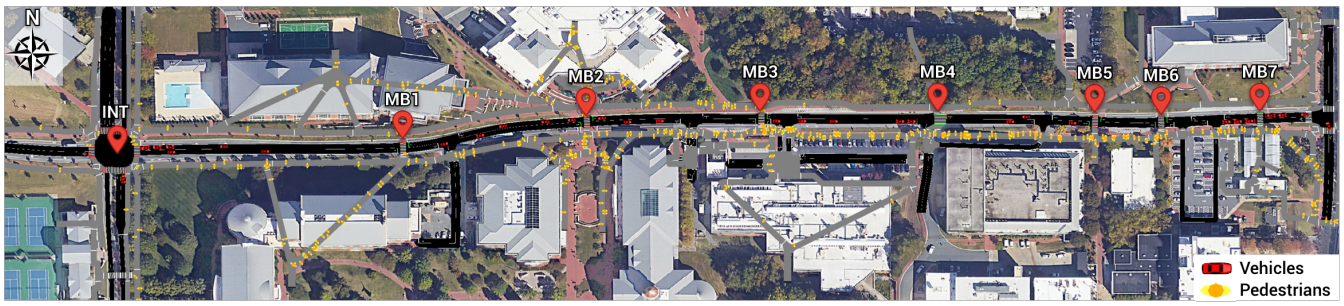

Fig. 1: The Craver Road corridor with an intersection (INT) that contains one primary traffic signal and four signalized crosswalks, along with seven mid-block signalized crosswalks (MB1–MB7) that control pedestrian–vehicle interactions.

spite their widespread deployment and clear advantages over fixed-time control, these systems struggle to capture the inherent complexity and stochasticity of urban traffic. Common challenges include unpredictable variations in traffic volume and queue propagation through multiple intersections [19]. In response to these limitations, computational techniques that learn adaptive policies without relying on explicit rules or models, particularly machine learning and deep reinforcement learning (DRL), offer promising alternatives for traffic signal control [20]–[25]. In single-agent settings, DRL has been applied to optimize intersection signal timing by selecting appropriate phase [26] or its duration [27] based on observed traffic conditions. While multi-agent scenarios explore approaches where agents either collaborate to coordinate signals across a network [28] or compete to prioritize movement at a traffic signal with higher demand [15]. These methods demonstrate improved performance over fixed-time and conventional approaches [29], achieving, for instance, reductions in average travel time and queue lengths [30]. Yet, it is worth noting that the majority of existing studies remain primarily focused on optimizing vehicle-centric outcomes such as reducing queue, decreasing stop frequency, and increasing throughput [16].

However, real-world urban environments include both vehicles and pedestrians. Integrating pedestrian dynamics into ATSC systems introduces several complex challenges such as ensuring pedestrian safety at crosswalks [14], [31] and balancing the needs of both vehicles and pedestrians to prevent excessive delays for one group while prioritizing the other [32]. Recognizing these complexities, recent studies have extended RL-based traffic signal control to incorporate pedestrians. Several approaches have emerged in this direction: some introduce pedestrian-specific phases, i.e., eliminate vehicle-pedestrian interactions for improved safety [33], while others incorporate pedestrian-centric performance metrics directly into the reward function [34]. However, these studies typically rely on synthetic demand data or remain constrained to ideal road networks [8]. This leaves pedestrian-inclusive ATSC strategies relatively underexplored, particularly for complex urban corridors with multiple signalized crosswalks.

Our work addresses these limitations by proposing a DRL framework that jointly optimizes vehicular and pedestrian waiting times in a real-world corridor-level setting. Unlike earlier methods that either focused solely on vehicles or used synthetic demand on ideal networks, our approach employs a real-world urban network of eight traffic signals with real-world demand data for both vehicles and pedestrians. The work most similar to ours uses a real-world corridor and vehicle demand data but differs by adopting a multi-agent framework and relying on synthetic pedestrian demand [35]. Our approach demonstrates that a single-agent policy can effectively control an entire urban corridor while balancing the needs of both vehicles and pedestrians.

III. Methodology

A. Real-World Network and Traffic Data

Craver Road, shown in Figure 1, is a 750m corridor that serves as the primary arterial through the University of North Carolina, Charlotte's main campus. The campus covers 1.56 square miles and includes 85 buildings and approximately $34,000$ students, faculty, and staff [36]. To capture pedestrian demand data, we utilize Wi-Fi logs collected in September 2021 by the university's IT department. The Wi-Fi network comprises $2,492$ access points (APs) distributed across 82 buildings, with $88,409$ unique clients. The Wi-Fi log data captures communication events between Wi-Fi clients (e.g., smartphones and laptops) and APs. Each log entry indicates "when" (timestamp) and "where" (AP location) each client connected to the network. We processed this raw data using a generalized Wi-Fi processing framework based on a "Point-Line-Plane" hierarchical concept [37] with the following filtering assumptions:

- Client activities were aggregated at the building level; each client's AP sessions within the same building were merged to represent presence in that location.
- Clients identified as visitors or irregular commuters (detected for fewer than three days per month, constituting 11.81% of the dataset) were excluded, as our analysis focuses on typical campus travel patterns.
- To address individuals carrying multiple devices, we used K-means clustering to classify and remove "non-mobile" devices based on the mean and variance of their stationary ratio relative to total daily activity time [38].

