

Dynamic Load Balancing for EV Charging Stations Using Reinforcement Learning and Demand Prediction

Hesam Mosalli, Saba Sanami, Yu Yang, Hen-Geul Yeh, and Amir G. Aghdam

Abstract—This paper presents a method for load balancing and dynamic pricing in electric vehicle (EV) charging networks, utilizing reinforcement learning (RL) to enhance network performance. The proposed framework integrates a pre-trained graph neural network to predict demand elasticity and inform pricing decisions. The spatio-temporal EV charging demand prediction (EVC DP) dataset from Shenzhen is utilized to capture the geographic and temporal characteristics of the charging stations. The RL model dynamically adjusts prices at individual stations based on occupancy, maximum station capacity, and demand forecasts, ensuring an equitable network load distribution while preventing station overloads. By leveraging spatially-aware demand predictions and a carefully designed reward function, the framework achieves efficient load balancing and adaptive pricing strategies that respond to localized demand and global network dynamics, ensuring improved network stability and user satisfaction. The efficacy of the approach is validated through simulations on the dataset, showing significant improvements in load balancing and reduced overload as the RL agent iteratively interacts with the environment and learns to dynamically adjust pricing strategies based on real-time demand patterns and station constraints. The findings highlight the potential of adaptive pricing and load-balancing strategies to address the complexities of EV infrastructure, paving the way for scalable and user-centric solutions.

I. INTRODUCTION

The adoption of electric vehicles (EVs) is rapidly increasing, driven by a growing awareness of environmental issues and the need to reduce carbon emissions to fight climate change. Governments are offering incentives such as subsidies,

Hesam Mosalli, Saba Sanami, and Amir G. Aghdam are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada. Emails: hesam.mosalli@mail.concordia.ca, saba.sanami@mail.concordia.ca, amir.aghdam@concordia.ca. Yu Yang is with the Department of Chemical Engineering and Hen-Geul Yeh is with the Department of Electrical Engineering, California State University Long Beach. Emails: yu.yang@csulb.edu, henry.yeh@csulb.edu.

tax breaks, and significant investments in charging infrastructure, making the transition to EVs more appealing and practical for both individuals and businesses. This rise in EV usage is transforming the transportation sector, leading to a greater demand for a well-developed and accessible charging network to accommodate the increasing number of EV users [1], [2].

Given the increase in the number of EVs, many charging stations have been established across various regions in cities. However, some stations experience heavy demand during the week, while others remain underutilized. It is crucial to implement a smart pricing strategy that can help balance the network in real time. By reducing the price at low-demand stations, EV drivers can be encouraged to charge their vehicles at less crowded locations. This strategy effectively distributes the charging load across the network and helps avoid traffic congestion. This not only improves user convenience by reducing wait times and ensuring access to charging but also benefits service providers by optimizing the utilization of their infrastructure, ultimately leading to a more efficient and balanced charging ecosystem [3], [4], [5].

There is a rich body of research on EV charging infrastructure and the associated economics. Early studies have shown that charging prices are one of the critical determinants for users when selecting charging stations [6]. Dynamic pricing strategies have been explored extensively in recent years, with studies emphasizing their ability to manage energy load and optimize the performance of charging stations. Authors in [7] highlight the limitations of traditional pricing mechanisms, such as time-of-use (ToU) rates, which fail to adapt to real-time demand fluctuations. Their research proposes a dynamic pricing model addressing multiple conflicting objectives, including revenue genera-

tion, quality of service, and peak-to-average ratios, utilizing advanced algorithms like non-dominated sorting genetic algorithms (NSGA) II and III to find optimal trade-offs. Integrating machine learning and deep learning approaches, such as long short-term memory (LSTM), has also been increasingly investigated to enhance price optimization. Recent work in [8] provides a detailed analysis of the effect of electricity prices on EV charging behavior using a learning model incorporating a two-layer graph and temporal pattern attention. In addition to price optimization, demand prediction plays a crucial role in ensuring that charging infrastructure is adequately prepared for fluctuations in demand. Accurate demand prediction models help service providers anticipate peak usage periods and adjust their strategies accordingly [9]. Machine learning techniques, such as neural networks and gradient boosting, have been widely used for forecasting EV charging demand, providing valuable insights into usage [10], [11], [12], [13].

