38th International Electric Vehicle Symposium and Exhibition (EVS38) Göteborg, Sweden, June 15-18, 2025

EV Dynamic Wireless Charging Navigation in Electrified Transportation Network: A Shortest-Path-Based Reinforcement Learning Method

Chaoran Si[1], K. T. Chau[2*], Yunhe Hou[1], Wei Liu[2]

[1] *Department of Electrical and Electronic Engineering, The University of Hong Kong*

[2] *Research Centre for Electric Vehicles and Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China, Email: k.t.chau@polyu.edu.hk*

Executive Summary

Dynamic wireless charging (DWC) plays an important role in addressing range anxiety for electric vehicle (EV) owners and achieving extensive EV propulsion revolution. With real-time electricity prices and traffic conditions, DWC navigation is necessary for EVs to reduce the total cost. This paper first formulates a two-layer dynamic charging routing model to minimize EVs' travel and charging costs. Then, a shortest-path-based method is proposed to solve the model and extract low-dimensional features from real-time stochastic information. The optimal charging navigation strategy is finally obtained under unknown system uncertainties by feeding the system state containing extracted advanced features into the popular deep reinforcement learning algorithm, proximal policy optimization. Numerical results on a test traffic network demonstrate the effectiveness of the proposed method.

Keywords: Electric Vehicles, Inductive / Wireless Power Transfer, Optimal Charging Locations, Intelligent Transportation System for EVs, AI - Artificial Intelligence for EVs

1 Introduction

Wireless power transfer (WPT) technology has witnessed remarkable advancements over the past decades, enabling efficient and contactless energy transmission. Breakthroughs in magnetic resonant coupling [1], soft-switching techniques [2, 3], and multi-frequency configurations [4, 5] have significantly enhanced the reliability, efficiency, and flexibility of WPT systems. These technical improvements have laid a strong foundation for expanding WPT into diverse domains. WPT has found applications across various sectors beyond consumer electronics. In manufacturing and healthcare, WPT has enabled contactless heating [6] and actuation [7]. In industrial automation, wireless motor drives [8, 9], lighting systems [10], and industrial wireless networks [11] have been successfully deployed. Furthermore, innovations in impedance control [12], misalignment reduction [13], and long-range power transmission [14], have extended WPT's usability in complex environments. These applications collectively demonstrate WPT's growing maturity and readiness for large-scale integration [15].

Among the most promising frontiers for WPT is the electrification of transportation. Wireless charging offers an attractive alternative to plug-in methods, particularly for electric vehicles (EVs), by enabling automation, reducing wear, and enhancing user convenience [16−18]. With the rise of smart mobility, researchers have

explored efficient voltage balancing [19], magnetic quasi-resonant coupling [20], frequency modulation [21], and cyber-secure protocols [22] to ensure robust wireless EV charging solutions. One of the most transformative applications is dynamic wireless charging (DWC), which allows EVs to be charged while in motion. DWC systems overcome the limitations of short driving ranges and long charging times by providing energy on the go [23−26]. Recent innovations include quasi-omnidirectional charging systems [23], frequency-autonomous controls [24], and sizing of inductive power pads [25], which enable seamless power delivery during vehicle transit. The integration of magnetic extenders [14] and energy-encrypted transmitters [27] further enhances the adaptability and safety of DWC infrastructures.

The evolution of DWC reflects a broader shift toward sustainable, uninterrupted electric mobility. As EV adoption accelerates, DWC not only offers a scalable solution to charging infrastructure gaps but also plays a pivotal role in achieving zero-emission transportation goals [28]. It is therefore foreseeable that dynamic wireless charging systems (DWCS), particularly those based on charging lanes, will become an essential component of the electrified transportation network (ETN) in future smart cities. To ensure EV owners enjoy efficient and convenient travel and charging experiences, it is crucial to investigate user-oriented dynamic charging navigation problems that integrate route planning with real-time charging decisions in a complex urban environment.

Existing studies have focused on EV charging navigation to static charging stations. For example, [29] proposes a rapid-charging navigation system for EVs, which is modular to protect drivers' privacy. Crowdsensing and queue theory are utilized in [30], aiming at reducing travel costs and addressing waiting time at charging stations. [31] presents a distributed algorithm for the joint routing and charging scheduling problem, which reduces computational complexity and protects user privacy. [32] proposes a two-stage algorithm to optimize EV routing and charging while reducing computational complexity by decomposing the problem into linear programming problems. However, due to the characteristics of DWCS, dynamic and static charging routing models are significantly different in terms of charging facility modeling, path constraints, and objective functions [33]. These differences require researchers to build new models and solution methods tailored specifically to the characteristics of DWC.

On the other hand, under uncertain charging prices and traffic conditions, real-time EV charging navigation is a challenging sequential decision-making problem with uncertainty. Traditional optimization approaches rely heavily on prior knowledge about these uncertainties [34, 35], which is often not available or accurate in practice. Reinforcement learning (RL) methods that do not rely on explicit models have emerged as effective tools for handling sequential decisions, especially in environments characterized by uncertainty. Therefore, some work in recent years has applied it to EV applications in smart grids, including the EV charging navigation problem [36, 37]. However, directly using raw high-dimensional data can make it difficult for the agent to learn the underlying patterns and relationships necessary for effective decision-making [38]. Existing RL algorithms are desired to be combined with efficient feature extraction methods.

Based on these gaps, we formulate a two-layer dynamic charging routing model to minimize EVs' travel and charging costs. Then we prove that the lower-layer problem can be decomposed into two single-source shortest-path (SP) problems. Accordingly, an SP-based method is proposed to efficiently solve the model while extracting low-dimensional features from stochastic information. The extracted features are then fed into the RL network, and proximal policy optimization (PPO) is utilized to adaptively learn optimal charging and routing strategies for EVs without prior system uncertainty data.