For vehicle demand data, we analyzed four video recordings (total 21 minutes) captured at different times at the Craver Road intersection (INT). The observed flow was converted to hourly rates, resulting in an average headway of 18

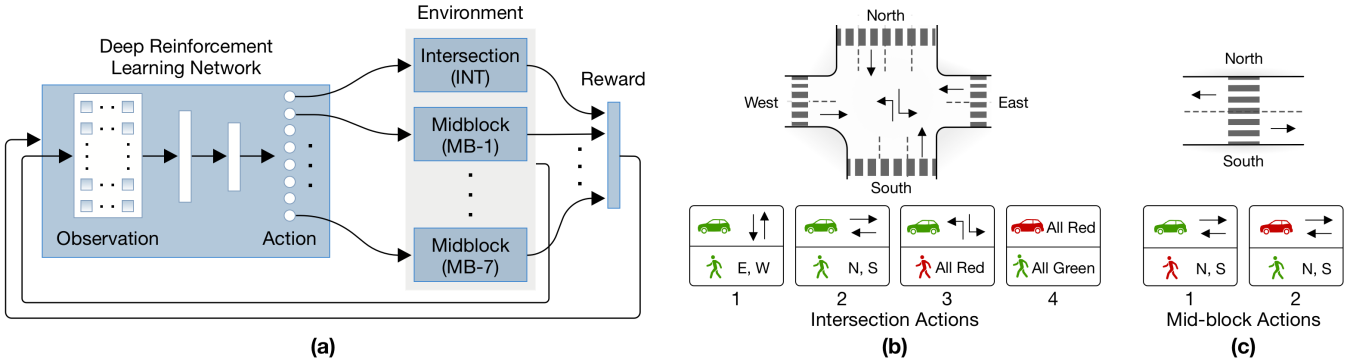

Fig. 2: (a) Deep reinforcement learning framework for corridor-level traffic control. (b) Intersection signal configurations controlling vehicle and pedestrian movements, including dedicated left turns (choice 3) and all-pedestrian phases (choice 4). (c) Two mid-block signal configurations allowing either vehicle movement (choice 1) or pedestrian crossing (choice 2).

seconds between vehicles. We mapped both pedestrian and vehicle demand to SUMO [39] trip definitions, creating a network-wide real-world demand of 2223 pedestrians/hr and 202 vehicles/hr. Of these pedestrians, 990 (44%) have trips that cross the corridor. For trip generation, we defined origin-destination pairs using Traffic Analysis Zones (TAZs). Pedestrian origin-destination pairs were derived from Wi-Fi building visit data, and vehicle pairs from movement records. In each trip, start and end points were assigned to specific edges within TAZs, with departure times based on observed timestamps.

B. Markov Decision Process

We formulate the Adaptive Traffic Signal Control (ATSC) as a sequential decision-making problem modeled as a partially observable Markov Decision Process represented by the tuple $(S, A, T, R, \Omega, O, \gamma)$. Here, S denotes the set of environment states, A represents the set of possible actions, $T : S \times A \times S \to [0, 1]$ is the probabilistic state transition function, and $R : S \times A \to \mathbb{R}$ defines the reward function. Due to partial observability in real-world traffic systems, Ω represents the set of observations, $O : S \times A \times \Omega \to [0, 1]$ defines the observation probability function, and $\gamma \in [0, 1)$ is the discount factor to balance immediate and future rewards.

State: Traffic representation significantly impacts ATSC performance [40]. Our observation fuses vehicle and pedestrian occupancy data in areas neighboring each traffic signal. Vehicles are detected within a 15–100m vicinity while pedestrians are detected within 5–10m. To capture both spatial and temporal dynamics, we stack occupancy information over the action duration (each action step encompasses N simulation timesteps). At action step t, the observation $o_t \in \Omega$ is formed by stacking N vectors from the previous action interval:

$$o_t = [v_1, v_2, \ldots, v_N], \quad \text{with}$$
$$v_k = \left[\phi_{t-1}, \{v_{i,\cdot}(k)\}_{i=1}^{M}, \{p_{i,\cdot}(k)\}_{i=1}^{M}\right],$$

for $k = 1, 2, \ldots, N$, where:

- ϕ_{t-1} is the signal phase during the previous action,
- $v_{i,\cdot}(k) = \big(v_{i,\text{in}}(k), v_{i,\text{inside}}(k), v_{i,\text{out}}(k)\big)$ denotes the vehicle occupancy vector at traffic signal i at the k-th simulation timestep; here, $v_{i,\text{in}}(k)$ represents occupancy from lanes approaching the traffic signal (incoming), $v_{i,\text{inside}}(k)$ from lanes within the controlled area, and $v_{i,\text{out}}(k)$ from lanes exiting (outgoing),
- $p_{i,\cdot}(k) = \big(p_{i,\text{in}}(k), p_{i,\text{out}}(k)\big)$ denotes the pedestrian occupancy vector at traffic signal i at the k-th simulation timestep; here, $p_{i,\text{in}}(k)$ represents occupancy from pedestrians approaching the crosswalk, and $p_{i,\text{out}}(k)$ from those leaving,
- M is the number of controlled traffic signals.

This spatio-temporal formulation provides a comprehensive snapshot of evolving traffic conditions.

