2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Mixed Traffic Control and Coordination from Pixels

Michael Villarreal[1], Bibek Poudel[1], Jia Pan[2], Weizi Li[1]

Abstract— **Traffic congestion is a persistent problem in our society. Previous methods for traffic control have proven futile in alleviating current congestion levels leading researchers to explore ideas with robot vehicles given the increased emergence of vehicles with different levels of autonomy on our roads. This gives rise to mixed traffic control, where robot vehicles regulate human-driven vehicles through reinforcement learning (RL). However, most existing studies use precise observations that require domain expertise and hand engineering for each road network's observation space. Additionally, precise observations use global information, such as environment outflow, and local information, i.e., vehicle positions and velocities. Obtaining this information requires updating existing road infrastructure with vast sensor environments and communication to potentially unwilling human drivers. We consider image observations, a modality that has not been extensively explored for mixed traffic control via RL, as the alternative: 1) images do not require a complete re-imagination of the observation space from environment to environment; 2) images are ubiquitous through satellite imagery, in-car camera systems, and traffic monitoring systems; and 3) images only require communication to equipment. In this work, we show robot vehicles using image observations can achieve competitive performance to using precise information on environments, including ring, figure eight, intersection, merge, and bottleneck. In certain scenarios, our approach even outperforms using precision observations, e.g., up to 8% increase in average vehicle velocity in the merge environment, despite only using local traffic information as opposed to global traffic information.**

I. Introduction

Traffic congestion is a prevalent challenge in modern society, causing delays, gridlocks, and substantial economic losses. Traditional traffic management methods such as traffic lights, stop signs, and ramp meters have proven insufficient in alleviating the current level of congestion [1], [2]. As more vehicles with varying degrees of autonomy are introduced into our transportation system, the idea of mixed traffic control, which involves the use of robot vehicles (RVs) to regulate human-driven vehicles (HVs), is gaining popularity as a potential solution. Studies have shown the effectiveness of this approach in stabilizing traffic on roads of different configurations, including ring and figure-eight roads [3], merge and bottleneck roads [4], intersections [4], [5], [6], [7], [8]. Among various control methods for mixed traffic, reinforcement learning (RL) has emerged as a promising tool, as it can handle the complex behaviors of mixed traffic without using predefined models or heuristics [9].

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

[2]Jia Pan is with the Department of Computer Science at the University of Hong Kong, China jpan@cs.hku.hk

Existing studies of mixed traffic control via RL predominantly uses precise traffic conditions as policy input [4], [5], [10], [11], [12]: the RV receives both exact global information such as network throughput and travel time as well as exact local information such as nearby vehicles' positions and velocities. While effective, *using precise observations necessitates completely re-designing the observation space across different road environments* [3], [4], [5], [13], which requires costly hand-engineering and domain expertise. For example, the figure-eight environment (Fig. 1B) uses all vehicles' positions and velocities, while the bottleneck environment (Fig. 2) uses averaged position and velocity of HVs and RVs in combination with network outflow.

In practice, for RV to obtain accurate global information, road sensor infrastructure is needed for data collection. Overhauling current road networks for this purpose requires substantial expenses. To receive precise local information, RV needs to establish vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. V2I again needs augmented infrastructure with a multitude of sensors. V2V, on the other hand, requires HVs to broadcast precise traffic information and engage in constant communication, which is difficult to achieve. An alternative to avoid these pitfalls is using image observations (instead of precise observations), a modality commonly seen in robotics research [14], [15], [16], [17], but rarely in mixed traffic control.

In this work, we use bird's-eye view images centered on RV as input to RL policies for mixed traffic control (see Fig. 1). The images have generic resolutions and only capture local traffic information. Our approach enjoys several benefits. First, using images as input enables end-to-end training, thus *avoiding the need for manually designing observation spaces*. The process of capturing image observations can be repeated over different road environments. Second, using images can *enable RVs to generalize to new environments as it omits the global information of road networks*. This feature is particularly useful since the V2I support (which is required to gain the global information) could vary significantly in different areas. Third, HVs are relieved from V2V communication—*a setting greatly enhances the practicality of mixed traffic control*. Fourth, imagery about traffic conditions is ubiquitous. Satellite imagery can capture traffic in both cities and rural areas, where communication is sparse. Our road infrastructure is equipped with ubiquitous camera systems [18], [19], which provide real-time images of traffic. Furthermore, modern cars' cameras can capture 360° view of the surroundings, which can be used to develop image observations in real-time [20], [21], [22], [23]. The effectiveness of our approach is demonstrated via comprehensive

979-8-3503-8457-4/24/$31.00 ©2024 IEEE

experiments. In summary, our contributions are as follows.

- We use image observations as policy input for RVs in mixed traffic control. These images are both generic and local: they only record the local surroundings of the RVs. In contrast, global information is needed by the RVs for control when using precise observations.
- We demonstrate the same-level performance of our approach to using precise observations on various road environments, including a ring road, a figure-eight road, and a merge scene.
- We achieve improved performance in several cases compared to using precise observations, e.g., an 8% increase in average vehicle velocity in the merge environment.

