2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 27 - October 1, 2021. Prague, Czech Republic

Automated Type-Aware Traffic Speed Prediction based on Sparse Intelligent Camera System

Xiaoyang Xie[1], Kangjia Shao[2], Yang Wang[3], Fei Miao[4], Desheng Zhang[1]

2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-1714-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IROS51168.2021.9636559

Abstract— **Many essential services for autonomous vehicles, e.g., navigation on high-quality maps, are designed based on the understanding of traffic conditions, e.g., travel time/speed on road segments, traffic flow, etc. However, most existing traffic condition models lack the consideration of the differentiation for vehicles with different types (e.g., personal vehicles or trucks) and thus they cannot satisfy some type-specific services, e.g., traffic-condition-based routing for autonomous vehicles with different types. To address this challenge, we design a novel vehicular mobility based sensing model called mDrive to predict the travel speed on the road segments, which is targeted for different types of vehicles by utilizing the camera data obtained from the traffic cameras equipped in the road intersections only, without any in-vehicle GPS devices. mDrive addresses the type-aware traffic speed prediction problem with sparse sensors based on three correlations: (1) the spatial correlation of travel speed on the connected road segments; (2) the temporal correlation of travel speed on the consecutive time slots; (3) the type correlation of different vehicular types' speed on the same road segment. We implement mDrive on traffic camera data from the Chinese city Suzhou and evaluate it by using the detailed GPS data from personal vehicles, taxis, and trucks, with road contextual data as ground truth. The experiment show mDrive outperforms state-of-the-art methods by reducing 6.2% mean relative error on average for all types of vehicles.**

I. Introduction

The traffic conditions on road segments are fundamental to many mobility-on-demand applications for autonomous vehicles, including navigation[1], lane change maneuver[2] for personal autonomous vehicles, etc. Most autonomous vehicles make decisions based on data collected from vehicles' sensors. For example, [3] provides real-time vehicle and pedestrian tracking for autonomous vehicles based on their equipped Lidars and cameras. However, due to the limitation of sensors' sensing ranges, vehicles may not obtain the global information for the system-level decision making. Autonomous vehicular applications based on V2X communication may address this issue, such as [4] providing global cruising control for vehicles based on the sharing information through V2X communication. Therefore, global traffic information is essential to autonomous vehicles.

The urban scale traffic condition prediction has been widely studied and could be extended to the applications for autonomous vehicles through V2X technology. In particular, many approaches have been designed based on stationary sensors, such as electronic toll stations[5], loop detectors[6]. However, due to the lack of vehicle type information, most of these works are type-agnostic, i.e., modeling the mobility of vehicles without considering the distinct features of the vehicular types. Traffic conditions on the same road segment might be different for vehicles with different types, because of different vehicular types' mobility features, e.g., driving behaviors (e.g., taxis vs. trucks), different sizes of vehicles (e.g., light-duty personal vehicles vs. heavy-duty trucks), and different traffic regulations on the road segments. To verify this, we calculate the travel speed for all types of vehicles on each road segment in a Chinese city, Suzhou, with one month of vehicular GPS data as Ground Truth. Because it is hard to find two vehicles with the exact same speed on a road segment in the original granularity, we coarsen speed into multiple levels and calculate how many roads that all vehicles have the same speed level for every 15 minutes. We show the result in Figure 1. We found there are less than 30% of roads where all vehicles have the same speed level every 15 minutes. Thus, most type-agnostic traffic condition models might have inaccurate results for some services because they fail to consider the difference in vehicle types, e.g., navigation or anomaly detection, leading to extra costs for vehicles of specific types.

