

Received 26 March 2024, accepted 27 April 2024, date of publication 30 April 2024, date of current version 7 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395430

RESEARCH ARTICLE

Transfer Learning-Based Deep Reinforcement Learning Approach for Robust Route Guidance in Mixed Traffic Environment

DONGHOUN LEE 

Department of Artificial Intelligence, Sejong University, Gwangjin-gu, Seoul 05006, Republic of Korea

e-mail: donghoun.lee@sejong.ac.kr

This work was supported by the Faculty Research Fund of Sejong University in 2023.


ABSTRACT A previously developed Deep Reinforcement Learning-based Vehicle Routing (DRL-VR) algorithm aims to be used for providing the shortest Origin-Destination (OD) travel time path in dynamic traffic environment. However, several issues may still arise regarding uncertainty associated with mixed traffic conditions coexisting Automated Vehicles (AV) and Human-driven Vehicles (HV), particularly in a wide-area urban road network. To develop a robust and interoperable route guidance algorithm based on the DRL approach, this study proposes Transfer learning-based deep reinforcement Learning Algorithm for Route Guidance (TLARG). It is an extended framework on the previous approach by incorporating transfer learning scheme that enables the DRL model of TLARG to converge even in a wide-area urban road network. The TLARG is evaluated in terms of OD travel time based on diverse OD trips with different urban road networks, including narrow- and wide-area road networks. This research conducts several evaluation studies based on microscopic traffic simulation experiments. The simulation result shows that the TLARG enables the agent to complete its OD trips not only with flexible routes but also with reductions in travel time depending on given traffic situations irrespective of network type. Furthermore, it demonstrates that the robustness of the proposed approach by measuring the error of Estimated Time of Arrival (ETA) for various OD trips in different urban road networks under the mixed traffic conditions. Such findings suggest that the TLARG has great potential to enhance the punctuality of mobility service by providing robust route guidance, even in the era of coexisting AVs and HVs.

INDEX TERMS Deep reinforcement learning, mixed traffic condition, origin-to-destination travel path, route guidance algorithm, transfer learning.

I. INTRODUCTION

As one of the ongoing efforts for continuous advancement in the global mobility industry, there has been a dedicated focus on the development of automated driving technology to provide a safer and more efficient mobility service implemented with future mobility systems. Such automated driving technology is being integrated into diverse forms of mobility services, including ride-sharing, car-hailing, and Demand

Responsive Transit (DRT) services with flexible routes as well as public transit buses operating on fixed routes [1], [2], [3], [4]. However, since the current automated driving technology is not yet in a mature stage to cover all the traffic situations [5], the advent of using Automated Vehicle (AV)-based mobility services during the initial phase of increasing penetration rates of AVs is expected to result in potential challenges such as traffic disruptions caused by AV-involved event occurrences [6]. For instance, a commercially available AV at the present stage suffers from a longer headway distance than that of Human-driven Vehicle (HV) when

The associate editor coordinating the review of this manuscript and approving it for publication was Eyuphan Bulut .

actuating the function of adaptive cruise control [7], [8], which amplifies the speed fluctuation in downstream region and affects upstream traffic flow, often resulting in a capacity drop [9]. In addition, secondary incidents or accidents may also occur since the Operational Design Domains (ODD) of existing AVs are limited to specific predetermined Dynamic Driving Tasks (DDT) depending on their levels of driving automation [10], which can lead to changes in the human drivers' driving behaviors [11]. Moreover, the AVs tend to operate at slower speeds compared to HVs due to their conservative design for accident prevention [12], which may contribute to frequent traffic congestions.

In the context of dealing with such challenges, implementing a reliable mobility service in the mixed traffic conditions coexisting with AVs and HVs requires the development of technology for providing a robust route guidance service. For instance, an Origin-Destination (OD) travel path that can bypass road links within the scope of influences, including unexpected congestions caused by the AVs, is one of the most effective methods in uprating the service reliability by reducing the variability of OD travel time [13]. Most of conventional route guidance methodologies have considered Dijkstra algorithm or A* algorithm based real-time traffic information using the latest traffic information [14], [15], [16], based on the traffic data obtained from Intelligent Transportation Systems (ITS), such as inductive loop detector or vision-based vehicle detection system. These approaches iteratively determine their optimal paths from their current locations to destinations given the real-time traffic information by modeling the travel costs of each road link associated with time-varying variables. However, using the real-time traffic information does not guarantee to anticipate the potential changes in traffic flow patterns, which substantially affect the consequences of the route guidance algorithms. As a result, the sequential en route diversion using the latest traffic information is highly likely to be suboptimal [17].

More recently, there have been enormous efforts to enhance the qualities of route guidance services by the advanced route guidance methods combined with the use of predictive traffic information. A previous research showed the benefits of incorporating predictive data on future traffic situations in terms of OD travel time [18]. Likewise, a hierarchical route planning algorithm employing a time-varying graph model has been proposed [19], which could find the effects of potential dynamics at future traffic states by using the information on traffic prediction. In addition, a Deep Learning (DL)-based proactive approach for predicting the future state of the traffic network has been developed to redirect vehicles from traffic congestion using vehicle data collection system in the field of ITS [20]. Similarly, a DL-based fine-grained path planning algorithm has also been considered in [21], which executes a gridded path planning based on the use of traffic prediction information. However, despite the extensive development of traffic prediction models utilizing state-of-the-art DL algorithms, there are still

prediction errors in the prediction models. Moreover, most of the existing algorithms using DL-based traffic prediction information have focused on recurrent congestions rather than non-recurrent congestions, though the latter requires more attention than the former [22]. Furthermore, since there remains uncertainty in predicting future traffic states due to the non-recurrent congestions caused by the instability of AVs, such routing algorithms using the prediction information still hold a high probability of not being an optimal solution.

Nowadays, Deep Reinforcement Learning (DRL)-based route guidance algorithm is gaining attention as a candidate to uprate the performance of mobility system associated with uncertain traffic conditions. The conventional approaches to the DRL-based route planning algorithms in the field of Vehicle Routing Problems (VRP) have focus on determining the optimal sequences for multiple vehicles to visit the locations of passengers or customers [23], [24], [25]. In contrast, there have been only few studies related to the DRL-based approach for OD route guidance taking into account dynamic traffic conditions [26], [27], [28], which basically considers an agent's OD travel path as a consequence of sequential decision-making processes on selecting road links in a target service area. The sequential decision-making process is mathematically formulated as Markov Decision Process (MDP), which is represented by a tuple of state, action, transition probability, reward, and discount factor. In order to describe the current dynamic traffic situation, a previous research formulated the state of MDP using multiple traffic variables for the first time [26]. The traffic variables involved in the state representation include the number of vehicles, average speed, and length of road link the DRL agent vehicle travels. The previous study inspired another research that developed a DRL-based cooperative route planning algorithm with the prioritization of urgent vehicles [27]. Similar to the previous one, the state definition of the MDP also incorporated one dynamic traffic variable, such as the number of vehicles on the road link where the agent is located, which assumed that the real-time traffic data could be obtained from Cooperative ITS (C-ITS). However, these previous algorithms are still doubtful of feasibility in terms of OD route guidance. Neither the existing ITS nor C-ITS facilities can detect the entire vehicles on a road since both of them mainly cover a single point or local area. Every single vehicle passing through a road link in a target area needs to have in-vehicle C-ITS devices if the state input would be measured by using the C-ITS facilities on the road. Moreover, even if all the state variables could be identified given the current traffic situation, they will not be able to provide the OD route guidance service. For instance, most of the conventional DRL-based route guidance algorithms generate an OD travel path from sequential local paths for en route diversion. Each local path indicates the consequence of DRL agent's action given the current state. However, since the existing approaches cannot specify the state of MDP in future dynamic traffic conditions, it still

shows limitations on generating the sequential local paths. Therefore, it is imperative to address such problems associated with the generation of OD travel path in the research field of DRL-based route guidance algorithm.

To deal with the related issues, a preliminary study of developing a pioneering framework on predictive traffic information-based Deep Reinforcement Learning-based Vehicle Routing (DRL-VR) algorithm has been proposed [28]. It was the first study that allows the DRL-VR algorithm to generate the OD travel path by involving a predictive traffic state representation in the state of MDP. The DRL-VR was design to be used to provide the shortest OD travel path in dynamic traffic environment. Nevertheless, the previous study still has some drawbacks on providing the OD travel path in the mixed traffic condition coexisting AVs and HVs, particularly in a wide-area urban road network. The DRL-VR already showed its benefits in uncertain traffic conditions caused by non-recurrent congestions compared with those of existing algorithms, including Dijkstra, A*, and DRL-based conventional routing algorithm without predictive representation. However, it could not be applied to different places not explicitly experienced during training process due to the state and reward function with a low interoperability. Especially when reaching the agent's destination in a wide-area urban road network where the AVs and HVs coexist, the complexity of sequential actions on each decision-making exponentially increases as the traffic uncertainty increases. Consequently, it is highly likely to observe the divergence from the reward model used in the DRL-VR, otherwise the travel path generated by the trained DRL model does not guide the agent vehicle to the desired destination. These issues are the primary motivations of the present study.