To the best of our knowledge, our work is the first to perform extensive experimentation on various road environments to demonstrate the feasibility of using image observations for RL-based mixed traffic control in alleviating traffic congestion.

II. Related Work

A significant portion of training RVs via RL with images focuses on individual vehicle driving but not controlling entire traffic [24], [25]. For example, images are used with vision transformers to learn an effective driving policy [26], to train RVs to drive in simulation [27], [28], [29], [30], or to prevent crashes by capturing the RVs' surroundings [31]. While these studies apply RL and images to RVs, they do not concentrate on traffic control.

Wu et al. [3] pioneer mixed traffic control using RL. They show the effectiveness of training an RV on smoothing out stop-and-go waves on a ring road. Further tests are conducted by Vinitsky et al. [4] on additional environments, including merge, bottleneck, and intersection scenarios. Recently, Wang et al. [6] manage to scale up mixed traffic control to real-world, complex intersections while controlling and coordinating hundreds of vehicles. While significant advancements have been made in mixed traffic control, all these studies use precise observations as inputs to the RL policy. These precise observations include both global and local traffic conditions, such as environment outflow and vehicle position and velocity.

Our work replaces these precise observations with image observations in the training of RL policies for mixed traffic control. This shift to image observations has several benefits. First, it eliminates the requirement for manual design of the observation space for different traffic scenarios, making it a more flexible solution. Second, it can leverage existing traffic infrastructure as image data is readily available from sources such as satellite imagery, traffic monitoring systems, and vehicle surround-view cameras.

Prior research use bird's-eye view (BEV) images (which is our image observations) with RVs. For example, significant research develop BEV images for 3D object detection and segmentation using modern cars' multi-camera vision systems [20], [21], [22]. Another work by Huang et al. [23] develops a framework for real-time BEV image perception using onboard vehicle chips. This research allows using modern vehicles' camera systems for mixed traffic control at the pace needed to take actions.

Several studies explore mixed traffic control with images. However, there are key differences between their efforts and this project. A prior work presents a decentralized method for training RVs with images [32], while another shows human driving can be positively augmented using an RL controller trained on local images [33]. Both studies only concern the ring and/or figure eight environments and the focus is not alleviating traffic congestion in varied road environments.

III. Methodology

We introduce our problem formulation and then outline the details (observation and reward) for our five road environments. We also provide details on precise observations [3], [4] and compare them to image observations.

A. Preliminaries

We model the RL problem as a Partially Observable Markov Decision Process (POMDP) represented by a tuple $(\mathscr{S}, \mathscr{A}, \mathscr{P}, \mathscr{R}, p_0, \gamma, T, \Omega, \mathscr{O})$ where: $\mathscr{S}$ is the state space; $\mathscr{A}$ is the action space; $\mathscr{P}(s'|s,a)$ is the transition probability function; $\mathscr{R}$ is the reward function; p_0 is the initial state distribution; $\gamma \in (0,1]$ is the discount factor; T is the episode length (horizon); Ω is the observation space; and $\mathscr{O}$ is the probability distribution of retrieving an observation $\omega \in \Omega$ from a state $s \in \mathscr{S}$. At each timestep $t \in [1,T]$, a robot vehicle (RV) uses its policy $\pi_\theta(a_t|s_t)$ to take an action $a_t \in \mathscr{A}$, given the state $s_t \in \mathscr{S}$. The RV's environment provides feedback on action a_t by calculating a reward r_t and transitioning the agent into the next state s_{t+1}. The RV's goal is to learn a policy π_θ that maximizes the discounted sum of rewards, i.e., return, $R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i$. We use Proximal Policy Optimization [34], a model-free, on-policy algorithm, to learn the optimal policy.

B. Road Environments

We train RVs using RL on five road environments (ring, figure eight, intersection, merge, and bottleneck) shown in Fig. 1 and Fig. 2 using image observations. These environments originate from FLOW [3], a RL framework for traffic management. We give a brief environment description, compare the differences between image observations and precise observations, and provide the reward functions.

1) Ring: The ring environment is a widely used benchmark in traffic control [3], [9]. A single-lane circular road environment consisting of 22 vehicles with 21 human-driven vehicles (HVs) and one RV. For 3000 warmup timesteps, the 22 vehicles act as HVs. During this warmup, subtle perturbations from imperfect human driving behavior can amplify leading some vehicle standstill. This situation is stop-and-go traffic and acts as a wave backpropagating continually through the ring. After the warmup period, the RV begins taking actions for 3000 timesteps with the goal to dampen and prevent these waves, thus increasing overall average vehicle velocity.