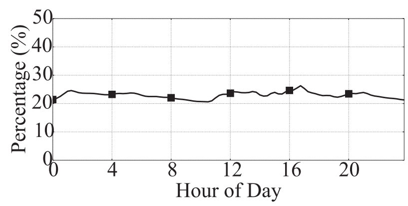

Fig. 1: % of road segments having Same Speed

There exist some works to apply the prediction model for travel speed prediction based on the data of each vehicle type individually. For example, some methods utilize the spatial correlations among road networks to infer the missing sensor data based on the Graph Convolution Network (GCN)[7][8]. However, these models are designed for one specific vehicle type only and miss out on the correlation between different vehicle types on the same road segment. For instance, for the prediction of the speed of taxis on a road segment without any taxi data, the existing work may use taxi data on other road segments or historical data. However, we argue that the speed of other vehicle types on the same road segment or surrounding segments might be more helpful based on their correlations. Hence, in this paper, we integrate data from multiple vehicle types for travel speed prediction. We argue

[1]Xiaoyang Xie and Desheng Zhang are at Rutgers University, USA {xx88, Desheng}@cs.rutgers.edu
[2]Kangjia Sha is at University of Science and Technology of China, Suzhou, China SA517285@mail.ustc.edu.cn
[3]Yang Wang is the corresponding author, at University of Science and Technology of China, Hefei, China angyan@ustc.edu.cn
[4]Fei Miao is at University of Connecticut, USA fei.miao@uconn.edu

that although the mobility features of different vehicle types are different, the speed of one vehicle type on a road segment may be used for predicting the speed of another vehicle type.

In this paper, we design a traffic condition prediction model called mDrive to predict the type-aware travel speed on citywide road segments by utilizing the intelligent traffic camera system in Suzhou. Figure 2 shows the heatmap of the distribution of the intelligent traffic cameras in Suzhou and an example of the captured image from a camera. We found the distribution of traffic cameras is sparse and each camera captures the real-time information of vehicles in the intersections, such as plate numbers, vehicle types, timestamps, locations, etc.

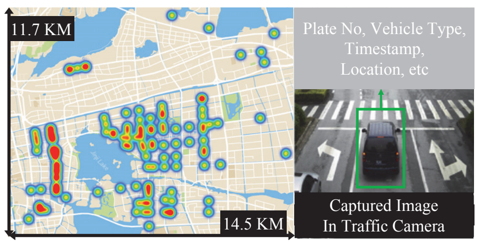

Fig. 2: Heatmap of Cameras in Suzhou

Given the data from traffic cameras, mDrive predicts the future travel speed through the combination of Graph Convolution Network (GCN) and Gated Recurrent Unit (GRU) layers to address three correlations. The detail of the workflow is shown in follows: ***Step 1:*** we utilize an Expectation–Maximization (EM) model to estimate the average type-specific travel speed on road segments with the observation from camera data; ***Step 2:*** mDrive feeds the estimated travel speed of partial road segments for a time slot into a GCN module to learn the **spatial correlation** between adjacent road segments and the **type correlation** between different vehicle types; ***Step 3:*** given travel speed of a sequence of time slots, mDrive utilizes the GRU module to learn the **temporal correlation** between consecutive time slots. With benefits from these three correlations, mDrive predicts the future travel speed on road segments for each vehicle type at the next time slot. The contributions of this paper are:

- To the best of our knowledge, we conduct the first study to infer the type-aware travel speed in a city based on data collected from the traffic camera system as infrastructure to assist current or future autonomous driving applications such as navigation, complementary to applications based in-car sensors. Our work advances the state-of-the-art traffic condition sensing methods in two aspects: 1) our method provides the type-aware travel speed prediction for different vehicular types; 2) our method utilizes the existing infrastructures and captures the sparse traces of vehicles without GPS devices, potentially alleviating the privacy issue.
- We design a novel traffic condition prediction model called mDrive to predict the real-time type-aware travel speed in a city for different vehicular types. mDrive utilizes GCN module to infer the travel speed on road segments of a city for each specific type based on a set of correlations we carefully model. With the completed inferred travel speed of previous time slots, mDrive utilizes the GRU module to predict the future road segment travel speed for multiple vehicular types. The key novelty of mDrive is to take advantage of the combination of GCN and GRU to learn three correlations, i.e., spatial correlation, type correlation, and temporal correlation.
- We implement mDrive based on the real-world data obtained from a Chinese city, Suzhou, which captures around 9.3 million daily camera records. We evaluate mDrive with the ground truth obtained from 10 thousand taxicabs, 3 thousand personal vehicles, and 3 thousand trucks in Suzhou. Compared to the state-of-the-art methods, our mDrive outperforms them for all vehicular types by reducing the mean relative error (MRE) by 6.2% on average.