Action: The agent's action is composed of two independent components—intersection and mid-block actions—both derived from a policy parameterized by $\theta \in \mathbb{R}^d$, which takes as input the current observation o_t to produce logits that are split into distribution parameters for the two actions:

- *Intersection Action*: The agent selects one of four mutually exclusive phase configurations as shown in Figure 2 (b), i.e., $a_t^{int} \in \{1, 2, 3, 4\}$, with j denoting the chosen configuration. This selection is drawn from a Categorical distribution:

$$P(a_t^{int} = j \mid o_t, \theta) = \text{Categorical}\Big(j; \mathbf{p}^{int}(o_t, \theta)\Big),$$

  where $\mathbf{p}^{int}(o_t, \theta) \in \Delta^3$ is the probability vector over the four phase configurations.
- *Mid-Block Actions*: For each of the seven mid-block signals, the agent independently selects a binary action $a_{t,i}^{mb} \in \{1, 2\}$, each modeled as a Bernoulli distribution:

$$P(a_{t,i}^{mb} = b \mid o_t, \theta) = \text{Bernoulli}\Big(b; \mu_i^{mb}(o_t, \theta)\Big),$$

  where $\mu_i^{mb}(o_t, \theta)$ is the probability for signal i and $b = 1$ indicates permission for vehicle flow, i.e., Mid-block choice 1 shown in Figure 2 (c).

The overall action $a_t = \text{concat}(a_t^{int}, a_{t,1}^{mb}, \ldots, a_{t,7}^{mb})$ is an 8-component vector. Restricting the intersection action to pre-defined safe configurations inherently enforces safety (preventing conflicting pedestrian–vehicle greens) and reduces the exploration space for more sample-efficient learning. For additional safety, a 4-timestep mandatory yellow phase is automatically introduced for all signals via an

internal mechanism before switching from a green to a red signal. Note that while direct control over the yellow phase duration is not permitted, the duration of any other phase can be controlled by repeatedly selecting the same action. Each action lasts for 10 simulation steps ($N = 10$).

Reward: To simultaneously minimize wait times for vehicles and pedestrians, we propose the *Exponentially Increasing Maximum Wait Aggregated Queue (EI-MWAQ)* reward function. This reward is designed to reflect real-world behavior, where very long wait times lead to disproportionately high driver and pedestrian frustration. It builds on the Maximum Wait Aggregated Queue (MWAQ) [41], which uses the product of queue length and maximum waiting time to approximate worst-case delay, but introduces two modifications. First, penalties from all mid-block crossings are aggregated using an L_2-norm. Second, the final penalty values are obtained by normalizing these aggregate delays and applying an exponential function. This results in a penalty that grows rapidly as queue lengths and wait times increase. For the intersection, we compute:

$$Q_{\text{veh}}^{\text{int}} = \frac{N_{\text{veh}}^{\text{int}} \cdot W_{\text{veh}}^{\text{int}}}{8|D|}, \quad Q_{\text{ped}}^{\text{int}} = \frac{N_{\text{ped}}^{\text{int}} \cdot W_{\text{ped}}^{\text{int}}}{10|D|},$$

where $N_{\text{veh}}^{\text{int}}$ and $N_{\text{ped}}^{\text{int}}$ denote the counts of waiting vehicles and pedestrians at the intersection, $W_{\text{veh}}^{\text{int}}$ and $W_{\text{ped}}^{\text{int}}$ their respective maximum waiting times, and $|D|$ the number of incoming directions. For each mid-block signal i, we compute:

$$Q_{\text{veh}}^{\text{mb}}(i) = \frac{N_{\text{veh}}^{\text{mb}}(i) \cdot W_{\text{veh}}^{\text{mb}}(i)}{8|D_{\text{mb}}|},$$
$$Q_{\text{ped}}^{\text{mb}}(i) = \frac{N_{\text{ped}}^{\text{mb}}(i) \cdot W_{\text{ped}}^{\text{mb}}(i)}{10},$$

where $|D_{\text{mb}}|$ is the number of incoming directions. These per-signal values are aggregated across all 7 mid-block signals using the L_2 norm:

$$Q_{\text{veh}}^{\text{mb}} = \left\| \left(Q_{\text{veh}}^{\text{mb}}(i)\right)_{i=1}^{7} \right\|_2, \quad Q_{\text{ped}}^{\text{mb}} = \left\| \left(Q_{\text{ped}}^{\text{mb}}(i)\right)_{i=1}^{7} \right\|_2.$$

The amplified penalties are obtained by applying the exponential function:

$$R_{\text{veh}}^{\text{int}} = \exp\left(Q_{\text{veh}}^{\text{int}}\right), \quad R_{\text{ped}}^{\text{int}} = \exp\left(Q_{\text{ped}}^{\text{int}}\right),$$
$$R_{\text{veh}}^{\text{mb}} = \exp\left(Q_{\text{veh}}^{\text{mb}}\right), \quad R_{\text{ped}}^{\text{mb}} = \exp\left(Q_{\text{ped}}^{\text{mb}}\right).$$

The final reward is given by:

$$\boxed{R = -\left(R_{\text{veh}}^{\text{int}} + R_{\text{ped}}^{\text{int}} + R_{\text{veh}}^{\text{mb}} + R_{\text{ped}}^{\text{mb}}\right)},$$

which is clipped within the range $[-10^5, 0]$ for numerical stability. Vehicles are considered waiting below a speed of $0.2\,\text{m/s}$ and pedestrians below a speed of $0.5\,\text{m/s}$, and the normalization constants 8 and 10 are empirically chosen. Additionally, both the state and reward statistics are updated at each action step using a Welford Normalizer [42], [43].

