applied sciences

MDPI

Article

Optimal Placement of Charging Stations in Road Networks: A Reinforcement Learning Approach with Attention Mechanism

Jiaqi Liu, Jian Sun and Xiao Qi *

Key Laboratory of Road and Traffic Engineering, Ministry of Education, College of Transportation Engineering, Tongji University, Shanghai 201804, China
* Correspondence: qixiao@tongji.edu.cn

Abstract: With the aim of promoting energy conservation and emission reduction, electric vehicles (EVs) have gained significant attention as a strategic industry in many countries. However, the insufficiency of accessible charging infrastructure remains a challenge, hindering the widespread adoption of EVs. To address this issue, we propose a novel approach to optimize the placement of charging stations within a road network, known as the charging station location problem (CSLP). Our method considers multiple factors, including fairness in charging station distribution, benefits associated with their placement, and drivers' discomfort. Fairness is quantified by the balance in charging station coverage across the network, while driver comfort is measured by the total time spent during the charging process. Then, the CSLP is formulated as a reinforcement learning problem, and we introduce a novel model called PPO-Attention. This model incorporates an attention layer into the Proximal Policy Optimization (PPO) algorithm, enhancing the algorithm's capacity to identify and understand the intricate interdependencies between different nodes in the network. We have conducted extensive tests on urban road networks in Europe, North America, and Asia. The results demonstrate the superior performance of our approach compared to existing baseline algorithms. On average, our method achieves a profit increase of 258.04% and reduces waiting time by 73.40%, travel time by 18.46%, and charging time by 40.10%.

Keywords: location selection; reinforcement learning; attention mechanism; proximal policy optimization

Citation: Liu, J.; Sun, J.; Qi, X. Optimal Placement of Charging Stations in Road Networks: A Reinforcement Learning Approach with Attention Mechanism. *Appl. Sci.* **2023**, *13*, 8473. https://doi.org/10.3390/app13148473

Academic Editor: Andreas Sumper

Received: 4 July 2023
Revised: 18 July 2023
Accepted: 20 July 2023
Published: 22 July 2023

1. Introduction

In many countries, electric vehicles (EVs) have emerged as a vital strategic industry to promote energy conservation and emission reduction [1–3]. To enable the widespread adoption of electric vehicles and alleviate range anxiety (i.e., concerns about insufficient energy storage for a trip), the provision of accessible charging infrastructure is crucial [3–5]. While the number of charging stations has increased, it remains insufficient to meet the growing charging demands. Optimizing the location and deployment of charging stations, known as the charging station location problem (CSLP) [6], is a critical approach to address this challenge. The CSLP is an application of the facility location problem, aiming to select suitable locations from a candidate set to optimize various objectives. Numerous studies have focused on the CSLP as a long-standing problem [7].

However, the optimal placement of charging stations in road networks is not without its challenges. Influenced by a myriad of factors, including road network topology, existing charging infrastructure, traffic patterns, and charging duration, determining appropriate placements is complex [6]. Additionally, existing methods often neglect practical constraints, such as the need for fairness in charging station distribution and the requirement for automated decision making in identifying viable new charging station locations. Traditional regional optimization methods, which predominantly consider isolated junctions or parking lots, are insufficient in the face of widespread charging demands across the road network [7].

Moreover, prevalent greedy algorithms fail to account for the intricate spatiotemporal relationships between the road network and charging demand [8].

In our research, we address these challenges head-on. Our approach emphasizes coverage and fairness in the objective function for determining charging station locations, while also considering the benefits of station placement and driver comfort. We assess the benefits of charging station placement by considering the coverage of charging stations and network nodes, while driver comfort is quantified by the total time spent by drivers during the charging process. Fairness is achieved by balancing the average coverage of charging stations across the network. We reformulate the CSLP as a reinforcement learning problem, introducing a novel algorithm, PPO-Attention, that extends the Proximal Policy Optimization algorithm [9] by integrating a policy network with a multi-head attention layer.

In summary, our contributions are as follows:

- We propose an optimal model for the placement of charging infrastructure that balances coverage and fairness of charging station locations while considering driver comfort.
- We formulate the CS placement problem as a reinforcement learning problem and introduce a novel reinforcement learning model, PPO-Attention. This model enhances the policy network of the Proximal Policy Optimization algorithm by incorporating an attention layer with two attention heads.
- We collect data from multiple cities and regions worldwide and evaluate the performance of our algorithm using these datasets. The results demonstrate the effectiveness and efficiency of our method, surpassing existing baseline algorithms.

2. Related Works

2.1. Charging Station Location Problem

In recent years, the charging station location problem (CSLP) has garnered considerable attention from researchers, leading to a multitude of studies exploring different aspects and perspectives [4,6,10–22]. Notable contributions include a deployment framework proposed by Zhao et al. [16], which considers existing competitors in the planning of PEV fast-charging stations. Xie et al. [18] introduced a two-stage data-driven method for determining CS station locations on highways.

Researchers have also focused on optimizing the objectives in CSLP. Liu et al. [14] adopted a model to minimize drivers' discomfort, while Liu et al. [15] designed a model to minimize CO_2 emissions. Other considerations, such as brand preferences [20] and the total number of charging stations [23], have also been taken into account. However, the existing studies primarily focus on optimizing the location of charging stations, while overlooking the number of chargers at each station and the fairness of their distribution among users.

Various algorithms have been applied to address CSLP. Choi et al. [21] proposed a large-scale charging station concept solved using the K-means algorithm. Genetic algorithms [13,22], Bayesian optimization [20], and greedy algorithms [5] have also been utilized for determining optimal charging station locations. Given that CSLP is an NP-hard problem, many approximation algorithms have been employed. However, as problem size and constraints increase, computational efficiency becomes a significant limitation for these methods. Von Bahr et al. [6] presented a reinforcement learning approach to address CSLP, demonstrating its potential advantages in scalability and computational efficiency.

However, in the implemented studies, the fairness issue in charging station location planning has been scarcely addressed. Furthermore, reinforcement learning, as a promising approach for solving optimization problems, still has ample room for exploration in the context of CSLP. This paper aims to consider elements overlooked in previous research and explore the potential for problem solving using advanced reinforcement learning algorithms.

2.2. Reinforcement Learning and Attention Mechanism

Reinforcement learning (RL) is a domain of artificial intelligence concentrating on how an intelligent agent should behave in an environment to maximize cumulative reward. In recent years, a substantial amount of research has been dedicated to exploiting RL to tackle complex optimization problems [24,25]. The intrinsic flexibility of RL, its capacity to handle high-dimensional and continuous spaces, and its proficiency in balancing exploration and exploitation render it a unique advantage in solving optimization challenges [26–28]. RL algorithms, such as Q-Learning, Deep Q-Network (DQN), and Proximal Policy Optimization (PPO) [9], have found applications in logistics, supply chain management, resource allocation, and other optimization-intensive areas [24,26,29].

The attention mechanism, an essential cognitive function in humans, has recently been integrated into computer vision research.

In the machine learning community, attention mechanism has emerged as a powerful technique in neural network models, particularly in sequence modeling [30]. This architecture enables neural networks to discover interdependencies and correlations within variable numbers of inputs.

Consequently, the attention mechanism has become a common component of neural architectures and finds applications in various tasks, such as image caption generation, text classification, machine translation, action recognition, image-based analysis, speech recognition, and recommendation systems [31]. Apart from performance improvements, attention mechanism also provides interpretability, addressing the lack of interpretability faced by deep learning, which has practical and ethical implications. While the extent to which attention mechanism can reliably explain deep networks remains a subject of debate, it offers intuitive explanations to some degree [30].