2 Problem Formulation

2.1 Two-Layer Optimization Problem

The UETN can be represented by the directed graph $\mathcal{G}=(\mathcal{N},\mathcal{E})$, where $\mathcal{N}$ is the set of nodes (traffic junctions) and $\mathcal{E}\subset\{(i,j)\,|\,(i,j)\in\mathcal{N}\times\mathcal{N}, i\neq j\}$ is the set of edges (roads). Adjacent nodes are always bidirectionally connected in the studied network, meaning that if $(i,j)\in\mathcal{E}$, then $(j,i)\in\mathcal{E}$. Let $\mathcal{E}^{\text{ch}}\subset\mathcal{E}$ denote the set of edges with charging lanes. Fig. 1 shows the configuration of an edge $(m,n)\in\mathcal{E}^{\text{ch}}$.

By utilizing real-time information from power and transportation systems (i.e., charging prices and average velocity on edges), the EV manages to minimize both travel and charging costs. Enumerating the charging lanes in the set $\mathcal{E}^{\text{ch}}$, we formulate a two-layer optimization problem to be solved at the beginning of each step t. A charging lane (m,n) is selected at the upper layer, and the SP from the current node loc_t to the ending

node n_{end} passing the selected charging lane (m,n), the corresponding optimal travel cost $C^{\text{tr},*}_{mn,t}$, and the battery level e_m at the entry m of the charging lane (m,n), are determined at the lower layer.

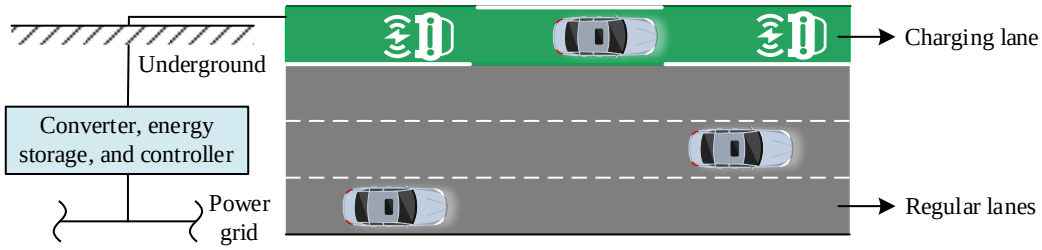

Figure 1: Configuration of an edge with charging lane

Upper-Layer Problem (ULP):

$$\min_{(m,n)\in\mathcal{E}^{\text{ch}}} \; C^{\text{tr},*}_{mn,t} + C^{\text{ch}}_{mn,t} \tag{1}$$

$$\text{s.t.} \quad C^{\text{ch}}_{mn,t} = \rho^{\text{ch}}_{mn}[\tau_t]\cdot e^{\text{ch}}_{mn} \tag{2}$$

$$e^{\text{ch}}_{mn} = \min\left(e^{\max} - e_m, e^{\text{ch,max}}_{mn,t}\right) \tag{3}$$

$$e^{\text{ch,max}}_{mn,t} = \frac{d_{mn}}{v_{mn}[\tau_t]}\,p^{\text{ch}} \tag{4}$$

Lower-Layer Problem (LLP):

$$C^{\text{tr},*}_{mn,t} = \min_{\{x_{ij},y_{ij}\}} \underbrace{\alpha\rho^{e}\left(\sum_{(i,j)\in\mathcal{E}} d_{ij}(x_{ij}+y_{ij}) + d_{mn}\right)}_{\text{Travel distance cost}} + \underbrace{q^{t}\left(\sum_{(i,j)\in\mathcal{E}} \frac{d_{ij}}{v_{ij}[\tau_t]}(x_{ij}+y_{ij}) + \frac{d_{mn}}{v_{mn}[\tau_t]}\right)}_{\text{Travel time cost}} \tag{5}$$

$$\sum_j x_{ij} - \sum_j x_{ji} = \begin{cases} 1 & i = loc_t \\ -1 & i = m \\ 0 & otherwise \end{cases} \tag{6}$$

$$\sum_j y_{ij} - \sum_j y_{ji} = \begin{cases} 1 & i = n \\ -1 & i = n_{\text{end}} \\ 0 & otherwise \end{cases} \tag{7}$$

$$e_m = e_t - \alpha\sum_{(i,j)} d_{ij}x_{ij} \tag{8}$$

$$x_{ij}, y_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{E} \tag{9}$$

Constraint (2) defines the charging cost $C^{\text{ch}}_{mn,t}$ on charging lane (m,n), where $\rho^{\text{ch}}_{mn}[\tau_t]$ is the predicted value of the fluctuating charging price on edge (m,n) at time τ_t and e^{ch}_{mn} is the charging energy on edge (m,n). Constraint (3) requires e^{ch}_{mn} to be the smaller of the EV battery consumption $e^{\max} - e_m$ and the max charging capacity $e^{\text{ch,max}}_{mn}$ of edge (m,n). As calculated in (8), e_m is the battery level when the EV arrives at node m, the entry of the charging lane (m,n). Constraint (4) gives the definition of $e^{\text{ch,max}}_{mn}$, where p^{ch} is the DWC power and $v_{mn}[\tau_t]$ is the predicted value of the random vehicle speed on edge (m,n) at time τ_t. Constraint (6) and (7) ensure a connected route from current node loc_t to the ending node passes the selected charging lane (m,n). The binary decision variables x_{ij} and y_{ij} indicate the driving route before and after charging, respectively.

2.2 Shortest-Path-Based Solution

Theorem: Given the selected charging lane (m,n) and the values of $\{v_{ij}[\tau_t]\}_{(i,j)\in\mathcal{E}}$, **LLP** is equivalent to two single-source SP problems with nonnegative edge weight.