II. Related Work

Autonomous Vehicle Assistant System: Many works on urban sensing have been conducted to provide various vehicular services, e.g., autonomous vehicle assistant. To improve driving safety and efficiency, many of these works are designed based on in-vehicle sensors, e.g., [2] provides a lane change maneuver in real-time dynamic traffic conditions for autonomous vehicles based on the model predictive control method. In contrast, some works utilize stationary infrastructures and V2X technology, e.g., [9] presents an architecture of the cybernetic transportation systems and an automated global planner for autonomous vehicles to provide door-to-door transportation services.

Traffic Condition for Autonomous Vehicles: Traffic condition prediction is a classic topic in the transportation domain, which has been widely studied. Some works for traffic condition prediction are designed based on data from the sensors in transportation infrastructure, e.g., [5] utilizes the data from the electrical toll collections to estimate the travel condition on the highway system. The work on this topic could be extended to providing global traffic information for autonomous vehicles, e.g., [10] designs a control policy to make navigation decisions based on the traffic environment information, such as positions, velocities, and lane numbers, which is shared among autonomous vehicles through V2X communication.

Summary: Even with great advance, none of these works above provide traffic condition sensing for each specific type of vehicles. Compared with the existing works, **our work mDrive** leverages intelligent traffic cameras to predict the travel speed of road segments for different vehicular types. To our best knowledge, mDrive is the first work to estimate the type-aware fine-grained travel speed for the different types of vehicles only utilizing the data from the traffic camera system.

III. Design

We design mDrive to predict travel speed on road segments based on the pervasive GCN models[7][8]. We choose the

GCN model because GCN is well fitted to the topological traffic data on the road network. In this section, we first introduce the preliminary of mDrive and then present the data flow of mDrive followed by the design of mDrive.

TABLE I: Data Description of Cameras in Suzhou

# of Cameras	# of Intersections	# of Road Segments	Daily Volume	Daily Records
95	3467	6930	200 K	2 M
Format				
Vehicle ID	Date&Time	Lon	Lat	Vehicle Type

A. Preliminary

1) Data Description: We have access to the dataset collected by the traffic camera system in Suzhou as shown in Table I through our local collaborators. This system is used for the surveillance of traffic violations for vehicles on the road, such as drive through a stop signal. The system utilizes high-speed cameras to capture the situation of intersections with high frequency and applies sophisticated technology to reduce the error of camera data. When a vehicle is passing an intersection equipped with a camera, this vehicle is detected, and then the corresponding record is captured, uploaded, and recognized with high accuracy. The used data fields of records are shown in Table I, including the vehicle ID, date&time, longitude, latitude, and vehicle type. In particular, the traffic camera could capture the snapshot of many objects including vehicles, motorcycles, and pedestrians. For the vehicle types, we only include three vehicle types, i.e., taxis, trucks, and personal vehicles because we have the ground truth of these types based on a sample of GPS data.

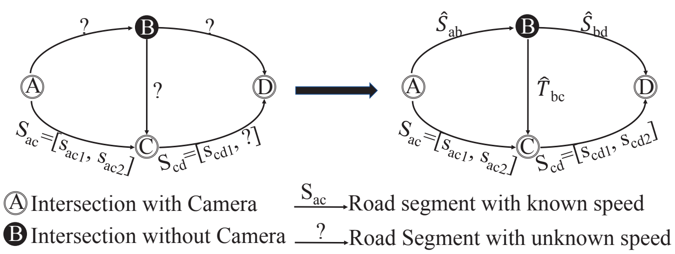

Fig. 3: Scenario of Travel Speed Prediction

2) Problem Definition: To better illustrate the travel speed prediction, we first give the formal definitions.

