CatCharger: Deploying In-motion Wireless Chargers in a Metropolitan Road Network via Categorization and Clustering of Vehicle Traffic

Li Yan, Haiying Shen, *Senior Member, IEEE, Member, ACM* Juanjuan Zhao, Chengzhong Xu, *Fellow, IEEE* Feng Luo, Chenxi Qiu, Zhe Zhang and Shohaib Mahmud

***Abstract*—In metropolitan areas with heavy transit demands, electric vehicles (EVs) are expected to be continuously driving without recharging downtime. Wireless Power Transfer (WPT) provides a promising solution for in-motion EV charging. Nevertheless, previous works are not directly applicable for the deployment of in-motion wireless chargers due to their different charging characteristics. The challenge of deploying in-motion wireless chargers to support the continuous driving of EVs in a metropolitan road network with the minimum cost remains unsolved. We propose *CatCharger* to tackle this challenge. By analyzing a metropolitan-scale dataset, we found that traffic attributes like vehicle passing speed, daily visit frequency at intersections (i.e., landmarks) and their variances are diverse, and these attributes are critical to in-motion wireless charging performance. Driven by these observations, we first group landmarks with similar attribute values using the entropy minimization clustering method, and select candidate landmarks from the groups with suitable attribute values. Then, we use the Kernel Density Estimator (KDE) to deduce the expected vehicle residual energy at each candidate landmark and consider EV drivers' routing choice behavior in charger deployment. Finally, we determine the deployment locations by formulating and solving a multi-objective optimization problem, which maximizes vehicle traffic flow at charger deployment positions while guaranteeing the continuous driving of EVs at each landmark. Trace-driven experiments demonstrate that *CatCharger* increases the ratio of driving EVs at the end of a day by 12.5% under the same deployment cost.**

***Index Terms*—Vehicle wireless charging, charger deployment, mobile data analysis, kernel density estimation**

I. Introduction

Electric Vehicle (EV) industry has been burgeoning in recent years because of the quick depletion of fossil fuels [1], [2]. In countries like China, India and USA, governments are establishing new policies to replace gasoline-based vehicles with electric ones [3]–[5]. Due to the limit of battery capacity, the driving range of most EVs is still quite limited (e.g., 100 miles) [2]. Hence, most EVs must be recharged frequently during service time and their recharge time generally takes more than 30 minutes. However, to fulfill metropolitan transit demands, EVs, especially public service EVs, are expected to be continuously operable without recharging downtime [6]. Driven by this expectation, multiple Wireless Power Transfer (WPT) techniques for in-motion EV charging have been proposed [6]. Specifically, the wireless chargers are deployed in certain road segments to serve as charging lanes. As long as an EV drives through a charging lane, its State-of-Charge (SoC) can be charged dynamically [7], [8]. However, a grave challenge remains unsolved: *how to determine the deployment plan of in-motion wireless chargers (i.e., charger locations and charger lengths) for a metropolitan road network that minimizes the deployment cost while maintaining the continuous operability of EVs on the roads.* By operability, we mean that an EV's SoC is maintained above some level throughout its driving.

Many works have been proposed for the optimal deployment of plug-in charging stations [9]–[15]. Some of them [9]–[11] focus on developing models to infer EVs' charging demands and aim to deploy charging stations to maximally meet the inferred charging demands of the EVs with the minimum deployment cost. However, the inferred demands output by the methods may not reflect the real situation with sufficiently high accuracy. The other works focus on developing traffic flow models based on fine-grained analysis of historical traffic data. The deployment positions of charging stations are determined to maximally cover the traffic flows with the minimum deployment cost under several constraints (e.g., battery capacity, traffic flow) [12]–[15].

However, these works are not directly applicable for the deployment of in-motion wireless chargers due to the different charging characteristics between plug-in chargers and in-motion wireless chargers. Specifically, the charged amount of energy of an in-motion wireless charger is determined by the driving speed that the EVs pass through the charger and the charging lane length. The slower speed an EV passes through the charging lane, the lower deployment cost is required for fully charging the EV, and vice versa. However, such charging characteristics are not considered in the previous works for deploying plug-in charging stations.

It is preferable to deploy in-motion wireless chargers at locations with relatively slower EV passing speed and higher vehicle visit frequency to reduce the deployment cost and increase the service ability of the chargers [16]. Also, the deployment plan of in-motion wireless chargers must take into account multiple sources of vehicle traffic that constitute the public transit services in a metropolitan city, such as taxicabs, buses, and customized transit vehicles (e.g., UberPool) [17]. Although there have been some researches on the optimal deployment of wireless charging lanes on small-scale road networks (mostly no more than 20 road segments) with synthetic traffic [7], [18]–[28], the challenge of deploying wireless

An earlier version of this paper was presented at the INFOCOM 2017 and was published in its Proceedings (DOI: 10.1109/INFOCOM.2017.8057019).

charging lanes in a metropolitan road network with different sources of traffic and many roads has not been studied, which however is much more formidable.

In this paper, we propose *CatCharger* to tackle this challenge. Based on the metropolitan-scale charging demands reflected by multiple sources of EVs, *CatCharger* uses Categorization and clustering of vehicle traffic attributes (i.e., driving speed, visit frequency) to determine the deployment positions of in-motion wireless Chargers. Specifically, we first analyze a metropolitan-scale mobility dataset collected in Shenzhen, China, which includes the status (e.g., timestamp, GPS position, speed) of 14262 buses, 15610 taxicabs, 12386 customized transit vehicles in every 30 seconds from July 1 to July 31, 2015. Since intersections determine traffic volumes of the road segments they connect [29], [30], we represent the road network as a directed graph, and extract intersections (i.e., landmarks) as candidate locations for deploying chargers. The traffic statistics we measured are vehicles' average passing speed at each landmark, which determines the charging lane length required for fully charging an EV [7], and average daily vehicle visit frequency at each landmark, which determines the potential service ability of the charging lane. We observed that the traffic attributes of the landmarks are widely distributed, and the traffic attributes at some landmarks have quite high variance. Moreover, we also observed that vehicle flow rate at a landmark (i.e., number of vehicles driving through the landmark per unit time [31]) is important for maximizing the charging capacity of deployed chargers.

To group the landmarks with similar average passing speed and visit frequency into their respective clusters, we adopt an entropy-based algorithm [32] which could effectively cluster items with similar attributes together. Then we extract candidate landmarks from these clusters. Landmarks in each group are ranked base on their average daily visit frequency and the estimated lane length required for fully charging an EV. To ensure the extracted landmarks have constantly high vehicle visit frequency and slow vehicle passing speed, we also consider the variance of the two traffic attributes in ranking. Consequently, the rank of a landmark represents its suitability for deploying a charging lane. Some groups are not suitable for deployment because of either too high vehicle passing speed or too low vehicle visit frequency. From groups suitable for charging lane deployment, we select landmarks with the highest ranking scores as candidate deployment locations.

To guarantee the continuous operation of EVs in a metropolitan road network, the EVs need to have a certain level of residual energy when arriving at each landmark. Therefore, we aim to first infer the EVs' expected level of residual energy at each landmark given a certain deployment plan of charging lanes, and then find the optimal deployment of charging lanes. Specifically, since the EVs' mobility across the landmarks is not uniformly distributed, we first use the Kernel Density Estimator (KDE) [33] to infer the EVs' probability of reaching each landmark in the road network. Then, we use the inferred probabilities to estimate the expected residual energy of EVs at each landmark. Moreover, we also consider EVs' routing choice behavior given a certain charger deployment plan. Finally, a multi-objective optimization problem is formulated and solved, which aims at minimizing the entire cost of deployment, maximizing the vehicle traffic flow at the landmarks with chargers and ensuring that the residual energy of EVs would remain above a certain level at each landmark. The main difference between our approach and the state-of-the-art wireless chargers deployment methods is that we take into account multiple real-world constraints in formulating the multi-objective optimization problem and the proposed approach is able to obtain the effective wireless charger deployment solution for metropolitan road networks.

In summary, our contributions include:

(1) We comprehensively study a metropolitan-scale, long-term mobility dataset consisting of multiple vehicles for collecting traffic statistics relevant to in-motion wireless charging, which serves as the foundation of *CatCharger*.

(2) We propose *CatCharger*, which utilizes categorization and clustering of landmarks by their traffic flow attributes (i.e., vehicle passing speed, vehicle visit frequency), a KDE based traffic model, and EV drivers' routing choice behavior to determine the landmarks for deploying in-motion wireless chargers and the charger length. It minimizes the total deployment cost, maximizes the vehicle traffic flow at the landmarks with chargers, and meanwhile ensures the continuous driving of EVs at each landmark.

(3) We have conducted extensive trace-driven experiments to show the effectiveness of *CatCharger* in supporting continuous operability of EVs on a metropolitan road network. Compared with previous methods, *CatCharger* increases the ratio of operable EVs at the end of a day by 12.5% under the same charger deployment cost.

To our knowledge, *CatCharger* is the first work to handle the in-motion wireless charger deployment in a metropolitan scenario with various sources of EVs. The remainder of our paper is organized as follows. Section II presents literature review. Section III presents the result of our metropolitan-scale dataset measurement. Section IV gives the details of *CatCharger* design. Section V presents trace-driven evaluations. Section VI gives conclusion with future work plan.

II. Related Work

Plug-in charging station deployment. Many works have been proposed for the optimal deployment of plug-in charging stations. Some of them [9]–[11] focus on inferring EVs' charging demands and deploy charging stations to maximally meet the EVs' charging demands with the minimum deployment cost. For example, Bae *et al.* [9] proposed to utilize the M/M/s queueing theory and the fluid dynamic model for estimating the EVs' charging demand and deploying charging stations. Zheng *et al.* [10] considered the life cycle cost of charging stations, and formulated an optimization problem to maximize the number of charged EVs with the minimum deployment cost. Eisel *et al.* [11] considered EV drivers' charging preference in determining charging station deployment.

Meanwhile, some charging station deployment methods that aim to maximally cover EV traffic flows by taking into account multiple constraints have been proposed [12]–[15]. Lam *et al.* [12] summarized the deployment of charging stations as a vertex cover problem, and proposed several methodologies to solve the problem. Wang *et al.* [13] designed and solved a multi-objective optimization problem that takes into account multiple EV mobility constraints (e.g., driving range, traffic

volume) to maximally cover the charging demand of EVs. Sánchez-Martín *et al.* [14] takes into account the distribution of EVs' parking events and their parking time length to deploy charging stations with the minimum deployment cost and meanwhile offer sufficient charging opportunities for the EVs. Yao *et al.* [15] formulated an optimal charging station deployment method for a 20-node road network. These works are not directly applicable for the deployment of in-motion wireless chargers due to the different charging characteristics between plug-in chargers and in-motion wireless chargers.