Proof: Let

$$w_{ij} = \alpha \rho^e d_{ij} + q^t \frac{d_{ij}}{v_{ij}\left[\tau_t\right]} \tag{10}$$

be the weight of edge (i, j), for each $(i, j) \in \mathcal{E}$. Since all the components are non-negative, we have $w_{ij} \geq 0$. Then **LLP** can be decomposed into the following two sub-problems:

$$\min_{x_{ij}} \sum_{(i,j)\in\mathcal{E}} w_{ij} x_{ij}, \quad \text{s.t. (6)}, \; x_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{E} \tag{11}$$

$$\min_{y_{ij}} \sum_{(i,j)\in\mathcal{E}} w_{ij} y_{ij}, \quad \text{s.t. (7)}, \; y_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{E} \tag{12}$$

which are two single-source shortest-path problems in the same direct graph $\mathcal{G}$ with non-negative edge weight w_{ij}. The single sources for (11) and (12) are node loc_t and node n_{end}, respectively. Both of the problems can be solved by Dijkstra's algorithm.

Based on the above theorem, an SP-based method is proposed to solve **LLP** before the EV arrives at the charging lane. Let the sequence

$$L_{m,t} = \left(L_{m,t}^1, L_{m,t}^2, ..., L_{m,t}^{end}\right) \tag{13}$$

with $L_{m,t}^{end} = m$, denote the optimal path from the current location $L_{m,t}^1 = loc_t$ to the start of the charging lane (m, n) determined at step *t*. Therefore, the battery level of the EV upon its arrival at the charging lane (m, n) can be calculated as follows:

$$e_m = e_t - \alpha \sum_{(i,j)\in L_{m,t}} d_{ij} \tag{14}$$

2.3 Markov Decision Process Formulation

With collected uncertainty data updated, DWC navigation can be formulated online via a finite Markov decision process (MDP) with unknown transition probability. The event-driven control is adopted since the EV selects the next route when it arrives at a node. Let t denote the step of the MDP and τ_t be the time at the beginning of step t.

1) State: The system state at step t is given by:

$$\mathbf{s}_t = \left(loc_t, SoC_t, \tau_t, \left\{v_{loc_t,j}\left[\tau_t\right]\right\}_{(loc_t,j)\in\mathcal{E}}, \left\{C_{mn,t}^{\text{tr},*}\right\}_{(m,n)\in\mathcal{E}^{\text{ch}}}, \left\{C_{mn,t}^{\text{ch}}\right\}_{(m,n)\in\mathcal{E}^{\text{ch}}}, \left\{L_{m,t}\right\}_{m\in\mathcal{M}^{\text{ch}}}\right) \tag{15}$$

2) Action: Given the system state at step t, the EV takes an action by selecting a charging lane (m, n) in $\mathcal{E}^{\text{ch}}$:

$$a_t = (m, n) \tag{16}$$

and taking a step en route $L_{m,t}$ to the selected charging lane.

3) Transition process: The state transition of the first two elements of $\mathbf{s}_t$ is controlled by the action a_t:

$$loc_{t+1} = L_{a_t,t}^2 = L_{m,t}^2 \tag{17}$$

$$SoC_{t+1} = SoC_t - \frac{\alpha d_{loc_t loc_{t+1}}}{e^{\max}} \tag{18}$$

where $L_{m,t}^2$ is the second node on the optimal path $L_{m,t}$ as defined in (13). The state transition of other elements in $\mathbf{s}_t$ are not only controlled by the action a_t, but also influenced by the uncertain data. It is challenging to learn the system uncertainties and establish the analytical transition model from $\mathbf{s}_t$ to $\mathbf{s}_{t+1}$. Therefore, RL is adopted in Section 3 to learn the state transition probability implicitly.

4) Reward: The reward at step t corresponding to state $\mathbf{s}_t$ and action a_t can be defined as:

$$r_t(\mathbf{s}_t, a_t) = \begin{cases} -w^{+}_{loc_t loc_{t+1}} & loc_t \neq m \\ -C^{\mathrm{ch}}_{mn,t} - w^{+}_{mn} - dist^{-}_{n,t} & loc_t = m \,\&\, loc_{t+1} = n \end{cases} \tag{19}$$

where loc_{t+1} is derived from (17). When the EV has not arrived at m, the start of the target charging lane, the reward is the negative equivalent weight of edge (loc_t, loc_{t+1}), i.e., the travel cost; when the EV has arrived at m, the reward is the sum of the negative charging cost on edge (loc_t, loc_{t+1}) and the negative travel cost from node loc_t to the ending node (using the data obtained at step t).

3 Proposed RL-Based Method

3.1 Overview of the Scheme

PPO is an online policy gradient-based RL algorithm that has gained significant attention in recent years. It strikes a favorable balance between sample complexity, simplicity, and wall time. In the absence of prior data on system uncertainties, PPO updates the policy parameters based on the observed rewards and states. The policy learned by PPO determines the probability distribution of actions, which the agent uses to choose the next action to take. The overall scheme of the proposed PPO-based method is shown in Fig. 2.