Although dynamic pricing has been studied, there remains a knowledge gap in addressing network balancing that takes into consideration overutilization and underutilization of charging stations, especially in densely packed urban areas where the interaction between multiple stations is crucial. This paper introduces a novel graph-based reinforcement learning-based approach to optimize network balancing in EV charging stations. The primary objective of this research is to enhance load distribution by dynamically adjusting pricing strategies in near real time. To achieve this, we represent the charging station network as a graph. By formulating the problem as a deep Q-learning (DQL) task, the model leverages graph neural networks (GNN) to understand interdependencies between stations, addressing the challenges of optimizing price elasticity to balance demand, reduce over-utilization, and improve overall network efficiency. The DQL-based model learns an optimal pricing strategy that dynamically adapts to changes in charging demand and reduces overload events.

The organization of the paper is as follows. Section II defines the preliminaries and presents the problem statement. The proposed method is discussed in Section III. The implementation of the method and simulation results are outlined in

Section IV. Finally, Section V offers the conclusion of the paper.

II. PRELIMINARIES AND PROBLEM STATEMENT

In a standard RL framework, an agent learns to maximize cumulative reward through interactions with an environment. RL problems are commonly modelled as Markov decision processes (MDPs), which are defined by a tuple (S, A, T, R) . Here, S represents the set of possible states of the environment, A denotes the set of actions available to the agent, T is the state transition function describing how actions affect future states, and R is the reward function that evaluates the desirability of each state-action pair. At each time step t , the agent observes the current state $s_t \in S$, selects an action $a_t \in A$, and receives a reward $R(s_t, a_t)$ from the environment. The objective of RL is to learn a policy $\pi(a|s)$ that maximizes the expected cumulative reward, known as the return, defined as:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

where $\gamma \in [0, 1]$ is a discount factor that controls the importance of future rewards relative to immediate rewards.

In this work, we employ deep Q-learning (DQL), a value-based RL approach that leverages deep neural networks to approximate the Q-value function $Q(s, a)$. The Q-value function estimates the expected return of taking action a in state s and following the policy thereafter. The optimal Q-value function $Q^*(s, a)$ satisfies the Bellman equation:

$$Q^*(s, a) = \mathbb{E} \left[R(s, a) + \gamma \max_{a'} Q^*(s', a') \mid s, a \right],$$

where the agent learns this function to derive an optimal policy $\pi^*(s) = \arg \max_a Q^*(s, a)$. To improve stability, particularly in large, complex environments, DQL incorporates experience replay and a target network. Experience replay stores past experiences in a buffer and samples them randomly to break the correlation between consecutive experiences, enhancing training stability. A target network, which is updated less frequently, provides a stable reference for Q-value updates, further aiding convergence. Double Q-learning, a

variant of DQL, addresses overestimation bias by using two Q-networks, one for action selection and the other for evaluation.

In the current balancing problem in EV charging networks, the RL agent aims to manage charging demand by adjusting prices at each station and directing users to underutilized stations to achieve balanced load distribution. Each charging station is represented as a node in a graph, with edges denoting adjacency between stations. The objective of this problem is to balance the utilization of resources across the network by adjusting prices at individual stations in a way that encourages or discourages demand, thereby achieving equitable load distribution across the network.

The problem formulation leverages the MDP framework, where the state $s \in S$ represents the real-time status of the network, including the utilization of each station—measured as the ratio of current load to capacity—and the average utilization within the network, reflecting local load distribution. Additionally, the state incorporates demand projections derived from a pre-trained price elasticity model, allowing the agent to forecast the impact of price adjustments at each station. The action space A consists of possible price adjustments at each station, which influence demand by encouraging users to select underutilized stations and deterring the use of heavily utilized ones. The transition function $T(s'|s, a)$ accounts for both the spatial dependencies captured by the network structure and the elasticity of demand to price changes, which is modelled through a GNN. The GNN provides spatially-aware demand predictions, enabling the agent to anticipate the effects of localized price adjustments on broader network dynamics.