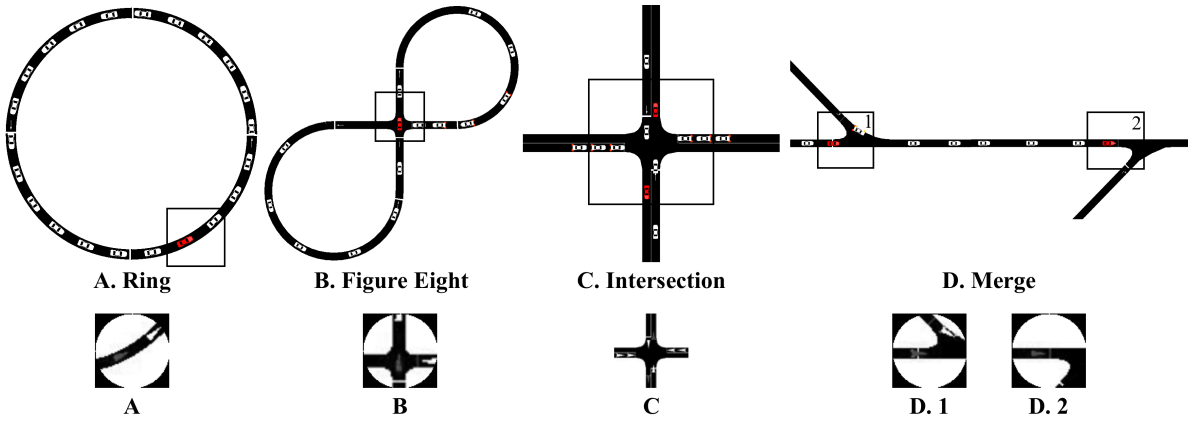

Fig. 1. We experiment on five mixed traffic control environments (bottleneck shown in Fig 2), with image observations presented beneath them. Robot vehicles (RVs) are red, while human-driven vehicles (HVs) are white. With image observations, HVs are cyan to provide contrast from the white background. We use static, grayscale, 84×84 images centered over RVs (or intersection) that provide only local information. Merge and bottleneck are multi-agent, while the other three are single agent.

Our observation is a gray-scale image of dimensions 84×84 pixels, centered on a single RV as shown in Fig. 1. To simulate limited visibility, the image is masked by a circle with a radius corresponding to 28.75 meters in real world. Precise observations are a vector of the RV's velocity, the difference between the leading vehicle's velocity and the RV's velocity, and the difference between the leading vehicle's position and the RV's position. This precise observation space has been used to produce state-of-the-art performance [3]. The action space is the continuous acceleration $[-1,1]$ m/s^2. The reward function encourages high average velocity and small control actions (acceleration) through a weighted combination:

$$\mathrm{r} = \frac{1}{n}\sum_i v_i - \alpha * |a_{RV}|, \tag{1}$$

where $n = 22$, v is vehicle velocity, α is four (chosen empirically), and a_{RV} is the RV's acceleration.

2) Figure Eight: The figure eight environment simulates an intersection in a closed loop with 14 vehicles–13 HVs and one RV. From its shape and the number of vehicles in the environment, queues form among cars trying to cross the intersection. This causes a environment-wide decrease in average vehicle velocity. The RV's objective, over 1500 timesteps, is to increase all vehicle average velocity.

Our observation is the same as of the ring environment (see Sec. III-B.1) with an example given in Fig. 1B. The masked circle's dimension corresponds to a 21.25 m radius. Precise observations are the velocities and positions of all vehicles within the figure-eight environment [3]. This complete information reflects that the state space is used. As our observations are local, i.e., images centered on the RV, learning an optimal policy includes additional, challenging steps of perception and representation learning compared to precise observations. The action space is the continuous acceleration $[-3,3]$ m/s^2. The reward function aims to increase all vehicles' velocity in the environment:

$$\mathrm{r} = \frac{max(\|v_{des} * \mathbb{1}^k\|_2 - \|v_{des} - V_{\mathrm{all}}\|_2, 0)}{\|v_{des} * \mathbb{1}^k\|_2}, \tag{2}$$

where v_{des} (desired velocity) is 10 m/s (chosen empirically) and k is the total number of vehicles in the environment.

3) Intersection: Intersection is an idealized two-way stop where east/westbound traffic flow (500 vehicles/hour) is less than north/southbound traffic flow (1333 vehicles/hour). This flow difference causes east/westbound queues as it would otherwise be unsafe to cross the intersection. RVs are placed in the north/south directions with a penetration rate of 20%. The RVs take actions, over 400 timesteps, to minimize queue formation and increase average vehicle velocity along the east/west directions. This environment allows for studying mixed traffic control in directions absent of RVs since RVs control only the north/south directions.

Our observation is an 84×84, grayscale image (shown in Fig. 1) taken solely at the intersection's center. No circle mask is applied to the images. The image dimensions correspond to $50\mathrm{m} \times 50\mathrm{m}$. Precise observations include global and local traffic information, and consider a user-defined number of vehicles closest to the intersection [4]. Specifically, precise observations contain each vehicle's velocity, each vehicle's position distance to the intersection, each vehicle's edge number (this identifies if the vehicle is in east/west traffic or in north/south traffic), each edge's density, and each edge's average vehicle velocity. Our observation is centered on the intersection for a fair comparison as precise observations collect information using the center as a focal point. However, we envision image observations being collected from the RVs or a fusion between RVs and infrastructure. Our RVs face a more difficult learning task due to their limited radius observations preventing them from inferring environment-wide information. Actions taken are defined by the continuous acceleration $[-7,7]$ m/s^2. The reward function penalizes both vehicle delay and vehicle standstills in traffic:

$$\mathrm{r} = -\frac{t * \sum((V_{\max} - V_{\mathrm{all}})/V_{\max})}{\mathrm{n} + \mathrm{eps}} - (\mathrm{gain} * ss_n), \tag{3}$$

where t is current timestep, V_{max} is a vector of intersection's speed limit, V_{all} is a all vehicle velocity, n is number of vehicles, eps prevents zero division, $gain$ is 0.2, and ss_n is the number of standstill vehicles. Given the reward is negative, the RVs' goal is to minimize delay and vehicle standstills.