Category	Parameter	Value
PPO	Learning Rate (α)	1×10^{-4}
	Discount Factor (γ)	0.99
	GAE Estimation (λ)	0.95
	Clip Parameter (ϵ)	0.2
	Value Function Coeff.	0.5
	Update Frequency	1024
	K-epochs	4
Policy	Architecture	MLP
	Hidden Layers (Actor)	[512, 256, 128, 64, 32]
	Hidden Layers (Critic)	[512, 256, 128, 64, 32]
Simulation	Crossing Width	4 m
	Sidewalk Width	4 m
	Time Step (Δt)	1 second
	Action Duration	10 steps
	Warmup Timesteps	100-250
	Episode Horizon	600 steps
	Vehicle Control Model	IDM [44]
	Pedestrian Control Model	Stripping [45]
	Vehicle Speed Limit	50 km/hr
	Pedestrian Walking Speed	2.78 m/s

TABLE I: Key parameters for PPO, policy, and simulation.

IV. Experiments

A. Setup

We conduct all simulations using SUMO [46]. The reinforcement learning policy is trained using Proximal Policy Optimization (PPO) [47] with Generalized Advantage Estimation (GAE) [48]. Training occurs across 24 parallel actors over 6×10^6 simulation timesteps on an Intel Core $\text{i9} - 14900\text{KF}$ processor and an NVIDIA RTX A5000 GPU. We implement a multi-layer perceptron policy architecture with separate networks for actor and critic. During training, each episode consists of a warmup period (randomly selected between 100 and 250 timesteps) during which all signals operate on fixed-time control, followed by a 600-timestep episode horizon. To ensure robust policy learning, we randomly scale both pedestrian and vehicle demand between $1\times$ and $2.25\times$ the original demand for each episode. SUMO's dynamic routing behavior introduces additional variability by adapting vehicle and pedestrian routes based on current traffic conditions. A comprehensive list of simulation and training parameters is provided in Table I.

B. Benchmarks

We evaluate our approach against two traffic control strategies that represent common real-world implementations.

1) Unsignalized: In this benchmark, we implement all mid-block locations (MB1-MB7) as unsignalized crosswalks, closely matching the current real-world setup of the corridor. At these unsignalized crosswalks, pedestrians have the right-of-way as specified by the Uniform Vehicle Code [49]. This right-of-way behavior is implemented in our simulation using

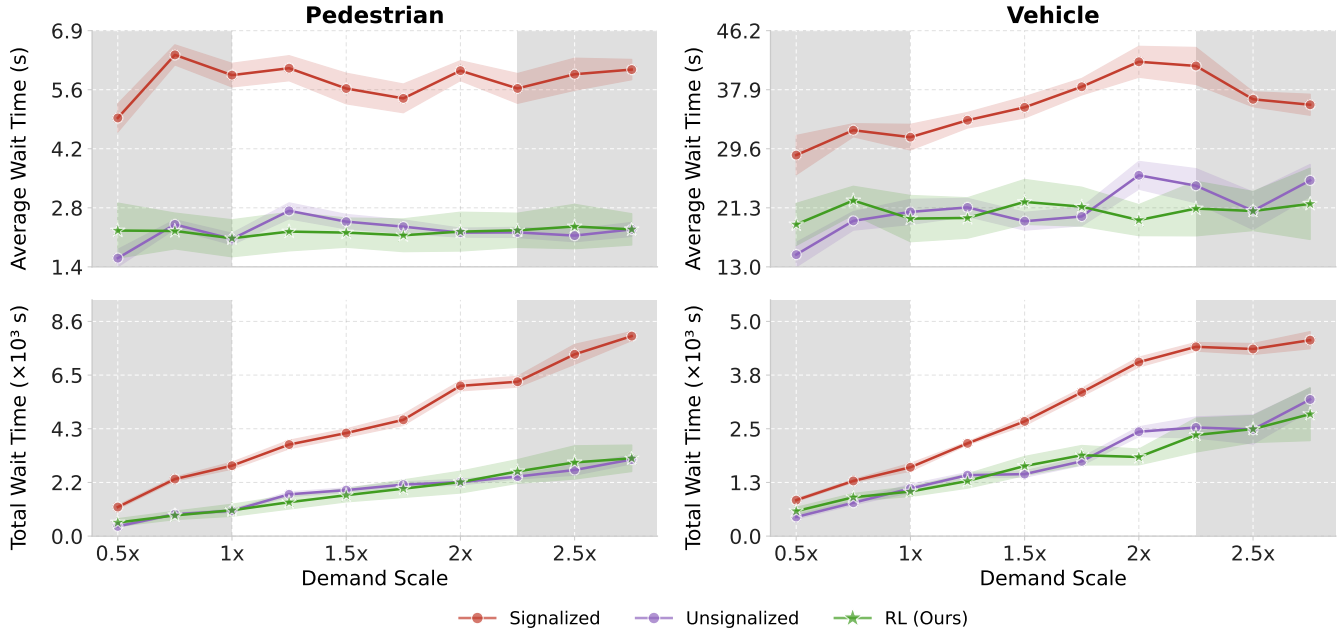

Fig. 3: Performance comparison between three traffic control approaches at various demands. The figure shows wait times for both pedestrians (left) and vehicles (right), measured as average wait time per pedestrian/vehicle (top) and total wait time for pedestrians/vehicles (bottom). Our RL agent consistently outperforms the fixed-time signal control (Signalized) across all demand levels, reducing average vehicle wait times by up to 52% (from 42.6 to 20.5 seconds per vehicle at 2× demand) while decreasing pedestrian wait times by up to 67% (from 6.0 to 2.0 seconds per pedestrian at 2× demand). Despite providing protected crossing phases through signalization, our RL approach achieves pedestrian wait times comparable to or slightly better than unsignalizing mid-block crosswalks (Unsignalized), while delivering approximately 20% lower average vehicle wait times at higher demands. The gray-shaded areas to the left and right indicate demand levels unseen during training ($< 1\times$ and $> 2.25\times$), where our approach generalizes effectively with consistent performance improvements across both low and high demands. All values are averaged across 10 independent simulation runs with a total of 600 runs.