- **Road Segments:** A road network consists of the intersections and the road segments connecting these intersections, e.g., the graphs in Figure 3. The road segment is the smallest unit for a road network, e.g., the edges AB, BD, AC, CD, BC in Figure 3. One pair of intersections is connected by several consecutive road segments.
- **Observation**: An observation in this paper is defined as a pair of two continuous records of the same vehicles collected by the traffic cameras at two intersections. When a vehicle passes two intersections equipped with traffic cameras, one observation is generated.
- **Routes**: A route is a sequence of connected road segments that link two intersections, e.g., route AB-BD in Figure 3.

Figure 3 shows a scenario of our travel speed prediction. In the scenario, cameras can only obtain the observations set O_{AD}, O_{AC}, O_{CD} because only intersections A, C, D are equipped with cameras, where O_{AD} represents the observations collected in intersection A and D, and so on. Besides, the collection of records from these equipped intersections with cameras may lack the data of one specific vehicular type during a time slot, e.g., the travel speed of vehicular type 2 on the road segment *CD* is missing. Furthermore, if the vehicle drives through *A*, *B*, and *D*, the observation of *AD* only provides the aggregated travel time. Note that if the vehicle drives through *A*, *C*, and *D*, then there will be two observations *AC* and *CD*. Hence, we utilize an Expectation Maximization (EM) based module on these aggregated data to obtain speed for partial road segments, and then feed them to the GCN module.

B. Design of mDrive

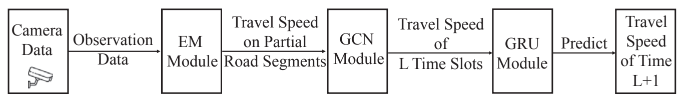

Fig. 4: Data Flow of mDrive

1) Data Flow: **Input:** As shown in Figure 4, we first preprocess *camera data* to obtain observations and then input them into the EM module to infer the travel time for partial road segments for each time slot. Here each time slot is for 15 minutes as we partition one day into 96 time slots. Then EM module estimates the travel time of each time slot for all vehicle types on the road segments that are involved in the selected route. For the estimated travel time for a time slot, we calculate type-aware travel speed and feed them into our GCN module. The output of the GCN module is the travel speed on all road segments for the same time slot. **Output:** Finally, we feed the travel speed of road segments for L consecutive time slots into the GRU module to predict the *travel speed of road segments in time slot* $L+1$.

2) EM based Travel Time Inference Model: The EM-based model[5] estimates the travel time by three steps, i.e., initialization, E-step, and M-step. In the **initialization**, the travel time of one observation in the historical camera data is divided into travel time for each road segment on the shortest route between two intersections of the observation. This division is based on the proportion of road segment's lengths in the total length of the shortest route between these two intersections. For each road segment involved in the historical data, we obtain its travel time distribution. In the **E-step**, the EM model utilizes a route inference method to obtain the travel time spent on each road segment in the selected route. It first samples travel time based on the distribution for all possible road segments between two intersections. Then, it selects the inferred route as the route whose sum of travel time is closest to the travel time of the new observation. Then we resample the travel time to road segments in the route under the constraint that their sum is equal to the travel time of the observation. We address this sampling process by using a fast simulation algorithm[11] for the multinormal distribution. In the **M-step**, given a duration, e.g., 15 minutes, we group the estimated travel time of the same road segments obtained from the observations collected within this duration and update the current distribution by the parameter updating methods in [5].

We run the EM model iteratively to obtain the travel time distribution, and then calculate the average speed on involved road segments. With the EM module, we obtain travel speed for a part of road segments. However, due to the complexity of the road network in a city and sparse issues, many road segments still lack travel speed data. To address this issue, we use the GCN module for the completion of the inference for travel speed.