Optimal deployment of wireless chargers for EVs. Several works have been proposed for the optimal deployment of wireless charging lanes on road networks. Jang *et al.* [7] formulated an optimization problem, which considers battery capacity and charging lane length as constraints, to deploy wireless charging lanes to maintain the SoC of buses on a single determined route with the minimum cost. He *et al.* [18] proposed two pricing models and formulated a mathematical program to optimize the deployment of wireless charging tolls. Ko *et al.* [19] designed a mathematical optimization model to allocate the in-motion wireless chargers and determine buses' battery size given specific bus driving routes. Riemann *et al.* [20] proposed a mixed-integer nonlinear program model to maximize the captured traffic flow of deployed in-motion wireless chargers through applying the stochastic user equilibrium to describe EVs' route choice. Fuller *et al.* [21] considered various combinations of charging power and EV driving range, and formulated and solved a flow-based set covering problem to determine the number of wireless charging infrastructures required in California. Chen *et al.* [22] developed a user equilibrium model for describing the equilibrium flow distribution across a road network, and formulated a mathematical program to optimize the charging lane deployment. Hwang *et al.* [23] proposed a Particle Swarm Optimization (PSO) method to solve a mathematical model that optimizes the economical allocation of charging lanes, given the battery size and multi-route environment. Chen *et al.* [24] further studied the deployment problem of both stationary and in-motion wireless chargers through considering different scenario requirements. Liu *et al.* [25] proposed a deterministic model and a robust model to address the problem of optimizing the charging lane locations for a real-world bus system that consists of 8 bus lines. Bi *et al.* [26] proposed a novel multi-objective optimization model framework based on life cycle assessment (LCA) to solve the deployment problem of in-motion wireless chargers in a multi-route electric bus system. Manshadi *et al.* [27] proposed a decentralized optimization framework to address the impact of wireless charging on electricity and transportation networks. Li *et al.* [28] designed a bi-objective model considering both traffic delay and charger utilization rate to optimize the deployment of wireless chargers on urban road networks with traffic signals. These works are established on small-scale road networks (mostly no more than 20 road segments) with synthetic traffic and cannot handle the metropolitan-scale deployment of wireless charging lanes.

III. Metropolitan-Scale Dataset Analysis

A. Dataset Description and Data Processing System

Our datasets for Shenzhen record the status (e.g., timestamp, position) of vehicles for one year (Jan 1 – Dec 31, 2015), with a recording period less than 30 seconds, which include:

(1) **Taxicab Dataset.** It is collected by the Shenzhen Transport Committee, which records the status (e.g., timestamp, position, speed) of 15,610 taxicabs. The daily size of the uploaded data is around 2GB.

(2) **Bus Dataset.** It is also collected by the Shenzhen Transport Committee, which records the status of 14,262 buses (e.g., timestamp, GPS position).

(3) **Dada bus Dataset.** It is provided by the Dada Bus corporation (a customized transit service similar to UberPool), which records the status (e.g., timestamp, position, speed) of 12,386 reserved service buses.

(4) **Road Map.** The road map of Shenzhen is obtained from OpenStreetMap [34]. According to the municipal information of Shenzhen [35], we use a bounding box with coordinate $(lat = 22.4450, lon = 113.7130)$ as the south-west corner, and coordinate $(lat = 22.8844, lon = 114.5270)$ as the north-east corner, which covers an area of around 2,926km^2, to crop the road map data.

For efficient management of such large datasets, we utilized a 34 TB Hadoop Distributed File System (HDFS) [36] on a cluster consisting of 10 nodes, each of which is equipped with 16 cores and 64 GB RAM. For data processing, we utilized Apache Spark [37], which is a fast in-memory cluster computing system, alongside the Hadoop cluster.

B. Important Issues

There are two main issues that need to be addressed in handling the charging lane deployment challenge:

(1) **Reducing charging lane length.** On the one hand, the length of a charging lane should be sufficiently long to ensure that the passing EVs can be fully charged. On the other hand, the charging lane length needs to be minimized to reduce the deployment cost. The objective of selecting the optimal locations for charging lane deployment is non-trivial.

(2) **Reducing the number of deployed charging lanes.** The number of deployed charging lanes should be able to support the continuous driving of EVs on roads and meanwhile minimized to save cost. The determination of charging lane locations considering the above objectives is also non-trivial.

1) Vehicle Velocity at Charging Lanes Matters: The amount of energy transmitted to an EV from a wireless charging lane (E) equals: $E = L \cdot r/v$, where L denotes the length of the charging lane, r denotes its energy supply rate, and v denotes the vehicle's speed passing through the charging lane. Since EVs with different battery capacities may pass a charging lane with various speeds, to ensure that any EV can be charged certain amount of energy after it passes a charging lane with a speed slower than a certain value (average vehicle passing speed in this paper), we can manually specify an expected minimum charge amount threshold E_{min} (e.g., the 50%, 80%, or 100% of the EVs' maximum battery capacity). That is, any EV can be charged with at least E_{min} if it passes through the charging lane with a speed slower than the average vehicle passing speed at the charging lane. A larger E_{min} enables the charging lanes to maintain higher SoC levels in application, but requires higher cost (i.e., longer charging lane length) and is limited by technology issues [23], and vice versa. Thus, the value of E_{min} should be adjusted according to city planner's expectations. Therefore, when a landmark i

with average passing speed $\overline{v}_i$ is chosen to deploy a charging lane, its length is determined to meet the above condition:

$$L_i = \frac{E_{min}}{r}\overline{v}_i. \quad (1)$$

Note that $\frac{E_{min}}{r}$ is a constant, so the charging lane length (L_i) is directly determined by vehicle average passing speed ($\overline{v}_i$). Therefore, the positions with the slowest EV passing speed are more suitable for charging lane deployment because they can fully charge passing EVs with shorter charging lanes (i.e., lower deployment cost) [7], [8], [16].

2) Vehicle Visit Frequency and Multi-source Vehicle Traffic Matter: To keep the continuous driving of EVs in the city road network, the deployed charging lanes must offer sufficient charging opportunities to most of the EV traffic. Therefore, the determination of charging lane positions must fully consider the EVs' visit frequency at the landmarks. Moreover, considering that the EV-based public transit system consists of multiple sources of vehicle traffic, we must consider all sources of public service vehicle traffic in determining the deployment of charging lanes to avoid bias. Our datasets can reflect the charging demand of multiple sources of public service EVs.

C. Dataset Analysis

We represent the road network with a directed graph, in which vertices represent landmarks (i.e., intersections) and edges represent road segments [38]. The movement records of a vehicle are continuous. If a vehicle stops at a location for more than 10 minutes, we presume that it has ended its previous trajectory. Thus, the movement records of the vehicle are split into separate trajectories. Considering that vehicles can only change movement direction at intersections, we normalize the original GPS positions to their respective nearest landmarks (in Euclidean distance) as in previous methods [39]. Specifically, we define vehicle trajectory as:

Definition 1. *A vehicle n_i's trajectory is a sequence of time-ordered spatial positions, Tr_i : $\{(\boldsymbol{p_0}, t_0), (\boldsymbol{p_1}, t_1), \ldots, (\boldsymbol{p_m}, t_m)\}$, where each position is represented by a latitude and a longitude $\boldsymbol{p_j} = (lat_j, lon_j)$.*

Through measurement, we found that the range and the average of vehicle visit frequency at a landmark are $[0veh veh/day, 96,637veh/day]$ and $3,840veh/day$, and the range and the average of vehicle passing speed in a landmark are $[0km/h, 142km/h]$ and $20km/h$. Figure 1 shows the Cumulative Distribution Function (CDF) of average vehicle passing speed and average vehicle visit frequency per day of each landmark. Figure 2 plots the density distribution of vehicle passing speed with respect to (w.r.t.) vehicle visit frequency to illustrate the distribution of positions with both slow vehicle passing speed and high vehicle visit frequency. In Figure 1, we see that the landmarks with vehicle visit frequency higher than $10^4veh/day$ only take less than 25% of all the landmarks, and the landmarks with vehicle passing speed less than $60km/h$ take up about 80% of all the landmarks. In Figure 2, we can see the landmarks with both low vehicle passing speed ($60km/h$) and high vehicle visit frequency ($10^4veh/day$) take up a small portion within the red square circle. Additionally, even for the landmarks with high average vehicle visit frequency, their actual vehicle visit frequency may vary a lot. Considering that the charging lane length is determined after deployment, a landmark with a relatively more stable vehicle visit frequency is more suitable for deploying wireless charging lanes since there will be continuous flows of EVs passing through them (e.g., landmarks nearby train station, airport). Therefore, we also measured the variance of the vehicle visit frequency of the landmarks in the square circle of Figure 2. The measurement results are illustrated in Figure 3. We can see that although the standard deviation of vehicle visit frequency at around 80% of the landmarks is lower than 1,000, the standard deviation of vehicle visit frequency at the other 20% landmarks can be as high as 10,000 in the worst case. Even for some landmarks with extremely high average vehicle visit frequency, their actual vehicle visit frequency can vary significantly. This means that the variance (standard deviation) of vehicle visit frequency of the landmarks needs to be considered in measuring the suitability of deploying wireless charging lanes.

In addition, we also measured the variance of vehicles' passing speed at the landmarks. The results are illustrated in Figure 4. We can see that the variances of vehicles' passing speed differ a lot in different regions. More than 40% of the positions have a variance of vehicle passing speed higher than $20km/h$, and the variance can be as high as $50km/h$. It means that if we solely determine the charging lane length by vehicles' average passing speed at these positions, the deployed charging lane may not be able to fully charge most vehicles passing through the positions. However, simply deploying charging lanes with the maximum possible length to ensure all the vehicles can be fully charged is unrealistic due to high deployment cost. Therefore, in addition to average vehicle passing speed, we need to also consider the variance of vehicle passing speed at the potential charging positions. The above observations motivate us to find an innovative method to properly extract candidate charging lane placement positions considering the diversity in vehicle passing speed and visit frequency, and their distribution in different regions. The details will be elaborated in Section IV-C.

Average vehicle flow rate of a landmark is defined as the average number of vehicles driving through the landmark per unit time [31]. From the definition of average vehicle flow rate of a landmark, it equals to the product of average vehicle density and average vehicle passing speed on the landmark. A landmark with a high vehicle flow rate means that there are many vehicles that pass through the landmark per unit time, and the vehicles can pass the landmark with a relatively high velocity (i.e., no traffic congestion). Thus, it is usually a good indicator on how well a charging facility can serve EVs. Therefore, in addition to vehicle visit frequency, we also measured the average vehicle flow rate of all the landmarks. We consider the landmarks in the red square in Figure 2 as potential candidate landmarks (i.e., landmarks with vehicle visit frequency higher than $10^4veh/day$ and vehicle passing velocity lower than $60km/h$). We also measured the average vehicle flow rate of the candidate landmarks. The CDF of the measurement results are illustrated in Figure 5. The black curve is the measurement results of all the landmarks. The red dashed curve is the measurement results of candidate landmarks suitable for deploying in-motion wireless chargers. We can see that the black curve is about to reach 1 at a vehicle flow rate of $1000/h$ (i.e. almost all landmarks have a vehicle flow rate that is less than $1000/h$). About 80% of all the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3121756, IEEE Internet of Things Journal

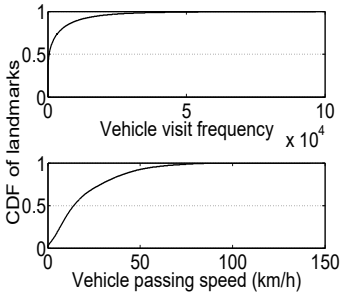


Fig. 1: Average vehicle passing speed & daily vehicle visit frequency.

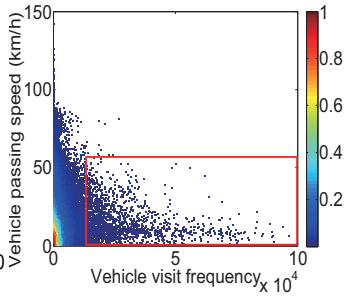


Fig. 2: Density scatter of vehicle passing speed w.r.t. vehicle visit frequency.

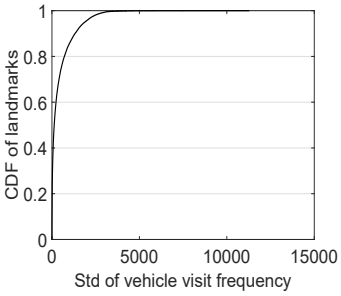


Fig. 3: Variance of vehicle visit frequency of landmarks.