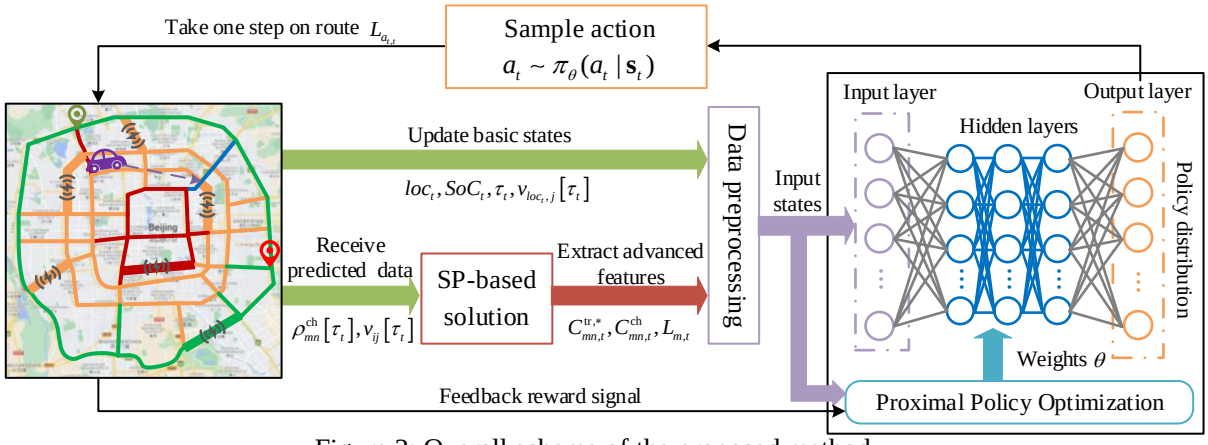

Figure 2: Overall scheme of the proposed method

At each step t, the smart EV observes the basic states and receives updated predicted data from power and transportation systems (the green arrows in Fig. 2). The SP-based solution for **LLP** extracts advanced feature states from the collected predicted data (the red arrow in Fig. 2). After necessary data preprocessing (cf. the next section), basic states and advanced features are together fed into the deep neural network (the purple arrow in Fig. 2). PPO updates the network weights (the blue arrow in Fig. 2) based on the reward signal and processed states. The action to be taken, i.e., the charging lane to be passed through, is then selected by sampling from the probability distribution generated by the policy network. Following the optimal route to the selected charging lane, the smart EV travels to the next node. The data from power and transportation systems will be updated and the process will be repeated until the EV arrives at the entry of a designated charging lane.

3.2 SP-PPO Algorithm

The PPO algorithm alternates between sampling data through interaction with the environment, and optimizing a surrogate objective function using stochastic gradient ascent. To avoid stepping too far with each update and accidentally causing performance collapse, specialized clipping in the objective function is given as follows:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(r_t(\theta)\hat{A}_t, clip\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t\right)\right] \tag{20}$$

where the expectation $\hat{\mathbb{E}}_t[...]$ indicates the empirical average over a finite batch of samples, and ε is the clipping hyperparameter. $r_t(\theta)$ is defined as the probability ratio of the current policy to the old policy, i.e.,

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid \mathbf{s}_t)}{\pi_{\theta_{old}}(a_t \mid \mathbf{s}_t)} \tag{21}$$

$\hat{A}_t$ is an estimator of the advantage function, and here we use the Generalized Advantage Estimate (GAE) to obtain it [39]. The total loss is combined with the value function error L_t^{VF} and the exploration term S to enable parameter sharing and ensure sufficient exploration, which is written as:

$$L_t^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t\left[L^{CLIP}(\theta) - c_1 L_t^{VF}(\theta) + c_2 S\left[\pi_\theta\right](\mathbf{s}_t)\right] \tag{22}$$

where c_1 and c_2 are coefficients. The proposed SP-PPO algorithm is summarized in Algorithm 1.

Algorithm 1 SP-PPO Algorithm

1: Initialize policy and value function parameters θ
2: **for** episode = $1: M$ **do**
3: Randomly pick up n_{start} and n_{end}, satisfying $n_{\text{start}} \neq n_{\text{end}}$ and $n_{\text{start}}, n_{\text{end}} \in \mathcal{N} - \mathcal{M}^{\text{ch}}$
4: Generate the initial state $\mathbf{s}_0$
5: **while** EV not in $\mathcal{M}^{\text{ch}}$ **do**
6: Select a charging lane as action a_t from the policy network and get corresponding route $L_{m,t}$
7: Take a step along route $L_{m,t}$, observe reward r_t, receive updated data, and generate new state $\mathbf{s}_{t+1}$
8: Store transition $(\mathbf{s}_t, a_t, r_t, \mathbf{s}_{t+1})$ in buffer Ξ
9: **if** Ξ is full **do**
10: Calculate loss function according to (22)
11: Update θ via stochastic gradient ascent
12: Release Ξ
13: **end if**
14: **end while**
15: **end for**

Algorithm 1 follows the general training structure of PPO, and our major work is further illustrated as follows. In line 3, the locations of the starting node and ending node are randomly selected among the non-charging nodes, which provides diversity to training samples. In line 4 and line 7, the advanced features in both the initial state $\mathbf{s}_0$ and the subsequent state $\mathbf{s}_{t+1}$ are generated by the SP-based solution. It should be noted that before feeding the updated and extracted features into the network, necessary data prepossessing, especially data alignment, must be performed. For example, since the current node loc_t may be adjacent to a different number of nodes, $\{v_{loc_t,j}[\tau_t]\}_{(loc_t,j)\in\mathcal{E}}$ can have variant lengths while iterating over the state. The length of the optimal route $L_{m,t}$ depends not only on the locations of the starting and ending nodes but also on the EV movement. Therefore, appropriate padding or truncating needs to be applied for input data alignment. Suppose the length of $\{v_{loc_t,j}[\tau_t]\}_{(loc_t,j)\in\mathcal{E}}$ is set to a and the first b elements of $L_{m,t}$ are selected, the dimension of the input state should be $(b+2)|\mathcal{E}^{\text{ch}}|+a+3$.

4 Performance Evaluation

4.1 Implementation Details

To evaluate the effectiveness of the proposed method, we simulate it on the main road network of Beijing, China, which includes 41 nodes and 12 charging lanes (comprising 6 bi-directional segments equipped with wireless charging, as illustrated in Fig. 3). Specifically, red segments have limits of either 50 km/h or 60 km/h, orange roads allow speeds up to 80 km/h, green ones permit 90 or 100 km/h, and blue routes accommodate up to 120 km/h. The travel speeds on these roads are sampled from truncated normal distributions whose means and standard deviations are proportional to their speed limits.