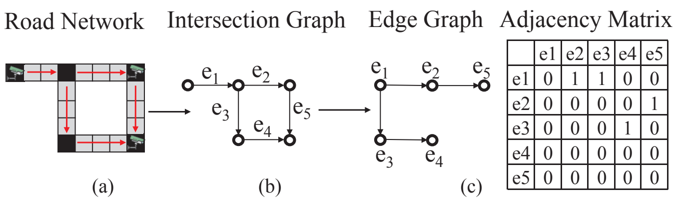

Fig. 5: Example of Graph Construction

3) Architecture of Neural Network: We show an example of the road network in Figure 5(a). In Figure 5(a), the gray grid represents the road segment; the black grid represents the intersection. This road network example is directly represented by a directed graph $G = (V,E)$ as shown in Figure 5(b), where the vertex set V will be referred to as the intersections of the road network, and the edge set E will be referred to as the road segments of the road network. As shown in the Figure 5(b), there are 5 edges (road segments) in the graph, i.e., $e_1, e_2, \ldots, e_5$. However, our final objective is to estimate travel speed for all the edges. To fit the graph into the mDrive, we convert the road graph into an edge graph $G = (E,A)$, where E is the road segment set and A is the adjacency matrix that indicates the connectivity of road segments. Figure 5(c) shows an example of the edge graph and the adjacency matrix. In this edge graph, e_1 and e_2 are connected because it is feasible for a vehicle to travel from e_1 to e_2 through the intersection between them. In the adjacency, $a_{ij} = 1$ indicates e_i and e_j are connected, and *vice versa*. Figure 6 shows the design of GCN and GRU modules. **(1) GCN Module:** In each time slot, we use the edge adjacency matrix $A \in R^{n\times n}$, and the speed matrix $H \in R^{n\times K}$ as **inputs**, where H_l is the input speed matrix in time slot l; n is the number of road segments; K is the number of vehicular types. Each entry in the input matrix H_l is the travel speed of a road segment for one specific vehicle. In particular, if the speed on one road segment is unknown, its corresponding entry is empty in the matrix. In detail, we apply a variant GCN version of ChebNet[12] to improve efficiency. For the adjacency matrix, we obtain the Laplacian variant, i.e., $\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}$, where $\hat{A} = A + I$; I is the identify matrix for A; $\hat{D}$ is the degree matrix of $\hat{A}$. Therefore, one GCN layer could be defined as a function $f(H_l, A) = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H_lW)$, where σ is the activation function, W is the weight matrix for GCN. For a speed matrix, H_l the GCN module input W_l into 3 GCN layers to obtain one matrix $\hat{H}_l$, which is the estimated speed matrix for time slot i. Let I_i be the indicator that the speed of the i_{th} road segment is labeled if $I_i = 1$. The final matrix $\hat{H}_l$ follows a requirement that $\hat{h}_{ij} = h_{ij}$ if $I_i = 1$. For this objective, we utilize a least square error function as the loss function of mDrive, which

$$Loss(H_l, \hat{H}_l) = \sum_i^n I_i \sum_j^k \cdot (h_{ij} - \hat{h}_{ij})^2 \tag{1}$$

When we provide the input matrices, we remove part of known speed and use them as labels in the loss computing. Note we use GPS data as separate ground truth data for evaluation only. We train the GCN layers via the back-propagation to learn the correlations among all types.

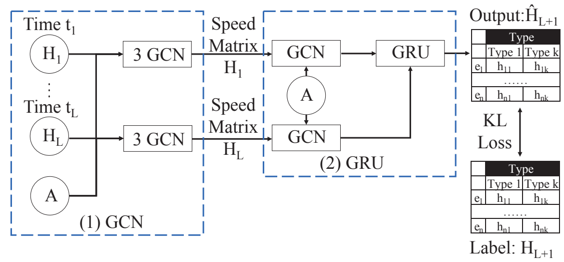

Fig. 6: Architecture of Neural Network

(2) GRU Module: In the previous L time slots, mDrive applies the GCN module to obtain the estimated speed matrix mentioned above for each time slot. Given these previous outputs from previous L time slots, mDrive utilizes one GCN layer and one GRU layer (i.e., the TGCN cell in [8]) to study the correlation of travel speed between consecutive time slots on the same road segments for different types of vehicles. In particular, the GRU layer receives the two inputs, i.e., the output matrix from the current GCN layer and the output matrix from the GRU layer from the previous time slot. The output of the final GRU layer is a speed matrix $\hat{H}_{L+1} \in R^{n\times K}$. The detail is shown in Figure 6, where each row in the output matrix $\hat{H}_{L+1}$ is a vector containing K speed, which are the travel speed of all the vehicular types on the road segment. The parameters of mDrive are learned to minimize a loss function based on the L1 loss.