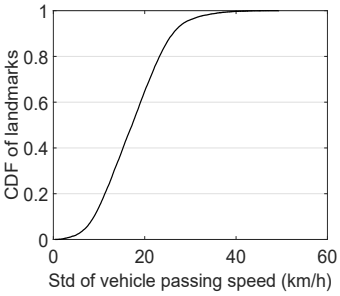


Fig. 4: Variance of vehicle passing speed of landmarks.

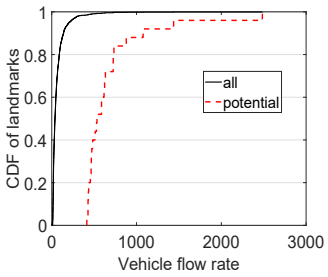


Fig. 5: Distribution of vehicle traffic flow rate of landmarks.

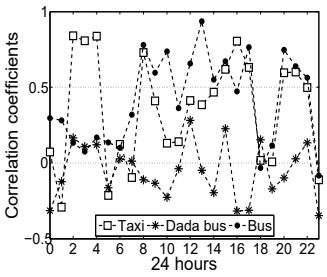


Fig. 6: Correlation between each vehicle source and public transit traffic.

landmarks have a vehicle flow rate less than $125/h$. About 60% have a vehicle flow rate less than $62/h$. The trend of the result is similar to that in Figure 1.

From the measurement results of candidate landmarks (i.e., red dashed curve in Figure 5), we can see that its general trend is similar to that of the black curve but shifts to the right significantly. This means that the potential candidate landmarks typically have much higher average vehicle flow rate than other landmarks. Specifically, the CDF is about to reach 1 at a vehicle flow rate of $2500/h$. About 80% of all candidate landmarks have a vehicle flow rate less than $750/h$. About 60% of all candidate landmarks have a vehicle flow rate less than $625/h$. We can see that although all the candidate landmarks have a much higher vehicle flow rate than other landmarks, their vehicle flow rates still vary a lot. We must consider deploying in-motion wireless chargers to the positions with the highest vehicle flow rate to ensure the total charging capability of the deployed chargers. In Section IV-D3, we will explicitly explain how we consider the average vehicle flow rates of candidate landmarks in determining the locations to deploy wireless charging lanes.

To illustrate that it is necessary to consider multiple sources of public service vehicle traffic characteristics in determining charging lane deployment, we further analyze the mobility characteristics of each single vehicle source to reflect the possible bias against the actual total public service vehicle traffic. As long as a vehicle changes its position in its trajectory, we define the change of position as an activity. We count the number of activities of each source of vehicles during each hour throughout a day for one month, and calculate the daily average number of activities during each hour over all the days. Then, we use the Pearson correlation coefficient [40] to measure the correlation between each source of vehicles and the total traffic activity of public service vehicles (i.e., combined traffic of bus, taxi and Dada bus). Figure 6 shows the result. We can see that during the hours before dawn (i.e., 00:00∼06:00), the correlation between the activity of taxis and the public service vehicle activities is higher than those of bus and Dada bus. This means that during these hours, the taxis are the primary public service vehicles driving on the road. This is because that most of the other public service vehicles are out of service during the time period. Starting from 07:00, the activity of buses is more correlated with the public service vehicle traffic than the others. This is because that the buses are back to service starting from this time. Since Dada buses are driven by pre-determined pick-up requests which are not as regular as other public service vehicles. Therefore, the correlation between Dada buses and the public service vehicle traffic is relatively lower during all hours. The exceptions are at around 12:00, 13:00 and 18:00, where the activities of Dada buses are similarly correlated with the public service vehicle traffic compared with the others. It means that during these hours, Data buses serve as a supplement to the scheduled public service demands like buses. Therefore, it is necessary to consider multiple sources of public service vehicle traffic characteristics in determining charging lane deployment.

Considering that the length of a vehicle trajectory determines the vehicle's energy consumption [33], we measured the lengths of all the trajectories, of which distribution is shown in Figure 7. We can see that most of the trajectories are shorter than 10,000 meters, and the distribution of the trajectory lengths does not follow a parametric distribution. To describe this distribution and use it for the inference of the vehicles' probability of reaching each landmark in the road network, we use the KDE model to fit the Probability Density Function (PDF) of the non-parametric distribution. Figure 7 shows the fitting result generated by the KDE model, which is a function of the trajectory length. Given the length of a trajectory, we first estimate an EV's residual energy upon the completion of the trajectory. Then, we use the obtained PDF of the trajectories to infer the expected residual energy of EVs at each landmark on the road network given a certain charging lane deployment. Finally, by formulating and solving an optimization problem that aims to minimize the charging lane deployment cost while maintaining the continuous driving of EVs, we can determine the final charging lane deployment plan. In Section IV-D1 and IV-D3, we will elaborate the details.

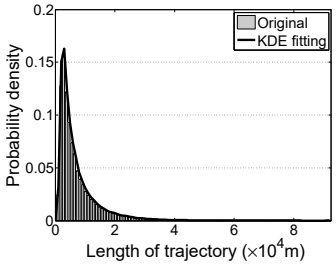


Fig. 7: Distribution of trajectory lengths & estimation of KDE.

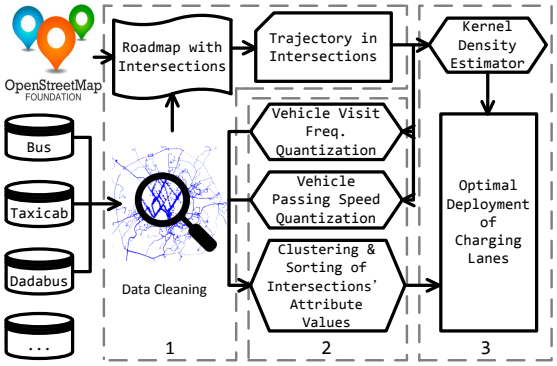


Fig. 8: Framework of CatCharger.

IV. System Design of CatCharger

A. Framework of CatCharger

CatCharger consists of the following three stages as shown in the three dashed boxes in Figure 8:

1. **Vehicle mobility normalization** (Section IV-B). First, we apply *Data Cleaning* to filter out erroneous records with duplicated GPS positions and timestamps, and positions out of the bounding box, etc. Then, we use *OpenStreetMap* to extract the intersections and generate the *Roadmap with Intersections*. Finally, we normalize the positions to their respective nearest landmarks (in Euclidean distance) as in previous methods [39] to obtain the *Trajectory in Intersections*.
2. **Charging lane location candidate extraction** (Section IV-C). Based on the data output from the first stage, we apply *Vehicle Visit Frequency Quantization* and *Vehicle Passing Speed Quantization* to obtain the traffic attribute values of each intersection. Then, we apply *Clustering & Sorting of the Intersections' Attribute Values* to extract the intersections with both high vehicle visit frequency and short required charging lane length.
3. **Charging lane location determination** (Section IV-D). We first feed the lengths of the trajectories to the *Kernel Density Estimator* (KDE) to infer the vehicles' probability of reaching each landmark in the road network. Then we formulate and solve an optimization problem to determine the charging lane positions and lane lengths for the *Optimal Deployment of Charging Lanes*.

B. Vehicle Mobility Normalization

Original vehicle movement data contain erroneous records with duplicated GPS positions and timestamps, and positions out of the bounding box, etc., a data cleaning step is necessary. Specifically, considering that vehicles can only change movement direction at intersections, we normalize each position record to its nearest landmarks in Euclidean distance, as in previous methods [39].

C. Charging Lane Location Candidate Extraction

According to previous studies [41], [42] and our data analysis results, vehicle visit frequency of the landmarks varies a lot in different regions. For example, compared to the landmarks distributed in industrial areas, the vehicle visit frequency of the landmarks distributed in downtown areas is much higher. When deploying the charging lanes, in addition to covering the landmarks frequently visited by vehicles, we must also make sure that the EVs driving in all the areas can obtain sufficient charging opportunities and maintain their continuous operability. To this end, we propose to extract landmarks with the most outstanding traffic attribute values from each area as candidate landmarks for deploying charging lanes. Specifically, we first cluster the landmarks with similar attribute values (i.e., average vehicle passing speed, daily vehicle visit frequency) to the same group. This is because that the landmarks that have similar traffic attribute values generally have similar suitability in offering charging opportunities for EVs (i.e., charging lane deployment), and are located in diverse areas. Then, we rank the groups by their attribute values and choose the groups with the most outstanding attribute values for deploying charging lanes. Finally, to ensure that there are candidate landmarks in each area, from each group, we further select landmarks with the most outstanding attribute values from each area. Thus, the selected landmarks are the candidate landmarks, which are relatively more suitable for deploying charging lanes and located in diverse areas.

1) Categorization of Original Mobility Data: As explained in Section III-C, vehicle passing speed and vehicle visit frequency are two important traffic attribute values that determine the suitability of landmarks in deploying charging lanes. However, it is non-trivial to determine the similarity between the traffic attribute values of different landmarks. For example, given three landmarks: $A(50km/h, 10000veh/day)$, $B(50km/h, 9000veh/day)$ and $C(100km/h, 10000veh/day)$. If we simply use a distance metric to measure their similarity (e.g., Euclidean distance), compared with B, C is more similar to A because the distance between A and C is 50, while the distance between A and B is 1000. However, A is far different with C regarding the suitability of placing the charging lanes because the low average passing speed of A can save much more charging lane length than that of C, while A is more similar to B because they require the same charging lane length and both have high vehicle visit frequency.

To solve this problem, we use different categorization intervals to represent the original attribute values of speed (v) and visit frequency (f) into their respective attribute ranges, which are marked with *attribute IDs*. Specifically, we use $5km/h$ as the speed categorization interval based on previous deployment setting of charging lane length interval [7], and use 1,000 as the visit frequency categorization interval considering the total number of vehicles (>50,000 vehicles). Thus, the original attribute values are represented with attribute IDs as follows:

$$\begin{aligned} v :&\{0, 0 \sim 5km/h\}, \{1, 5 \sim 10km/h\}, \ldots, \\ f :&\{0, 0 \sim 1000veh/day\}, \{1, 1000 \sim 2000veh/day\}, \ldots. \end{aligned} \tag{2}$$

where each attribute range is represented as <attribute ID, description>. For example, suppose that the average vehicle passing speed and the daily average vehicle visit frequency of a landmark are $3km/h$ and $1500veh/day$, respectively, the attribute IDs of the landmark will be represented as $\{0, 1\}$.