SUMO's pedestrian interaction model [50], where vehicles must yield to pedestrians in two specific scenarios:

- Vehicles and pedestrians share the same road
- Vehicles pass through designated pedestrian crossings

The unsignalized approach represents a pedestrian-prioritized baseline that minimizes pedestrian delay at mid-block locations but may increase vehicle delay and conflicts (right-of-way negotiations) between vehicles and pedestrians. As shown in Figure 4(a), the conflicts become increasingly critical at higher traffic volumes, where unsignalized crosswalks experience up to 28.1 conflicts on average. The intersection (INT) remains signalized.

2) Signalized: This benchmark implements fixed-time control for both the intersection (INT) and mid-block (MB1-MB7) crosswalks. The signal timings are based on real-world observations and standard traffic engineering practices:

INT: Operates on a 5-phase cycle with 90-second green periods alternating between N-S and E-W through vehicle movements. Each direction change includes a 4-second yellow and 2-second all-red transition period. Consistent with the real-world implementation, the signal timing does not include dedicated left-turn phases. Complementary pedestrian crossings activate simultaneously with their corresponding vehicle phases (i.e., when N-S vehicle movement is green, the E-W pedestrian crosswalks are also green, and vice versa). We derived these timings through manual observation of video footage captured at different times of day.

MB1-MB7: Operate on a 62-second cycle with phases set according to guidelines from FHWA's Manual on Uniform Traffic Control Devices (MUTCD) [51] and Traffic Signal Timing Manual (TSTM) [52]:

- Pedestrian Phase (MUTCD 4I.06): 16 seconds consisting of a 7-second minimum interval followed by a 9-second clearance interval calculated as:

$$\text{Clearance time} = \frac{\text{Crosswalk length}}{\text{Walking speed}} = \frac{32 \text{ ft}}{3.5 \text{ ft/s}} \approx 9 \text{ s}$$

- Vehicle Phase (TSTM 6.6.3, MUTCD 4F.17): 46 seconds consisting of 40-second green time (64% of the split distribution), a 4-second yellow change interval, and a 2-second red clearance interval.

The signalized approach represents a fully-controlled safety-oriented baseline with dedicated signal phases ensuring rule compliance and eliminating vehicle-pedestrian right-of-way conflicts. Our RL approach also uses the fully-controlled setup but replaces the fixed signal timing cycles with adaptive timings controlled by the policy.

C. Results

Our reinforcement learning (RL) based approach effectively resolves the safety-efficiency trade-off in traffic control by providing the safety benefits of signalized crossings while achieving wait times comparable to or better than unsignalized crossings. As shown in Figure 3, the approach consistently outperforms *Signalized* across all demand levels, reducing average vehicle wait times by up to 52% and

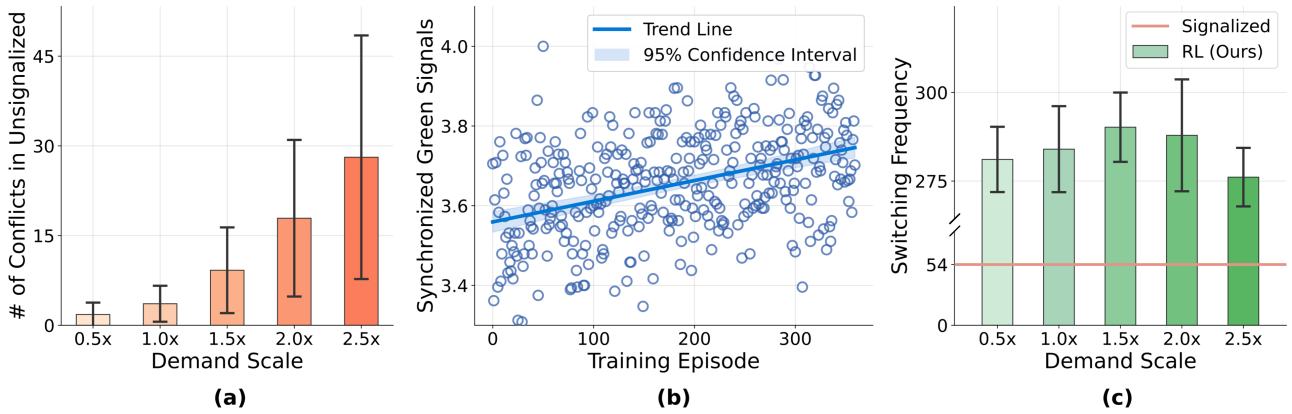

Fig. 4: (a) Vehicle-pedestrian right-of-way conflicts in unsignalized mid-block crosswalks increase substantially with traffic demand, from 1.8 conflicts at 0.5x demand to 28.1 at 2.5x demand. The error bars show standard deviation values rising at higher demand levels, indicating safety outcomes become both worse and more unpredictable as traffic volumes increase. (b) Emergence of traffic signal coordination behavior during training. The average number of mid-block traffic lights simultaneously set to green for vehicular flow shows an upward trend (from approximately 3.55 to 3.75) indicating that the RL agent gradually learns to coordinate multiple signals. (c) Total signal phase switches after the warmup period across all signals in the network. RL agent demonstrates adaptive switching, with more than fivefold increase in switching frequency compared to fixed-time signalized control. Plots (a) and (c) data are averaged over 10 runs.

pedestrian wait times by up to 67%. The approach also demonstrates strong generalization capabilities, maintaining these performance advantages at both low (0.5x, 0.75x) and high (2.5x, 2.75x) demands that were unseen during training.