IV. Evaluation

In this section, we show the evaluation based on vehicular GPS data from different vehicular types as Ground Truth in Suzhou with different factors, including vehicular types, time intervals of the day, and density of cameras. In particular, we evaluate mDrive by some metrics with state-of-the-art methods as baselines.

TABLE II: Vehicle Data in Suzhou

Vehicular Types	# of Vehicles	Sampling Rate	Daily Record
Personal Vehicle	3 K	Every 10 seconds	4 M
Taxicabs	4 K	Every 30 seconds	6 M
Truck	2 K	Every 15 seconds	5 M

A. Evaluation Setting

1) Training Setting: We employ a cluster with 3 large servers, each with 1 TB of memory, 80 cores, and 8 Nvidia 1080Ti GPUs to process data and implement mDrive. We utilize one-month data of the intelligent traffic camera data (shown in Table I) from the Suzhou Industrial Park for the experiment. During the training phase, we choose 70% of road segments whose mobility features are obtained as the input and 30% of these road segments as the target for training. We coarsen the road speed into several levels with every 10 km/h for each level as well as we did in Figure1.

2) Ground Truth: We utilize the vehicular dataset from Suzhou with the same period of camera data for experiments, including different vehicular types also shown in Table II. Here the camera data is preprocessed by our local collaborator and includes the spatial-temporal information of vehicles captured by the cameras. We utilize a map-matching algorithm on these GPS data to assign them to the road segment for the calculation of the travel time on each road segment. In particular, we utilize the location information of traffic cameras in Suzhou to emulate the camera data for those vehicles.

3) Metrics: We use Kolmogorov-Smirnov Goodness-of-Fit test (K-S Test)[13] to evaluate the estimated travel time distribution of EM module to the distribution obtained by the ground truth. Here we test the null hypothesis, which verifies if the travel time distribution of a road segment follows a specified distribution and the metric is defined as $K-S=\frac{N_f}{N_g}*100$, where N_f is the number of road segments that accept the null hypothesis of a vehicular type, and N_g is the number of road segments that accept the null hypothesis of the ground truth. In addition, we study the Mean Relative Error (MRE) between our estimated travel time and the ground truth, which is defined as $MRE=\frac{\sum_i^m |y_{e_i}|\hat{y}_{e_i}}{\sum_i^m y_{e_i}}$, where e_i is the road segment i, y_{e_i} is the ground truth, and $\hat{y}_{e_i}$ is the estimated travel time.

4) Baselines: To evaluate the performance of mDrive, we compare mDrive with the following state-of-the-art methods. History Average Speed method (HA), which is a straightforward method that uses the average historical travel speed on road segments directly to predict speed on the next time slot. Autoregressive Integrated Moving Average method (ARIMA) [14], which is a classic time series analysis method that is fitted to better understand time-series data to predict future traffic speed. Support Vector Regression method (SVR) [6], which is based on the Support Vector Machine (SVM) model for speed prediction. The kernel function used in the baseline is the linear kernel.

B. Goodness-of-Fit Result for Travel Time from EM module

We study the Goodness-of-Fit of the inferred travel time distribution by EM module and the ground truth for each vehicular type, which is to count the number of road segments that accept the null hypothesis. We show the performances of three vehicular types in Suzhou in Figure 7. The x-axis is the time of day and the y-axis is K-S defined in section IV-A.3. The higher the curve is, the closer the estimated travel time of the vehicular type to the ground truth. We found that personal vehicles have the best performance in the three vehicular types. In particular, during the daytime, the K-S value of personal vehicles is very large, and in the evening rush hour, it increases to about 90%. This is because, during the daytime and the rush hours, more vehicles are captured by traffic cameras, which increases the number of road segments covered by the observations. Similarly, because there are more taxis in the daytime compared to that of trucks, in the comparison of the other two types, the K-S value of taxis is better than that of trucks from 7 AM to 9 PM. With the travel time inferred by the EM module, we could obtain the speed matrix for GCN and GRU modules. Therefore, we evaluate the goodness-of-fit for the EM module to see the accuracy.