2) Clustering of Landmarks: Based on the categorized attribute IDs, we propose to cluster the landmarks into different groups by the similarity of their attribute IDs. We use entropy, which measures categorical disorder (i.e., dissimilarity of attribute IDs within a group) [32] for clustering. Let's take the attribute of vehicle passing speed as an example. Suppose

a group has two landmarks with attribute IDs $\{0\}$ and $\{1\}$, respectively. F is a discrete random variable representing an attribute (e.g., average vehicle passing speed), $A(F)$ is the set of the attribute IDs of F in a group (e.g., 0, 1), and $p(f)$ is the probability function of F, namely the ratio of the attribute ID in the group (e.g., 0.5). The entropy of the attribute $H(F)$ within the group is defined as:

$$H(F) = -\sum_{f \in A(F)} p(f)\log_2(p(f)), \tag{3}$$

where $-\log_2(p(f))$ measures the dissimilarity of the attribute in the group. The entropy of the two landmarks in the example is $\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$. Higher dissimilarity between two landmarks' attribute IDs leads to a larger entropy. Since each landmark has two attributes, the entropy of a cluster C_i can be calculated as the sum of the entropies of the two attributes:

$$H(C_i) = H_i(F_0) + H_i(F_1). \tag{4}$$

Suppose all candidate landmarks LM are clustered into k clusters $\mathcal{C} = \{C_0, \ldots, C_{k-1}\}$. To measure the quality of the clustering, we use the weighted sum of the entropies of all clusters as the expected entropy resulted from the clustering. The weight for each cluster is calculated as $\frac{|C_i|}{|LM|}$, where $|\cdot|$ means the number of landmarks in the set. Thus, the expected entropy is calculated by:

$$\overline{H}(\mathcal{C}) = \sum_{k=0}^{k-1} \frac{|C_i|}{|LM|} H(C_i). \tag{5}$$

Given a set of landmarks for clustering, we first find all possible clustering arrangements, and then choose the one with the minimum expected entropy. As a result, the optimal clustering strategy renders clusters whose member landmarks have the least dissimilar attribute IDs between each other. Unfortunately, such a clustering strategy is difficult to execute because it is NP-complete [43]. Then, *CatCharger* instead follows a heuristic method introduced in [32] to approximate the best solution. The steps of the landmark clustering are as follows:

(i) *Initialization:* To cluster landmarks into k groups, we must start with k most dissimilar landmarks. But directly extracting such k landmarks from the entire set of landmarks is non-trivial. To handle this problem, we take a sample S from the set of landmarks LM ($|S| \ll |LM|$). In S, we enumeratively calculate the entropy generated by each pair of landmarks, and place the two landmarks that generate the maximum entropy in two clusters (C_0, C_1) as the two starting clusters. Then, the remaining $k-2$ starting landmarks will be incrementally found as the ones that are most dissimilar with the already determined ones.

(ii) *Incremental clustering:* After the initialization, the remaining $|LM| - k$ landmarks will be clustered to the respective starting landmark that renders the minimum total expected entropy (Equation (5)) one by one.

The major problems with such heuristic clustering include: i) how to select the sample S, ii) how to determine the number of clusters k, and iii) incrementally clustering the landmarks may deteriorate the clustering quality. For i), we randomly select $\gamma\%$ (e.g., 10%) of the landmarks from every functional region of Shenzhen, and combine them as the sample because each region needs several charging positions to support the EV traffic. For ii), within the sample, we follow the algorithm developed in [44] to find the most suitable k that results in the maximum difference in entropy changing rate, of which complexity is $O(|S|^2)$. As for iii), we repeat the clustering steps (in which landmarks are randomly picked) for several times and choose the result with the minimum entropy.

3) Extracting Top Ranked Landmarks from Clusters: Note the required length of charging lane i (L_i) can be calculated by Equation (1) based on the average passing speed of a landmark. Since the shorter charging lane a landmark requires, and the higher vehicle visit frequency the landmark has, the more suitable it is for placing a charging lane. In Section III-C, we also verified that the variance (standard deviation) of the vehicle visit frequency and the variance (standard deviation) of the vehicle passing speed of the landmark are two important factors that will influence how stable the landmark can provide charging service to EVs once it is equipped with a wireless charging lane. Generally, the less variance of vehicle visit frequency and the less variance of vehicle passing speed a landmark has, the more stable a wireless charging lane can serve many EVs and fully charge the EVs at this landmark. Therefore, we need to extract landmarks that have short charging lane length, high vehicle visit frequency, and small variance of vehicle visit frequency and vehicle passing speed. Then, we define the rank of a landmark $\boldsymbol{lm_i} \in C_j$ as:

$$R(\boldsymbol{lm_i}) = \frac{\log(\frac{\bar{f}_i}{\sigma_i^f})}{L_i \sigma_i^v}, \tag{6}$$

where $\bar{f}_i$ is the average vehicle visit frequency at $\boldsymbol{lm_i}$, σ_i^f is its standard deviation, and σ_i^v is the standard deviation of vehicle passing speed at $\boldsymbol{lm_i}$. Thus, the larger $\bar{f}_i$ and the smaller L_i that $\boldsymbol{lm_i}$ has, and meanwhile the smaller σ_i^v and σ_i^f the landmark has, the higher rank it will have. We use logarithmic value of $\bar{f}_i$ because $\bar{f}_i$ is generally much larger than L_i. To ensure the suitability of selected landmarks, we need to remove landmarks with low ranks. For this purpose, we calculate the average rank of each group, and then remove groups with ranks lower than a threshold. Next, we order the landmarks in each group in decreasing order of the rank. In one group, if there are several landmarks in one region, we remove the low-rank landmarks. Finally, we select the top ranked $\varepsilon\%$ (e.g., 10%) of the landmarks from each group, and use them as the candidates for charging lane deployment, which are denoted as $\widetilde{LM} = \{\boldsymbol{lm_0}, \boldsymbol{lm_1}, \ldots, \boldsymbol{lm}_{|\widetilde{LM}|}\}$.

D. Charging Lane Location Determination

To determine the deployment plan on the selected candidate locations, we first use the KDE, which is fed with vehicle mobility, to infer the EVs' expected residual energy at each landmark given that certain landmarks are installed with charging lanes. Then, we formulate an optimization problem that aims to minimize the total cost of deployment while ensuring that the EVs can have a certain level of expected residual energy when they arrive at each landmark. This residual energy level enables an EV to move to its nearest charging lane.

1) Inferring Expected Residual Energy: KDE can be used to describe the vehicles' probability of reaching a landmark on the road network given a source landmark. Also, the residual energy of a vehicle is a function of the distance from the vehicle's source landmark to the destination landmark. Then, the expected residual energy of a vehicle at a landmark in the road network can be calculated. We present the details below.

Since vehicles' mobility patterns imply their traffic at certain locations [33], we feed the vehicles' trajectories into a KDE model to infer the PDF of the distribution of the trajectory lengths as in Equation (7), namely the trip lengths that need to be supported.

$$\hat{f}_h(d) = \frac{1}{mh}\sum_{k=0}^{m-1} K(\frac{d-d_k}{h});\ -\infty < d < \infty, \tag{7}$$

where m is the number of sample trajectories, d_k is the length of the k^{th} trajectory, and h is the smoothing parameter influencing the estimation accuracy of the KDE and is determined according to the MISE criterion [45]. $K(\cdot)$ is the kernel function whose value decays with the increasing of d. It is set to the Gaussian function as in Equation (8) based on [46], [47].

$$K(\frac{d-d_k}{h}) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(d-d_k)^2}{2h^2}\right]. \tag{8}$$

According to the state-of-the-art EV energy consumption model [48], the energy consumption of a taxicab (E_c) is primarily determined by air drag (E_{air}) and rolling resistance (E_{roll}). Therefore, the consumption rate is:

$$\begin{aligned}\Delta E_c &= \Delta E_{air} + \Delta E_{roll} \\ &= c_a v^2 \Delta l + c_e \kappa g \Delta l\end{aligned} \tag{9}$$

where c_a is the air drag coefficient determined by vehicle front surface area; v is the driving speed; Δl is the distance that the taxicab has moved; c_e is the rolling resistance coefficient; κ is the taxicab's mass; and g is the gravity acceleration.

According to Equation (1), any EV can be at least charged to the expected charge amount threshold E_{min} if it drives through a charging lane with a speed slower than the landmark's average vehicle passing speed. Given an EV starting from a charger, based on Equation (9), its residual energy (i.e., SoC) at a location, which is d distance away from the charger through the shortest route, can be estimated as [48]:

$$E_r^d = E_{min} - \sum_{n=0}^{N^R-1}(c_a v_n^2 + c_e \kappa g) l_n, \tag{10}$$

where N^R is the number of road segments of the shortest route, and v_n and l_n are the speed limit and length of the n^{th} road segment, respectively. Then, the EVs' SoC at the location can be represented as:

$$SoC(d) = \begin{cases} E_r^d/E_0, & \text{if } E_r^d \geqslant 0 \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Thus, given a binary integer x_i to denote whether a candidate landmark $\boldsymbol{lm_i} \in \widetilde{LM}$ is installed with a charging lane or not, the expected SoC of EVs at a landmark $\boldsymbol{lm_j} \in LM$ in the road network is:

$$\overline{SOC}_j = \sum_{i=0}^{|\widetilde{LM}|-1} \hat{f}(d_{i,j}) SOC(d_{i,j}) x_i, \tag{12}$$

where $d_{i,j}$ is the shortest route distance from $\boldsymbol{lm_i}$ to $\boldsymbol{lm_j}$.

2) Describing Drivers' Routing Choice Behavior: It has been confirmed that EV drivers are more likely to choose routes with charging facilities to mitigate range anxiety [49]. Therefore, there exists a mutual interaction between the location of charging facilities and the resultant network traffic flow. It is necessary to consider EV drivers' routing choice behavior in determining the charger deployment. According to previous studies on EV drivers' routing choice behavior [20], [22], [27], the travel time cost of the route to the drivers' destination and the benefit brought by charging facilities are two main factors that affect their decision of choosing whether to charge on the way or drive directly to the destination. Specifically, an EV driver's probability of choosing a candidate route can be described with a multinomial logit model [20], [22], [27]:

$$P_u^w = \frac{\exp(\alpha \frac{1}{T_u^w} + \beta y_u^w)}{\sum_{k \in U^w} \exp(\alpha \frac{1}{T_k^w} + \beta y_k^w)}, \forall u \in U^w, w \in W, \tag{13}$$

where P_u^w is the probability of choosing the route u among all the candidate routes between the origin-destination (O-D) pair w; T_u^w is the travel time cost of the route u between the O-D pair w; y_u^w is the binary variable indicating the availability of charging lanes on the route u, $y_u^w = 1$ if there is at least one charging lane in u (i.e., $\sum_{\boldsymbol{lm_i} \in u} x_i \geqslant 1$), $y_u^w = 0$ otherwise; U^w is the set of all feasible routes of the O-D pair w reflected in all the historical trajectory data; W is the set of all possible O-D pairs on the road network; α and β are the scaling parameters which describe the routing decision sensitivity in terms of travel time cost and the availability of charging lanes, respectively. In practice, α and β should be calibrated by using survey data. In this study, we follow the settings of these parameters as recommended in [20]: $\alpha = 0.1$ and $\beta = 0.8$. According to Equation (13), the longer travel time cost a route has, the lower probability an EV driver will choose the route, and vice versa. This is consistent with driver's expectation of minimizing the travel time cost.

Next, we elaborate the calculation of the travel time cost of a candidate route, T_u^w. According to [24], T_u^w consists of the driving time of normal road segments (t_u^d), the driving time of charging lanes (t_u^c) and the waiting time at the charging lanes (t_u^w) if there are charging lanes on the route:

$$T_u^w = t_u^d + y_u^w (t_u^c + t_u^w). \tag{14}$$

The driving time of normal road segments included in the route u can be calculated as:

$$t_u^d = \sum_{n=0}^{N_u^R-1} \frac{l_n}{v_n}, \tag{15}$$

where N_u^R is the number of road segments of the route u, v_n and l_n are the speed limit and length of the n^{th} road segment, respectively.