4) Merge: The merge environment contains a highway and two merging on-ramps. We expand on the original

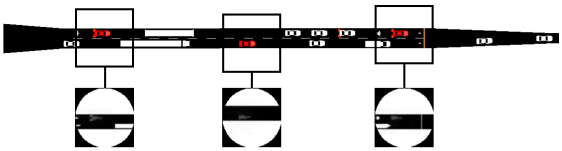

Fig. 2. Bottleneck environment with heterogeneous human-driven traffic. We add motorcycles (behind leftmost and rightmost RVs), public buses (in front of leftmost RV), semi-trucks (right of public bus), and delivery trucks (diagonally behind the rightmost RV) alongside regular passenger vehicles.

environment [4], which only contains the one, right-side on-ramp. The highway and on-ramps respective flows can create stop-and-go waves along the highway and congest the on-ramps, reducing the average velocity and outflow (vehicles/hour). The RVs' goal, for 750 timesteps, is to minimize such wave formation and increase average vehicle velocity at a 10% penetration rate.

Our observation is a stack of five images, each of size 84×84 (shown in Fig. 1 D.), centered on the RVs. We observe at most five RVs. If there are less than five RVs present, the remaining stack is padded with black images; if more, the extra RVs are treated as HVs. The image dimensions correspond to 41.25 m in real world. Precise observations are a vector of the following [4]: the velocities of the following and leading vehicle for each RV; the difference in positions between the RV and the following and leading vehicles; and the velocity of each RV. HVs on the on-ramp are observed only if they are following vehicles or have merged onto the highway. The action space is the continuous acceleration $[-1.5, 1.5]$ m/s^2. The reward function is:

$$\mathrm{r} = \text{Eq. } 2 - \alpha \sum_{i \epsilon RVs} \max[h_{\max} - h_i(t), 0], \tag{4}$$

where h_{max} is empirically set to one and $h_i(t)$ is the headway (the time distance between two consecutive vehicles) of an RV at t. The latter half's objective is to penalize small headways between a RV and a HV to discourage traffic bunching, potentially causing stop-and-go waves.

5) Heterogeneous Bottleneck: We also experiment on a heterogeneous bottleneck environment (Fig. 2). The original bottleneck environment [4], [11] has only four-door passenger vehicles and simulates vehicles experiencing capacity drop [35] on a bridge where an environment's outflow significantly decreases after the environment inflow surpasses a threshold. The capacity drop comes from the lanes decreasing from $4 \times l$ to $2 \times l$ to l (where l is a scaling factor and is one for our work). We expand the environment to include heterogeneous HVs comprised (percentage of HVs) of four-door passenger vehicles (70%), semi-trucks (10%), motorcycles (10%), delivery trucks (5%), and public buses (5%) to better reflect bridge traffic. RVs are only four-door passenger vehicles. The penetration rate of the RV is 10%. The RVs' objective is improved outflow in 1000 timesteps with 40 prior warmup timesteps.

Our observation consists of 15 stacked images, each of size 84×84 (shown in Fig. 1), as a maximum of 15 RVs are placed in the environment. If there are less than 15 RVs, the remaining stack is filled with black images; if more than 15 RVs, additional RVs are treated as HVs. The image dimensions correspond to a circle with a radius of 25 m in real world. Precise observations (collected on user-defined road segments) contain: mean positions and velocities of HVs, mean positions and velocities of RVs, and environment outflow over the last twenty seconds. This observation is difficult to design, consisting of macroscopic and microscopic traffic statistics. Global information is considered given segments are examined rather than individual vehicles, which contrasts with out observations containing only local information. Unlike previously defined environments where the action is an acceleration range, here the action space is the RVs' velocity $[0.01, 23]$ m/s. The reward's objective is to increase the outflow of the environment and reduce the frequency of capacity drops:

$$\mathrm{r} = o_{10}, \tag{5}$$

where o_{10} is outflow over the last 10 seconds.

IV. Experiments and Results

A. Experiment Setup

We train RVs using Proximal Policy Optimization [34], with default hyperparameters from RLlib [36]. HVs are operated by Intelligent Driver Model (IDM) [37] with stochastic noise in the range $[-0.2, 0.2]$ added to account for heterogeneous driving behaviors. RVs are trained for 200 episodes.

Trained policies are evaluated for 10 rollouts, and results are presented as averages. The policies are convolutional neural networks with filters (formatted as [out channels, kernel size, stride]) of $[16, 8, 4]$, $[32, 4, 2]$, and $[256, 11, 1]$ followed by two fully-connected layers. Experiments are conducted using i9-13900k CPU with 64GB RAM.