Additionally, the dynamic charging prices are also modeled using truncated normal distributions with different statistical characteristics for each charging lane. The considered EV is assumed to have an 80 kWh battery, where the initial state of charge SoC^{ini} is randomly sampled from a uniform distribution in the range of 0.4 to 0.6. The wireless charging power p^{ch} on each charging lane is taken as 30 kW [40]. The starting and ending locations n_{start} and n_{end} are drawn randomly from the set of all nodes excluding the charging lane network. The probabilistic parameters used in the experiments are listed in Table I. We assign $\alpha = 0.15$ kWh/km [37] and

set $\rho^{e} = 0.7673$ CNY/kWh, which reflects the average charging price applicable to non-local residents in Beijing. Moreover, based on the statutory minimum hourly wage in Beijing as of July 1, 2023, the EV user's time value q^{t} is estimated at 25.3 CNY/hour.

Figure. 3: Main road network of Beijing city with charging lanes equipped

Table 1: Probabilistic parameters

	Distribution	Boundary
Speeds on red segments (km/h)	$v_{\text{red}} \sim N(0.5\overline{v}_{\text{red}}, (0.15\overline{v}_{\text{red}})^2)$	$0 < v_{\text{red}} \le \overline{v}_{\text{red}}$
Speeds on orange segments (km/h)	$v_{\text{orange}} \sim N(0.5\overline{v}_{\text{orange}}, (0.15\overline{v}_{\text{orange}})^2)$	$0 < v_{\text{orange}} \le \overline{v}_{\text{orange}}$
Speeds on green segments (km/h)	$v_{\text{green}} \sim N(0.5\overline{v}_{\text{green}}, (0.15\overline{v}_{\text{green}})^2)$	$0 < v_{\text{green}} \le \overline{v}_{\text{green}}$
Speeds on blue segments (km/h)	$v_{\text{blue}} \sim N(0.5\overline{v}_{\text{blue}}, (0.15\overline{v}_{\text{blue}})^2)$	$0 < v_{\text{blue}} \le \overline{v}_{\text{blue}}$
Dynamic charging prices	$\rho_{mn}^{\text{ch}} \sim N(0.5c, (0.15c)^2)$	$0.3 \le c \le 0.7$
Initial SOC of the EV	$SoC^{\text{ini}} \sim U(0.4, 0.6)$	$0.4 \le SoC^{\text{ini}} \le 0.6$

In RL-based methods, the deep neural network consists of three fully connected layers of 128 units with the ReLU activation function. We set the discount factor $\gamma = 0.99$, the learning rate $\eta = 0.0003$, the batch size $|B| = 64$, and other hyperparameters as in [41]. All experiments are implemented in Python 3.9 and executed on a workstation configured with an Intel Core i9-13900K CPU and an NVIDIA GeForce RTX 4090 GPU. The deep learning models are developed using the PyTorch library, ensuring compatibility with CUDA acceleration for efficient training and inference.

4.2 Benchmark Case Studies

In this section, we benchmark the performance of various methods in solving the DWC navigation problem of EVs. The notation and explanation of each method are as follows.

(1) SP-PPO: the method proposed in this paper;
(2) SP-DQN: the SP-based deep Q-learning (DQN) method;
(3) SP-A2C: the SP-based advantage actor-critic (A2C) method;
(4) Regular PPO: the PPO method adopting high-dimensional raw data as network inputs;
(5) Regular DQN: the DQN method adopting high-dimensional raw data as network inputs;
(6) SP-RHO: the SP-based rolling horizon optimization (RHO), which solves the two-layer optimization problem in an online event-triggered manner but only takes the computed action for the current timestep.

4.2.1 Training Process

Given $q^t = 25.3$ CNY/hour, the learning curves of the first 40,000 episodes of the RL-based methods are shown in Fig. 4, in which each method converges to its optimal value. Note that across different q^t configurations, the learning curves display consistent patterns.

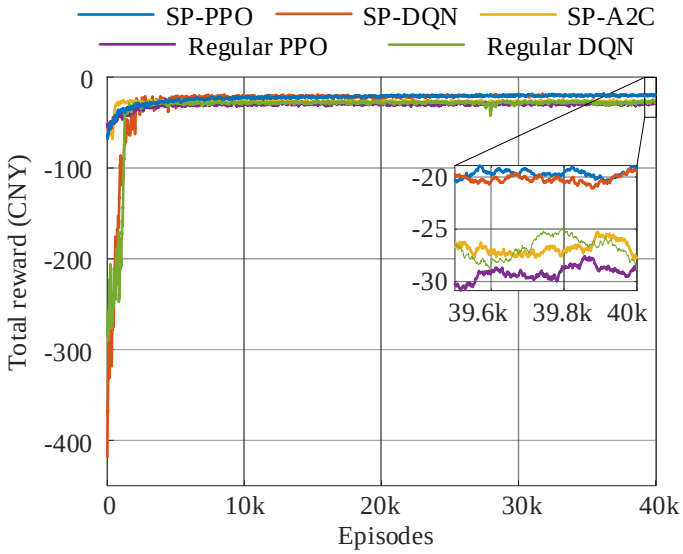

Figure. 4: Smoothed learning curves of the first 40,000 episodes of the RL-based methods

In Fig. 4, the SP-PPO and SP-DQN methods converge to better solutions than their regular versions, demonstrating the advantage of integrating SP-based solutions into RL algorithms. For the three SP-based RL methods, SP-PPO and SP-DQN converge to almost the same point and outperform SP-A2C. However, if we look at the entire learning curve, the training process of SP-PPO is quite stable while SP-DQN starts learning from a very poor policy.

4.2.2 **Testing Results**

The larger the value of q^t, the greater the urgency of time the EV owner has. After training under different q^t settings, the performance of the various methods on the testing set is compared in Fig. 5.