Pedestrian Wait Times. At 1x demand, our approach achieves an average wait time per pedestrian of 2.1 seconds compared to 5.9 seconds for *Signalized* and 2.1 seconds for *Unsignalized*. This represents a 65% reduction compared to *Signalized* while matching *Unsignalized*. As demand increases, our approach maintains its efficiency: at 2x demand, 2.0 seconds versus 6.0 seconds for *Signalized* (67% reduction) and 2.2 seconds for *Unsignalized*. The total pedestrian wait time shows a similar pattern (up to 67% reduction), with our approach consistently outperforming both *Signalized* and *Unsignalized* across the demands. Even as pedestrian demand increases, our agent consistently maintains efficient crossing opportunities without compromising safety (no pedestrian-vehicle conflicts).

Vehicle Wait Times. At 1x demand, our approach achieves an average wait time per vehicle of 20.8 seconds compared to 31.3 seconds for *Signalized* and 20.7 seconds for *Unsignalized*. This represents a 34% reduction compared to *Signalized* while maintaining comparable performance to *Unsignalized*. This pattern continues at higher demands: at 2x demand, 20.5 seconds versus 42.6 seconds for *Signalized* (52% reduction) and 25.9 seconds for *Unsignalized*. The total vehicle wait time at 2x demand shows similar improvements: 1.16 hours for *Signalized*, 0.55 hours for our approach, and 0.56 hours for *Unsignalized*.

Generalization to Unseen Demands. As shown in the shaded (gray) regions of Figure 3, our approach maintains consistent performance advantages for both pedestrians and vehicles at both low demand (up to 0.5x) and high demand (up to 2.75x), reducing wait times by up to 67% for pedestrians and up to 39% for vehicles.

To understand the observed wait time reductions, we examine the internal behaviors of our policy and identify two emergent phenomena that explain its advantage over baselines:

Signal Coordination. Our policy autonomously learns to coordinate mid-block crosswalks, similar to the "green wave" effect observed in other RL traffic systems [53], [54]. The action space models each mid-block signal independently using a Bernoulli distribution (as described in Section III), with no built-in mechanism for coordination. Yet as shown in Figure 4(b), the RL agent learns to increase the number of mid-block traffic signals simultaneously set to vehicle green phase. The trend line shows this coordination increasing from approximately 3.55 to 3.75 synchronized green signals over the course of training—a significant shift when considering each data point represents an average over 1440 actions taken across 24 parallel actors. The coordination of signal timing reduces the number of stops and enables more efficient vehicular flow through the corridor.

Adaptive Switching Frequency. As shown in Figure 4(c), our policy exhibits more than fivefold increase in switching frequency with an average of 284 switches compared to 54 in signalized control. Additionally, the policy demonstrates adaptability across different demand scales, with standard deviations ranging from 9.23 (at 2.5x demand) to 16.42 (at 2.0x demand). Despite the policy not being explicitly incentivized to switch more (or less) often in the reward, it learned that demand-responsive and generally more frequent switching enables the system to minimize waiting times.

V. Conclusion and Future Work

In this work, we introduced a deep reinforcement learning framework for corridor-level adaptive traffic signal control that jointly optimizes for pedestrian and vehicular efficiency. We applied this framework to a real-world urban network using real-world pedestrian and vehicle demand data. Our trained policy reduces wait times significantly for both pedestrians (up to 67%) and vehicles (up to 52%) compared to traditional fixed-time signals, while generalizing to both lower and higher traffic volumes not seen during training.

While our approach demonstrates significant improvements, several limitations exist. First, under high demand, we observe vehicle queue propagation upstream when signals are closely spaced (a back-spill effect). This is likely a result of our modeling assumption, i.e., we assume signals operate independently. Second, the high switching frequency exhibited by our policy, while reducing wait times, breaks traffic flow continuity and may increase vehicle energy consumption because of more frequent stop-and-go events. Third, our evaluation was limited to a university campus setting with its specific pedestrian and vehicle patterns; generalization to heterogeneous traffic conditions, different corridor geometries including irregular intersections, and diverse urban environments remains to be validated. Fourth, while our baselines include common real-world approaches (fixed-time control and unsignalized crosswalks), direct comparison with other RL-based controllers proved challenging: existing RL methods either exclude pedestrian considerations entirely or require separate agents at each intersection with distributed training infrastructure and communication protocols, whereas our approach uses a single policy controlling all eight signals through centralized training.

Future work could address these limitations by explicitly modeling correlations between adjacent signals, incorporating traffic flow continuity or energy consumption metrics into the reward function, and evaluating performance across diverse urban settings with varying pedestrian-vehicle ratios and geometric configurations. Additional future directions include mixed-traffic control [55], [56], adversarial robustness evaluation [57], and real-world traffic disruption analysis [58], [59].

ACKNOWLEDGMENTS

This research is supported by NSF IIS-2153426 and ECCS-2332210. The authors also thank NVIDIA and the Tickle College of Engineering at University of Tennessee, Knoxville for their support.