The driving time of charging lanes consists of the EVs' waiting time before charging and charging time. Let λ_i denote the arrival rate of EVs at the charging lane located at landmark $\boldsymbol{lm_i}$ (i.e., the number of EVs arriving at $\boldsymbol{lm_i}$ for charging per unit time), which is actually the vehicle flow rate of $\boldsymbol{lm_i}$. Let μ_i denote the service rate of the charging lane located at landmark $\boldsymbol{lm_i}$ (i.e., the number of EVs that the charging lane can charge per unit time), which is calculated as $\mu_i = \overline{v}_i / L_i$ [7], [23], where $\overline{v}_i$ is the average vehicle passing speed at $\boldsymbol{lm_i}$ and L_i is the planned charging lane length determined by Equation (1). Thus, an EV's charging time at the charging lane is:

$$t_u^c = 1/\mu_i. \tag{16}$$

The utilization ratio of the charging lane is $\rho_i = \lambda_i / \mu_i$. According to the M/M/1 queuing theory [50], the EVs' waiting time at the charging lane is:

$$t_u^w = \begin{cases} \frac{\rho_i/\mu_i}{1-\rho_i}, & \text{if } \rho_i < 1 \\ \rho_i, & \text{otherwise.} \end{cases} \quad (17)$$

3) Formulating Optimization Problem: Our objective is to minimize the total deployment cost through properly selecting landmarks from $\widetilde{LM}$ to install charging lanes while ensuring that at each landmark, the expected residual energy of an EV is higher than a threshold η (e.g., 20%). Note that the threshold is determined to ensure that the EVs have sufficient SoC to drive from charging lane to charging lane, which can be adjusted based on the specific charging coverage requirements of the city. Meanwhile, we must ensure that the total charging capacity of the deployed chargers is able to support the maximum power consumption rate of all the EVs. According to Equation (9), we can derive the battery consumption rate for each EV as $\phi = \frac{\Delta E_c}{\Delta t} = c_a v^3 + c_e \kappa g v$. Hence, the battery consumption rate depends on the speed limit of every road segment. That is, as the speed limit v increases, the battery consumption rate increases. To derive the maximum battery consumption rate ϕ_{max}, we use the maximum speed limit v_{max} of the entire road map. In Section III-C, we have identified that vehicle traffic flow rate varies a lot at different landmarks. Therefore, in addition to the above objective of minimizing the charger deployment cost, we also aim to maximize the average vehicle traffic flow rate covered by the deployed wireless charging lanes. Recall that the deployed charging facility may affect the distribution of EV traffic [20], [22], [27], [49], we propose to consider the EV drivers' routing choice behavior in the calculation of vehicle traffic flow covered by the charger deployment. Finally, the optimization problem can be formulated as below:

$$\text{minimize} \quad \sum_{\boldsymbol{lm_i} \in \widetilde{LM}} \omega_0 x_i L_i, \quad (18)$$

$$\text{maximize} \quad \sum_{\boldsymbol{lm_i} \in \widetilde{LM}} x_i \sum_{w \in W} \sum_{u \in U_w^i} \bar{f}_u^i P_u^w, \quad (19)$$

$$\text{subject to} \quad \overline{SOC}_j \geqslant \eta, \forall\, \boldsymbol{lm_j} \in LM, \quad (20)$$

$$x_i L_i \leqslant L_i^{max}, \forall\, \boldsymbol{lm_i} \in \widetilde{LM}, \quad (21)$$

$$\mathcal{C} \sum_{\boldsymbol{lm_j} \in \widetilde{LM}} x_i \geqslant \phi_{max} N_v, \quad (22)$$

$$x_i \in \{0,1\}, \forall\, \boldsymbol{lm_i} \in \widetilde{LM} \quad (23)$$

where ω_0 is a constant representing the cost of deploying a unit length of charging lane, $\mathcal{C}$ is the charging rate of one charger, $\bar{f}_u^i$ is the average vehicle flow rate (i.e., average vehicle visit frequency) at $\boldsymbol{lm_i}$, which is caused by the vehicles that drive through route u, recall that U_w^i is the set of historical routes between O-D pair w that pass through landmark $\boldsymbol{lm_i}$, W is the set of all possible O-D pairs on the road network, and N_v is the total number of EVs driving on the road network. This problem tries to minimize the total deployment cost of the in-motion wireless chargers (Equation (18)), and maximize the average vehicle traffic flow rate covered by the deployed in-motion wireless chargers considering the routing choice behavior of drivers (Equation (19)) with three constraints: i) the expected residual energy of an EV is no less than a threshold η (Equation (20)), ii) each individual charging lane cannot exceed the maximum road segment length (denoted by L_i^{max}), which is consistent with the reality (Equation (21)) [23], [24] and iii) the total charging capacity of the deployed chargers is able to support the maximum battery consumption rate of all the EVs (Equation (22)).

Note the reason we filter candidate landmarks by their attribute values of vehicles' passing speed and visit frequency is that *CatCharger* does not consider the landmarks that require a too long charging lane to fully charge an EV or have low vehicle visit frequency. Therefore, the binary integers for the non-candidate landmarks are 0, namely $x_i = 0, \forall\, \boldsymbol{lm_i} \in LM \setminus \widetilde{LM}$. Given source landmark $\boldsymbol{lm_i}$ and destination landmark $\boldsymbol{lm_j}$, the coefficient $\hat{f}(d_{i,j})SoC(d_{i,j})$ in Equation (12) is determined. Therefore, we can use a constant θ_{ij} to represent $\hat{f}(d_{i,j})SoC(d_{i,j})$. As a result, the final multi-objective optimization problem is actually a classic Multi-objective Integer Programming (MIP) problem. Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) is effective for solving MIP problem with a lower computational complexity than traditional multi-objective genetic local search algorithm [51]. Therefore, we employ MOEA/D to solve our formulated optimization problem. Generally, MOEA/D first uses the Tchebycheff approach [51] to decompose the MIP problem into several optimization sub-problems. Then, MOEA/D solves the sub-problems simultaneously and maintains the population of the best solutions to each sub-problem during the evolution of solutions until the stop criteria has been reached.

4) Objective Transformation and Normalization: The first objective of the proposed model is to minimize the overall deployment cost of the chargers, while the second objective is to maximize the average vehicle traffic flow captured by the deployed chargers. Therefore, it is necessary to make the following transformation:

$$\begin{aligned} F(\boldsymbol{x}) &= [\min f_1(\boldsymbol{x}), \max f_2(\boldsymbol{x})]^T \\ &= \min[f_1(\boldsymbol{x}), -f_2(\boldsymbol{x})]^T \end{aligned} \quad (24)$$

where $\boldsymbol{x} = \{x_i | \boldsymbol{lm_i} \in \widetilde{LM}\}$ is the vector of binary decision variables of all the candidate landmarks. $f_1(\boldsymbol{x}) = \sum_{\boldsymbol{lm_i} \in \widetilde{LM}} \omega_0 x_i L_i$ and $f_2(\boldsymbol{x}) = \sum_{\boldsymbol{lm_i} \in \widetilde{LM}} x_i \sum_{w \in W} \sum_{u \in U_w^i} \bar{f}_u^i P_u^w$ are the two objective functions: Equation (18) and Equation (19), respectively. To make f_1 and f_2 comparable within the same scale, we normalize them as $\overline{f_i} = \frac{f_i - f_i^{\min}}{f_i^{\max} - f_i^{\min}}, \forall i = 1, 2$ where $f_i^{\max}$ and $f_i^{\min}$ are the maximum and minimum values of f_i, respectively. These values are obtained by solving the optimization problem with each single objective function as the optimization goal.

V. Performance Evaluation

A. Experiment Settings

The experiments are driven by the datasets introduced in Section III-A. Based on the datasets, we used Apache Spark 1.5.2 [37] to develop a trace-driven simulator and simulate the movement and energy consumption of the EVs. We compared the performance of *CatCharger* with two representative methods: random placement (denoted by *Random*) as the baseline, and a traditional charging station deployment method that aims to maximally cover traffic flows (denoted by *MaxFlow*) [15]. In addition, we also evaluate the performance of a variance of *CatCharger* (denoted by *CatCharger+*), which considers the variances of vehicle visit frequency and vehicle passing speed

TABLE I: Table of experiment parameters.

Parameters	Setting	Source
Charging rate $\mathcal{C}$	150 kW	Chen *et al.* [24], [52]
Charger unit price ω_0	\$500/m	Chen *et al.* [24]
Air drag coefficient c_a	0.3	Kurczveil *et al.* [48]
Rolling resistance coefficient c_e	0.01	Kurczveil *et al.* [48]
Mass of a taxicab κ	2,020 kg	Tian *et al.* [53]
Gravity acceleration g	9.8 m/s^2	Tian *et al.* [53]
Battery capacity of an EV E_0	5kWh – 10kWh	Tian *et al.* [54]
Ratio for selecting landmarks from regions γ	10%	Author's assumption
Ratio for selecting top ranked landmarks ε	10%	Author's assumption
Residual energy (SoC) threshold η	20%	Author's assumption
Expected minimum charge amount threshold E_{min}	80%	Author's assumption
Scaling parameters of drivers' choice behavior α and β	$\alpha = 0.1$ and $\beta = 0.8$	Riemann *et al.* [20]

in extracting the candidate landmarks, and the maximization of vehicle traffic flows in determining the landmarks for deploying in-motion wireless chargers.

In simulation, the battery capacities of the EVs follow a uniform distribution ranging from 5kWh to 10kWh [24]. We suppose every vehicle starts driving with full energy in battery at the beginning of a day. The energy supply rate of a charging lane is 150kW [7], [24], [52]. The unit price of a charging lane is $\$500/m$ [7], [24]. In *CatCharger*, the length of a charging lane is calculated by Equation (1). According to [23], [24], the length of a charging lane cannot exceed the maximum road segment length at its scheduled deployment landmark, which ranges from 100.4m to 926.7m based on the map information extracted from OpenStreetMap [34]. Since *Random* and *MaxFlow* do not have methods to determine the charging lane length, we suppose they deploy a maximally 500m-long charging lane (maximum length in *CatCharger*) at each charging landmark, which can charge 50% SoC for the EVs with a battery capacity smaller than 10kWh and a passing speed slower than 15km/h. For fair comparison, the deployment cost in *Random* and *MaxFlow* is the same as *CatCharger*. In *Random*, the locations for placing charging lanes are chosen randomly from the collection of landmarks. *MaxFlow* is for charging station deployment and we use it for charging lane deployment. We choose the landmark that covers the most traffic sequentially until the deployment cost is reached. *MaxFlow* is a traffic flow based method. Since traffic flow based methods can more accurately estimate the charging demands than the charging demand based methods [12], we do not include a charging demand based method for comparison. In landmark categorization (Section IV-C1), the speed interval and the frequency interval are $5km/h$ and 1,000, respectively. In clustering initialization (Section IV-C2), the ratio for selecting landmarks from every administrative region of Shenzhen, γ, is 10%. In candidate position extraction (Section IV-C3), the ratio of the top ranked landmarks, ε, is 10%. The threshold of expected residual energy, η, is set to 20%. The expected minimum charge amount threshold E_{min} is set to 80% of the EVs' maximum battery capacity. As for the scaling parameters of drivers' choice behavior (i.e., α and β), we follow the settings as recommended in [20]: $\alpha = 0.1$ and $\beta = 0.8$. The parameters are listed in Table I.