The main objective of this study is to develop a robust and interoperable DRL-based route guidance algorithm used for the near future coexisting AVs and HVs in a wide-area urban network. To accomplish the research goal, this study proposes Transfer learning-based deep reinforcement Learning Algorithm for Route Guidance (TLARG). It is an extended framework on the previously developed DRL-VR by incorporating transfer learning scheme that enables the DRL model of TLARG to converge even in a wide-area urban road network. The TLARG is evaluated in terms of OD travel time based on diverse OD trips with different urban road networks, including narrow- and wide-area urban road networks. This research conducts several evaluation studies based on microscopic traffic simulation experiments considering various traffic scenarios involved with non-recurrent congestions caused by AVs. The evaluation studies include case study, performance review, and comparison study. The case study explores the characteristics of the TLARG with respect to different network configurations and OD trips. For the generalized performance of the proposed algorithm, the performance review performs simulation experiments multiple times to assess its overall performance based on statistical results. Lastly, the comparison study conducts a case-by-case

comparison regarding the performance differences among the models of TLARG. Such numerical studies will demonstrate that the robustness and interoperability of the proposed approach with respect to various OD trips in different urban road networks under the mixed traffic conditions.

The contribution of this research can be highlighted as follows:

- This research develops an extended framework for DRL-based route guidance algorithm by incorporating transfer learning scheme, which can provide more reliable route guidance service than previous ones, particularly in diverse OD with mixed traffic conditions.
- This study proposes a modified MDP formulation to allow the transfer learning scheme to be applied for different urban road networks, even in a wide-area road network.
- This research demonstrates the robustness of advanced TLARG model by comparing between predicted and actual Estimated Time of Arrival (ETA) for diverse OD trips both in narrow- and wide-area urban road networks under uncertain traffic conditions caused by the advent of AVs.

The remainder of this paper is organized as follows. Section II provides the detailed descriptions on the TLARG. Section III describes the details of evaluation approaches used in numerical studies. Section IV presents the results and analyses of the numerical studies. Then, this paper ends with the concluding remarks in the last section.

II. METHODOLOGY

There are several key elements to implement the TLARG for providing a reliable OD route guidance in the mixed traffic conditions. The following subsections provide the details of vital components to be considered for the route guidance service in the proposed algorithm.

A. USE OF PREDICTIVE TRAFFIC INFORMATION

The TLARG basically follows the concept of the previously proposed DRL-VR algorithm [28], which incorporates a predictive traffic state representation in the MDP. Unlike the conventional approaches concerning DRL-based route guidance algorithms, which only produce the local travel path for en route diversion, the TLARG expresses the time-varying variables as the predictive representation in the state variables of the MDP formulation. Therefore, it enables the TLARG to generate the OD travel path by specifying the corresponding values of the state variables for any given time, even before reaching a specific state. Furthermore, using the predictive information on traffic dynamics allows the TLARG to model the sequential decision-making process under uncertainty associated with future traffic condition, particularly in the mixed traffic.

For the predictive state representation used in the DRL-based route guidance algorithm, the TLARG consist of two fundamental functions, including *traffic prediction*

and *route guidance*. The former provides traffic prediction information using a DL-based model, while the latter uses the output values generated by the traffic prediction to produce the OD route guidance data based on a DRL model. Note that any sophisticated DL and DRL models can be used for those functions in the proposed framework. For the sake of convenience, this study considers that the traffic prediction and route guidance functions are implemented with Graph WaveNet and Prioritized Experience Replay-based Double-Deep Q-Network (PDDQN) model, respectively [29], [30], [31].

Fig. 1 shows the details of the learning methods on the traffic prediction and route guidance model used in the TLARG. The traffic prediction model provides predicted speed values of each road section by considering spatiotemporal characteristics on historical traffic flows on a target road network. As shown in Fig. 1, L spatiotemporal layers are considered in the traffic prediction model to capture spatial and temporal features of the traffic flows on the target road network. Each layer has building block with two gating mechanism-based Temporal Convolutional Networks (TCN) and one Graph Convolutional Network (GCN). The TCN plays a vital role in extracting temporal features based on Dilated Causal Convolution Neural Network (DCCNN) [32]. The GCN is used for identifying the latent spatial dependencies using learnable parameters without a priori knowledge, based on Diffusion

Convolutional Recurrent Neural Network (DCRNN) [18]. The TCN shows a notable reduction in computing time and the ability to consider long-range sequence data based on dilated causal convolutions, which allows it to have a large receptive field without using a prohibitively large number of parameters. On the other hand, the GCN captures the characteristics of each node and the relationships between neighboring nodes based on node embedding and adaptive adjacency matrix, where it regards a road link as a node. Such approach enables the prediction model to consider both the node's own features and the features of its neighbors when making predictions.

The concept of building block in the traffic prediction model is identical to the original Graph WaveNet, excluding loss function. The TLARG computes the training and validation losses based on Root-Mean-Square Errors (RMSE). The loss function L in the prediction horizon I can be expressed as follows:

$$L(\hat{X}^{(t+1):(t+I)}; \theta_p) = \frac{1}{IJK} \sqrt{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{X}_{jk}^{t+i} - X_{jk}^{t+i})^2}, \quad (1)$$

where \hat{X} indicates the predicted output vector, θ_p denotes the learnable parameter sets in the traffic prediction model of the TLARG, J represents the number of traffic detectors in the target road network, K describes the number of features associated with each traffic detector, and X is the desired output vector.

As described in Fig. 1, the predicted information of the traffic prediction module interacts with *environment* in the route guidance model, which allows the DRL to specify agent's *state* s given traffic situation. Based on the agent's current state, the environment also plays a crucial role to determine the *reward* r corresponding to the *action* a determined by the policy π_θ of Q-network parameterized with θ as well as its *next state* s' . The experience transitions (s, a, r, s') are transmitted to replay buffer, which is well known for effectively reducing the impact of sequential dependencies by eliminating the temporal correlation between consecutive experiences in the off-policy DRL model. Still, the agent's experiences do not contribute equally to learning. Moreover, the learning environment often shows sparse rewards when the agents have various OD trips, especially in a wide-area road network. Hence, the route guidance model considers the PER to address such challenges by prioritizing and replaying rare and informative experiences, which can maintain a more consistent and effective training trajectory.

The learning method used in the route guidance model is DDQN [30], which aims to reduce the discrepancy between the predicted and target Q-values by minimizing the Temporal Difference (TD) error δ , as formulated in (2).

$$\delta = r + \gamma Q_{\theta^-}(s', \underset{a'}{\operatorname{argmax}} Q_{\theta}(s', a')) - Q_{\theta}(s, a), \quad (2)$$

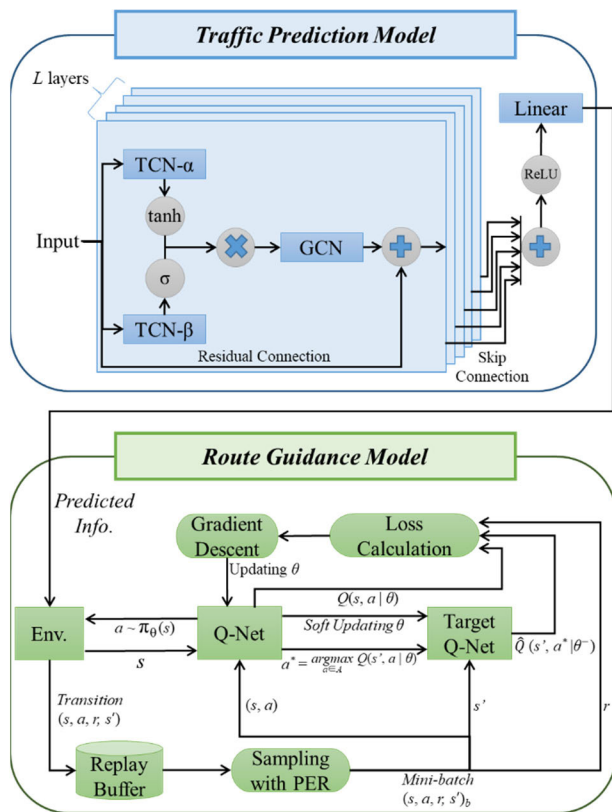


FIGURE 1. The learning methods of traffic prediction and route guidance model in the TLARG.

where γ is the discount factor, θ^- represents the learnable parameters of the target Q-network, a' describes the action taken in the next state s' , and θ indicates the learnable parameters of the online Q-network. θ^- is periodically updated with the copy of θ for the online Q-network, while θ is adjusted using a learning rule combined with the PER method, as described in (3) and (4).

$$\theta \leftarrow \theta + \eta_g w \delta \frac{\partial Q_\theta(s, a)}{\partial \theta}, \quad (3)$$

where η_g denotes the learning rate used in the route guidance model, w indicates the importance sampling weight associated with the PER algorithm.

$$w = \left(\frac{u^\alpha}{\sum_b u_b^\alpha} B \right)^{-\beta}, \quad (4)$$

where u^α describes the priority of given transition with the prioritization exponent parameter α , b represents the mini-batch size, B indicates the size of the replay buffer, and β refers to the importance sampling exponent parameter. More detailed descriptions of the parameter values used in the traffic prediction and route guidance model are provided in B. TUNING WORK of III. DATA DESCRIPTION.