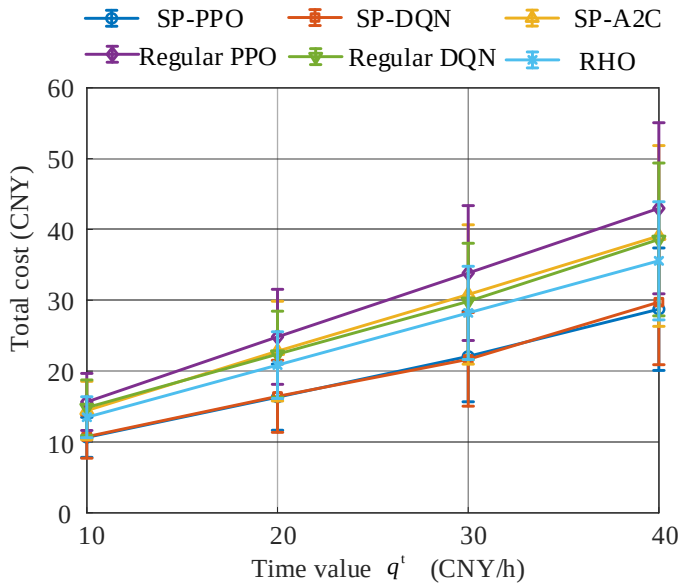

Figure. 5: Total cost vs. time value for different methods

Compared to the regular PPO method, SP-PPO reduces the mean and standard deviation of cost by approximately 33% and 25%, respectively; similarly, SP-DQN reduces the mean and standard deviation of cost by about 26% and 20%, respectively, compared to the regular DQN method. The performance gap between regular RL methods and SP-based RL methods demonstrates the advantage of adopting SP-based solutions to

extract effective features and reduce the dimensionality of the state space for RL algorithms. Compared to the SP-RHO method, SP-PPO exhibits a 21% reduction in the mean value of cost, with the standard deviation remaining largely unchanged. The advantage of employing PPO to learn the dynamics of system uncertainties is evident in the significant performance gap observed. For the three SP-based RL methods, the proposed SP-PPO has almost the same mean reward as SP-DQN, but the standard deviation of reward is slightly smaller, indicating that it is more stable and reliable. SP-A2C performs relatively poorly and may not be suitable for this numerical setting.

4.3 Adaptability to Dynamic Charging Prices

In Fig. 6, we change the mean of charging price distribution at each charging lane every 200,000 timesteps. In Stage 1, SP-PPO starts training with randomly initialized parameters, and after almost the entire period of training, the total reward finally rises to about −19 CNY. At the beginning of Stage 2, the total reward sharply drops to −22.5 CNY with a sudden charging price change and converges to −21 CNY at a much faster speed than in Stage 1. Similar patterns appear in Stage 3, verifying that the proposed SP-PPO is robust enough to adapt to real-time charging price changes after pre-training with initial parameters.

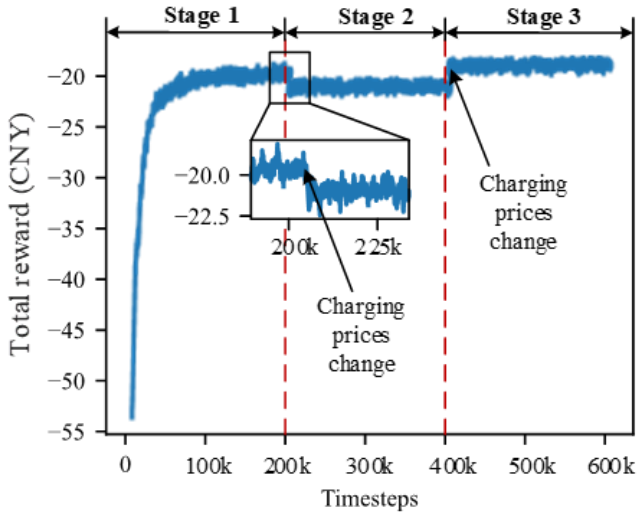

Figure. 6: Learning curve of SP-PPO method under dynamic charging prices

4.4 Real-World Validation on Large-Scale Network

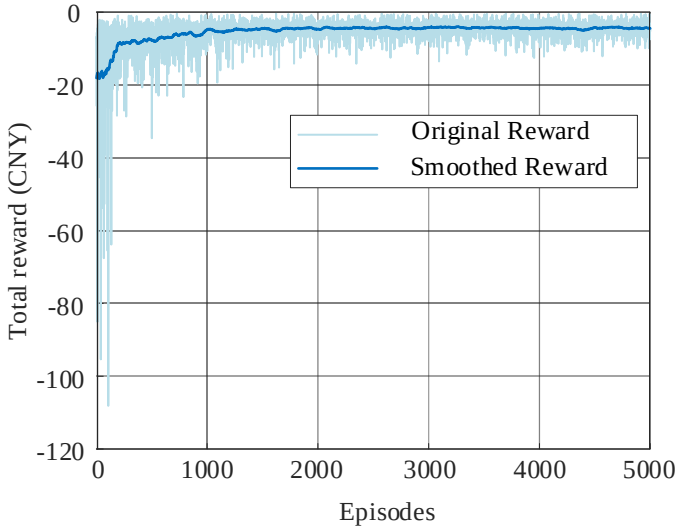

Figure. 7: Learning curve of SP-PPO method using real-world data

To examine the real-world effectiveness of the proposed method, a simulation environment is built using actual road layouts and EV-related data. A central urban region in Beijing, consisting of 506 intersections and 1,348

roads, is selected as the ETN under study [42]. The area is assumed to include 19 charging lanes, with charging prices aligned with charging piles based on time-of-use rates [43]. Real-time traffic data [44] for five weekdays (Dec 7, 8, 10, 11, and 14, 2015) are used, where the first four days serve as training data and the last day for testing.

As shown in Fig. 7, SP-PPO converges within 5,000 episodes, achieving a reward of −4.25 CNY; on a test set of 10,000 episodes, the average reward and standard deviation are −4.30 CNY and 1.87 CNY, respectively. The training process and testing results are both quite satisfactory, which verifies the scalability of the proposed SP-PPO on large-scale networks.