In recent years, the attention mechanism has been successfully applied to reinforcement learning [32,33], yielding promising results. However, whether this approach can be effectively employed in optimization problems remains an open question and a subject worthy of further exploration and discussion. This paper aims to investigate the performance of a reinforcement learning method combined with an attention mechanism in the context of optimizing the layout of charging stations.

3. Preliminaries

This section introduces some basic concepts for the problem, including three parts: charging station, reinforcement learning algorithm, and attention mechanism.

3.1. The Charging Station Location Problem

The distribution of electric vehicle (EV) charging stations on real-world networks necessitates the introduction of a definition for the road network. In this context, we consider a directed graph as the representation of the city's road network. Let $G = (V, E)$ represent the entire road network, where V denotes the set of vertices and E represents the set of edges. The vertices and edges correspond to Points of Interest (POIs), such as road junctions or charging stations, and the roads that connect these nodes in V, respectively. The coordinates of a vertex v are denoted by $\tau(v)$.

This study takes into account the scenario where multiple charger plugs are available in a single charging station. An EV charging station is denoted as s, and its location is specified as $v(s)$. The status of chargers with different types in a station is recorded as $t(s) = (t_1, t_2, \ldots, t_m)$, where m represents the number of charger plug types and t_i denotes the number of chargers of the i-th type. The set of all EV charging stations is denoted as S.

A charging plan is defined as $p = \{s_1, s_2, \ldots, s_n\}$, which is a combination of selected charging stations. The collection of all possible charging plans is denoted as $\mathcal{P}$. The objective of this study is to find the optimal plan p^* from $\mathcal{P}$ that maximizes benefits or minimizes costs in the real-world context.

3.2. The Markov Decision Process and Reinforcement Learning

Reinforcement learning is a general framework for sequential decision making under uncertainty. The reinforcement learning problem is often represented by a Markov Decision Process (MDP). A standard MDP is defined as follows:

- State space $\mathcal{S}$;
- Action space $\mathcal{A}$;
- Initial observation distribution $\rho : \mathcal{S} \to \mathbb{R}$;
- Transition distribution $p : \rho : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$;
- Reward function $\imath : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$.

The agent makes decisions and takes actions according to the policy π and the current observation. The goal of the agent is to find the optimal policy π^* maximizing expected γ-discounted cumulative reward, called the value function V^π. Thus, we have:

$$V^\pi(s) \stackrel{\text{def}}{=\!=} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_t \sim \pi(a_t|s_t), s_{t+1} \sim P(s_{t+1}, a_t)] \tag{1}$$

$$Q^\pi(s, a) \stackrel{\text{def}}{=\!=} R(s, a) + \gamma E_{s' \sim P(s'|s,a)} V^\pi(s') \tag{2}$$

The optimal action-value function $Q^* = \max_\pi Q^\pi(s)$ satisfies the Bellman Optimality Equation:

$$Q^*(s, a) \stackrel{\text{def}}{=\!=} E_{s' \sim P(s'|s,a)} \max_{a' \in A} [R(s, a) + \gamma Q^*(s', a')] \tag{3}$$

In our experimental investigations, we adopt Proximal Policy Optimization (PPO) [9], a policy gradient method for reinforcement learning. PPO demonstrates superior efficiency and reliability compared to Trust Region Policy Optimization (TRPO) due to its utilization of first-order optimization techniques exclusively. Within the PPO framework, two primary variants exist: PPO-Penalty and PPO-Clip. PPO-Penalty approximates a KL-constrained update, akin to TRPO, but introduces a penalty term in the objective function to address the KL-divergence without imposing a hard constraint. Additionally, the penalty coefficient is automatically adjusted during training to ensure proper scaling. In contrast, PPO-Clip takes a different approach by omitting the KL-divergence term from the objective function and dispensing with explicit constraints. Instead, it relies on specialized clipping methods within the objective function to discourage substantial deviations between the new and old policies.

For the current project, we employ PPO-Clip, as it offers convenience and ease of implementation, making it a more feasible choice compared to PPO-Penalty. The central concept behind PPO-Clip is the clipping surrogate objective, defined as follows:

$$L^{PPO}(\theta) = E_t[min(r_t(\theta)A_t, clip(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t)] \tag{4}$$

where ϵ is a hyperparameter. The first term inside the min is L^{PPO}. The second term, $clip(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t$, modifies the surrogate objective by clipping the probability ratio, which removes the incentive for moving it outside of the interval $[1 - \epsilon, 1 + \epsilon]$.

The PPO-Clip is described in Algorithm 1.

Algorithm 1: PPO-Clip

Input: Initial policy parameters θ_0, initial value function parameters ϕ_i

for k=0,1,2,$\cdots$ **do**

Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment. Compute rewards-to-go $\hat{R}_t$

Compute advantage estimates, $\hat{A}_t$ (using any method of advantage estimation) based on the current value function V_{ϕ_k}

Update the policy by maximizing the PPO-Clip objective: $\theta_{k+1} = \arg\max_\theta \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi\theta_k}(a_t, s_t), g\left(\epsilon, A^{\pi\theta_k}(a_t, s_t)\right)\right)$, typically via stochastic gradient ascent with Adam.

Fit value function by regression on mean-squared error: $\phi_{k+1} = \arg\min_\phi \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \min\left(V_\phi(s_t) - \hat{R}_t\right)^2$, typically via some gradient descent algorithm

end

3.3. The Attention Mechanism

The attention mechanism, drawing inspiration from the selective concentration observed in human cognition [30], functions as a weighted message-passing algorithm among nodes in a graph. It has demonstrated impressive performance in the domain of Natural Language Processing (NLP) and Transformer Models. Figure 1 illustrates the fundamental principle of the attention mechanism. It involves the mapping of a query q and a collection of key-value (k-v) pairs, which results in an output obtained through a weighted summation of the values. These weights not only represent the similarity between the query and the corresponding key, but also serve as the weights assigned to the message value v received by a node from its neighboring nodes.

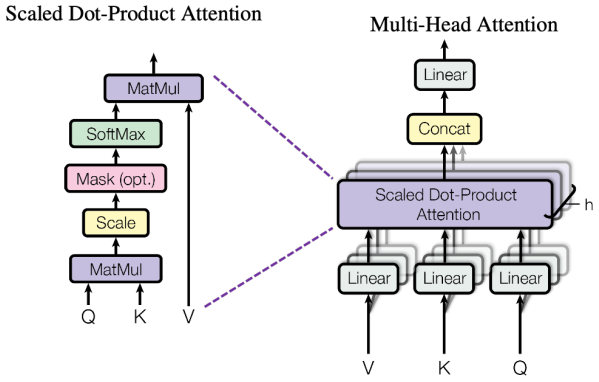


Figure 1. Schematic diagram of the multi-head attention mechanism.

To be more specific, we define the query $q_i \in \mathbb{R}^{d_k}$, key $k_i \in \mathbb{R}^{d_k}$, and value $v_i \in \mathbb{R}^{d_v}$, where d_k denotes the dimensionality of the query q_i and key k_i, and d_v represents the dimensionality of the value v_i. These values are computed by projecting the embedding h_i using the following equations:

$$q_i = W^Q h_i \tag{5}$$

$$k_i = W^K h_i \tag{6}$$

$$v_i = W^V h_i \tag{7}$$

where W^Q, W^K, and W^V are matrices of dimensions $d_k \times d_h$, $d_k \times d_h$, and $d_v \times d_h$, respectively.