B. INTEROPERABLE MDP EXPRESSION

The TLARG follows the concept of using the use of predictive information on the road traffic network, which enables the DRL-based route guidance algorithm to incorporate the predictive traffic state representation in the MDP. Therefore, the TLARG can be used for providing the OD travel path even in the mixed traffic condition. However, the TLARG still needs to consider the modification of the MDP involved in the previously developed DRL-VR algorithm, which cannot address the complexity associated with diverse OD trips. Moreover, the conventional expression of the MDP in the DRL-VR algorithm requires a more flexible approach to generate the OD route guidance data in spaces not explicitly represented during training.

With these backgrounds, the TRLARG introduces an interoperable MDP expression, which can be applied to any arbitrary spatial configuration, even in different OD trips. Key variables of the MDP formulation in the TLARG are illustrated in Fig. 2.

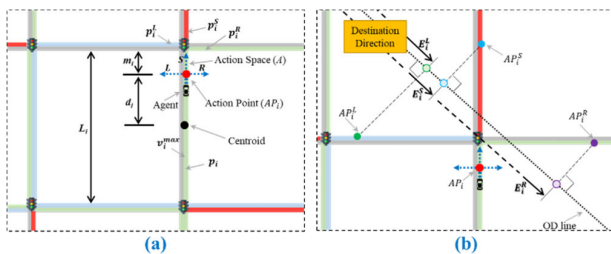


FIGURE 2. Overview of key variables associated with the MDP formulation of the TLARG: (a) state and (b) distance reward.

The route choice of DRL agent is based on an action space (A), where actions are encoded as discrete values: *Right-turn*(R) with 0, *Go-straight*(S) with 0.5, and *Left-turn*(L) with 1. The action is going to be determined when the agent reaches Action Point AP_i at time step i , as shown in (a) of Fig.2. The timing of AP_i is calculated by the decision area d_i , which depends on the minimum distance m_i , as expressed in (5) and (6).

$$m_i = \frac{(V_i^{\max})^2}{2a_{dec}^{\max}} + P_i \tau, \quad (5)$$

where V_i^{\max} and P_i indicate the maximum speed limit and predicted speed of the road section where the agent is located at time step i , a_{dec}^{\max} represents the maximum deceleration rate, and τ denotes the perception-reaction time for the lane change to follow the route choice.

$$d_i = \frac{L_i}{2} - m_i, \quad (6)$$

where L_i is the length of the road section where the agent vehicle travels at time step i .

Regarding the state definition, the TLARG considers future traffic dynamics. Unlike the existing DRL-based route guidance algorithms, all of the variables involved in the state definition can be specified for any given time before the agent actually faces the dynamic traffic environment, even not in places the agent has not experienced during the training process. The TLARG formulates the state for a link at time step i as (7).

$$s_i = [L_i, V_i^{\max}, P_i, D_i, C_i, p_i^R, p_i^S, p_i^L], \quad (7)$$

where C_i represents the agent's moving direction that is identical to the previous action choice at time step $i-1$, $p_i^{R,S,L}$ describes the predicted speed values of subsequent road links connected to the road link where the agent is located at time step i , and D_i indicates the relative distance from the current position to destination compared to the OD-distance at time step i , as shown in (8).

$$D_i = \frac{E_{AP_i \rightarrow D}}{E_{O \rightarrow D}}, \quad (8)$$

where $E_{AP_i \rightarrow D}$ represents the Euclidean distance between AP_i and destination, and $E_{O \rightarrow D}$ indicates the Euclidean distance between origin and destination.

Note that D_i and C_i are newly incorporated into the state definition of the TLARG compared to that of the DRL-VR algorithm. They enable the agent to have diverse OD trips in road networks with different configurations. Furthermore, no matter how much the penetration rates of in-vehicle C-ITS devices with respect to HVs are in the road network where the AV-based mobility service is implemented, it is not feasible for all the vehicles to directly communicate with the C-ITS facilities for sharing information on the non-recurrent congestions caused by the AVs. Thus, the source data required for the traffic variables in the state definition of the TLARG can be obtained in real-time based on the assumption for widespread ITS detectors in the service area.

The objective of modeling the TLARG with the MDP framework directly relates to the reward function. It plays a pivotal role to obtain the optimal policy for maximizing the expected return, which is the expected cumulative rewards over the course of an episode. To generate a robust route guidance data under traffic uncertainty associated with the mixed traffic condition, the reward function r_i consists of four parts, including distance reward $r_{i,distance}$, time reward $r_{i,time}$, prediction reward $r_{i,prediction}$, and trip completion reward $r_{i,completion}$. The details of the reward function is as follows:

$$r_i = r_{i,distance} + r_{i,time} + r_{i,prediction} + r_{i,completion}, \quad (9)$$

where

$$r_{i,distance} = 1 - \frac{2(E_i^a - \min(E_i^R, E_i^S, E_i^L))}{\max(E_i^R, E_i^S, E_i^L) - \min(E_i^R, E_i^S, E_i^L)}, \quad (10)$$

$$r_{i,time} = clip(2 - \frac{TT_{i+1} - TT_i}{\frac{m_i}{V_{i+1}^{max}} + \frac{L_{i+1} - m_{i+1}}{V_{i+1}^{max}}}, -1, 1), \quad (11)$$

$$r_{i,prediction} = clip(1 - \left| 1 - \frac{1}{TT_{i+1} - TT_i} \left(\frac{m_i}{P_i} + \frac{L_{i+1} - m_{i+1}}{P_{i+1}} \right) \right|, -1, 1), \quad (12)$$

$$r_{i,completion} = \begin{cases} T, & \text{if desired destination} \\ -T, & \text{otherwise} \end{cases} \quad (13)$$

where $E_i^{R,S,L}$ indicates the Euclidean distances between the agent's destination to several virtual points that are the orthogonal projection of $AP_i^{R,S,L}$ to the OD line, as depicted in (b) of Fig. 2, and E_i^a describes the Euclidean distances between the destination to a virtual point determined by an action a at AP_i , which corresponds to one of $E_i^{R,S,L}$; TT_i represents the travel time from the agent's origin to AP_i ; T means the target value for the agent's trip completion in the terminal state.

The TLARG requires several intrinsic rewards to effectively improve the policy based on the MDP framework since there are sparse rewards in dealing with the route guidance problem, particularly in interacting with uncertain traffic environments. Hence, the TLARG involves both intrinsic and extrinsic rewards, where the intrinsic rewards are $r_{i,distance}$, $r_{i,time}$, and $r_{i,prediction}$, while the extrinsic reward is $r_{i,completion}$. Note that the most distinctive characteristic of the reward function used in the TLARG is modifying $r_{i,distance}$, which is inversely proportional to the Euclidean distance between the destination to a virtual point determined by a route choice a at AP_i . Such modification allows the TLARG to guide the DRL agent for its trip completion. In addition, it is designed to range from -1 to 1 , which is the same scale with other intrinsic rewards.

It is straightforward that the time reward function $r_{i,time}$ intends to minimize the agent's OD travel time, which is the

ultimate goal of using the route guidance algorithm. On the other hand, the prediction reward $r_{i,prediction}$ is designed to enhance the reliability of route guidance service by reducing the variability of ETA. Such approach plays an important role in learning the optimal policy since an acceptable level of errors between the expected and actual travel time directly determines whether the service user is satisfied with the quality of route guidance service.

Finally, the TLARG incorporates the extrinsic reward employing trip completion reward $r_{i,completion}$. A substantial positive reward is assigned when the agent complete its trip, which means that the agent arrives its desired destination link when it reaches terminal state. It is useful to monitor the performance of the TLARG in the training process when the agents have diverse OD pairs.