5 Conclusions

This paper studies the DWC navigation problem under unknown uncertainties for individual EVs in UETNs deployed with charging lanes. A two-layer dynamic charging routing model is formulated to minimize the travel and charging costs for EVs. Then, an SP-based method is proposed to solve it efficiently. To overcome the curse of dimensionality, low-dimensional solutions containing stochastic traffic and electricity information are incorporated as part of the system state. Finally, an event-triggered MDP is constructed, and PPO is utilized to learn the system dynamics and generate the optimal DWC navigation strategy. Numerical results demonstrate the effectiveness of the proposed method. Compared with directly using raw data as RL input and the traditional rolling-horizon optimization algorithm, the proposed method reduces costs by 32−35% and 20−22%, respectively. Its adaptability to dynamic charging prices and scalability on large-scale networks are also analyzed, showing the potential for practical operation in the real world.

Acknowledgments

This work was supported in part by a grant under Project 17206222 and a grant under Project T23-701/20-R from the Hong Kong Research Grants Council, Hong Kong Special Administrative Region, China.

References

[1] Z. Zhang, K. T. Chau, C. Liu, F. Li, and T. W. Ching, *Quantitative analysis of mutual inductance for optimal wireless power transfer via magnetic resonant coupling*, IEEE Transactions on Magnetics, ISSN 1941-0069, 50(11), 1-4.

[2] W. Liu, K. T. Chau, C. H. T. Lee, W. Han, X. Tian, and W. H. Lam, *Full-range soft-switching pulse frequency modulated wireless power transfer,* IEEE Transactions on Power Electronics, ISSN 1941-0107, 35(6), 6533-6547.

[3] Y. Zhang, X. Li, S. Chen, and Y. Tang, *Soft switching for strongly coupled wireless power transfer system with 90° dual-side phase shift,* IEEE Transactions on Industrial Electronics, ISSN 1557-9948, 69(1), 282-292.

[4] W. Liu, K. T. Chau, C. H. T. Lee, C. Jiang, W. Han, and W. H. Lam, *Multi-frequency multi-power one-to-many wireless power transfer system*, IEEE Transactions on Magnetics, ISSN 1941-0069, 55(7), 1-9.

[5] C. Jiang, K. T. Chau, W. Liu, C. Liu, W. Han, and W. H. Lam, *An LCC-compensated multiple-frequency wireless motor system*, IEEE Transactions on Industrial Informatics, ISSN 1941-0050, 15(11), 6023-6034.

[6] W. Han, K. T. Chau, Z. Zhang, and C. Jiang, *Single-source multiple-coil homogeneous induction heating*, IEEE Transactions on Magnetics, ISSN 1941-0069, 53(11), 1-6.

[7] H. Zhang, Z. Li, and C. K. Lee, *Transmitter adaptation and wireless power control for capsule endoscopy*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 39(4), 4884-4894.

[8] C. Jiang, K. T. Chau, C. H. T. Lee, W. Han, W. Liu, and W. H. Lam, *A wireless servo motor drive with bidirectional motion capability*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 34(12), 12001-12010.

[9] S. Li, K. T. Chau, W. Liu, J. Guo, C. Liu, and Y. Hou, *A wireless permanent-magnet brushless dc motor using contactless feedback and autonomous commutation*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 40(7), 10140-10153.

[10] C. Jiang, K. T. Chau, Y. Y. Leung, C. Liu, C. H. T. Lee, and W. Han, *Design and analysis of wireless ballastless fluorescent lighting*, IEEE Transactions on Industrial Electronics, ISSN 1557-9948, 66(5), 4065-4074.

[11] W. Ejaz, M. Naeem, and S. Zeadally, *On-demand sensing and wireless power transfer for self-sustainable industrial Internet of Things networks*, IEEE Transactions on Industrial Informatics, ISSN 1551-3203, 17(10), 7075-7084.

[12] H. Pang, F. Xu, W. Liu, C. K. Tse, and K. T. Chau, *Impedance buffer-based reactance cancellation method for CLC-S compensated wireless power transfer*, IEEE Transactions on Industrial Electronics, ISSN 1557-9948, 71(7), 6894-6906.

[13] Z. Yan, B. Song, Y. Zhang, K. Zhang, Z. Mao, and Y. Hu, *A rotation-free wireless power transfer system with stable output power and efficiency for autonomous underwater vehicles*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 34(5), 4005-4008.

[14] W. Liu, K. T. Chau, H. Wang, and T. Yang, *Long-range wireless power drive using magnetic extender*, IEEE Transactions on Transportation Electrification, ISSN 2332-7782, 9(1), 1897-1909.

[15] W. Liu, K. T. Chau, X. Tian, H. Wang, and Z. Hua, *Smart wireless power transfer — opportunities and challenges*, Renewable and Sustainable Energy Reviews, ISSN 1364-0321, 180(2023), 113298.

[16] C. Qiu, K. T. Chau, T. W. Ching, and C. Liu, *Overview of wireless charging technologies for electric vehicles*, Journal of Asian Electric Vehicles, ISSN 1348-3927, 12(1), 1679-1685.

[17] A. Ahmad, M. S. Alam, and R. Chabaan, *A comprehensive review of wireless charging technologies for electric vehicles*, IEEE Transactions on Transportation Electrification, ISSN 2332-7782, 4(1), 38-63.

[18] Z. Xue, W. Liu, C. Liu, and K. T. Chau, *Critical review of wireless charging technologies for electric vehicles*, World Electric Vehicle Journal, ISSN 2032-6653, 16(2), 65.

[19] J. Guo, K. T. Chau, W. Liu, Y. Hou, and W. L. Chan, *Pulse magnitude modulation and token rotation-based voltage balancing method of multilevel inverter for wireless power transfer*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 40(3), 4652-4663.

[20] W. Liu, K. T. Chau, C. H. T. Lee, C. Jiang, W. Han, and W. H. Lam, *Wireless energy-on-demand using magnetic quasi-resonant coupling*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 35(9), 9057-9069.