REFERENCES

[1] 2023 urban mobility report. Technical report, Texas Department of Transportation, 2023. Accessed: 2025-02-15.

[2] Secretary of Transportation, Pete Buttigieg. Budget highlights 2022. Technical report, U.S. Department of Transportation, 2021. Accessed: 2025-02-15.

[3] Inrix 2024 global traffic scorecard: Employees & consumers returned to downtowns, traffic delays & costs grew. https://inrix.com/press-releases/2024-global-traffic-scorecard-us/, 2024. Press Release. Accessed: 2025-02-15.

[4] Yi Zhao and Zong Tian. An overview of the usage of adaptive signal control system in the united states of america. *Applied Mechanics and Materials*, 178:2591–2598, 2012.

[5] Xingmin Wang, Zachary Jerome, Zihao Wang, Chenhao Zhang, Shengyin Shen, Vivek Vijaya Kumar, Fan Bai, Paul Krajewski, Danielle Deneau, Ahmad Jawad, et al. Traffic light optimization with low penetration rate vehicle trajectory data. *Nature communications*, 15(1):1306, 2024.

[6] Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.

[7] Yang Xiao, Jun Liu, Jiawei Wu, and Nirwan Ansari. Leveraging deep reinforcement learning for traffic engineering: A survey. *IEEE Communications Surveys & Tutorials*, 23(4):2064–2097, 2021.

[8] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. *ACM SIGKDD explorations newsletter*, 22(2):12–18, 2021.

[9] Governors Highway Safety Association. Pedestrian traffic fatalities by state: 2022 preliminary data. Technical report, Governors Highway Safety Association, 2023.

[10] National Highway Traffic Safety Administration. Traffic safety facts: 2021 data. Technical Report DOT HS 813 375, U.S. Department of Transportation, 2023.

[11] National Safety Council. Injury facts: Pedestrians. Technical report, National Safety Council, 2023. Analysis of NHTSA Fatality Analysis Reporting System (FARS) data.

[12] Charles V Zegeer, J Richard Stewart, Herman Huang, and Peter Lagerwey. Safety effects of marked versus unmarked crosswalks at uncontrolled locations: analysis of pedestrian crashes in 30 cities. *Transportation research record*, 1773(1):56–68, 2001.

[13] Kay Fitzpatrick, Vichika Iragavarapu, Marcus Brewer, Dominique Lord, Joan G Hudson, Raul Avelar, and James Robertson. Characteristics of texas pedestrian crashes and evaluation of driver yielding at pedestrian treatments. 2014.

[14] Yi Zhang, Kaizhou Gao, Yicheng Zhang, and Rong Su. Traffic light scheduling for pedestrian-vehicle mixed-flow networks. *IEEE Transactions on Intelligent Transportation Systems*, 20(4):1468–1483, 2018.

[15] Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, Hua Wei, Huichu Zhang, Yong Li, Kai Xu, and Zhenhui Li. Learning phase competition for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1963–1972, 2019.

[16] Ammar Haydari and Yasin Yılmaz. Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):11–32, 2020.

[17] PB Hunt, DI Robertson, RD Bretherton, and M Cr Royle. The scoot on-line traffic signal optimisation technique. *Traffic Engineering & Control*, 23(4), 1982.

[18] Arthur G Sims and Kenneth W Dobinson. The sydney coordinated adaptive traffic (scat) system philosophy and benefits. *IEEE Transactions on vehicular technology*, 29(2):130–137, 1980.

[19] Markos Papageorgiou, Christina Diakaki, Vaya Dinopoulou, Apostolos Kotsialos, and Yibing Wang. Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12):2043–2067, 2003.

[20] Dipti Srinivasan, Min Chee Choy, and Ruey Long Cheu. Neural networks for real-time traffic signal control. *IEEE Transactions on intelligent transportation systems*, 7(3):261–272, 2006.

[21] Wade Genders and Saiedeh Razavi. Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142*, 2016.

[22] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE transactions on intelligent transportation systems*, 21(3):1086–1095, 2019.

[23] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2496–2505, 2018.

[24] Yiming Bie, Yuting Ji, and Dongfang Ma. Multi-agent deep reinforcement learning collaborative traffic signal control method considering intersection heterogeneity. *Transportation Research Part C: Emerging Technologies*, 164:104663, 2024.

[25] Alexander Genser, Michail A Makridis, Kaidi Yang, Lukas Ambühl, Monica Menendez, and Anastasios Kouvelas. Time-to-green predictions for fully-actuated signal control systems with supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 25(7):7417–7430, 2024.

[26] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems*, 11(7):417–423, 2017.

[27] Li Li, Yisheng Lv, and Fei-Yue Wang. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica*, 3(3):247–254, 2016.

[28] Elise Van der Pol and Frans A Oliehoek. Coordinated deep reinforcement learners for traffic light control. *Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016)*, 8:21–38, 2016.

[29] Patrick Mannion, Jim Duggan, and Enda Howley. An experimental review of reinforcement learning algorithms for adaptive traffic signal control. *Autonomic road transport support systems*, pages 47–66, 2016.

[30] Xiaoyuan Liang, Xunsheng Du, Guiling Wang, and Zhu Han. Deep reinforcement learning for traffic light control in vehicular networks. *arXiv preprint arXiv:1803.11115*, 2018.