C. TRANSFER LEARNING SCHEME

Fig.3 depicts the system architecture of the TLARG, which mainly consists of traffic prediction module and route generation module. The traffic prediction module collects real-time traffic data obtained from the ITS detectors, which is not only stored in a historical traffic database with information related to the corresponding road section but also used for predicting the future traffic states of each road section.

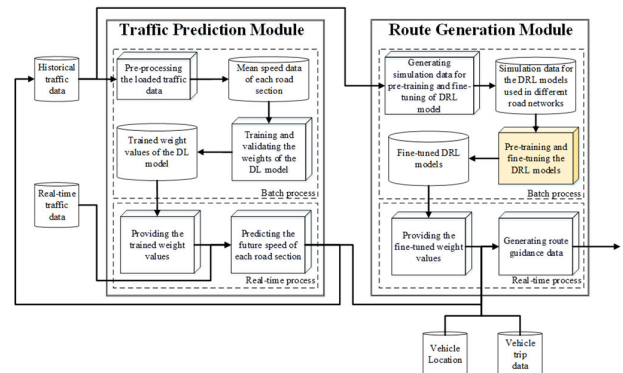


FIGURE 3. The framework of proposed approach.

Based on the historical traffic data, the traffic prediction module trains and validates the modified Graph Wavenet model in batch process. Simultaneously, it uses the trained DL model to predict the future speed of each road section based on the real-time traffic data. The predicted output vectors for the target area will be used to update the state variables of the DRL model involved in the route generation module when generating route guidance data in real-time process.

The historical traffic data is also used to generate simulation data for training the PDDQN model in the route generation module. The training process of the DRL model is conducted with the transfer learning, as highlighted with the yellow one in Fig. 3. The transfer learning-based training approach includes a pre-training and fine-tuning process. The DRL model can be fine-tuned through the transfer learning process and be used to generate guidance data for OD travel

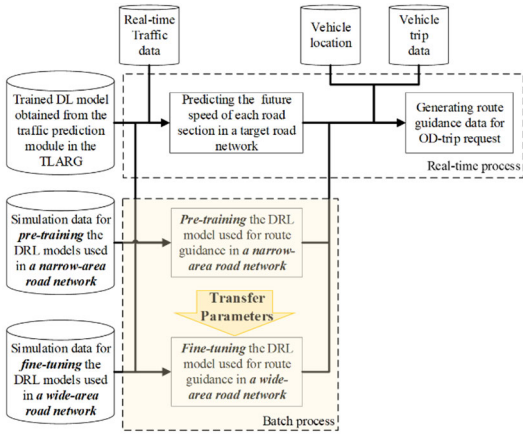


FIGURE 4. The process on transfer learning of TLARG.

path. More detailed explanations of the operational flow associated with the transfer learning in the TLARG are shown in Fig. 4.

With a batch process, the route generation module first pre-trains the DRL model based on the generated simulation data in a narrow-area road network, which is easy to obtain a fast convergence rate. However, the target area for route guidance service is generally more complex and wider than the service area used in the pre-training process. Moreover, it requires more explorations to find the optimal policy, which may tend to face challenges in achieving prompt model convergence when training the DRL model from scratch. To accelerate the exploration process by providing a starting policy that already captures some useful strategies, the route generation module uses the pre-trained model to fine-tune the DRL model. The fine-tuning process on a target area can be formulated as follows:

$$\theta_{\pi}^{fine-tuned} = \theta_{\pi}^{pre-trained} + \eta_t \cdot \nabla_{\theta_{\pi}} E_{\tau \sim \pi_{\theta_{\pi}}} [\sum_{t=0}^T R'(s_t, a_t)], \quad (14)$$

where $\theta_{\pi}^{fine-tuned}$ and $\theta_{\pi}^{pre-trained}$ represent the learnable parameters of the Q-networks associated with the route generation module in the fine-tuning and pre-training processes, η_t indicates the learning rate used in the fine-tuning process, and $R'(\cdot)$ describes the output of the reward function used in the target area.

It is worth noting that the transfer learning-based training approach can facilitate the effectiveness and efficiency of the learning process [34], which allows the TLARG to be applied for the route guidance service in a wide-area road network. Finally, the fine-tuned DRL model will be used for producing route guidance data in real-time process when requesting an OD travel path in a service area.

III. DATA DESCRIPTION

To explore the effect of employing the TLARG on the route guidance service in different urban road network

configurations with mixed traffic environments, this study conducts microscopic traffic simulation experiments based on the Simulation of Urban MObility (SUMO) [35]. Several experimental scenarios are necessary to describe the uncertain traffic conditions caused by the AV-involved congestions. In addition, the tuning works on determining the hyperparameters used in the traffic prediction and route generation modules are also required to perform the evaluation studies for the performance of the TLARG. The details are discussed in the following subsections.

A. EXPERIMENTAL SCENARIO

The simulation experiments are conducted in different road network configurations to analyze the robustness and interoperability of the proposed algorithm. Suppose that a route guidance service provides an agent vehicle with a global travel path from its origin to destination. Then, the global travel path can be composed of diverse combinations of sub-origins and sub-destinations given a target area. Therefore, the partial trips between sub-origins and sub-destinations can be regarded as the OD travel paths associated with the target area.

The number of OD travel paths determined by the possible combinations of sub-origins and sub-destinations increase as the target area increases. To increase the complexity of finding the optimal solution, the simulation experiment considers three types of target areas, as shown in Fig. 5. The target area includes 3×3 , 5×5 , and 8×8 grid-shaped urban road networks. Both of study sites consider two major eastbound and northbound traffic demands as well as two minor westbound and southbound traffic demands. In addition, there are four lanes in each link of the study sites, where the maximum speed limit is 50 km/h, the lengths of road links L_1 and L_2 are 200 m and 300 m, respectively. Every

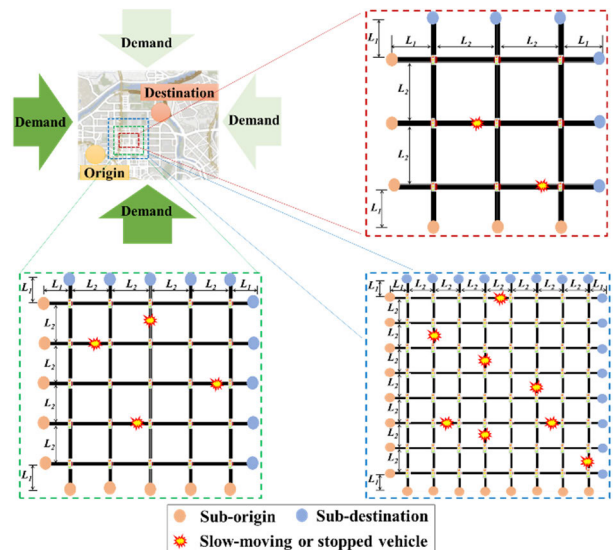


FIGURE 5. Experimental simulation scenarios with different road network configurations.

single intersection is signalized intersection with a four-phase signal plan. An adequate coordination between eastbound consecutive signals is applied to the simulation experiments, so that it can prevent the massive eastbound traffic flow from inducing traffic queue spillback at intersections.

The asymmetric inbound traffic flow is intended to describe the characteristics of traffic flow in urban area during peak hours [36]. Every single experiment generates the inbound traffic flows with Gaussian random variables in order to mimic the stochastic daily traffic demands. With the trip generation, the dynamic traffic assignment is conducted based on SUMO DUArouter tool, which can assign the traffic flow with different time intervals [37]. This suggests that it is more appropriate to describe the day-to-day variations in traffic demand based on heterogeneously loaded traffic flow throughout the study site, rather than other traffic assignment tools, such as JTRrouter, OD2trips, MARouter, and DFrouter. Moreover, in order to generate the near future traffic environment that involves non-recurrent congestions associated with low penetration rates of AVs, a few slow-moving vehicles are assigned to the study sites. Furthermore, several randomly distributed stopped vehicles are intentionally involved in the simulation experiments, which generates queue discharge flow reduction, capacity drop, and shock wave propagation [38], [39], [40], often resulting in unexpected traffic delays. Consequently, there will be traffic prediction errors, which highly affects the performance of the route guidance service. It is expected that the DRL agent either avoids the abnormal congestions by en route diversion or sticks to initial OD travel path despite additional time required for passing through the congested road. Apart from the route choices, the initial route guidance may show take a reliable OD travel time path that does not involve any congested roads. Hence, the performance of the TLARG can be evaluated appropriately in such experimental scenarios that describe the mixed traffic conditions in the near future.

B. HYPERPARAMETER TUNING

There have been numerous C-ITS implementation projects in Europe and USA [41]. However, it is still deployed with a pilot project at low levels of penetration [42]. Besides, additional C-ITS infrastructures should be required when the penetration rate of AVs will significantly increase in the far future. In contrast, the data observation on the traffic flow in the near future is still going to be monitored and be detected by the legacy ITS sensors. Hence, this study assumes that the real-time traffic data can be obtained from the widespread ITS detectors installed at each road link in the study sites.