[21] A. Babaki, S. V. Zadeh, A. Zakerian, and G. A. Covic, *Variable-frequency retuned WPT system for power transfer and efficiency improvement in dynamic EV charging with fixed voltage characteristic*, IEEE Transactions on Energy Conversion, ISSN 1558-0059, 36(3), 2141-2151.

[22] H. Wang, N. Tashakor, W. Jiang, W. Liu, C. Q. Jiang, and S. M. Goetz, *Hacking encrypted frequency-varying wireless power: Cyber-security of dynamic charging*, IEEE Transactions on Energy Conversion, ISSN 1558-0059, 39(3), 1947-1957.

[23] W. Han, K. T. Chau, C. Jiang, W. Liu, and W. H. Lam, *Design and analysis of quasi-omnidirectional dynamic wireless power transfer for fly-and-charge*, IEEE Transactions on Magnetics, ISSN 1941-0069, 55(7), 1-9.

[24] Z. Hua, K. T. Chau, H. Pang, and T. Yang, *Dynamic wireless charging for electric vehicles with autonomous frequency control*, IEEE Transactions on Magnetics, ISSN 1941-0069, 59(11), 1-5.

[25] G. R. Nagendra, G. A. Covic, and J. T. Boys, *Sizing of inductive power pads for dynamic charging of EVs on IPT highways*, IEEE Transactions on Transportation Electrification, ISSN 2332-7782, 3(2), 405-417.

[26] Z. Zhang and K. T. Chau, *Homogeneous wireless power transfer for move-and-charge*, IEEE Transactions on Power Electronics, ISSN 1941-0107, 30(11), 6213-6220.

[27] W. Liu, K. T. Chau, C. H. T. Lee, C. Jiang, and W. Han, *A switched-capacitorless energy-encrypted transmitter for roadway-charging electric vehicles*, IEEE Transactions on Magnetics, ISSN 1941-0069, 54(11), 1-6.

[28] K. T. Chau, *Pure electric vehicles*. In *Alternative fuels and advanced vehicle technologies for improved environmental performance – towards zero carbon transportation*, (Ed.) R. Folkson, ISBN 978-0-8570-9522-0, Woodhead Publishing, 2014, 655-684.

[29] Q. Guo, S. Xin, H. Sun, Z. Li, and B. Zhang, *Rapid-charging navigation of electric vehicles based on real-*

time power systems and traffic data, IEEE Transactions on Smart Grid, ISSN 1949-3061, 5(4), 1969-1979.

[30] H. Yang, Y. Deng, J. Qiu, M. Li, M. Lai, and Z. Y. Dong, *Electric vehicle route selection and charging navigation strategy based on crowd sensing*, IEEE Transactions on Industrial Informatics, ISSN 1941-0050, 13(5), 2214-2226.

[31] X. Tang, S. Bi, and Y. J. A. Zhang, *Distributed routing and charging scheduling optimization for Internet of Electric Vehicles*, IEEE Internet of Things Journal, ISSN 2327-4662, 6(1), 136-148.

[32] C. Yao, S. Chen, and Z. Yang, *Joint routing and charging problem of multiple electric vehicles: A fast optimization algorithm*, IEEE Transactions on Intelligent Transportation Systems, ISSN 1524-9050, 23(7), 8184-8193.

[33] Z. Tan, F. Liu, H. K. Chan, and H. O. Gao, *Transportation systems management considering dynamic wireless charging electric vehicles: Review and prospects*, Transportation Research Part E: Logistics and Transportation Review, ISSN 1366-5545, 163(2022), 102761.

[34] C. Liu, M. Zhou, J. Wu, C. Long, and Y. Wang, *Electric vehicles en-route charging navigation systems: Joint charging and routing optimization*, IEEE Transactions on Control Systems Technology, ISSN 1558-0865, 27(2), 906-914.

[35] X. Shi, Y. Xu, Q. Guo, H. Sun, and W. Gu, *A distributed EV navigation strategy considering the interaction between power system and traffic network*, IEEE Transactions on Smart Grid, ISSN 1949-3061, 11(4), 3545-3557.

[36] D. Qiu, Y. Wang, W. Hua, and G. Strbac, *Reinforcement learning for electric vehicle applications in power systems: A critical review*, Renewable and Sustainable Energy Reviews, ISSN 1364-0321, 173(2023), 113052.

[37] J. Jin and Y. Xu, *Shortest-path-based deep reinforcement learning for ev charging routing under stochastic traffic condition and electricity prices*, IEEE Internet of Things Journal, ISSN 2327-4662, 9(22), 22571-22581.

[38] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, ISBN 0-26203-924-9, Cambridge, MIT Press, 2020.

[39] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, *High-dimensional continuous control using generalized advantage estimation*, arXiv preprint arXiv:1506.02438, 2015.

[40] P. Machura, V. D. Santis, and Q. Li, *Driving range of electric vehicles charged by wireless power transfer*, IEEE Transactions on Vehicular Technology, ISSN 1939-9359, 69(6), 5968-5982.

[41] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, *Stable-baselines3: Reliable reinforcement learning implementations*, The Journal of Machine Learning Research, ISSN 1532-4435, 22(1), 12348-12355.

[42] X. Zhan, S. V. Ukkusuri, and P. S. C. Rao, *Dynamics of functional failures and recovery in complex road networks*, Physical Review E, ISSN 2470-0053 96(5), 052301.

[43] Charge it, https://www.bjev520.com, accessed on 2025-04-11.

[44] Network state data, https://github.com/zhanzxy5/VSR-dataset, accessed on 2025-04-11.

Presenter Biography

Chaoran Si received the B.Eng. and B.M. degrees from Tianjin University, Tianjin, China, in 2018, and the M.Eng. degree from Zhejiang University, Hangzhou, China, in 2021. She is currently pursuing a Ph.D. degree at the Department of Electrical and Electronic Engineering at the University of Hong Kong. Her research interests include power-transportation system, wireless charging of electric vehicles, and deep reinforcement learning.