Subsequently, the compatibility $u_{ij} \in \mathbb{R}$ is calculated through dot product [30] as follows:

$$u_{ij} = \begin{cases} \frac{q_i^T k_j}{\sqrt{d_k}} & \text{if } i \text{ is adjacent to } j \\ -\infty, & \text{otherwise} \end{cases} \tag{8}$$

The attention weights a_{ij} are obtained by applying a softmax function:

$$a_{ij} = \frac{e^{u_{ij}}}{\sum_j' e^{u_{ij'}}} \tag{9}$$

Consequently, the message node i receives, denoted as h_i', is the convex combination of the message values v_j:

$$h_j' = \sum_j s_{ij} v_j \tag{10}$$

Multi-head Attention: In scenarios involving large and complex datasets, multi-head attention has been observed to deliver improved performance [30]. This approach allows each node in the graph to receive different types of messages from different neighbors. Let M denote the number of attention heads, and h_{im}' denote the resultant vectors for $m \in 1, \ldots, M$. These vectors are then projected back to a single d_h-dimensional vector using parameter matrices W_m^O of dimensions $d_h \times d_v$. Therefore, the final multi-head attention value for node i is a function of $h_1, \ldots, h_n$ through h_{im}':

$$MHA_i(h_1, \ldots, h_m) = \sum_{m=1}^{M} W_m^O h_{im}' \tag{11}$$

4. Problem Definition and Modeling

This section presents a detailed introduction to the programming model of CSLP, encompassing the objective function and constraints.

The problem of charging station location can be framed as an optimization problem, where the objective function $gain(p)$ is defined to identify the optimal plan p^* that best fulfills our expectations. The objective function $gain(p)$ in our study comprises three components: the profit term $profit(p)$, the cost term $cost(p)$, and the fairness term $fairness(p)$. Each term within the objective function will be discussed in detail below.

4.1. Profit Function

For a station s with m charger types, let c_i denote the available charging power of the charger t_i, and $C(s)$ is the whole capacity of the charging station s. $C(s)$ is computed as: $C(s) = \sum_{i=1}^{m} t_i c_i$.

Intuitively, a charging station with a larger capacity should serve more electric vehicles and have a wider range of influence. Hence, the influential radius $r(s)$ is defined to indicate the distance within which the charging station attracts electric vehicles. The decay of influence with increasing distance is described using a Gaussian function:

$$r(s) = r_{max} e^{-C(s)/2} \tag{12}$$

where r_max is the maximal influential radius of the charging station s.

The service scope for CS s $scope(s)$ and the coverage for the vertex v $cov(v)$ are defined as follows:

$$scope(s) = |\{vs. \in V | d(v,s) \leq r(s)\}| \tag{13}$$

$$cov(v) = |\{s \in S | d(v,s) \leq r(s)\}| \tag{14}$$

where $d(v,s)$ represents the Euclidean distance between charging station s and vertex v. The value of $scope(s)$ for charging station s corresponds to the number of vertices within

the scope of station s. Similarly, the coverage $cov(v)$ for vertex v indicates the number of charging stations within the influential radius of vertex v.

When the number of charging stations is not infinite, we anticipate that the limited number of charging stations can provide services to a greater number of electric vehicles. Additionally, for a single vertex v, having more choices is considered advantageous. Consequently, the profit function is defined as:

$$profit(p) = \Big(\frac{1}{|V|}\sum_{v\in V} cov(v) + \frac{1}{|S|}\sum_{s\in S} scope(s)\Big) \tag{15}$$

4.2. Cost Function

In our study, we define the cost of a charging station plan, denoted as p, in terms of the overall time expended during the charging process for drivers. This comprises three key components: travel time (t_1), charging time (t_2), and waiting time (t_3). Travel time encapsulates the cumulative duration required for all electric vehicles within the road network to reach a charging station. Charging time, on the other hand, refers to the aggregate time spent by all electric vehicles during the charging process itself. Finally, waiting time represents the collective duration that electric vehicles are queued for charging. We employ the sum total of these times as a proxy for drivers' comfort in our analysis. The implication here is that a reduction in total time correlates with an enhancement in driver comfort.

Let $demand(v)$ represent the charging demand of the junction v; then, which charging station the vehicles at the junction v are heading to need to be determined. Thus, the function of attraction for CS s and junction v $f_{att}(v,s)$ is defined as follows:

$$f_{att}(v,s) = \frac{w_1}{dis(v,s)} + w_2 C(s) \tag{16}$$

where w_1 and w_2 are the weight coefficients.

For a given junction v and all the possible CS s, the larger $f_{att}(v,s)$ is, the more likely the EVs in the junction v going to s. Furthermore, we use the softmax function to calculate the possibility of going one CS for junction v:

$$att(v,s_k) = \frac{e^{f_{att}(v,s_k)}}{\sum_{s\prime\in S} e^{f_{att}(v,s\prime)}} \tag{17}$$

$\sigma(v,s)$ is used to record whether EVs at junction v are heading to s:

$$\sigma(v,s) = \begin{cases} 1 & if d(v,s) \le r(s) \text{ and } s = \arg_{s'\in S}\max att(v,s') \\ 0, & otherwise \end{cases} \tag{18}$$

Then, for the junction v, the total travel time of all EVs t_1 induced by a plan p on the road network is:

$$t_1 = \sum_{v\in V}\sum_{s\in p} \frac{\sigma(v,s) demand(v) dis(v,s)}{\bar{V}} \tag{19}$$

where $\bar{V}$ is a constant, representing the average speed the EV drive on the road.

Then, the charging time t_2 is modeled and calculated. The total capacity of CSs is $C(S)$, and the expected charging time of all CSs is $\frac{1}{C(S)}$. Thus, the charging time for all EVs is:

$$t_2 = \sum_{s\in p}\sum_{v\in V} \frac{\sigma(v,s) demand(v)}{C(s)} \tag{20}$$

For waiting time t_3, it is first modeled as an $M/D/1$ queue [34], where M denotes the coming event of EV follows a Poisson process, D denotes the service time is a deterministic

function, and "1" means only one queue for a station. Then, the expected waiting time t_3 can be calculated using the Pollaczek–Khintchine formula [34] as follows:

$$t_3 = \sum_{s \in p} W(s) D(s) \tag{21}$$

where $W(s) = \frac{\rho(s)}{2\mu(s)(1-\rho(s))}$,$\rho(s) = \frac{D(s)}{\mu(s)} < 1$, and $D(s) = \sigma(v,s) * demand(v)$.

Finally, the total time cost is calculated as follows:

$$cost(p) = t_1 + t_2 + t_3 \tag{22}$$

4.3. Fairness Function

In this subsection, we design a fairness function. We believe that in any given area of the urban road network, the average service that each electric vehicle (EV) receives should be approximately balanced. Therefore, we define the average number of charging stations matched at each intersection as the evaluation benchmark, and employ the Mean Square Error (MSE) to quantify fairness. The average number of charging stations matched at a single vertex is represented as:

$$scope_{ave}(V) = \frac{1}{|V|} \sum_{v \in V} scope(v) \tag{23}$$

Furthermore, the fairness of the plan p is measured as follows:

$$fairness(p) = \frac{1}{|V|} \sum_{v \in V} (scope(v) - scope_{ave}(V))^2 \tag{24}$$

Finally, we obtain the objective function $gain(p)$:

$$gain(p) = c_1 profit(p) - c_2 cost(p) + c_3 fairness(p) \tag{25}$$

where c_1, c_2, and c_3 represent the weight coefficients assigned to different terms, satisfying $c_1 \geq 0$, $c_2 \geq 0$, $c_3 \geq 0$, and $c_1 + c_2 + c_3 = 1$.

4.4. Problem Definition

Having formulated the objective function, the problem is then defined as a constrained non-linear integer optimization problem to find the optimal plan p^*. The entire model can be described by Equations (27)–(29):

$$p^* = \arg_{p \in \mathcal{P}} \max gain(p) \tag{26}$$

s.t.

$$\sum_{s \in p} f(s) \leq B \tag{27}$$

$$\sum_{s \in p} \sigma(v,s) \leq 1, \forall vs. \in V \tag{28}$$

$$\rho(s) \leq 1, \forall s \in p \tag{29}$$

Equation (27) denotes that the limitation of the financial cost by a fixed budget B, Equation (28) ensures one node just chooses one charging station, and Equation (29) is used to make the waiting time well-defined. $f(s)$ is the total cost of installing one new charging station.