We use the movement records mentioned in Section III-A for performance evaluation. Below, Figure 9 to Figure 12 demonstrate the metrics of the vehicles under different hours on July 15, 2015. Figure 13 to Figure 15 demonstrate the metrics of vehicles in multiple days: January 12 (Monday), March 10 (Tuesday), May 13 (Wednesday), July 16 (Thursday), September 18 (Friday), November 21 (Saturday) and December 13 (Sunday) in 2015. These days are representative because they are unrelated to each other, belong to 4 different seasons, and cover weekdays and weekends [39]. Specifically, we measured the following metrics:

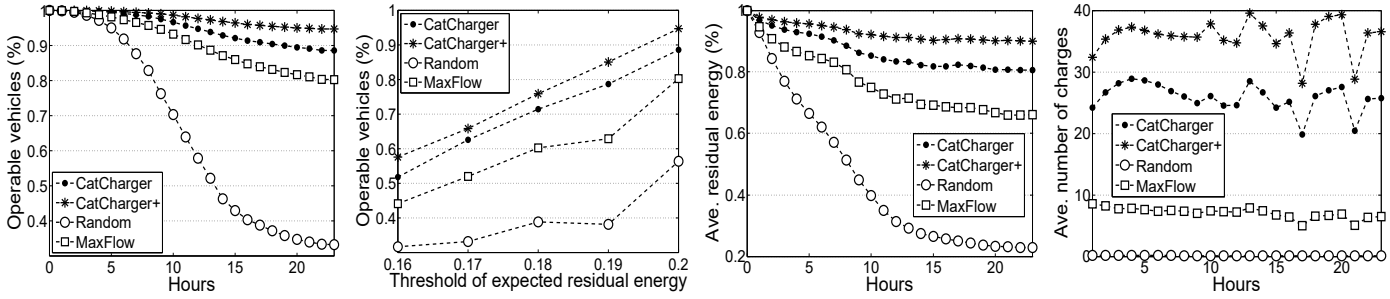
- *Average ratio of operable vehicles*. The average ratio of vehicles that have SoC above 0%. We measure this ratio under different deployment costs to compare the ability of supporting EVs' operability and cost efficiency of different methods.
- *Average residual energy of vehicles*. The vehicles' average amount of energy (in percentage) left in the EVs' batteries. We measure it to compare the level of energy that different methods can maintain.
- *Average number of charges of vehicles*. The average number of charges that the EVs receive per hour. We measure it to compare the methods' ability in offering charging opportunities to EVs.
- *Performance in distributing energy supply overhead*. The average amount of energy (in logarithmic scale) transferred per charging lane per hour. We measure it to compare the charging overhead generated by different methods. Meanwhile, we also measure the average number of charges (in logarithmic scale) occurred per charging lane per hour. We measure it to compare the energy supply opportunity generated by different methods. In addition, we also measure the CDF of the energy supply overhead over all charging lanes. We measure it to compare the balance of energy supply overhead of different methods.
- *Quality of solution*. The distance between the found solution and the optimal solution. Suppose P' is a set of points uniformly distributed along the optimal solution (i.e., Pareto Front), and A is the found solution, the average distance from P' to A, called the D-metric [51], is defined as

$$D(A, P') = \frac{\sum_{v \in P'} d(v, A)}{|P'|}. \tag{25}$$

where $d(v, A)$ is the minimum Euclidean distance between v and the points in A. We measure $D(A, P')$ to illustrate the quality of the solution to the proposed optimization problem.

B. Experimental Results

Based on the traffic data extracted from the 1-year long dataset mentioned in Section III-A, from total 26,036 landmarks, *CatCharger+* chose 930 landmarks to deploy charging lanes, *CatCharger* chose 922, while *Random* and *MaxFlow* chose 228. Since *CatCharger* and *CatCharger+* place most of the charging lanes at positions with short required lane lengths, while *Random* and *MaxFlow* use the same deployment cost and set the length of each charging lane to the longest length



(a) Operable vehicles over time. (b) Operable vehicles versus threshold. (c) Average residual energy. (d) Avg number of charges of vehicles.

Fig. 9: Performance in supporting EV charging demands.

in *CatCharger* and *CatCharger+*, so they result in much fewer charging lanes.

1) Average Ratio of Operable Vehicles: Figure 9(a) shows the average ratios of operable vehicles (SoC>0%) in each hour in a day during the month. Figure 9(b) shows the ratios of operable vehicles resulted from different residual energy thresholds. In both figures, the ratios follow: *CatCharger+*>*CatCharger*>*MaxFlow*>*Random*.

From Figure 9(a), we can see that at the beginning of a day (00:00∼06:00), the ratio of operable vehicles remains high and drops slowly. This is because that most vehicles are not driving during this time period. In *Random*, starting from 06:00, the ratio drops significantly faster than that before 06:00 and drops almost linearly with time. This means that many vehicles deplete energy almost linearly with time and are not charged at all. After 15:00, the ratio of operable vehicles drops much slower and remains at around 30%. This is primarily because that only taxicabs are serving after this time and get random charges, while most of the buses and Dada buses become out of service and do not consume energy any more. In comparison, the ratio of operable vehicles in *MaxFlow* is more stable and only reduces by around 15% between 06:00 and 14:00, This is because that *MaxFlow* aims to maximize the traffic flow covered by the charging lanes, so some vehicles sharing the same routes can remain operable during this period of time. But the operability of EVs cannot be guaranteed because *MaxFlow* does not aim to maintain the SoC of EVs. After 15:00, the ratio of operable vehicles also drops much slower and remains at around 30% as in *Random* due to the same reason. More than 90% of the EVs in *CatCharger* remain operable at the end of the day. This is because that in *CatCharger*, the landmarks selected to deploy charging lanes can be frequently visited by EVs and recharge the passing EVs with relatively shorter charging lane length. Meanwhile, the optimization problem in *CatCharger* ensures that the EVs can maintain their SoC at any landmarks in the road network. Therefore, *CatCharger* can support the highest ratio of operable EVs compared with the other methods.

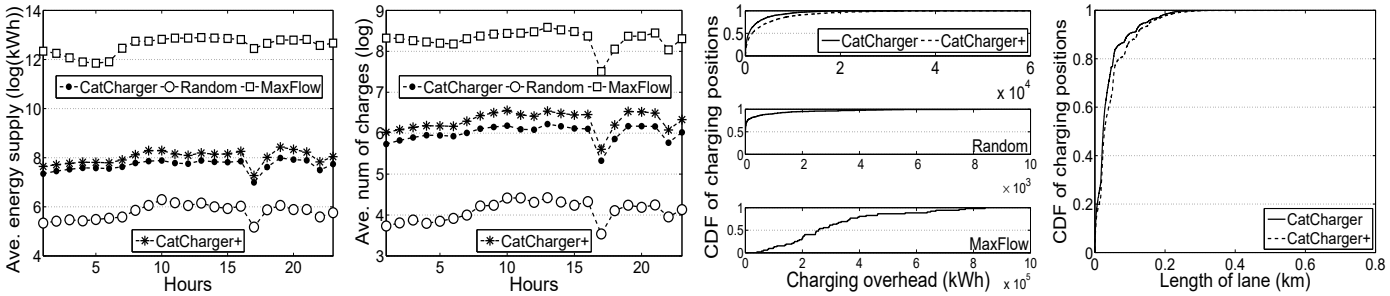
Figure 9(b) shows that the ratio of operable vehicles of *CatCharger* increases linearly with the increase of the threshold of expected residual energy, since a higher residual energy guarantee increases the probability that an EV can be continuously operable. A higher expected residual energy threshold results in a higher deployment cost. As a result, as the allowed deployment cost increases, the ratio of operable vehicles in *Random* and *MaxFlow* increases. The increase rate of *CatCharger* is higher than *Random* and *MaxFlow*, which means that *CatCharger* can more effectively plan the positions and lengths to maintain the highest ratio of operable vehicles given a deployment budget.

Since *CatCharger+* further considers the variances of vehicle visit frequency and vehicle passing speed, and the maximization of vehicle traffic flow at the chargers, the deployed chargers can fully recharge most EVs with their lane length since the EVs' passing speed at the charger landmarks does not vary a lot, and are free from vehicle traffic congestion. Therefore, *CatCharger+* can maintain the highest ratio of vehicles operable by the end of a day (Figure 9(a)) and under different residual energy thresholds (Figure 9(b)).

2) Average Residual Energy of Vehicles: Figure 9(c) shows the average residual energy of vehicles under different hours in a day. The results follow *CatCharger+*>*CatCharger*>*MaxFlow*>*Random*. The relationship between the methods and the changing of this metric are similar to those in Figure 9(a) due to the same reasons.

3) Average Number of Charges of Vehicles: Figure 9(d) shows the average number of charges of all vehicles under different hours throughout a day. The results follow *CatCharger+*>*CatCharger*>*MaxFlow*>*Random*. We can see that *CatCharger+* and *CatCharger* offer the most and second most charging opportunities to vehicles due to the same reasons as in Section V-B1. Note that the vehicles in *Random* almost receive no charging opportunities the whole day.

4) Performance in Distributing Energy Supply Overhead: Figure 10(a) shows the average energy supply overhead (i.e., amount of transferred energy) per landmark (in logarithmic scale) under different hours in a day. Figure 10(b) shows the average number of charges per landmark (in logarithmic scale) under different hours in a day. In both figures, the results follow *MaxFlow*≫*CatCharger+*>*CatCharger*>*Random*. We can see that the supply overhead in *Random* is the lowest. This is because that most EVs receive no charge from the charging lanes. In comparison, the supply overhead in *MaxFlow* is much higher. This means that *MaxFlow* is more effective in capturing the landmarks that EVs most frequently visit. However, since *MaxFlow* do not aim to maintain the SoC of EVs at any landmark in the road network, the charging lanes generally concentrate on popular areas. Moreover, since *MaxFlow* does not consider the passing speed, the cost of a single charging lane in *MaxFlow* is higher, which results in fewer charging positions. In contrast, *CatCharger* is able to deploy more shorter charging lanes with the same cost since it considers vehicle passing speed. In addition, *CatCharger* considers vehicle visit frequency, so the EVs in *CatCharger*



(a) Energy supply overhead. (b) Num of charges of charging lanes. (c) Distribution of supply overhead. (d) Distribution of charging lane length.

Fig. 10: Performance in distributing energy supply overhead.

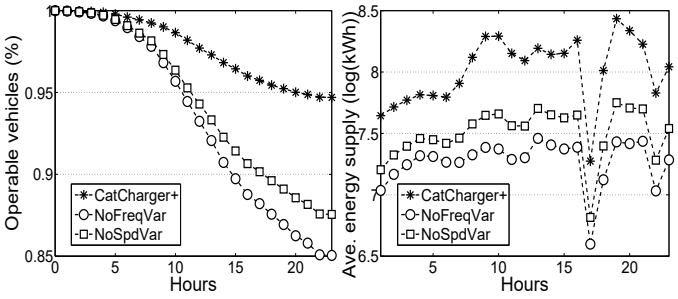


Fig. 11: Impact of components on the ratio of operable vehicles.

Fig. 12: Impact of components on average energy supply overhead.

receive more charges at more landmarks, which results in its lower average energy supply overhead over all the landmarks. *CatCharger+* further avoids deploying chargers at the landmarks with variant vehicle passing speed an vehicle visit frequency, and enables more EVs to receive recharge than *CatCharger*. Therefore, it achieves the highest results in both Figure 10(a) and Figure 10(b).

Figure 10(c) shows the CDF of the energy supply overhead of all the charging lanes. Figure 10(d) shows the CDF of the charging lane lengths in *CatCharger* and *CatCharger+*, and *Random* and *MaxFlow* are not included since they have the same charging lane length. From Figure 10(c), we can see that the energy supply overhead of the charging lanes in *CatCharger* is more evenly distributed than the others. In Figure 10(d), we see that most of the charging lanes in *CatCharger* and *CatCharger+* have lengths shorter than 0.1km due to the constraint 21. This result is consistent with the results in [7], [23], [24]. According to Figure 10(c), more than 80% of the charging lanes suffer from energy supply overhead less than 5,000kWh in *CatCharger*. Most of the charging lanes have no energy supply overhead in *Random*. In comparison, the supply overhead of more than 75% of the charging stations in *MaxFlow* is higher than 200,000kWh, which is due to the same reasons as explained for Figure 10(a) and Figure 10(b). These results verify the superior performance of *CatCharger* and *CatCharger+* in evenly distributing the energy supply overhead of the charging lanes.