This study generates the training, validating, and testing datasets of the proposed algorithm based on 30 days of observation data obtained from the experimental scenarios. The observation data for the first 24 days are considered as the training dataset, while the rest of data for the following consecutive 3 days and the remaining 3 days are applied to the validating and testing of the traffic prediction model and

route guidance model involved in the TLARG, respectively. Every single day has 4 hours of simulation runtime based on the observation data with a 5-minute resolution, which means that the unit time interval for the traffic prediction model corresponds to 5 minutes. The values of hyperparameters used in the traffic prediction model of the TLARG are provided in Table 1.

TABLE 1. The values of hyperparameters involved in the traffic prediction model of TLARG.

Parameter	Value
Road Configuration	3X3, 5X5, 8X8
The number of ITS detectors	48, 120, 288
Prediction horizon	1 (hour)
The number of features for ITS detector	2
Historical observation	1 (hour)
The number of epochs	100, 200, 300
Learning rate	0.001 → 0.0001
Weight decay	0.0001
Dropout rate	0.5
Mini-batch size	256
The number of spatiotemporal layers	4
The number of building blocks	3
The number of channels for residual connections	32
The number of channels for skip connections	256

The traffic prediction module of the TLARG trains the traffic prediction model with the hyperparameter values in batch process. With the trained traffic prediction model, the traffic prediction module provides the traffic prediction values, which are transmitted to the historical traffic database and the route generation function, in order to specify the state variables of the MDP in the route guidance model. More specifically, the prediction values can be utilized for pre-training and fine-tuning the route guidance model in batch process as well as inferencing in real-time process.

Similar to the traffic prediction model used in the traffic prediction module of the TLARG, the route guidance model involved in the route generation module of the proposed algorithm also requires to determine the specific values of its

TABLE 2. The values of hyperparameters involved in the route guidance model of TLARG.

Parameter	Value
Road Configuration	3X3, 5X5, 8X8
The number of episodes	1500, 2500, 9000
Learning rate	0.001
Discount factor	0.99
Trip Completion reward	100
Perception-reaction time (s)	1.3
Maximum deceleration rate (m/s ²)	4.5
The number of neurons in hidden layers	256, 512, 128
Soft update frequency	100, 200, 600
The size of replay buffer	5000, 20000, 80000
The size of mini-batch	128
Prioritization exponent	0.6
Importance sampling exponent	0.4
Increment step of importance sampling exponent	200, 300, 1000

hyperparameters. The values of hyperparameters involved in the route guidance model of TLARG are shown in Table 2. The route generation module of the TLARG pre-trains the route guidance model with the hyperparameter values in batch process. The pre-trained route guidance model is used for transfer learning in fine-tuning other route guidance models of different target areas. The fine-tuned models are deployed to mobility service areas where a robust OD route guidance service is required.

In order to verify the effectiveness of imposing the TLARG, this study considers three types of TLARG-based models, including *Baseline Model (BM)*, *Basic Transfer Model (BTM)*, and *Advanced Transfer Model (ATM)*. The BM indicates the route guidance model trained from scratch without using transfer learning given a target area. The BTM represents a basic TLARG model. It is the fine-tuned model derived from the pre-trained with BM, where the target area of BTM is wider than that of BM. Likewise, the ATM describes an advanced TLARG model, which is the fine-tuned model obtained from the pre-trained with BTM. The ATM is fine-tuned on a wider-area urban road network than the one the BTM has been pre-trained on. For instance, when there are 3×3 , 5×5 , and 8×8 urban road network scenarios, the ATM model of the 8×8 network scenario can be fine-tuned based on the BTM pre-trained from the 5×5 network scenario. Similarly, the BTM model of the 5×5 network scenario can be fine-tuned based on the BM pre-trained from the 3×3 network scenario.

Fig. 6 shows the trends in average rewards of different TLARG-based models with respect to given urban road networks, where the average reward is calculated by the mean of rewards in five consecutive episodes. It is easily observed that there is only the BM in the 3×3 road network scenario, where its average reward converges after approximately 500 episodes although the average reward does not get close to the optimal value in earlier episodes due to the exploration process. The overall trend is consistent with the finding of the previous study [28], which considered only 3×3 urban road network scenario. However, unlike the narrow-area road network scenario, the wide-area road network scenarios require much more episodes to reach exploitation, as shown in the second and third column of the BM in Fig. 6. It is also found that the DRL agent often fails to get to its desired destination, even in the exploitation process. As highlighted with orange circles in Fig. 6, this phenomenon becomes more easily observable as the network size increases.

Compared to the outcome of the BM in the 5×5 road network scenario, the BTM shows much less exploration processes. In addition, it does not suffer from the event associated with agent's undesired destination arrival. Such trend can also be seen in the 8×8 road network scenario. Unlike the BM, the BTM exhibits model convergence even though it shows some undesired events during exploration process.

The most remarkable outcome can be observed in the ATM, as shown in the last column of Fig. 6. The ATM has a few episodes for its exploration process, even in a wide-area

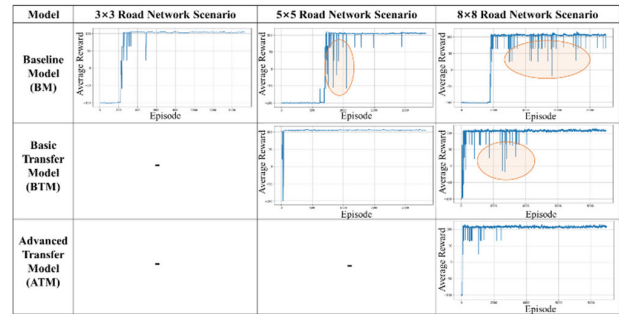


FIGURE 6. Trends in average rewards of different TLARG-based models with respect to given road networks.

road network. Moreover, it is also found that the agent shows its trip completion every episode after the exploration process. Such trends suggest that the convergence speed of using the TLARG is much faster than that of conventional one, particularly in a wider-area urban road network with mixed traffic condition.

It is worth noting that the robustness and interoperability of the TLARG are evaluated in wide-area road networks using 5×5 and 8×8 road network scenarios, rather than 3×3 road network scenario. Such scenarios cover the mixed traffic in urban area, where the agent faces never-before-seen traffic and diverse trip conditions. Moreover, the traffic demands, AV-involved congestions, and agent's OD considered in the training and validating process do not overlap with those used in the testing process. Thus, it is expected that this study can demonstrate the performance validation of the TLARG in diverse uncertain traffic environments.

Furthermore, each simulation experiment is conducted based on a specific computational environment: Python 3.8.10 platform on an Ubuntu 20.04 (Intel(R) Core(TM) i9-13900K CPU 32 cores with 5.80GHz processing, 64 GB RAM, and NVIDIA GeForce RTX 3060 12GB). Based on the code optimization and parallel computing techniques, it could observe that the average inference times for generating the OD paths in the three different road network scenarios were 12, 14, and 15ms, which are much less than one second. Consequently, it is no doubtful of feasibility for the real-time application of the proposed algorithm.

IV. RESULT AND ANALYSIS

This study conducts several numerical studies, including case study, performance review, and comparison study, based on the experimental traffic scenarios stated in the previous section. The case study analyzes the details of outcomes from different models in several specific experimental scenarios. The performance review provides the overall performances of each model with respect to diverse experimental scenarios. Lastly, the comparison results of models' performances are discussed in the comparison study.

A. CASE STUDY

Through the case study that describes several specific traffic conditions and agents' OD trips in different urban road

networks, the characteristics of each derived from the TLARG are explored in terms of OD travel path and travel time. In addition, since the existing DRL-based route guidance algorithms do not converge in a wide-area urban road network, the agents following the policy trained by the previous approaches do not tend to get to their destinations. Thus, the case study focuses on examining the performances of the TLARG-based models, rather than conventional DRL-based route guidance models.

Fig. 7 represents several examples of route guidance services with the TLARG-based models in different road networks with different traffic conditions and OD trips. There are two different OD trips in 5×5 and 8×8 road network scenarios, respectively. As shown in (a) of Fig. 7, there are four slow-moving and stopped vehicle-involved events in the 5×5 urban road network in order. One can observe that all agents guided by the proposed algorithm reaches their destinations, which implies that the modification of MDP expression enables the DRL agent to complete its OD trip. In addition, it is found that both the BM and BTM provide the OD travel paths avoiding the traffic interruptions induced by the unexpected congestions. The agent of following the route guidance generated by the BM takes the travel route from the left boundary link to the top boundary link of the road network. In contrast to the case of BM, the BTM guides the agent to pass through the study site using some links located inside the road network. Even though the agents have different OD travel paths, they spend similar amounts of OD travel time in reaching their destinations since they have not encountered the abnormal traffic interruptions caused by the AVs.