5. PPO-Attention Algorithm

In this section, the attention model architecture and PPO algorithm we use are discussed.

5.1. Reinforcement Learning Problem Formulation

In this subsection, the problem of charging station placement is modeled as a reinforcement learning problem with a single agent. In the reinforcement problem, state representation, action representation, and reward function design are introduced in detail. The attention-based policy network we propose will be introduced in the next subsection.

5.1.1. State Representation

The state of the agent contains four components: $s_i = \{s_{cp}, s_{coor}, s_{dem}, s_{price}\}$, where s_{cp},s_{coor},s_{dem}, and s_{price} represent the current charging plan p, the latitude and longitude coordinate, charging demand $deman(v)$, and the installation cost for each node $v \in V$ at episode i, respectively. For episode i, the state $s_i \in P \times \mathbb{R}^{|V|^2} \times \mathbb{R}^{|V|} \times \mathbb{R}^{|V|}$.

For episode i, the information of every vertex v is represented by a vector $\overrightarrow{v} = [plan(v_i), x_v, y_v, demand(v), price(v)]$, where $plan(v)$ denotes whether vertex v_i is set as the charging station, $plan(v_i) = 1$ meaning yes, 0 meaning otherwise, x_v,y_v are the coordinate of the vertex v, and $demand(v)$,$price(v)$ are the charging demand at vertex v and installation cost if v is set as charging station.

5.1.2. Action Representation

In this problem, discrete actions are denoted by the set of indices $\mathcal{A} = \{0, 1, 2, 3, 4, 5, 6\}$, representing "Create by benefit", "Create by demand", "Create by fairness", "Increase by benefit", "Increase by demand", "Increase by fairness", and "Relocate". The agent's action selection follows a sampling process from $\mathcal{A}$.

To create a new charging station (CS), three greedy strategies determine its position in the road network. "Create by benefit" selects the node with the lowest profit ($cov(v) + scope(v)$) to enhance node coverage. "Create by demand" chooses the node with the highest demand ($dem(v)$). A lookup table is created to determine charger configurations for feasible capacity demands, selecting the cheapest configuration for each demand. "Create by fairness" maximizes the fairness of the current plan ($fairness(p)$). Similarly, for increasing chargers, the same greedy strategies select a CS $s \in P$, adding one charger if the station has fewer than K chargers. In terms of charger relocation, the charging station s_{old} with the lowest benefit is identified. One of its chargers is then relocated to the charging station within the current plan p^i that exhibits the highest waiting and charging time. If s_{old} becomes empty, it is removed from p^i. This relocation strategy ensures the availability of charging services in situations where the supply falls short of the demand, without considering the costs associated with the relocation process.

5.1.3. Reward Function Design

The reward for the charging plan at the ith episode is computed using the proposed objective function $gain(p^i)$. Initially, the reward is set to 0 when $i = 0$, thus yielding:

$$reward^i = gain(p^i) \tag{30}$$

5.2. Attention Model Architecture

In the proposed approach, we utilize an attention model architecture to improve the algorithm's capacity to capture interdependencies between network nodes. This section provides a detailed description of the attention model, including the encoder and decoder components.

For the road network $G = (V, E)$, where each vertex v is represented by the vector $\overrightarrow{v}$, the desired charging plan p is a permutation of the vertices denoted as $\pi = (\pi_1, \ldots, \pi_n)$ in the reinforcement learning (RL) problem. A stochastic policy $po(\pi|s_i)$ is defined to select a solution π based on the state s_i. Typically, $po(\pi|s_i)$ is parameterized by θ as follows:

$$po_\theta(\pi|s_i) = \prod_{t=1}^{n} po_\theta(\pi_t|s_i, \pi_{1:t-1}) \tag{31}$$

When dealing with a continuous state space S, a neural network is commonly used to approximate the policy function $po(\pi|s_i)$, such as the Deep Q-Network (DQN), which outperforms previous methods such as Q-Learning. To achieve better performance and faster convergence, we propose a novel policy network based on the Encoder-Decoder Framework. The structure of the policy network is illustrated in Figure 2, which consists of two parts: the encoder and the decoder.

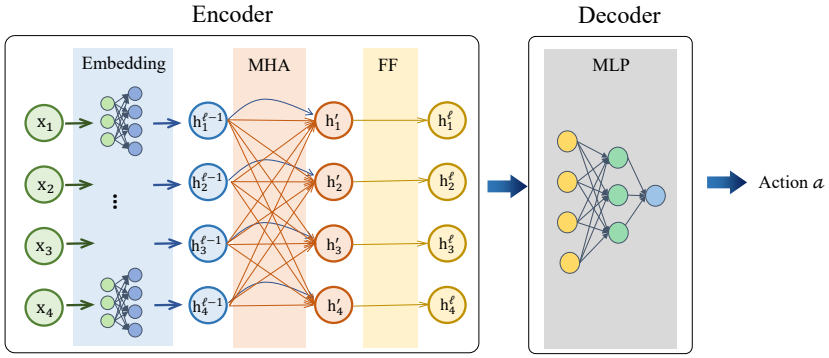


Figure 2. The encoder–decoder framework of our model.

5.2.1. Encoder

The encoder comprises two parts: the linear layer and the attention layer. The linear projection serves as the initial layer to embed all the original vertices, while the attention layer is responsible for passing weighted messages. Unlike the encoder in the Transformer Model [30], the encoder in our model does not include a positional encoding module.

First, for all the input vertices with d_x-dimensional representation, the linear layer embeds each node from d_x-dimensional to d_h-dimensional:

$$h_i^0 = W^X x_i + b^X \tag{32}$$

Next, the embedded information is passed to the attention layer, which may consist of N layers. Each attention layer consists of two sublayers: a multi-head attention (MHA) layer for exchanging weighted messages between vertices and a fully connected feed-forward (FF) layer. Empirical tricks such as skip-connection [35] and batch normalization (BN)[36] are utilized. Let h_i^ℓ denote the vertex embedding result produced by layer $\ell \in \{1, \ldots, N\}$. Therefore, we have:

$$\hat{h_\ell} = BN^\ell(h_i^{\ell-1} + MHA_i^\ell(h_1^{\ell-1}, \ldots, h_n^{\ell-1})) \tag{33}$$

$$h_i^\ell = BN^\ell(\hat{h_\ell} + FF^\ell(h_i)) \tag{34}$$

These equations describe the forward pass through the attention layer, where $h_i^{\ell-1}$ is the input to the layer, $\hat{h_\ell}$ is the intermediate result after the MHA layer, and h_i^ℓ represents the output after applying the FF layer. By stacking multiple attention layers, the encoder captures the dependencies and interactions between vertices in the road network.

5.2.2. Decoder

The decoder component employs a Multiple Layer Perceptron (MLP) to decode the information from the encoder. It consists of two linear layers. The final output of the decoder represents the policy network's output, which is a seven-dimensional vector $\overrightarrow{v}$ indicating the probabilities of different actions in the action set for the road network.

The decoder takes the encoded vertex information h_i^N from the encoder and processes it through the MLP layers to obtain the final action probabilities. The output $\overrightarrow{v}$ is computed as follows:

$$\overrightarrow{v} = \mathrm{MLP}(h_i^N) \tag{35}$$

The final action to be selected for each vertex is the one with the highest probability.