5) Impact of Variance of Vehicle Passing Speed and Visit Frequency: As discussed in Section IV-C3, the additional consideration of the variance of vehicle visit frequency (σ_i^f) and the variance of vehicle passing speed (σ_i^v) in Equation (6) can help extract the candidate landmarks for charging lane deployment with more stable vehicle visit frequency and vehicle passing speed, and then better guide the deployment of in-motion wireless chargers. To demonstrate the impact of these two components, we recalculated the score of each landmark without considering the variance of vehicle visit frequency (denoted as *NoFreqVar*), and without considering the variance of vehicle passing speed (denoted as *NoSpdVar*). Based on the new landmark scores, we redetermined the deployment of chargers, and measured the average ratio of operable vehicles. The measurement results are illustrated in Figure 11. In addition, we also measured the average energy supply overhead per landmark under different hours in a day. The measurement results are illustrated in Figure 12.

From Figure 11, we can see that *CatCharger+* increases the final ratio of operable vehicles by the end of the day by 11.8% when compared to *NoFreqVar* and 8.6% when compared to *NoSpdVar*, respectively. This is because that *NoSpdVar* selects some landmarks with high variance of vehicle passing speed to deploy chargers. At the landmarks where the vehicle passing speed is very high during some time, the deployed chargers can not fully charge the EVs after they pass by, which made only around 87.5% of the vehicles remaining operable by the end of the day. Similarly, *NoFreqVar* selects some landmarks with high variance of vehicle visit frequency to deploy chargers. At the landmarks where the vehicle visit frequency is very low during some time, the deployed chargers can not serve the charging demand of many EVs, which resulted in that there are only around 85% of the vehicles remaining operable by the end of the day. Since the lack of considering vehicle visit frequency causes more EVs to fail to charge than the lack of considering vehicle passing speed, the ratio of operable vehicles in *NoSpdVar* is a bit higher than that in *NoFreqVar*.

From Figure 12, we can see that *CatCharger+* increases the average energy supply overhead of all the chargers by 10% when compared to *NoFreqVar* and 7% when compared to *NoSpdVar*, respectively. This is due to the same reasons as explained in Figure 11. The consideration of the variances of vehicle passing speed and vehicle visit frequency enables the deployed chargers to serve more EVs. Therefore, the charging energy supply overhead per charger is increased in *CatCharger+*. These measurement results verify that the consideration of the variances of vehicle passing speed and vehicle visit frequency is effective in selecting landmarks that are more suitable for deploying the chargers.

6) Performance Evaluation in Multiple Days: To validate the effectiveness of our charger deployment method under different scenarios, we measured the ratio of operable vehicles and residual energy of the vehicles by the end of different days. Figure 13 shows the ratios of operable vehicles by the end of different days. Figure 14 shows the median, 5^{th} and 95^{th} percentiles of the residual energy (i.e., SoC) of all the vehicles

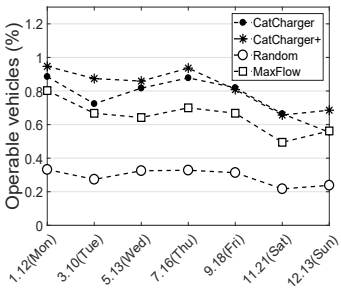


Fig. 13: Ratio of operable vehicles by the end of different days.

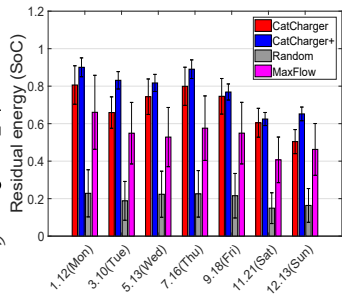


Fig. 14: Residual energy of vehicles by the end of different days.

by the end of different days. In these experiments, we assume that all the vehicles are fully charged at the beginning of a day. This assumption is reasonable because many previous studies have confirmed that most EVs are fully charged overnight at their home or dispatch center [55], [56].

From Figure 13, we can see that the ratios of operable vehicles generally follow: *CatCharger+*⩾*CatCharger*> *MaxFlow*>*Random* in different days. From Figure 14, we can see that the median residual energy of the vehicles follow: *CatCharger+*>*CatCharger*>*MaxFlow*>*Random* in different days. These results confirm that the charger deployment determined by our method can better support the continuous operability of EVs under various scenarios. We can also observe that the ratio of operable vehicles and the vehicles' SoC significantly drop on weekends, especially for *CatCharger+* and *CatCharger*. This is because that the traffic pattern on weekends is quite different from that on normal weekdays. One possible reason to the significant change of traffic pattern is that during weekends, the appearance pattern of passengers significantly changes. Some EV drivers (e.g., electric taxicab drivers) need to change their regular route to cover the changed passenger appearance pattern. If the charger deployment plan fully considers the drivers' routing behavior and place more charging lanes on the routes which the drivers are willing to drive through during both weekdays and weekends, the determined charger deployment plan may provide more charging opportunities to the EVs.

7) Impact of considering drivers' routing choice behavior: To confirm the impact of considering drivers' routing choice behavior on determining the charging lanes, we vary the sensitivity of our optimization problem towards the availability of chargers. Specifically, since the parameter β in Equation (13) describes how sensitive EV drivers are to driving a route equipped with charging facilities while making routing choices [20], we vary the value of β between 0.0 and 1.0 and measured the ratios of operable vehicles in different days. From Figure 15, we can see that increasing the value of β generally increases the ratio of operable vehicles in all days. This is because that according to Equation (13), a larger value of β will cause the route that has a shorter travel time cost to have higher weight of deploying chargers. We can also notice that increasing the value of β brings only about 12.5% improvement on the ratio of operable vehicles during weekdays (i.e., Jan 12, Mar 10, May 13, Jul 16 and Sep 18), but brings about 75% improvement on the ratio of operable vehicles during weekends (i.e., Nov 21 and Dec 13). This means that despite the significant change of traffic pattern on weekends, the drivers still prefer to drive the routes with a relatively lower travel time cost. Increasing the value of β enables the routes that are not frequently driven during weekdays to be covered with charging lanes, thereby significantly increases the charging opportunities of EVs during weekends.

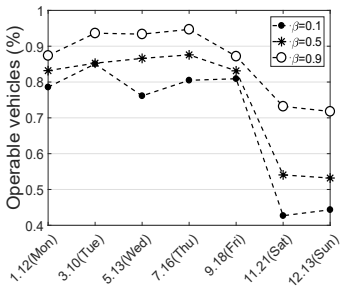


Fig. 15: Impact of considering drivers' routing choice behavior.

TABLE II: Table of D-metrics and computation time.

Generation	D-metric	Computation time (second)
100	0.575	197.766
200	0.440	818.938
300	0.437	1294.016
400	0.101	2029.203
500	0.179	2535.172
600	0.099	3423.562
700	0.097	4088.281
800	0.084	4478.438
900	0.077	5357.328
1000	0.088	6099.609

8) Quality of solution to the optimization problem: To illustrate the quality and efficiency of obtaining the deployment decisions, we measured the D-metric between the found solution and the optimal solution to the optimization problem and the computation time. The computer we used for solving the problems has a 2.50GHz Intel Core I5-7300HQ CPU and 32GB RAM. Since we do not know the actual Pareto Front of the optimization problem, we use an upper approximation of the Pareto Front for comparison. Specifically, we allow the decision variables x_i to take continuous values in [0, 1], which leads to a relaxed form of the problem. Then, we follow [51] to obtain the solutions to the relaxed problem (i.e., upper approximation). Finally, we measured the D-metric between the obtained Pareto Front under each generation and the upper approximation. Table II illustrates the D-metrics and computation time under different generations. We can see that along with the increase of the generations, the D-metric quickly converges to around 0.88, while the computation time almost increases linearly. Combining with the above experiment results, we conclude that the proposed method can generate near-optimal plans for deploying in-motion wireless chargers with feasible computation time.

VI. Conclusion

Previous plug-in station deployment methods are not applicable for the deployment of wireless charging lanes in metropolitan road networks due to different charging characteristics. Previous methods for deploying wireless charging lanes cannot handle the challenge in metropolitan scale. We propose *CatCharger* to tackle this problem. We analyzed a metropolitan-scale mobility dataset collected in Shenzhen, and the observations support the design of *CatCharger*. Using an entropy minimization based method, we conduct categorization and clustering on the intersections (landmarks), and extract the candidate positions for placing charging lanes that have low vehicle passing speed (hence short charging lanes) and high vehicle visit frequency (hence high covered traffic), and low variances of these two metrics. Then by using KDE

to model vehicle mobility and estimate the residual energy of EVs at a landmark, and considering EV drivers' routing choice behavior, we formulate a multi-objective optimization problem to minimize the total deployment cost, maximize the vehicle traffic flow at the landmarks with chargers, and meanwhile ensuring the continuous operability of the vehicles on roads. We conducted trace-driven experiments to verify the superior performance *CatCharger* over other methods. In the future, we plan to consider more human activities that affect the movement of public transit vehicles (e.g., pickup requests).

ACKNOWLEDGEMENTS

We would like to thank Ms. Xiaoying Li for her help in analyzing the dataset. This research was supported in part by National Natural Science Foundation of China (No.62103325, No.61802387, No. 61806192), The Science and Technology Development Fund, Macau SAR (File no. 0015/2019/AKP), National Natural Science Foundation of Shenzhen (No.JCYJ20190812153212464, JCYJ20190812160003719), Shenzhen Engineering Research Center for Beidou Positioning Service Improvement Technology, U.S. NSF grants NSF-2136948, NSF-1827674, CCF-1822965 and Microsoft Research Faculty Fellowship 8300751.

REFERENCES

[1] L. Yan, H. Shen, J. Zhao, C. Xu, F. Luo, and C. Qiu, "CatCharger: Deploying wireless charging lanes in a metropolitan road network through categorization and clustering of vehicle traffic," in *Proc. of INFOCOM*, 2017.

[2] V. Cirimele, M. Diana, F. Bellotti, R. Berta, N. El Sayed, A. Kobeissi, P. Guglielmi, R. Ruffo, M. Khalilian, A. La Ganga *et al.*, "The fabric ict platform for managing wireless dynamic charging road lanes," *IEEE TVT*, vol. 69, no. 3, 2020.

[3] "Shenzhen will electrify all taxicabs by 2020," http://tech.qq.com/a/20171227/022885.htm, 2020, accessed in November, 2020.

[4] "Roadmap for the electrification of public transportation in kolkata," http://shaktifoundation.in/report/roadmap-for-the-electrification-of-public-transportation-in-kolkata/, 2020, accessed in November, 2020.

[5] "How federal grants are accelerating the adoption of electrified mass transit," https://www.greentechmedia.com/articles/read/a-boost-for-electric-buses, 2020, accessed in November, 2020.

[6] P. Machura, V. De Santis, and Q. Li, "Driving range of electric vehicles charged by wireless power transfer," *IEEE TVT*, vol. 69, no. 6, 2020.

[7] Y. J. Jang, E. S. Suh, and J. W. Kim, "System architecture and mathematical models of electric transit bus system utilizing wireless power transfer technology," *Systems Journal*, 2015.

[8] S. Jeong, Y. J. Jang, and D. Kum, "Economic analysis of the dynamic charging electric vehicle," *TPE*, 2015.

[9] S. Bae and A. Kwasinski, "Spatial and temporal model of electric vehicle charging demand," *TSG*, 2012.

[10] Y. Zheng, Z. Y. Dong, Y. Xu, K. Meng, J. H. Zhao, and J. Qiu, "Electric vehicle battery charging/swap stations in distribution systems: comparison study and optimal planning," *TPS*, 2014.