On the other hand, when considering different traffic conditions and OD pairs on the same road network, different outcomes can be found in (b) of Fig. 7. Despite following the similar OD travel path as the previous OD path, significant differences are observed in the OD travel time for the OD trips. One can find that the agent encounters the traffic congestions caused by the AV-involved incident at the upper road link of the network when following the route guidance generated by the BM. It is also seen that the agent of using the BM needs approximately 27% additional time to get to its destination compared to the previous case in (a) of Fig. 7. In contrast, the agent following the policy of BTM does not face any abnormal traffic congestions until arrival. Moreover, the BTM reduces the OD travel time by nearly 36% compared with that of the BM, even in the identical traffic conditions. Such trend is also found in the 8×8 road network scenarios.

More detailed explanation on the effect of flexible route guidance for the proposed algorithm given dynamic traffic situations can refer to (c) and (d) of Fig. 7, where there are eight slow-moving or stopped vehicle-involved events in the 8×8 urban road network. As shown in (c) of Fig. 7, it is easily found that each agent completes their OD trips, even in the wide-area road network, which is consistent with the previous findings in the cases with (a) and (b). Such finding confirms

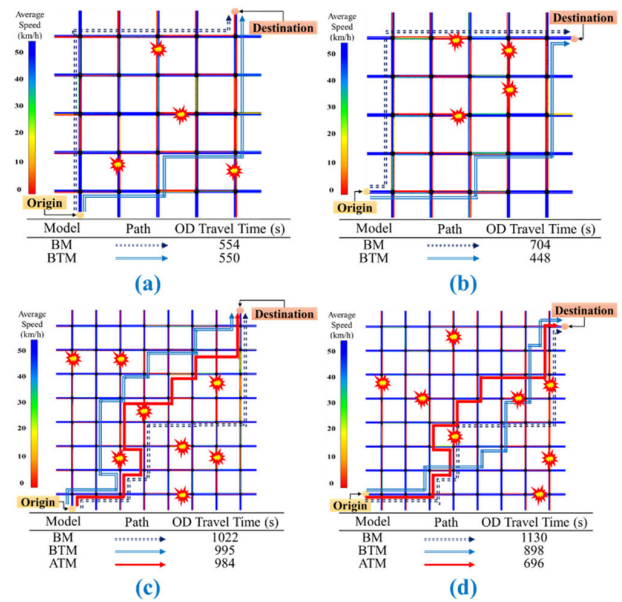


FIGURE 7. Examples of route guidance services with TLARG-based models in different road networks with different traffic conditions and ODs: (a) 5×5 road network with OD case 1 (b) 5×5 road network with OD case 2 (c) 8×8 road network with OD case 1 (d) 8×8 road network with OD case 2.

that the interoperable expression of the MDP can address the critical issues of the DRL-based approaches associated with the diverse OD trips in spaces not explicitly experienced during training. In addition, one can observe that each OD route consists of different combinations of links within the target area. Moreover, it is also found that all of the proposed models guide the agents to use their OD travel paths bypassing the AV-induced congested roads, where they show low OD travel times in the order of ATM, BTM, and BM.

Such similar trends can also be found in (d) of Fig. 7. It is shown that each model guides the agents to complete their OD trips. In addition, the magnitude of OD travel time is in the order of BM, BTM, and ATM. However, the results of the travel paths provided by each model show some differences in consequence. One can find that the TLARG-based models, including BTM and ATM, show flexible travel routes for the given OD trips and traffic conditions, except for the BM. This implies that the transfer learning of the TLARG contributes to improving the performance of DRL-based route guidance algorithm by the effective utilization of pre-trained models to accelerate learning in new tasks. Moreover, it is easily observed that the agents of using the BM and BTM pass through the AV-involved congested links, which highly affects the time of arrival, whereas the ATM bypasses the congested links. Consequently, the agent of using the ATM can reduce its OD travel time by approximately 23% compared to that of BTM. Such findings suggest that the proposed algorithm allows the DRL agent to complete its OD trips not only with flexible routes but also with reductions in OD travel time depending on given traffic situations.

B. PERFORMANCE REVIEW

Through the case study in the previous section, it could be found that the DRL-based route guidance models fine-tuned by using the TLARG provided the robust OD travel paths both in the 5×5 and 8×8 urban road networks with mixed traffic conditions. However, it still lacks sufficient evidence to generalize the effect of employing the TLARG, based solely on few specific cases. In other words, the overall performances of each model with respect to various traffic conditions and OD trips in different urban road networks have not yet been fully explored. Thus, this research considers different independent and identically distributed (i.i.d.) conditions to evaluate the generalized performance of the proposed algorithm. The following simulation experiments are generated by 500 i.i.d. random samples in each network scenario.

A statistical summary of OD travel times with each model in the two urban road networks is presented in Table 3, where the statistics represent the mean and standard deviation values of OD travel times in the i.i.d. cases. It is seen that the average and standard deviation values of BTM's OD travel times are less than those of BM's OD travel times in the 5×5 road network scenario. One can also observe that the mean value of OD travel times in the 8×8 road network scenario is the greatest in the order of BM, BTM, and ATM, while the standard deviation value of OD travel times in the wide-area network is the largest in the reverse order.

TABLE 3. Statistics of mean and standard deviation for OD travel time with each model in different road networks (unit: second).

Road Configuration	Model	Statistics
5X5 road network	BM	596.727 ± 66.142
	BTM	506.987 ± 51.624
8X8 road network	BM	1083.317 ± 68.915
	BTM	982.377 ± 91.847
	ATM	832.567 ± 110.296

Such statistical results represent that the DRL-based route guidance algorithms can benefit from the fine-tuned models, including BTM and ATM in terms of OD travel time, since the average values of OD travel times for the basic and advanced TLARG models are less than that of the baseline model. However, in the cases of 8×8 road network scenario, it is found that the standard deviation values of OD travel times for the BTM and ATM are greater than that of BM. Moreover, the ATM exhibits a larger standard deviation in OD travel time compared to the BTM. Hence, it is plausible to doubt that several exceptional instances could underestimate the basic and advanced TLARG models.

To further analyze the effect of imposing transfer learning in the TLARG, this comparison study introduces an additional performance metric, which is time saving. It is directly measured by the reduction in OD travel time for each i.i.d. case in 8×8 road network. The measurements of time savings

for the BTM and AM are formulated as (15) and (16).

$$S(BTM) = \left(1 - \frac{OD \text{ Travel Time of } BTM}{OD \text{ Travel Time of } BM}\right) \times 100(\%), \quad (15)$$

$$S(ATM) = \left(1 - \frac{OD \text{ Travel Time of } ATM}{OD \text{ Travel Time of } BM}\right) \times 100(\%), \quad (16)$$

where $S(BTM)$ and $S(ATM)$ indicate the percentage of relative time saving for the BTM and ATM compared with the BM.

Fig. 8 describes the ECDFs of time savings for the BTM and ATM in 8×8 road network. It is easily found that incorporating the transfer learning scheme in the TLARG always have positive outcomes since the minimum value of time savings for the BTM and ATM is equal to 0. In addition, it is observed that the BTM shows time savings of less than 5% in 40% of cases, and time savings of less than 10% in approximately 70% of cases. One can also observe that the BTM exhibits time savings ranging from 10 to 20% in nearly 15% of cases, and time savings larger than 20% in the rest of cases, with the largest time saving over 30%. On the other hand, it is found that the ATM generally provides better time savings than the BTM. The ATM shows time savings of more than 10% in 80% of cases. In addition, it is also observed that the ATM yields time savings larger than 20% in more than half of all cases. Furthermore, more than 22% of cases shows time savings greater than 30% with the ATM compared to the BTM, with the largest time saving close to 44%.

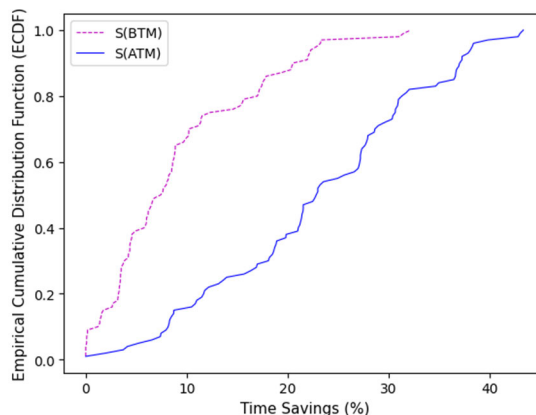


FIGURE 8. Empirical Cumulative Distribution Functions (ECDF) of Time savings for BTM and ATM in 8×8 road network.