B. Results

1) Ring: We train the RV on rings with circumference sampled uniformly from $[220, 270]$ m ring length ($[81.25, 100]$ density in Fig. 3). For testing, this range is extended to $[210, 290]$ m ring length ($[76, 104.5]$ density in Fig. 3). The results are shown in Fig. 3 LEFT. RVs trained using image observations (blue) prevents stop-and-go waves and achieve the same-level performance as RVs trained using precise observations (red) at all densities.

Fig. 3 MIDDLE shows the time-space diagram of all vehicles over an episode. The shockwaves from 200 to 300 seconds (one second equals 10 timesteps) reflect stop-and-go traffic. After 300 seconds, the image-trained RV takes actions by briefly accelerating and then stabilizes the traffic.

2) Figure Eight: In prior work [3], [4], the RV is trained only on a single, inner-loop radius [3] (inner-loop radius is used to calculate the overall environment length). We expand the scenario by training on the range $[20, 30]$ m ($[33, 49]$ density in Fig. 4 LEFT) and expand this range to $[18, 32]$ m ($[31, 54]$ density in Fig. 4 LEFT) during evaluation. The efficacy of a trained RV is measured by the average vehicle velocity at a particular traffic density.

Fig. 4 LEFT shows that an image-trained RV can achieve the same level of mixed traffic control as a precise-trained

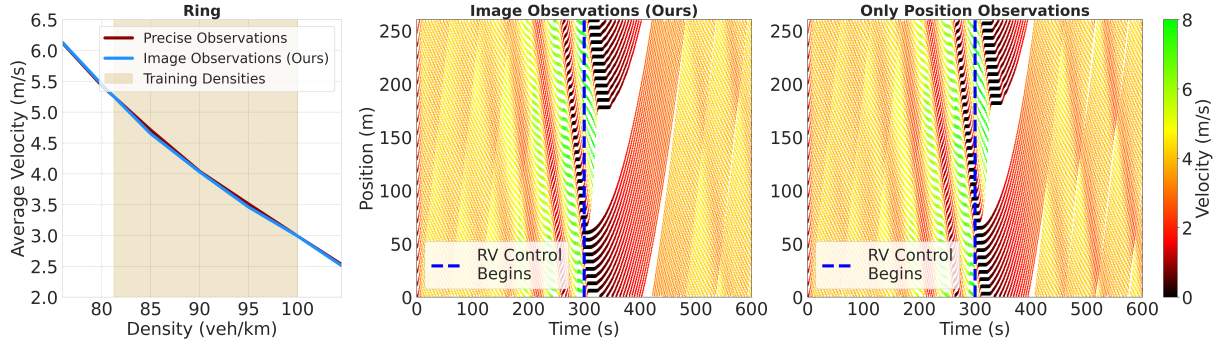

Fig. 3. LEFT: An RV using image observations prevents stop-and-go waves at all densities, same as an RV using precise observations. MIDDLE and RIGHT: Time-space diagrams showing stop-and-go waves (which form around 200 to 300 seconds) being alleviated after RVs start taking action. MIDDLE: An RV trained on image observations prevents stop-and-go waves similar to an RV trained on precise observations. RIGHT: An RV trained using only position information can also prevent stop-and-go waves. This gives further validity of using image observations without explicitly including the velocity information in preventing stop-and-go waves.

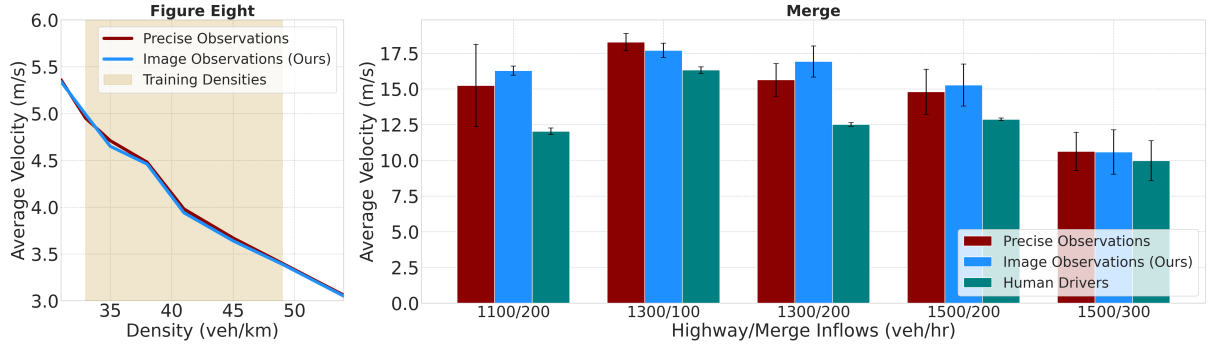

Fig. 4. LEFT: An RV using image observations achieves mixed traffic control comparable to an RV with precise observations in figure eight. RIGHT: Overall, RVs with image observations outperform RVs with precise observations in 1100/200, 1300/200, and 1500/200 by up to 8%.