5.3. Incorporation of an Attention-Based Policy Network into the PPO Algorithm

We have integrated the attention-based policy network into the Proximal Policy Optimization (PPO) algorithm, resulting in our PPO-Attention algorithm.

As described in Section 3.2, PPO is a highly regarded policy optimization method in the field of RL, known for its remarkable performance across diverse RL tasks. It utilizes policy gradient methods to update the parameters of the policy network based on advantage estimation and trust region constraints.

In our PPO-Attention algorithm, we exploit the attention mechanism to enhance the policy network's ability to capture the interdependencies among nodes in a road network. Specifically, the attention layer incorporated in the encoder section of the policy network assigns distinct weights to vertices, taking into account their significance and relevance for the charging station placement task. By doing so, the model becomes adept at identifying critical nodes and making informed decisions that incorporate the broader context of the road network.

The PPO-Attention algorithm follows the core framework of PPO while substituting the conventional policy network with the attention-based policy network described above. During the training process, the algorithm gathers trajectories through interactions with the environment and computes advantage estimates for each state–action pair. Subsequently, the policy network is updated based on these advantage estimates and a trust region constraint, which ensures that the policy updates remain within predefined bounds, thus preserving stability.

By incorporating attention mechanisms and reinforcement learning into the PPO-Attention algorithm, we can effectively optimize the placement of charging stations in road networks, considering various factors and capturing the interdependencies between network nodes.

6. Experiment and Analysis

In this section, we first introduce the datasets used for algorithm training and evaluation, the baseline algorithms employed for algorithm evaluation, and the evaluation criteria. We then provide a detailed analysis of the algorithm's performance on different datasets.

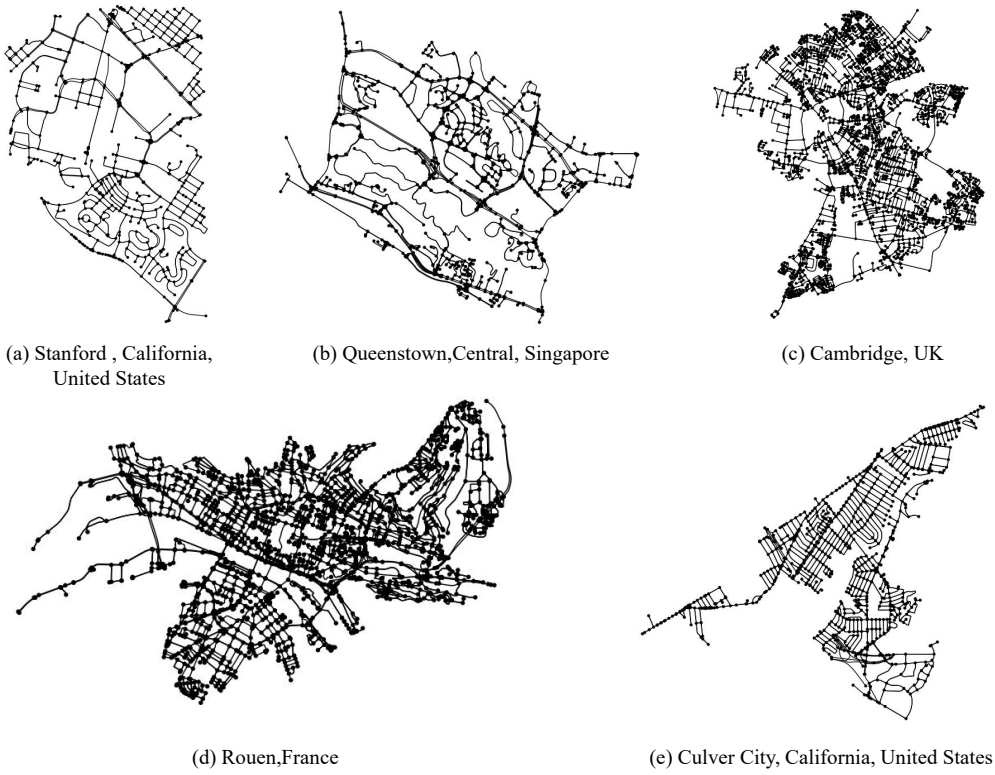
6.1. Dataset

To substantiate the efficacy of our approach in road networks, characterized by varying scales and geographical regions, we meticulously selected five distinct areas for algorithm training and evaluation: Stanford, California, United States; Queenstown, Central Region, Singapore; Cambridge, United Kingdom; Rouen, France; and Culver City, California, United States, which are shown in Figure 3. These datasets were used for algorithm training and evaluation.

- Road network data: The road network data for the aforementioned areas were obtained through Open Street Map [37]. We acquired information such as the coordinates (X_s, Y_s), road types (T_s), and number of lanes (N_s) for each node in the road network.
- Charging station data: Existing charging infrastructure data for these road networks were obtained from the Open Charge Map [38]. These data include the location (X_c, Y_c), type (T_c), number of chargers (N_c), charging capacity (Ca_c), charging cost (Co_c), and charger prices. In our calculations, publicly available charging station price data were used to estimate the costs associated with building new charging stations.

The scale of the road networks, the number of nodes, and the number of charging stations in each city are presented in Table 1. It can be observed that the Stanford road network has the fewest nodes (534), while the Cambridge, UK area has the largest coverage with 3233 nodes and 7121 edges. Culver City, California, United States, possesses the highest number of existing charging station infrastructure (113). The selection of road network data of varying scales allows for comprehensive evaluation of the algorithm's robustness and applicability.

During the data preprocessing stage, we matched the road network nodes with the charging station information for each area. Each charging station was assigned to the nearest road network node to simplify the problem-solving process. Additionally, we calculated the offline values of $r(s)$, $scope(s)$, and $cov(s)$ for each road node $s \in S$ based on Equations (12)–(14) to facilitate subsequent calculations. As described in Section 5.1.1, we considered all the nodes V in the road network, their topological relationships E, the charging demands Dem_v of all nodes, and the cost information $Price_v$ associated with building new charging stations as inputs to establish the observation space for reinforcement learning.



(a) Stanford , California, United States

(b) Queenstown,Central, Singapore

(c) Cambridge, UK

(d) Rouen,France

(e) Culver City, California, United States

Figure 3. The five road networks we selected for testing.

Table 1. Number of nodes and edges, and existing charging stations in different city road networks.

City	Nodes	Edges	Existing Charging Stations
Stanford	534	1178	13
Queenstown, Central, Singapore	1001	1008	3
Cambridge, UK	3233	7121	30
Rouen, France	2273	4824	29
Culver City, California, United States	863	2086	113

6.2. Baseline Algorithms

To assess the performance of our method meticulously, we conduct a comparative analysis against several established baseline algorithms:

- Benefit-first greedy algorithm: This algorithm employs a greedy approach that iteratively selects locally optimal solutions in the pursuit of a globally optimal solution. At each step of the solution process, the baseline algorithm focuses exclusively on maximizing the profit function when determining the charging station sites.
- Demand-first greedy algorithm: In contrast to the aforementioned Benefit-first Greedy Algorithm, this approach prioritizes charging demand when selecting charging station locations, employing a greedy strategy to achieve optimal solutions.
- Genetic algorithm: Widely recognized as a heuristic algorithm, the genetic algorithm emulates the natural evolutionary process to search for optimal solutions [39]. In our study, we enhance the traditional genetic algorithm and devised a Multi-Layer Perceptron (MLP) network for population encoding to address the problem at hand.

6.3. Implementation Details

For the entire optimization problem, we set the following parameters: $m = 3$, $B = 10^6$, $r_{max} = 1$ km. To balance the profit term, loss term, and fairness term, we set $c_1 = c_2 = c_3 = \frac{1}{3}$.