C. COMPARISON STUDY

In order to verify the performance differences among the models in each identical OD trip and traffic environment, the comparison study performs several Wilcoxon signed rank tests based on some paired comparisons, such as BM-BTM, BM-ATM, and BTM-ATM. The Wilcoxon signed rank test is conducted with one-sided hypothesis test with a significance level of 0.01. The hypothesis is set up by using the difference

TABLE 4. P-value on wilcoxon signed rank test for OD travel time with each paired comparison model.

H ₀ : median of $\delta_s = 0$ vs H ₁ : median of $\delta_s > 0$		
Paired Comparison	5X5 network scenario	8X8 network scenario
BM – BTM	$1.941e^{-18}$	$8.856e^{-18}$
BM – ATM	–	$2.839e^{-18}$
BTM – ATM	–	$2.845e^{-18}$

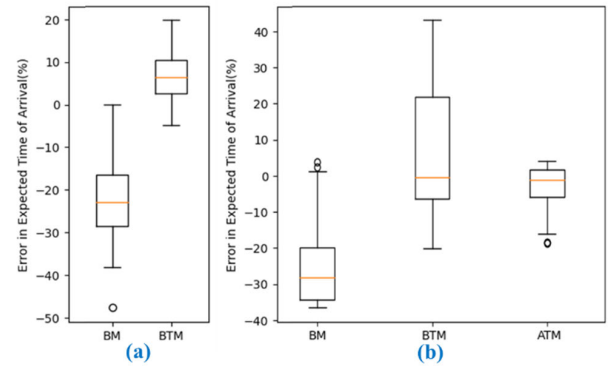
in OD travel time between paired comparison models in given traffic condition c , which is represented as δ_c . Thus, δ_c for s i.i.d. conditions is indicated as δ_s , where s is 500. Table 4 shows the results of the one-sided Wilcoxon signed rank tests. One can observe that there are p-values of much less than 0.01 in each scenario with the paired comparison models, which suggests that there is enough evidence to accept the alternative hypothesis. This implies that the BTM and ATM exhibit shorter OD travel times than the BM in the i.i.d. cases, which coincides with the previous research finding in *B. PERFORMANCE REVIEW*. Moreover, it is also found that the ATM provides shorter OD travel time paths compared to the BM and BTM even in a wide-area urban road network with mixed traffic conditions. In other words, the transfer learning of TLARG enables the advanced model to have a significant advantage over the conventional DRL approaches without retraining the model from scratch when a DRL-based route guidance algorithm is applied to certain urban area with uncertain traffic conditions.

One the other hand, it is still imperative to consider the service reliability of the route guidance service when applying the TLARG-based routing algorithm to the service area coexisting with AV-based mobility services, which is highly affected by the robustness of the TLARG for the changes in near future traffic conditions. Hence, the comparison study explores the robustness of the proposed algorithm by comparing between predicted and actual ETAs for the i.i.d. cases. The error between predicted and actual ETAs at the c^{th} i.i.d. case e_c is measured by (17).

$$e_c = \left(\frac{\text{predicted OD Travel Time}_c}{\text{actual OD Travel Time}_c} - 1 \right) \times 100(\%) \quad (17)$$

Fig. 9 shows the boxplots of errors in ETA with respect to different TLARG-based models in 5×5 and 8×8 road networks. As seen in (a) of Fig. 9, the median error for BM is much lower than that for BTM, whereas the absolute value of the median error for BM exceeds that for BTM by more than twice. Moreover, one can also find that the BTM exhibits a smaller Interquartile Range (IQR) than that of BM. Such results indicate that the BTM shows less variability of OD travel time compared with the BTM.

As shown in (b) of Fig. 9, the median error for BM in the wide-area road network is much lower than before, which suggests that it is not appropriate to apply the baseline model without the transfer learning to a wide range of service network. In addition, it is still observed that the BM often provides the travel paths which tend to underestimate

**FIGURE 9. Boxplots for errors in Expected Time of Arrivals (ETA) with different TLARG-based models in each road network: (a) 5×5 road network (b) 8×8 road network.**

the ETA. Furthermore, one can also observe that the BTM shows a larger IQR than before, though the median error for BTM is close to zero. This indicates that the BTM shows poor reliability of route guidance service due to increases in variability of OD travel times in the wide-area urban network with mixed traffic conditions.

The most exceptional outcome can be observed in the advanced TLARG model. It is easily found that there is the median error close to zero in the ATM. Moreover, it is also seen that the ATM exhibits a very small IQR compared to other models. Therefore, it suggests that more reliable route guidance service can be provided by using the ATM, rather than the BM and BTM, particularly in the wide range of service area with uncertain traffic conditions. In other words, it demonstrates that the robustness of the advanced TLARG model by measuring the error of ETA for diverse OD trips in the wide-area urban road network with the mixed traffic conditions.

V. CONCLUDING REMARKS

The main objective was to design a robust and interoperable DRL-based route guidance algorithm used for the near future coexisting AVs and HVs in a wide-area urban network. This research proposed the TLARG algorithm, which is an extended framework for the previously developed DRL-VR algorithm by incorporating transfer learning scheme. The use of predictive traffic information and interoperable MDP expression were considered for implementing the transfer learning scheme in the TLARG. The former was used for describing the time-varying dynamic variables as the predictive representation in the state variables of the MDP formulation. Therefore, it allowed the TLARG to generate the OD travel path by specifying the state variables for any given time, even before reaching a specific state. The latter was utilized for dealing with the complexity related to diverse OD trips in different road networks. Furthermore, it enabled the TLARG to generate the OD route guidance data even in spaces not explicitly experienced during training process. Consequently, the TLARG could provide more reliable route guidance service than previous one in a wide-area road

network, particularly in diverse OD trip with mixed traffic conditions.

This research conducted several evaluation studies based on microscopic traffic simulation experiments in different ranges of urban road network with uncertain traffic conditions caused by the AV-involved congestions. The evaluation studies performed case study, performance review, and comparison study based on using three types of TLARG-based models, including BM, BTM, and ATM. The case study explored that the characteristics of the TLARG-based models with respect to different service networks and agents' OD trips. Through the case study, it could observe that the TLARG enabled the agent to complete its OD trips not only with flexible routes but also with reductions in travel time depending on given traffic situations irrespective of network type. Such research findings suggest that the TLARG can be used for regenerating a new global path even if a given driving situation enforces the Automated Driving System (ADS) to satisfy its safety envelope at operational or tactical level when the AV encounters an imminent Object and Event Detection and Response (OEDR). In addition, unlike the case study that was based solely on few specific cases, the performance review analyzed the generalized performances of each TLARG-based model with respect to various traffic conditions and OD trips in different urban road networks. Based on the performance review, it could find that incorporating the transfer learning scheme in the TLARG always have positive outcomes. In addition, it is also found that the ATM generally provided better time savings than the BTM. The comparison study verified that the ATM provided shorter OD travel time paths compared to the BM and BTM, even in a wide-area urban road network with mixed traffic conditions. Furthermore, the comparison study also demonstrated that the robustness of the advanced TLARG model by measuring the error of ETA for diverse OD trips in the wide-area urban road network with the mixed traffic conditions. Hence, this study conclude that the TLARG has great potential to enhance the punctuality of mobility service by providing robust route guidance, even in the era of coexisting AVs and HVs.

There are several variations of TLARG that can be further extended in future research. One can enhance the performance of TLARG by replacing the RL-based traffic prediction model as well as the DRL-based route guidance model used in discrete action space with other advanced models, such as Graph Multi-Attention Network (GMAN), Spatio-Temporal Graph Attention Network (ST-GRAT), Conservative Q-Learning (CQL), and Implicit Q-Learning (IQL) [43], [44], [45], [46]. The choice of routes might also be considered to include U-turn that can increase the combinations of links for OD travel paths within the target area of mobility service. In addition, some Multi-Agent Reinforcement Learning (MARL) schemes will be incorporated in the TLARG to accelerate its convergence speed. Furthermore, additional analyses might also be conducted regarding the impact of changes in penetration rates of AVs on the proposed algorithm in future study.