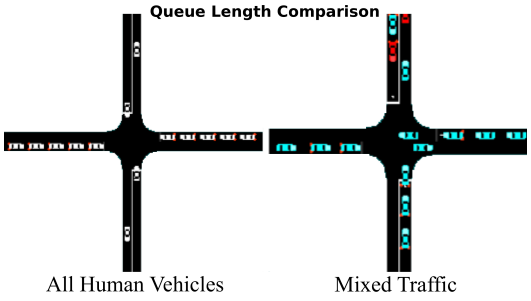

Fig. 5. Comparison between the queue lengths at the end of an episode between all human drivers (cyan in mixed traffic) and mixed traffic using image observations. RVs trained with image observations lessen east/westbound congestion by decreasing queue lengths by two vehicles.

RV. This same-level performance is in spite of the image-trained RV only receiving local information compared to the precise-trained RV having complete global/state information.

3) Intersection: We consider the average all vehicle velocity and east/westbound queue lengths. RVs with image observations attain 4.75 ± 0.02 m/s average vehicle velocity with a three vehicle queue length. RVs with precise observations obtain 5.90 ± 0.23 m/s average vehicle velocity with a three vehicle queue length. RVs with image observations can provide similar performance to RVs with precise observations in regard to queue length. The average vehicle velocity of RVs with image observations is less than RVs with precise observations. We believe this performance difference is due to precise global observations knowing exactly what edges vehicles are on and their corresponding velocities, which allow the RVs to know when HVs are at standstill in the east/west directions. Both RV types outperform HVs (3.50 ± 0.00 m/s average velocity; five vehicle queue length) in both evaluation metrics. Fig. 5 illustrates the development of queues with only HVs versus RVs with image observations. The southbound RV in the right image is slowing down momentarily, which allows east/westbound HVs to safely cross the intersection. Only HVs travel at velocities that cause queue development in the east/west directions.

4) Merge: Fig. 4 RIGHT presents our results. We evaluate merge using five combinations of highway/merge inflow rates (x-axis), {1100/200, 1300/100, 1300/200, 1500/200, 1500/300}. Merging on-ramps share inflow rates, and we compare average vehicle velocities (y-axis) at those inflows. This particular network setup and different inflow combinations that cause varying congestion levels have not been previously studied.

RVs with image observations outperform RVs with precise observations at 1100/200, 1300/200, and 1500/200 inflow rates by 7%, 8%, and 3%, respectively. RVs with precise observations outperform RVs with image observations in the 1300/100 inflows scenario, while both of them have similar performance in the 1500/300 inflows scenario. RVs with image observations provide the largest performance improvement with a 1.29 m/s–an 8% increase–over using precise observations in the 1300/200 scenario. The 1500/300 scenario is difficult to learn on (evidenced by both RV types improving performance over HVs the least)

as the inflow rates cause sufficient congestion inhibiting the RVs' potential to increase traffic flow from taking intelligent actions. Overall, RVs with image observations outperform RVs with precise observations.

5) Heterogeneous Bottleneck: We train on two inflows, $\{2300, 2500\}$, and compare outflow over the last 500 seconds. The different inflows allow for capturing different congestion levels that allow for improvement through mixed traffic control. At 2300 and 2500 inflows, RVs with image observations obtain 1497.60 ± 26.94 and 1506.96 ± 29.17 outflows, respectively, while RVs with precise observations obtain 1528.56 ± 49.26 and 1513.44 ± 24.90 outflows, respectively. Both outperform HVs at 2300 and 2500 inflows, which achieves 1448.64 ± 23.40 and 1447.20 ± 14.04 outflows, respectively. RVs with precise observations outperform RVs with image observations at both inflow rates. Although at 2500 inflow, RVs with image observations achieve an outflow close to RVs with precise observations. We hypothesize that RVs with precise observations outperform RVs with image observations because precise observations contain network-wide traffic information, while image observations only contain local traffic state information.

6) Only Position Observations: We conduct an additional experiment to test position-only observations in training RVs using precise information in the ring environment. The purpose is to analyze whether RVs can still be leveraged to alleviate traffic congestion given only static positional information, similar to positional inference using image observations. Precise observations change to a vector of the difference between the RV's position and the leading vehicle's position. Fig. 3 RIGHT shows the time-space diagram for this experiment. A RV with only position information can achieve the same level of performance as of using complete information (i.e., both position and velocity). This result solidifies our approach of using image observations without explicitly including the velocity information in preventing stop-and-go wave formation.

C. Limitations and Discussion

In this project, we do not assume our RVs to be fully autonomous vehicles with equipment and sensors to allow for complete control. Our RVs control only their acceleration (or velocity), which can be achieved by controlling the throttle signal through a control mechanism using images that do not require significant computational resources to process and receive actions from. This partial autonomy mechanism allows for a human, in a real-world setting, to still control other vehicular functions (such as changing lanes, turning, handling emergency situations, etc), while improving and coordinating traffic conditions. The signals being sent to the vehicle to control traffic conditions can be overwritten by the human driver, allowing for safety within the system.