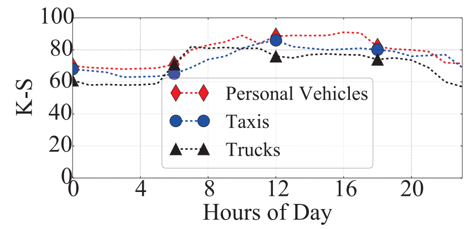

Fig. 7: K-S Test

C. Prediction Result on Travel Speed

In this section, we evaluate the performance of mDrive on the prediction of travel speed and compare it with three state-of-the-art methods, i.e., HA, ARIMA, and SVR, all of which provide the type-agnostic travel speed prediction.

1) Impact of Vehicular Types: We apply three baselines on three types of vehicles' data from Suzhou individually, and then compare them with the result of mDrive for each vehicular type. In Table III, we found the MRE of mDrive on all vehicular types is better than that of three baselines. mDrive reduces the MRE by around 1.5% on personal vehicles, 7.5% on taxis, 9.6% on trucks, compared to the average MRE of all baselines. This might be because the numbers of taxis and trucks in the city are smaller than that of personal vehicles. In this case, all three baselines only utilize the records of taxis or trucks; whereas mDrive considers the correlations among all three types, which may improve the accuracy of the travel time prediction.

TABLE III: Performance of Methods in terms of MRE

	Factor					
	Vehicular Types			Time-interval of Day		
	Personal Vehicles	Taxis	Trucks	Day Time	Night Time	Late Night
HA	0.404	0.412	0.415	0.52	0.4	0.408
SVR	0.41	0.415	0.416	0.44	0.4	0.408
ARIMA	0.405	0.429	0.42	0.5	0.395	0.5
mDrive	0.4	0.385	0.377	0.4	0.42	0.408

2) Impact of Time: We compare the performances of mDrive with three baselines on the data of all vehicles within three time-intervals in one day, i.e., the daytime (6 am to 6 pm), the nighttime (6 pm to 10 pm), and late-night (10 pm to 6 am) In Table III, we found the MRE of mDrive on daytime and late-night is better than that of three baselines. While at nighttime, mDrive is slightly worse than baselines. However, in the overall performance, mDrive still outperforms baselines by a 7.5% reduction on average.

3) Impact of Density: To evaluate the performance on regions with different camera densities, we randomly choose a portion of intersections in the Suzhou Industrial Park with percentages from 20% to 100% as active equipped intersections and utilize the corresponding data for evaluation. Then we apply mDrive on these densities and show the results in Figure 8. The metric used is the performance gain, which is the increased MRE of mDrive on a particular density compare to the MRE of mDrive on 100% density. Limited by the smaller number of cameras, the performance of mDrive on the 20% density increases the MRE by around 23% on average. Therefore, the density of cameras in a region is a very important factor in the performance of mDrive.

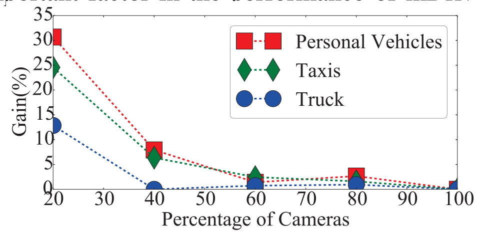

Fig. 8: Impact of Density

V. Discussion

Generalization for Autonomous Vehicles: This paper focuses on the type-aware travel speed prediction at road segment levels. However, our work could also be generalized to predict other traffic information, e.g., travel flow. The global traffic information could be extended to some topics for self-driving, such as mixed autonomy[15] by providing the global traffic condition information. Besides, the predicted traffic information could be utilized for mobility-on-demand services through V2X communication[16], e.g., intelligent traffic signal system and freight-Specific dynamic travel, etc.
Privacy Protection: We took two steps for privacy protections, i.e., 1) De-identification: All video data from traffic cameras is converted into text data, and all identifiable information, e.g., plate numbers are anonymized with a serial identifier by service providers. 2) Relative Locations: The utilized location information in the collected data is the location of cameras, instead of GPS coordinates of vehicles.
Ethics: All the data used in mDrive is legally collected by the service providers for traffic monitoring purposed by the Suzhou transportation department. In our analysis, we did not focus on any specific driver but only calculate the travel speed. Our work is to understand and improve traffic conditions, which in turn benefits all drivers.