REFERENCES

- [1] G. Ben-Dor, E. Ben-Elia, and I. Benenson, "Determining an optimal fleet size for a reliable shared automated vehicle ride-sharing service," *Proc. Comput. Sci.*, vol. 151, pp. 878–883, Jan. 2019.
- [2] Y. Zhou and M. Xu, "Robotaxi service: The transition and governance investigation in China," *Res. Transp. Econ.*, vol. 100, Sep. 2023, Art. no. 101326.
- [3] J. Bischoff, K. Führer, and M. Maciejewski, "Impact assessment of autonomous DRT systems," *Transp. Res. Proc.*, vol. 41, pp. 440–446, Oct. 2019.
- [4] S. Choi, D. Lee, S. Kim, and S. Tak, "Framework for connected and automated bus rapid transit with sectionalized speed guidance based on deep reinforcement learning: Field test in Sejong city," *Transp. Res. C, Emerg. Technol.*, vol. 148, Mar. 2023, Art. no. 104049.
- [5] Y. Wang, Z. Han, Y. Xing, S. Xu, and J. Wang, "A survey on datasets for the decision making of autonomous vehicles," *IEEE Intell. Transp. Syst. Mag.*, vol. 16, no. 2, pp. 23–40, Apr. 2024.
- [6] M. T. Ashraf, K. Dey, S. Mishra, and M. T. Rahman, "Extracting rules from autonomous-vehicle-involved crashes by applying decision tree and association rule methods," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2675, no. 11, pp. 522–533, Nov. 2021.
- [7] M. Makridis, K. Mattas, and B. Ciuffo, "Response time and time headway of an adaptive cruise control," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1677–1686, Oct. 2019.
- [8] M. Makridis, K. Mattas, A. Anesiadou, and B. Ciuffo, "OpenACC. An open database of car-following experiments to study the properties of commercial ACC systems," *Transp. Res. C, Emerg. Technol.*, vol. 125, Apr. 2021, Art. no. 103047.
- [9] M. Makridis, K. Mattas, B. Ciuffo, F. Re, A. Kriston, F. Minarini, and G. Rognelund, "Empirical study on the properties of adaptive cruise control systems and their impact on traffic flow and string stability," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 4, pp. 471–484, Apr. 2020.
- [10] S. Tak, S. Kim, H. Yu, and D. Lee, "Analysis of relationship between road geometry and automated driving safety for automated vehicle-based mobility service," *Sustainability*, vol. 14, no. 4, p. 2336, Feb. 2022.
- [11] E. Cascetta, A. Carteni, and L. Di Francesco, "Do autonomous vehicles drive like humans? A Turing approach and an application to SAE automation level 2 cars," *Transp. Res. C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103499.
- [12] J. A. Matute-Peaspan, A. Zubizarreta-Pico, and S. E. Diaz-Briceno, "A vehicle simulation model and automated driving features validation for low-speed high automation applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7772–7781, Dec. 2021.
- [13] M. A. P. Taylor, "Travel through time: The story of research on travel time reliability," *Transportmetrica B, Transp. Dyn.*, vol. 1, no. 3, pp. 174–194, Dec. 2013.
- [14] D.-D. Zhu and J.-Q. Sun, "A new algorithm based on Dijkstra for vehicle path planning considering intersection attribute," *IEEE Access*, vol. 9, pp. 19761–19775, 2021.
- [15] C. Wang, J.-S. Pan, H.-R. Xu, J. Jia, and Z.-Y. Meng, "An improved A* algorithm for traffic navigation in real-time environment," in *Proc. 3rd Int. Conf. Robot, Vis. Signal Process. (RVSP)*, Nov. 2015, pp. 47–50.
- [16] N. Sun, H. Shi, G. Han, B. Wang, and L. Shu, "Dynamic path planning algorithms with load balancing based on data prediction for smart transportation systems," *IEEE Access*, vol. 8, pp. 15907–15922, 2020.
- [17] M. G. H. Bell, "Hyperstar: A multi-path Astar algorithm for risk averse vehicle navigation," *Transp. Res. B, Methodol.*, vol. 43, no. 1, pp. 97–107, Jan. 2009.
- [18] K. Kim, M. Kwon, J. Park, and Y. Eun, "Dynamic vehicular route guidance using traffic prediction information," *Mobile Inf. Syst.*, vol. 2016, pp. 1–11, Jul. 2016.
- [19] Q. Song, D. Li, and X. Li, "Traffic prediction based route planning in urban road networks," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 5854–5858.
- [20] P. Perez-Murueta, A. Gómez-Espinosa, C. Cardenas, and M. Gonzalez-Mendoza, "Deep learning system for vehicular re-routing and congestion avoidance," *Appl. Sci.*, vol. 9, no. 13, p. 2717, Jul. 2019.
- [21] H. Wu, H. Zhou, J. Zhao, Y. Xu, B. Qian, and X. Shen, "Deep learning enabled fine-grained path planning for connected vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10303–10315, Oct. 2022.

- [22] M. Shaygan, C. Meese, W. Li, X. Zhao, and M. Nejad, "Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities," *Transp. Res. C, Emerg. Technol.*, vol. 145, Dec. 2022, Art. no. 103921.
- [23] M. Nazari, A. Oroojlooy, L. Snyder, and M. Takác, "Reinforcement learning for solving the vehicle routing problem," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9839–9849.
- [24] J. J. Q. Yu, W. Yu, and J. Gu, "Online vehicle routing with neural combinatorial optimization and deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3806–3817, Oct. 2019.
- [25] J. Zhao, M. Mao, X. Zhao, and J. Zou, "A hybrid of deep reinforcement learning and local search for the vehicle routing problems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 7208–7218, Nov. 2021.
- [26] S. Koh, B. Zhou, H. Fang, P. Yang, Z. Yang, Q. Yang, L. Guan, and Z. Ji, "Real-time deep reinforcement learning based vehicle navigation," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106694.
- [27] B. Hou, K. Zhang, Z. Gong, Q. Li, J. Zhou, J. Zhang, and A. de La Fortelle, "SoC-VRP: A deep-reinforcement-learning-based vehicle route planning mechanism for service-oriented cooperative ITS," *Electronics*, vol. 12, no. 20, p. 4191, Oct. 2023.
- [28] D. Lee, S. Tak, and S. Kim, "Development of reinforcement learning-based traffic predictive route guidance algorithm under uncertain traffic environment," *IEEE Access*, vol. 10, pp. 58623–58634, 2022.
- [29] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial-temporal graph modeling," 2019, *arXiv:1906.00121*.
- [30] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, 2016, pp. 2094–2100.
- [31] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," 2015, *arXiv:1511.05952*.
- [32] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
- [33] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.
- [34] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer learning in deep reinforcement learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13344–13362, Jun. 2023.
- [35] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiessner, "Microscopic traffic simulation using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2575–2582.
- [36] J. Zambrano-Martinez, C. Calafate, D. Soler, J.-C. Cano, and P. Manzoni, "Modeling and characterization of traffic flows in urban environments," *Sensors*, vol. 18, no. 7, p. 2020, Jun. 2018.
- [37] L. Urquiza-Aguiar, W. Coloma-Gómez, P. Barbecho Bautista, and X. Calderón-Hinojosa, "Comparison of SUMO's vehicular demand generators in vehicular communications via graph-theory metrics," *Ad Hoc Netw.*, vol. 106, Sep. 2020, Art. no. 102217.
- [38] X. Wu and H. X. Liu, "A shockwave profile model for traffic flow on congested urban arterials," *Transp. Res. B, Methodol.*, vol. 45, no. 10, pp. 1768–1786, Dec. 2011.
- [39] M. Ramezani and N. Geroliminis, "Exploiting probe data to estimate the queue profile in urban networks," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1817–1822.
- [40] M. Ramezani and N. Geroliminis, "Queue profile estimation in congested urban networks with probe data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 30, no. 6, pp. 414–432, Jun. 2015.
- [41] A. Kotsi, E. Mitsakis, and D. Tzani, "Overview of C-ITS deployment projects in Europe and USA," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.
- [42] S. Tak, K. Kang, and D. Lee, "Development of V2I2V communication-based collision prevention support service using artificial neural network," *J. Korea Inst. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 126–141, Oct. 2019.
- [43] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1234–1241.
- [44] C. Park, C. Lee, H. Bahng, Y. Tae, S. Jin, K. Kim, S. Ko, and J. Choo, "ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 1215–1224.
- [45] A. Kumar, A. Zhou, G. Tucker, and S. Levine, "Conservative q-learning for offline reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1179–1191.
- [46] I. Kostrikov, A. Nair, and S. Levine, "Offline reinforcement learning with implicit q-learning," 2021, *arXiv:2110.06169*.



DONGHOUN LEE was born in Seoul, South Korea, in 1988. He received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2011, and the M.S. and Ph.D. degrees in civil and environmental engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2014 and 2018, respectively. From 2018 to 2019, he was a Postdoctoral Researcher with the KAIST AI Mobility Laboratory. From 2019 to 2023, he was an Associate Research Fellow with the Mobility Transformation Department, Korea Transport Institute (KOTI). He is currently an Assistant Professor with the Department of Artificial Intelligence, Sejong University, Seoul. His research interests include advanced driver assistance systems, automated driving systems, deep learning, and deep reinforcement learning. He received the Best Student Paper Award from the IEEE Intelligent Vehicles Symposium, in 2015.

• • •