In our proposed PPO-Attention algorithm, we utilize the following parameters: $Update\ Steps = 512$, $Batch\ Size = 128$, $Learning\ Rate = 0.002$, and $\gamma = 0.8$. Moreover, the encoder MLP and decoder MLP both have a size of 64×64. The attention layer incorporates two heads, with $d_k = 32$. The total training timesteps for all three algorithms are set to 10^5.

For the genetic algorithm employed as a baseline, we encode and perform crossover operations on the population using an MLP with three linear layers, each consisting of 64 units. The parameters used in the genetic algorithm are as follows: $Number\ of\ Generations = 20$, $Population\ Size = 100$, $Crossover\ Rate = 0.8$, $Mutation\ Rate = 0.01$, $Mutation\ Factor = 0.001$, and $Maximum\ Global\ Steps = 5 \times 10^4$. The parameters for the profit-based greedy algorithm and demand-based greedy algorithm remain consistent with those of the PPO-Attention algorithm, except for the differences in policy selection.

All experiments were conducted on a platform equipped with an Intel Core i7-12700 CPU, NVIDIA GeForce RTX 3070 Ti GPU, and 32GB memory.

Meanwhile, a few indicators are used to evaluate the performance of the algorithm:

- The objective function $gain(p)$. The value of $gain(p)$ denotes the overall performance of the model. The higher the score is, the better.
- Benefit of the $gain(p)$: the sum of the profit term $profit(p)$ and fairness term $fairness(p)$ in the objective function, which represents the positive impact of the current solution. Higher values indicate better performance.
- Time cost $cost(p)$. $cost(p)$ contains following parts: travel time, which is the sum of travel times within the road network, charging time, which is the sum of the charging times within the road network, and waiting time, which is the sum of the waiting times occurring at all CS in the road network. Additionally, we calculated the longest travel time and longest queuing time throughout the entire process. For all the time terms, a lower score is better.

6.4. Performance Evaluation

In the conducted evaluation experiments, we undertake a comprehensive reconfiguration of the charging infrastructure by leveraging real-world charging station data while considering various constraints, including budgetary limitations. As the benchmark, we employ the existing layout of charging facilities, with all performance indicators normalized to 100%. Tables 2 and 3 present the performance outcomes of our PPO-Attention algorithm in comparison to several baseline algorithms across diverse urban road networks. Specifically, Table 2 showcases the results obtained for Stanford, California, United States;

Queenstown, Central Region, Singapore; and Cambridge, United Kingdom, while Table 3 encompasses the outcomes for Rouen, France, and Culver City, California, United States.

Table 2. Results are presented for the Stanford, Queenstown, and Cambridge datasets from the United Kingdom. Evaluation metrics favoring higher values are indicated with ↑, while those favoring lower values are labeled with ↓. The best scores are highlighted in bold.

Algorithm	Score ↑		Cost ↓				
	Gain	Benefit	Wait	Travel	Charging	Travel Max [min]	Wait Max [min]
Stanford							
Existing Charging	100.00	100.00	100.00	100.00	100.00	4.02	15.47
GA	228.09	142.71	57.06	77.93	72.33	3.20	8.77
Greedy_Benefit_first	329.98	183.08	35.84	64.76	66.41	2.37	7.17
Greedy_Demand_first	200.70	136.21	67.68	97.73	79.89	4.03	15.47
PPO-Attention (Ours)	**426.38**	**223.10**	**19.19**	**58.43**	**62.46**	**2.16**	**1.98**
Queenstown, Central, Singapore							
Existing Charging	100.00	100.00	100.00	100.00	100.00	8.35	-
GA	225.86	144.17	44.04	**75.65**	99.37	**7.73**	-
Greedy_Benefit_first	192.42	132.39	60.02	88.01	94.28	8.32	-
Greedy_Demand_first	192.42	132.39	60.02	88.01	94.28	8.32	-
PPO-Attention (Ours)	**253.10**	**133.87**	**14.85**	97.97	**13.45**	8.29	**19.47**
Cambridge, UK							
Existing Charging	100.00	100.00	100.00	100.00	100.00	5.18	819.75
GA	104.63	102.41	100.11	100.00	101.17	5.18	819.76
Greedy_Benefit_first	106.60	101.91	96.97	98.26	98.33	4.71	819.76
Greedy_Demand_first	106.60	101.87	96.90	98.31	98.18	4.70	819.76
PPO-Attention (Ours)	**141.00**	**113.84**	**84.34**	**89.34**	**96.31**	**4.17**	**796.26**

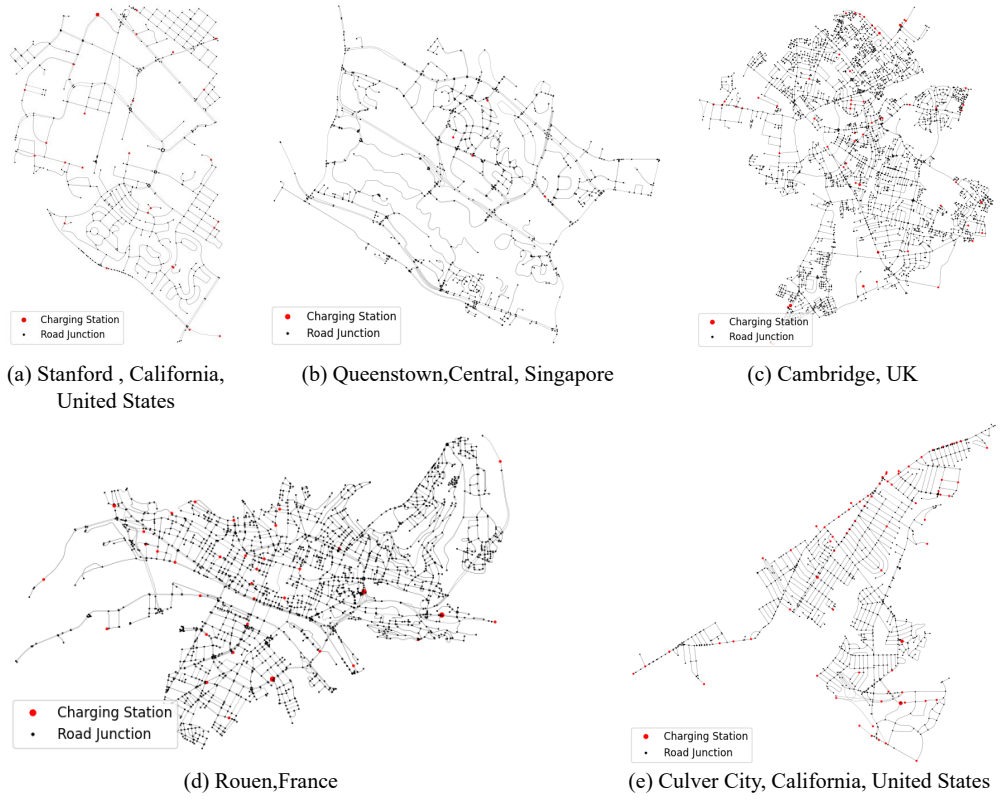
Table 2 reveals compelling findings for the Stanford region, wherein our algorithm achieves an astounding 426.38% increase in profit when contrasted with the established baseline configuration. This remarkable improvement markedly surpasses the outcomes attained by the other three baseline algorithms, which record profit increases of 228.09%, 329.98%, and 200.70%, respectively. Moreover, our devised solution effectively diminishes waiting time, travel time, and charging time by 80.81%, 41.57%, and 37.54%, respectively. Similarly, for the Queenstown, Central Region, Singapore area, our algorithm yields a remarkable profit escalation of 253.10% compared to the existing baseline, thus surpassing the performance of the other three baseline algorithms, which attain profit increments of 225.86%, 192.42%, and 192.42%, respectively. Furthermore, our proposed solution considerably reduces waiting time, travel time, and charging time by 85.15%, 2.03%, and 86.55%, respectively. Notably, it is evident that in this specific region, the genetic algorithm displays superior performance in terms of travel time and longest travel time, while our algorithm exhibits remarkable excellence across various other evaluation metrics. In the Cambridge, United Kingdom area, our algorithm's planning solution achieves a noteworthy comprehensive profit increase of 141.00%, surpassing the outcomes attained by the alternative baseline algorithms.