Transmitting image data in a V2V format is comparatively expensive to transmitting precise observations as images have higher dimensions. Drops/delays, resulting in data loss, when communicating with other vehicles is an inevitability [38], [39]. However, this issue can be mitigated by using existing image compression/decompression techniques, allowing significant image dimension reduction [40], [41], [42]. Our approach communicates with less vehicles, making the process easier, than precise observations as HVs are not communicated with. Additionally, our approach requires vehicles/infrastructure to be equipped with image sensors for proper implementation. While some vehicles/infrastructure may be too old or costly, advancements in image sensor technology within cars and transportation infrastructure have increased their prevalence and cost effectiveness.

Despite the generalizability of image observations, the reward functions used are still environment and task specific. A general purpose reward function for transportation environments is still an open problem given how task-specific environments can be. Thus, finding a general purpose reward function is out of scope; however, we believe finding such a general reward function is interesting to pursue in the future to further increase generalization.

V. Conclusion and Future Work

In this work, we demonstrate the ability of robot vehicles (RVs) to perform mixed traffic control using reinforcement learning (RL) policies trained on image observations. We examine RVs trained on image observations in the ring, figure eight, intersection, merge, and bottleneck environments. Additionally, we expand on the figure eight network lengths trained on, expand the merge environment and inflows trained on, and expand the bottleneck environment to include heterogeneous traffic and inflows trained on. We show that RVs trained on image observations have competitive performances to RVs trained on precise observations.

In the future, we aim to advance this study in several directions. First, we want to test our approach on more road networks, together with large-scale, long-term traffic simulations [43], [44]. This could involve combining multiple road networks together, where RVs must learn to perform multiple tasks concurrently. While simulating these scenarios is feasible, the increased task complexity presents a challenge to the RV's learning process. Second, we would like to incorporate additional generic information such as traffic state predictions [45] and vehicle trajectories into our observation space for potential improvement. Lastly, we want to explore the resilience aspect by taking adversarial attacks and image perturbations into account [46].

ACKNOWLEDGMENT

This research is supported by NSF IIS-2153426. The authors would also like to thank NVIDIA and the Center for Transportation Research (CTR) at the University of Tennessee, Knoxville for their support.

References

[1] P. Goodwin, "The economic costs of road traffic congestion," 2004.
[2] R. Arnott and K. Small, "The economics of traffic congestion," *American scientist*, vol. 82, no. 5, pp. 446–455, 1994.
[3] C. Wu, A. R. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 1270–1286, 2021.

[4] E. Vinitsky, A. Kreidieh, L. Le Flem, N. Kheterpal, K. Jang, C. Wu, F. Wu, R. Liaw, E. Liang, and A. M. Bayen, "Benchmarks for reinforcement learning in mixed-autonomy traffic," in *Conference on robot learning*. PMLR, 2018, pp. 399–409.

[5] Z. Yan and C. Wu, "Reinforcement learning for mixed autonomy intersections," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 2089–2094.

[6] D. Wang, W. Li, L. Zhu, and J. Pan, "Learning to control and coordinate mixed traffic through robot vehicles at complex and unsignalized intersections," *arXiv preprint arXiv:2301.05294*, 2023.

[7] M. Villarreal, D. Wang, J. Pan, and W. Li, "Analyzing emissions and energy efficiency in mixed traffic control at unsignalized intersections," in *IEEE Forum for Innovative Sustainable Transportation Systems (FISTS)*, 2024.

[8] D. Wang, W. Li, and J. Pan, "Large-scale mixed traffic control using dynamic vehicle routing and privacy-preserving crowdsourcing," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 1981–1989, 2024.

[9] F.-C. Chou, A. R. Bagabaldo, and A. M. Bayen, "The lord of the ring road: a review and evaluation of autonomous control policies for traffic in a ring road," *ACM Transactions on Cyber-Physical Systems (TCPS)*, vol. 6, no. 1, pp. 1–25, 2022.

[10] C. Wu, A. Kreidieh, E. Vinitsky, and A. M. Bayen, "Emergent behaviors in mixed-autonomy traffic," in *Conference on Robot Learning*. PMLR, 2017, pp. 398–407.

[11] E. Vinitsky, K. Parvate, A. Kreidieh, C. Wu, and A. Bayen, "Lagrangian control through deep-rl: Applications to bottleneck decongestion," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 759–765.

[12] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1475–1480.

[13] M. Villarreal, B. Poudel, and W. Li, "Can chatgpt enable its? the case of mixed traffic control via reinforcement learning," in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2023.

[14] S. Sinha, A. Mandlekar, and A. Garg, "S4rl: Surprisingly simple self-supervision for offline reinforcement learning in robotics," in *Conference on Robot Learning*. PMLR, 2022, pp. 907–917.

[15] Y. Zhu, A. Joshi, P. Stone, and Y. Zhu, "Viola: Object-centric imitation learning for vision-based robot manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 1199–1210.

[16] D. Shah, B. Osiński, S. Levine, *et al.*, "Lm-nav: Robotic navigation with large pre-trained models of language, vision, and action," in *Conference on Robot Learning*. PMLR, 2023, pp. 492–504.

[17] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, "Visual-locomotion: Learning to walk on complex terrains with vision," in *5th Annual Conference on Robot Learning*, 2021.