VI. Conclusion

We design a vehicle-type aware traffic condition prediction model, mDrive, which predicts travel speed based on sparse camera data. The key novelty of mDrive is to utilize three correlations to obtain a high spatial-temporal coverage of prediction for all types of vehicles. mDrive considers the difference between different vehicular types, which may be applied to future autonomous vehicles. The evaluation shows the overall performance of mDrive is better than that of the three baselines by reducing 6.2% MRE on average. We expect that the design and evaluation of mDrive will provide technical insights for various future autonomous applications, e.g., high-resolution type-aware vehicle navigation, and traffic context-aware cruising decision making.

Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments and suggestions. This work is partially supported by NSF 1849238, NSF 1932223, NSF 1951890, NSF 1952096, NSF 2003874, NSF 2047822, NSF 1932250, NSF 1952096 grants.

References

[1] S. Niijima, R. Umeyama, Y. Sasaki, and H. Mizoguchi, "City-scale grid-topological hybrid maps for autonomous mobile robot navigation in urban area," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020 - January 24, 2021*. IEEE, 2020, pp. 2065–2071. [Online]. Available: https://doi.org/10.1109/IROS45743.2020.9340990

[2] Z. Li, J. Jiang, and W. Chen, "Automatic lane change maneuver in dynamic environment using model predictive control method," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2020, Las Vegas, NV, USA, October 24, 2020 - January 24, 2021*. IEEE, 2020, pp. 2384–2389. [Online]. Available: https://doi.org/10.1109/IROS45743.2020.9341729

[3] A. Buyval, A. Gabdullin, R. Mustafin, and I. Shimchik, "Realtime vehicle and pedestrian tracking for didi udacity self-driving car challenge," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2064–2069.

[4] Z. Liu, H. Wei, H. Hu, C. Suo, H. Wang, H. Li, and Y.-H. Liu, "A synchronization approach for achieving cooperative adaptive cruise control based non-stop intersection passing," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 236–242.

[5] Y. Yang, F. Zhang, and D. Zhang, "Sharededge: Gps-free fine-grained travel time estimation in state-level highway systems," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 48, 2018.

[6] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Physica A: Statistical Mechanics and its Applications*, vol. 534, p. 120642, 2019.

[7] J. Hu, C. Guo, B. Yang, and C. S. Jensen, "Stochastic weight completion for road networks using graph convolutional networks," in *Proceedings of 34th Ieee International Conference on Data Engineering, Icde*, 2019.

[8] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[9] M. Vaca, J. E. Traver, V. Milanes, J. Perez, D. Gonzalez, and F. Nashashibi, "Automated global planner for cybernetic transportation systems," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2016, pp. 1–6.

[10] S. Han, J. Fu, and F. Miao, "Exploiting beneficial information sharing among autonomous vehicles," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2226–2232.

[11] Y. Cong, B. Chen, M. Zhou, *et al.*, "Fast simulation of hyperplane-truncated multivariate normal distributions," *Bayesian Analysis*, vol. 12, no. 4, pp. 1017–1037, 2017.

[12] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[13] M. A. Stephens, "Edf statistics for goodness of fit and some comparisons," *Journal of the American statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

[14] Z. Zhao, W. Chen, H. Yue, and Z. Liu, "A novel short-term traffic forecast model based on travel distance estimation and arima," in *2016 Chinese Control and Decision Conference (CCDC)*. IEEE, 2016, pp. 6270–6275.

[15] K. Jang, E. Vinitsky, B. Chalaki, B. Remer, L. Beaver, A. A. Malikopoulos, and A. Bayen, "Simulation to scaled city: zero-shot policy transfer for traffic control via autonomous vehicles," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, 2019, pp. 291–300.

[16] US-DOT, 2021. [Online]. Available: https://www.its.dot.gov/pilots/cv_pilot_apps.htm