Analogously, Table 3 showcases the impressive results for the Rouen, France and Culver City, California, United States regions, where our algorithm accomplishes comprehensive profit improvements of 248.67% and 221.04%, respectively, thereby outperforming the other baseline algorithms in each respective region.

Furthermore, Table 4 consolidates the average outcomes obtained across all datasets. Remarkably, the PPO-Attention algorithm, rooted in real-world charging station layouts, achieves a remarkable average profit increase of 258.04%, while concurrently reducing the average waiting time by 73.40%, travel time by 18.47%, and charging time by 40.10%. When juxtaposed with the alternative baseline algorithms, our PPO-Attention algorithm emerges as the definitive frontrunner across all evaluation metrics. To enhance comprehension, Figure 4 provides visual representations of the planning outcomes derived from our PPO-Attention algorithm for diverse urban charging station layouts. These visuals underscore the algorithm's ability to achieve a harmonious equilibrium by effectively considering various factors such as the prevailing distribution of charging stations, charging demand patterns, and budgetary constraints inherent to distinct urban road networks.

Table 3. Results on the Rouen, France, and Culver City, California, United States datasets. Evaluation metrics favoring higher values are indicated with ↑, while those favoring lower values are labeled with ↓. The best scores are highlighted in bold.

Algorithm	Score ↑		Cost ↓				
	Gain	Benefit	Wait	Travel	Charging	Travel Max [min]	Wait Max [min]
Rouen, France							
Existing Charging	100.00	100.00	100.00	100.00	100.00	7.74	-
GA	141.00	108.70	59.00	80.88	101.58	**5.34**	-
Greedy_Benefit_first	135.20	107.39	65.12	96.83	91.81	7.80	-
Greedy_Demand_first	113.19	104.85	100.00	100.27	93.29	7.74	-
PPO-Attention (Ours)	**248.67**	**139.55**	**8.29**	**78.03**	**65.48**	6.81	**147.60**
Culver City, California, United States							
Existing Charging	100.00	100.00	100.00	100.00	100.00	2.91	217.19
GA	117.17	108.41	100.36	100.00	101.62	2.91	217.19
Greedy_Benefit_first	134.20	105.09	53.28	96.52	94.67	2.88	93.60
Greedy_Demand_first	112.55	102.57	88.94	98.49	95.63	2.91	205.64
PPO-Attention (Ours)	**221.04**	**130.53**	**19.13**	**83.91**	**61.82**	**2.18**	**28.61**



(a) Stanford , California, United States (b) Queenstown,Central, Singapore (c) Cambridge, UK

(d) Rouen,France (e) Culver City, California, United States

Figure 4. The planning results for five cities from the PPO-Attention algorithm.

Table 4. Average results for all datasets. The best scores are highlighted in bold.

Algorithm	Gain Score	Benefit Score	Waiting Time	Travel Time	Charging Time
Existing Charging	100	100	100	100	100
GA	163.35	121.28	72.11	86.89	95.21
Greedy_Benefit_first	179.68	125.97	62.25	88.88	89.10
Greedy_Demand_first	145.09	115.58	82.71	96.56	92.25
PPO-Attention (Ours)	**258.04**	**148.18**	**26.60**	**81.53**	**59.90**

6.5. Advantages and Disadvantages of RL in CSLP

In this section, we aim to provide a balanced discussion on the advantages and disadvantages of deploying RL in the CSLP. Understanding the strengths and limitations of RL in this context is crucial to its effective application and continual development.

The advantages of RL in the CSLP are primarily characterized by its adaptability, scalability, and ability to balance between exploration and exploitation. RL's adaptability

allows for continual learning and adaptation to changing environments, such as variable EV adoption rates or alterations in road network complexity. This adaptability enables the model to optimize charging station placement as conditions change. Furthermore, RL algorithms demonstrate excellent scalability, being well-suited to managing large and complex road networks. As the size and complexity of the problem scale, RL algorithms can still find optimal solutions, making them a powerful tool for CSLP. Finally, RL's ability to balance between exploration (searching new possible charging station locations) and exploitation (leveraging knowledge of locations that are already known to be effective) results in a robust solution that can discover and capitalize on optimal charging station locations.

However, RL application in the CSLP also presents challenges related to data requirements, interpretability, and computational expense. RL algorithms often require substantial data for training. In the context of CSLP, obtaining sufficient and accurate data on traffic patterns, EV adoption rates, and driver behavior can be challenging. Additionally, the decision-making process of RL algorithms can be opaque, presenting difficulties in interpreting why certain locations were chosen for charging stations over others. Finally, RL, especially when combined with deep-learning structures, can be computationally expensive. This cost can be exacerbated by the complexity of the road networks and the continuous state and action spaces in the CSLP.

In conclusion, while RL poses certain challenges in its application to the CSLP, its advantages, particularly adaptability and scalability, render it a valuable tool in this context. With careful consideration and application of appropriate techniques, the disadvantages of RL can be effectively managed, affirming RL's potential as a promising approach to address the CSLP.

7. Conclusions

Efficient planning and layout of charging stations play a crucial role in improving charging efficiency, infrastructure utilization, and overall social benefits. In this study, we have addressed the charging station location problem by considering various factors, such as the benefits of the layout, driver waiting time, fairness of the distribution, and other constraints. By formulating the problem as a reinforcement learning task and leveraging the Proximal Policy Optimization (PPO) algorithm along with the attention mechanism, we have developed the PPO-Attention algorithm. Real-world data have been utilized for training and testing the proposed algorithm. Our experimental results demonstrate that the algorithm surpasses other baseline methods, and the novel charging station layout scheme leads to a significant improvement in overall social income.

It is important to note that certain parameters, such as actual installation cost, charging pile cost, and government budget, were unavailable through internet sources, which necessitated certain simplifications in the processing. This may introduce a degree of impact on the authenticity of the final results. In future research, we aim to enhance the model and algorithm, taking into account more realistic constraints and conducting extensive testing and application on diverse road networks.

Moreover, employing reinforcement learning algorithms to tackle optimization problems offers substantial advantages in terms of efficiency, scalability, and flexibility compared to conventional algorithms. In our future work, we will focus on further improving the computational efficiency and scalability of the algorithm for optimizing charging station layouts across various scales. Additionally, we plan to extend the application of this approach to address similar optimization problems related to baseline configuration layouts.