[18] S. o. C. Department of Transportation, "Caltrans pems," http://pems.dot.ca.gov/, 2022.

[19] C. S. ATMS, "Advanced traffic management system," https://coloradosprings.gov/traffic-and-transportation-engineering/page/traffic-management/, 2022.

[20] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Y. Qiao, and J. Dai, "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in *European conference on computer vision*. Springer, 2022, pp. 1–18.

[21] E. Xie, Z. Yu, D. Zhou, J. Philion, A. Anandkumar, S. Fidler, P. Luo, and J. M. Alvarez, "Mˆ 2bev: Multi-camera joint 3d detection and segmentation with unified birds-eye view representation," *arXiv preprint arXiv:2204.05088*, 2022.

[22] Y. Zhang, Z. Zhu, W. Zheng, J. Huang, G. Huang, J. Zhou, and J. Lu, "Beverse: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving," *arXiv preprint arXiv:2205.09743*, 2022.

[23] B. Huang, Y. Li, E. Xie, F. Liang, L. Wang, M. Shen, F. Liu, T. Wang, P. Luo, and J. Shao, "Fast-bev: Towards real-time on-vehicle bird's-eye view perception," *arXiv preprint arXiv:2301.07870*, 2023.

[24] Y. Shen, W. Li, and M. C. Lin, "Inverse reinforcement learning with hybrid-weight trust-region optimization and curriculum learning for autonomous maneuvering," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 7421–7428.

[25] B. Poudel, T. Watson, and W. Li, "Learning to control dc motor for micromobility in real time with reinforcement learning," in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 1248–1254.

[26] E. Kargar and V. Kyrki, "Vision transformer for learning driving policies in complex multi-agent environments," *arXiv preprint arXiv:2109.06514*, 2021.

[27] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[28] P. Cai, S. Wang, H. Wang, and M. Liu, "Carl-lead: Lidar-based end-to-end autonomous driving with contrastive deep reinforcement learning," *arXiv preprint arXiv:2109.08473*, 2021.

[29] Ó. Pérez-Gil, R. Barea, E. López-Guillén, L. M. Bergasa, C. Gomez-Huelamo, R. Gutiérrez, and A. Diaz-Diaz, "Deep reinforcement learning based control for autonomous vehicles in carla," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 3553–3576, 2022.

[30] Z. Zhang, A. Liniger, D. Dai, F. Yu, and L. Van Gool, "End-to-end urban driving by imitating a reinforcement learning coach," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 222–15 232.

[31] Z. Cao, E. Bıyık, W. Z. Wang, A. Raventos, A. Gaidon, G. Rosman, and D. Sadigh, "Reinforcement learning based control of imitative policies for near-accident driving," *arXiv preprint arXiv:2007.00178*, 2020.

[32] H. Maske, T. Chu, and U. Kalabić, "Large-scale traffic control using autonomous vehicles and decentralized deep reinforcement learning," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3816–3821.

[33] F. Wu and A. M. Bayen, "Cscrs road safety fellowship report: A human-machine collaborative acceleration controller attained from pixel learning and evolution strategies."

[34] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[35] M. Saberi and H. S. Mahmassani, "Empirical characterization and interpretation of hysteresis and capacity drop phenomena in freeway networks," *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, DC*, 2013.

[36] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, and I. Stoica, "Rllib: Abstractions for distributed reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 3053–3062.

[37] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

[38] B. Al-Hayani and H. Ilhan, "Efficient cooperative image transmission in one-way multi-hop sensor network," *The International Journal of Electrical Engineering & Education*, vol. 57, no. 4, pp. 321–339, 2020.

[39] S. M. Aziz and D. M. Pham, "Energy efficient image transmission in wireless multimedia sensor networks," *IEEE communications letters*, vol. 17, no. 6, pp. 1084–1087, 2013.

[40] D. K. Sonal, "A study of various image compression techniques," *COIT, RIMT-IET. Hisar*, vol. 8, pp. 97–102, 2007.

[41] Y. Choi, M. El-Khamy, and J. Lee, "Variable rate deep image compression with a conditional autoencoder," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3146–3154.

[42] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational autoencoder for low bit-rate image compression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 2617–2620.

[43] W. Li, D. Wolinski, and M. C. Lin, "City-scale traffic animation using statistical learning and metamodel-based optimization," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 200:1–200:12, 2017.

[44] K. Guo, W. Jing, L. Gao, W. Liu, W. Li, and J. Pan, "Long-term microscopic traffic simulation with history-masked multi-agent imitation learning," *arXiv preprint arXiv:2306.06401*, 2023.

[45] L. Lin, W. Li, and S. Peeta, "Efficient data collection and accurate travel time estimation in a connected vehicle environment via real-time compressive sensing," *Journal of Big Data Analytics in Transportation*, vol. 1, no. 2, pp. 95–107, 2019.

[46] Y. Shen, L. Zheng, M. Shu, W. Li, T. Goldstein, and M. C. Lin, "Gradient-free adversarial training against image corruption for learning-based steering," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021, pp. 26 250–26 263.