[11] M. Eisel, J. Schmidt, and L. Kolbe, "Finding suitable locations for charging stations," in *Proc. of IEVC*, 2014.

[12] A. Lam, Y.-W. Leung, and X. Chu, "Electric vehicle charging station placement: Formulation, complexity, and solutions," *TSG*, 2014.

[13] G. Wang, Z. Xu, F. Wen, and K. P. Wong, "Traffic-constrained multi-objective planning of electric-vehicle charging stations," *TPD*, 2013.

[14] P. Sanchez-Martin, G. Sanchez, and G. Morales-Espana, "Direct load control decision model for aggregated ev charging points," *TPS*, 2012.

[15] W. Yao, J. Zhao, F. Wen, Z. Dong, Y. Xue, Y. Xu, and K. Meng, "A multi-objective collaborative planning strategy for integrated power distribution and electric vehicle charging systems," *TPS*, 2014.

[16] O. C. Onar, J. M. Miller, S. L. Campbell, C. Coomer, C. White, L. E. Seiber *et al.*, "A novel wireless power transfer for in-motion ev/phev charging," in *Proc. of APEC*, 2013.

[17] L. Yan and H. Shen, "TOP: vehicle trajectory based driving speed optimization strategy for travel time minimization and road congestion avoidance," in *Proc. of MASS*, 2016.

[18] F. He, Y. Yin, and J. Zhou, "Integrated pricing of roads and electricity enabled by wireless power transfer," *TRC*, vol. 34, 2013.

[19] Y. D. Ko and Y. J. Jang, "The optimal system design of the online electric vehicle utilizing wireless power transmission technology," *IEEE TITS*, vol. 14, no. 3, 2013.

[20] R. Riemann, D. Z. Wang, and F. Busch, "Optimal location of wireless charging facilities for electric vehicles: Flow-capturing location model with stochastic user equilibrium," *TRC*, vol. 58, 2015.

[21] M. Fuller, "Wireless charging in california: Range, recharge, and vehicle electrification," *TRC*, vol. 67, 2016.

[22] Z. Chen, F. He, and Y. Yin, "Optimal deployment of charging lanes for electric vehicles in transportation networks," *TRB: Methodological*, vol. 91, 2016.

[23] I. Hwang, Y. J. Jang, Y. D. Ko, and M. S. Lee, "System optimization for dynamic wireless charging electric vehicles operating in a multiple-route environment," *IEEE TITS*, vol. 19, no. 6, 2017.

[24] Z. Chen, W. Liu, and Y. Yin, "Deployment of stationary and dynamic charging infrastructure for electric vehicles along traffic corridors," *TRC*, vol. 77, 2017.

[25] Z. Liu and Z. Song, "Robust planning of dynamic wireless charging infrastructure for battery electric buses," *TRC*, vol. 83, 2017.

[26] Z. Bi, G. A. Keoleian, and T. Ersal, "Wireless charger deployment for an electric bus network: A multi-objective life cycle optimization," *Applied energy*, vol. 225, 2018.

[27] S. D. Manshadi, M. E. Khodayar, K. Abdelghany, and H. Üster, "Wireless charging of electric vehicles in electricity and transportation networks," *IEEE TSG*, vol. 9, no. 5, 2017.

[28] M. Li, X. Wu, Z. Zhang, G. Yu, Y. Wang, and W. Ma, "A wireless charging facilities deployment problem considering optimal traffic delay and energy consumption on signalized arterial," *IEEE TITS*, 2019.

[29] L. Yan, H. Shen, Z. Li, A. Sarker, J. A. Stankovic, C. Qiu, J. Zhao, and C. Xu, "Employing opportunistic charging for electric taxicabs to reduce idle time," *ACM IMWUT*, vol. 2, no. 1, 2018.

[30] L. Yan, H. Shen, and K. Chen, "MobiT: Distributed and congestion-resilient trajectory based routing for vehicular delay tolerant networks," *IEEE/ACM ToN*, vol. 26, no. 3, 2018.

[31] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE TVT*, vol. 65, no. 12, 2016.

[32] D. Barbará, Y. Li, and J. Couto, "Coolcat: an entropy-based algorithm for categorical clustering," in *Proc. of CIKM*, 2002.

[33] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. of SIGKDD*, 2012.

[34] "Openstreetmap," http://www.openstreetmap.org/, 2020, accessed March, 2020.

[35] "Shenzhen wiki," https://en.wikipedia.org/wiki/Shenzhen, 2020, accessed March, 2020.

[36] "Apache Hadoop 2.7.3," http://hadoop.apache.org/, 2020, accessed in March, 2020.

[37] "Apache spark 1.5.2," http://spark.apache.org/, 2020, accessed March, 2020.

[38] K. Pandit, D. Ghosal, H. M. Zhang, and C.-N. Chuah, "Adaptive traffic signal control with vehicular ad hoc networks," *IEEE TVT*, vol. 62, no. 4, 2013.

[39] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proc. of GIS*, 2010.

[40] K. Pearson, "Note on regression and inheritance in the case of two parents," in *Proc. of the Royal Society of London*, 1895.

[41] X. Wang, Z. Zhou, F. Xiao, K. Xing, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis," *IEEE TMC*, vol. 18, no. 9, 2018.

[42] Q. Hu, S. Wang, X. Cheng, J. Zhang, and W. Lv, "Cost-efficient mobile crowdsensing with spatial-temporal awareness," *IEEE TMC*, 2019.

[43] R. G. Michael and S. J. David, "Computers and intractability: a guide to the theory of NP-completeness," *W. H. Freeman*, 1979.

[44] K. Chen and L. Liu, "The 'Best K' for entropy-based categorical data clustering," in *Proc. of SSDBM*, 2005.

[45] M. P. Wand and M. C. Jones, *Kernel smoothing*. CRC Press, 1994.

[46] R. Li, G. Rose, and M. Sarvi, "Using automatic vehicle identification data to gain insight into travel time variability and its causes," *JTRB*, 2006.

[47] R. Li, H. Chai, and J. Tang, "Empirical study of travel time estimation and reliability," *Mathematical Problems in Engineering*, 2013.

[48] T. Kurczveil, P. Á. López, and E. Schnieder, "Implementation of an energy model and a charging infrastructure in SUMO," in *Proc. of SUMO User Conference*, 2013.

[49] N. Jiang, C. Xie, J. C. Duthie, and S. T. Waller, "A network equilibrium analysis on destination, route and parking choices with mixed gasoline and electric vehicular flows," *EURO Journal on Transportation and Logistics*, vol. 3, no. 1.

[50] D. Gross, *Fundamentals of Queueing Theory*. John Wiley & Sons, 2008.

[51] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE TEC*, vol. 11, no. 6, 2007.

[52] M. Xylia, S. Leduc, P. Patrizio, F. Kraxner, and S. Silveira, "Locating charging infrastructure for electric buses in stockholm," *TRC*, vol. 78, 2017.

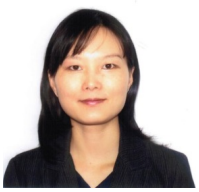
[53] Z. Tian, T. Jung, Y. Wang, F. Zhang, L. Tu, C. Xu, C. Tian, and X.-Y. Li, "Real-time charging station recommendation system for electric-vehicle taxis," *IEEE TITS*, vol. 17, no. 11, 2016.

[54] A. Gil and J. Taiber, "A literature review in dynamic wireless power transfer for electric vehicles: Technology and infrastructure integration challenges," in *Sustainable Automotive Technologies 2013*, 2014.

[55] Y. J. Jang, "Survey of the operation and system study on wireless charging electric vehicle systems," *TRC*, vol. 95, 2018.

[56] Z. Bi, G. A. Keoleian, Z. Lin, M. R. Moore, K. Chen, L. Song, and Z. Zhao, "Life cycle assessment and tempo-spatial optimization of deploying dynamic wireless charging technology for electric cars," *TRC*, vol. 100, 2019.

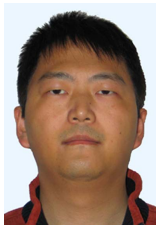
Li Yan received his Ph.D. degree in Computer Science from the University of Virginia, USA in 2019. He worked as a Postdoc in Senseable City Lab at Massachusetts Institute of Technology, USA from 2020 to 2021. Currently, he is an Associate Professor in the School of Cyber Science and Engineering at Xi'an Jiaotong University, China. His research interests include Data-driven Cyber-Physical Systems, Big Data Analytics and Mobile Computer Networks. He was the Best Transactions Paper Awardee of IEEE T-ITS and the Best-in-Session-Presentation Awardee of INFOCOM'2017.



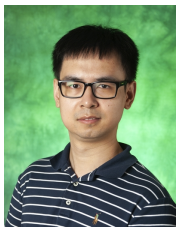
Haiying Shen received the BS degree in Computer Science and Engineering from Tongji University, China in 2000, and the MS and Ph.D. degrees in Computer Engineering from Wayne State University in 2004 and 2006, respectively. She is currently an Associate Professor in the Department of Computer Science at the University of Virginia. Her research interests include distributed computer systems, cloud and edge computing, machine learning, big data and cyber-physical systems. She is a Microsoft Faculty Fellow of 2010, and a senior member of the IEEE and the ACM.

Juanjuan Zhao received her MS from Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences in 2017, and became a Research Associate the same year. Her research interests include big data processing, streaming-data processing, spatio-temporal data mining.

Chengzhong Xu received his Ph.D. degree from the University of Hong Kong in 1993. He is currently a Chair Professor of Computer Science and the Dean of Faculty of Science and Technology, University of Macau. Prior to this, he was in the faculty of Wayne State University, USA and Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, China. His research interest is in cloud computing and data-driven intelligent applications. He has published 400+ papers and 100+ patents. He was the Best Paper Awardee or Nominee of ICPP'2005, HPCA'2013, HPDC'2013, Cluster'2015, GPC'2018, UIC'2018, AIMS'2019, and HPBD&Is'2020. He was the Chair of IEEE Technical Committee on Distributed Processing from 2015-2019. Dr. Xu is an IEEE Fellow.



Feng Luo currently is a professor in the School of Computing, Clemson University. He received his Ph.D. degree in computer science from the University of Texas at Dallas in 2004. He was a postdoctoral senior research associate at the Department of Pathology, University of Texas Southwestern Medical Center at Dallas. His research interests are big data analytics, bioinformatics, deep learning, and its applications. He has published 62 journals and 31 conference papers in these areas. Dr. Luo is a senior member of IEEE.



Chenxi Qiu received the BS degree in Telecommunication Engineering from Xidian University, China, in 2009 and the Ph.D. degree in Electrical and Computer Engineering at Clemson University in 2015. He worked as a Postdoc in the College of Information Science and Technology at Penn State University from 2016 to 2018. Currently, he is an Assistant Professor in The Department of Computer Science and Engineering at the University of North Texas, TX, United States. His research interests include cybersecurity, cyber physical systems, and cloud computing.

Zhe Zhang received her M.E. degree in Computer Engineering from the University of Virginia in 2019, and M.S. degree in Electrical Engineering from North Carolina State University in 2014. She currently works as a Research Engineer in Xi'an Jiaotong University. Her research interests include Data Mining and Cyber-Physical Systems.

Shohaib Mahmud received his B.Sc. Engineering Degree in Electrical Engineering from Bangladesh University of Engineering and Technology, Bangladesh in 2016. Currently he is a Ph.D. student in the Department of Computer Engineering at University of Virginia. His research interests include Cyber-Physical Systems, Machine Learning and Data driven Intelligent Transport Systems.