[31] Yi Zhang, Yicheng Zhang, and Rong Su. Pedestrian-safety-aware traffic light control strategy for urban traffic congestion alleviation. *IEEE Transactions on Intelligent Transportation Systems*, 22(1):178–193, 2019.

[32] Guangjie Han, Qi Zheng, Lyuchao Liao, Penghao Tang, Zhengrong Li, and Yintian Zhu. Deep reinforcement learning for intersection signal control considering pedestrian behavior. *Electronics*, 11(21):3519, 2022.

[33] Wanjing Ma, Yue Liu, and K Larry Head. Optimization of pedestrian phase patterns at signalized intersections: a multi-objective approach. *Journal of advanced transportation*, 48(8):1138–1152, 2014.

[34] Tong Wu, Pan Zhou, Kai Liu, Yali Yuan, Xiumin Wang, Huawei Huang, and Dapeng Oliver Wu. Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks. *IEEE Transactions on Vehicular Technology*, 69(8):8243–8256, 2020.

[35] Abhilasha Jairam Saroj, Yu Liang, Dalei Wu, Michael P Hunter, Mina Sartipi, et al. Pedestrian-involved traffic signal optimization using decentralized graph-based multi-agent reinforcement learning. Technical report, University of Tennessee at Chattanooga, 2024.

[36] University of North Carolina at Charlotte. About us. https://admissions.charlotte.edu/explore/. Accessed: February 25, 2025.

[37] Yuqiu Yuan, Lei Zhu, and Mohini Joshi. A hierarchical wi-fi log data processing framework for human mobility analysis in multiple real-world communities. *Travel Behaviour and Society*, 39:100985, 2025.

[38] Shiyu Zhang, Bangchao Deng, and Dingqi Yang. Crowdtelescope: Wi-fi-positioning-based multi-grained spatiotemporal crowd flow prediction for smart campus. *CCF Transactions on Pervasive Computing and Interaction*, 5(1):31–44, 2023.

[39] Daniel Krajzewicz, Georg Hertkorn, Christian Rössel, and Peter Wagner. Sumo (simulation of urban mobility)-an open-source traffic simulation. In *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002)*, pages 183–187, 2002.

[40] Liang Zhang, Qiang Wu, Jun Shen, Linyuan Lü, Bo Du, and Jianqing Wu. Expression might be enough: representing pressure and demand for reinforcement learning based traffic signal control. In *International Conference on Machine Learning*, pages 26645–26654. PMLR, 2022.

[41] Behrad Koohy, Sebastian Stein, Enrico Gerding, and Ghaithaa Manla. Reward function design in multi-agent reinforcement learning for traffic signal control. 2022.

[42] Shengyi Huang, Rousslan Fernand Julien Dossa, Antonin Raffin, Anssi Kanervisto, and Weixun Wang. The 37 implementation details of proximal policy optimization. *The ICLR Blog Track 2023*, 2022.

[43] Jeremiah Coholich. A bag of tricks for deep reinforcement learning, 2023. Accessed: 2025-02-21.

[44] Martin Treiber and Arne Kesting. *Traffic flow dynamics*, volume 1. Springer, 2013.

[45] Jakob Erdmann and Daniel Krajzewicz. Modelling pedestrian dynamics in sumo. *SUMO 2015-Intermodal Simulation for Intermodal Transport*, 28:103–118, 2015.

[46] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements*, 5(3&4), 2012.

[47] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[48] John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*, 2015.

[49] National Committee on Uniform Traffic Laws and Ordinances. *Uniform Vehicle Code: Millennium Edition*. National Committee on Uniform Traffic Laws and Ordinances, Alexandria, VA, 2000.

[50] DLR and contributors. SUMO Documentation: Pedestrians, 2025. Accessed: 2025-02-27.

[51] Federal Highway Administration. *Manual on Uniform Traffic Control Devices for Streets and Highways*. U.S. Department of Transportation, 11th edition, 2023.

[52] Peter Koonce et al. Traffic signal timing manual. Technical report, United States. Federal Highway Administration, 2008.

[53] Hua Wei, Chacha Chen, Kan Wu, Guanjie Zheng, Zhengyao Yu, Vikash Gayah, and Zhenhui Li. Deep reinforcement learning for traffic signal control along arterials. *Proceedings of the 2019, DRL4KDD*, 19, 2019.

[54] Xiao-Yang Liu, Ming Zhu, Sem Borst, and Anwar Walid. Deep reinforcement learning for traffic light control in intelligent transportation systems. *arXiv preprint arXiv:2302.03669*, 2023.

[55] Dawei Wang, Weizi Li, Lei Zhu, and Jia Pan. Learning to control and coordinate mixed traffic through robot vehicles at complex and unsignalized intersections. *The International Journal of Robotics Research*, page 02783649241284069, 2024.

[56] Michael Villarreal, Bibek Poudel, Jia Pan, and Weizi Li. Mixed traffic control and coordination from pixels. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4488–4494. IEEE, 2024.

[57] Bibek Poudel and Weizi Li. Black-box adversarial attacks on network-wide multi-step traffic state prediction models. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 3652–3658. IEEE, 2021.

[58] Bibek Poudel, Weizi Li, and Kevin Heaslip. Endurl: Enhancing safety, stability, and efficiency of mixed traffic under real-world perturbations via reinforcement learning. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

[59] Bibek Poudel, Weizi Li, and Shuai Li. Carl: Congestion-aware reinforcement learning for imitation-based perturbations in mixed traffic control. In *2024 IEEE 14th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 7–14. IEEE, 2024.