Author Contributions: Conceptualization, J.S. and X.Q.; Methodology, J.L.; Software, J.L.; Validation, X.Q.; Writing – original draft, J.L.; Writing – review & editing, J.S. and X.Q.; Visualization, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is sponsored by the Shanghai Committee of Science and Technology, China (Grant No. 22dz1203200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in our research can be obtained from Open Street Map (https://www.openstreetmap.org/) and Open Charge Map (https://openchargemap.org/).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yong, J.Y.; Ramachandaramurthy, V.K.; Tan, K.M.; Mithulananthan, N. A review on the state-of-the-art technologies of electric vehicle, its impacts and prospects. *Renew. Sustain. Energy Rev.* **2015**, *49*, 365–385. [CrossRef]
2. Padavala, P.; Inavolu, N.; Thaveedu, J.R.; Medisetti, J.R. Challenges in Noise Refinement of a Pure Electric Passenger Vehicle. *SAE Int. J. Veh. Dyn. Stab. NVH* **2021**, *5*, 45–64. [CrossRef]
3. Hazra, S.; Reddy, J.K. A Review Paper on Recent Research of Noise and Vibration in Electric Vehicle Powertrain Mounting System. *SAE Int. J. Veh. Dyn. Stab. NVH* **2021**, *6*, 3–22. [CrossRef]
4. Funke, S.Á.; Sprei, F.; Gnann, T.; Plötz, P. How much charging infrastructure do electric vehicles need? A review of the evidence and international comparison. *Transp. Res. Part D Transp. Environ.* **2019**, *77*, 224–242. [CrossRef]
5. Liu, Q.; Zeng, Y.; Chen, L.; Zheng, X. Social-aware optimal electric vehicle charger deployment on road network. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 5–8 November 2019; pp. 398–407.
6. von Wahl, L.; Tempelmeier, N.; Sao, A.; Demidova, E. Reinforcement Learning-based Placement of Charging Stations in Urban Road Networks. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 3992–4000.
7. Kchaou-Boujelben, M. Charging station location problem: A comprehensive review on models and solution approaches. *Transp. Res. Part C Emerg. Technol.* **2021**, *132*, 103376. [CrossRef]
8. Zhang, Y.; Liu, X.; Zhang, T.; Gu, Z. Review of the electric vehicle charging station location problem. In *Dependability in Sensor, Cloud, and Big Data Systems and Applications: 5th International Conference, DependSys 2019, Guangzhou, China, 12–15 November 2019, Proceedings 5*; Springer: Singapore, 2019; pp. 435–445.
9. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
10. Du, B.; Tong, Y.; Zhou, Z.; Tao, Q.; Zhou, W. Demand-aware charger planning for electric vehicle sharing. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1330–1338.
11. Gan, X.; Zhang, H.; Hang, G.; Qin, Z.; Jin, H. Fast-charging station deployment considering elastic demand. *IEEE Trans. Transp. Electrif.* **2020**, *6*, 158–169. [CrossRef]
12. Greene, D.L.; Kontou, E.; Borlaug, B.; Brooker, A.; Muratori, M. Public charging infrastructure for plug-in electric vehicles: What is it worth? *Transp. Res. Part D Transp. Environ.* **2020**, *78*, 102182. [CrossRef]
13. Krallmann, T.; Doering, M.; Stess, M.; Graen, T.; Nolting, M. Multi-objective optimization of charging infrastructure to improve suitability of commercial drivers for electric vehicles using real travel data. In Proceedings of the 2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Rhodes, Greece, 25–27 May 2018; IEEE: Piscataway, NJ, USA, 2018, pp. 1–8.
14. Liu, C.; Deng, K.; Li, C.; Li, J.; Li, Y.; Luo, J. The optimal distribution of electric-vehicle chargers across a city. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 261–270.
15. Liu, Q.; Liu, J.; Le, W.; Guo, Z.; He, Z. Data-driven intelligent location of public charging stations for electric vehicles. *J. Clean. Prod.* **2019**, *232*, 531–541. [CrossRef]
16. Zhao, Y.; Guo, Y.; Guo, Q.; Zhang, H.; Sun, H. Deployment of the electric vehicle charging station considering existing competitors. *IEEE Trans. Smart Grid* **2020**, *11*, 4236–4248. [CrossRef]
17. Zhang, H.; Hu, Z.; Xu, Z.; Song, Y. Optimal planning of PEV charging station with single output multiple cables charging spots. *IEEE Trans. Smart Grid* **2016**, *8*, 2119–2128. [CrossRef]
18. Xie, R.; Wei, W.; Khodayar, M.E.; Wang, J.; Mei, S. Planning fully renewable powered charging stations on highways: A data-driven robust optimization approach. *IEEE Trans. Transp. Electrif.* **2018**, *4*, 817–830. [CrossRef]
19. Erdinç, O.; Taşcıkaraoğlu, A.; Paterakis, N.G.; Dursun, I.; Sinim, M.C.; Catalao, J.P. Comprehensive optimization model for sizing and siting of DG units, EV charging stations, and energy storage systems. *IEEE Trans. Smart Grid* **2017**, *9*, 3871–3882. [CrossRef]
20. Bae, S.; Jang, I.; Gros, S.; Kulcsár, B.; Hellgren, J. A game approach for charging station placement based on user preferences and crowdedness. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 3654–3669. [CrossRef]
21. Choi, W. Placement of charging infrastructures for EVs using K-mean algorithm and its validation using real usage data. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2020**, *7*, 875–884. [CrossRef]
22. Zeng, L.; Krallmann, T.; Fiege, A.; Stess, M.; Graen, T.; Nolting, M. Optimization of future charging infrastructure for commercial electric vehicles using a multi-objective genetic algorithm and real travel data. *Evol. Syst.* **2020**, *11*, 241–254. [CrossRef]
23. Vazifeh, M.M.; Zhang, H.; Santi, P.; Ratti, C. Optimizing the deployment of electric vehicle charging stations using pervasive mobility data. *Transp. Res. Part A Policy Pract.* **2019**, *121*, 75–91. [CrossRef]

24. Wang, H.n.; Liu, N.; Zhang, Y.y.; Feng, D.w.; Huang, F.; Li, D.s.; Zhang, Y.m. Deep reinforcement learning: A survey. *Front. Inf. Technol. Electron. Eng.* **2020**, *21*, 1726–1744. [CrossRef]
25. Mazyavkina, N.; Sviridov, S.; Ivanov, S.; Burnaev, E. Reinforcement learning for combinatorial optimization: A survey. *Comput. Oper. Res.* **2021**, *134*, 105400. [CrossRef]
26. Cunha, B.; Madureira, A.M.; Fonseca, B.; Coelho, D. Deep reinforcement learning as a job shop scheduling solver: A literature review. In *Hybrid Intelligent Systems: 18th International Conference on Hybrid Intelligent Systems (HIS 2018), Porto, Portugal, 13–15 December 2018*; Springer: Cham, Switzerland, 2020; pp. 350–359.
27. Zhao, Z.; Lee, C.K.; Huo, J. EV charging station deployment on coupled transportation and power distribution networks via reinforcement learning. *Energy* **2023**, *267*, 126555. [CrossRef]
28. Li, Y.; Wang, J.; Wang, W.; Liu, C.; Li, Y. Dynamic pricing based electric vehicle charging station location strategy using reinforcement learning. *Energy* **2023**, *128*, 128284. [CrossRef]
29. Zhao, Z.; Lee, C.K.; Ren, J.; Tsang, Y.P. Optimal EV Fast Charging Station Deployment Based on a Reinforcement Learning Framework. *IEEE Trans. Intell. Transp. Syst.* 2023, *early access*. [CrossRef]
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
31. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
32. Niv, Y.; Daniel, R.; Geana, A.; Gershman, S.J.; Leong, Y.C.; Radulescu, A.; Wilson, R.C. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **2015**, *35*, 8145–8157. [CrossRef] [PubMed]
33. Iqbal, S.; Sha, F. Actor-attention-critic for multi-agent reinforcement learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 2961–2970.
34. Bhat, U.N. *An Introduction to Queueing Theory: Modeling and Analysis in Applications*; Birkhäuser: Boston, MA, USA, 2008; Volume 36.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
37. Bennett, J. *OpenStreetMap*; Packt Publishing Ltd.: Birmingham, UK, 2010.
38. Map, O.C. Open Charge Map—The global public registry of electric vehicle charging locations. *Open Charge Map* 2017. Available online: https://openchargemap.org/ (accessed on 12 December 2022).
39. Holland, J.H. Genetic algorithms. *Sci. Am.* **1992**, *267*, 66–73. [CrossRef]