III. METHODOLOGY

The primary objective is to ensure balanced utilization across the network, which involves minimizing load disparities among stations and preventing overloads at any one station. By managing utilization levels dynamically, the agent can reduce waiting times and prevent inefficiencies caused by uneven resource distribution. The proposed solution uses a DQL agent, enhanced with a GNN, to make price adjustments that achieve balanced

utilization across the EV charging network. The reward function is designed to incentivize balancedness within the network and penalize configurations where any station is near overload.

A. Reward Function Design

The reward function consists of two primary components: a balancedness term to minimize utilization variance within the network and an overload penalty to discourage near-capacity operation at any station,

$$R = \left[\sum_{i=1}^N \left(\frac{L_i}{C_i} - \frac{\bar{L}}{\bar{C}} \right)^2 + \lambda \sum_{i=1}^N \frac{1}{1 + e^{-\kappa(\frac{L_i}{C_i} - 0.9)}} \right]^{-1} \quad (1)$$

where N is the total number of stations, and C_i and L_i are the capacity and load of station i , i.e., the total number of charging piles and the number of occupied ones at station i , respectively. Also, \bar{L} and \bar{C} represent the network-wide averages for load and capacity. The balancedness term in the reward promotes the reduction of variance in utilization, ensuring an even load distribution among stations. The sigmoid penalty function (second term of (1)) remains near zero for utilizations up to 80% for $\kappa = 30$, but sharply increases as utilization approaches full capacity, thus discouraging overloads. Moreover, the inverse reward function effectively incentivizes the agent to minimize penalties by leveraging its steep sensitivity at low penalty values. This formulation ensures that even small reductions in penalties result in significant increases in reward, driving the agent toward near-optimal behavior. Additionally, the nonlinear scaling provided by the inverse reward formulation enables the agent to prioritize improvements where they matter most, avoiding disproportionate attention to penalties that are already large while encouraging consistent optimization across all penalty types.

B. Deep Q-Learning with GNN for Demand Prediction

The proposed DQL framework integrates a pre-trained Price-Adjusted Graph Neural Network (PAG) model [8] with a Multi-Layer Perceptron (MLP) to optimize electric vehicle (EV) charging

prices across a network of charging stations. Unlike traditional Q-learning approaches with simplistic environment simulations, this method leverages the PAG model to predict price elasticity, capturing the spatial and temporal interdependencies of charging demand across stations. This enables the agent to consider the effects of pricing decisions on network-wide utilization.

The PAG model acts as the environment step, predicting demand adjustments based on the current state of occupancy and pricing. The output of the PAG model, which represents price-adjusted station utilization, is fed into an MLP to form the Q-network. The Q-network consists of two fully connected layers and a final output layer that maps to the Q-values for all possible state-action pairs. This design allows the agent to predict expected rewards for pricing actions, supporting informed decision-making to minimize penalties and maximize network efficiency.

Training involves standard DQL techniques, including experience replay and a target network. Experience replay buffers past transitions for randomized sampling, reducing the correlations between updates and stabilizing the learning process. The target network, updated periodically, provides fixed Q-value targets, further enhancing training stability. The reward function encourages minimizing variance and overload penalties, promoting balanced utilization across stations while discouraging excessive demand that could lead to capacity constraints.

This approach enables adaptive, data-driven pricing strategies that respond dynamically to changing demand patterns across the network. By integrating price elasticity predictions into the reinforcement learning process, the proposed framework ensures efficient resource allocation, equitable station utilization, and the prevention of overloads, creating a robust solution for managing EV charging demand in large charging networks.

IV. EXPERIMENTAL TESTING OF THE METHOD

A. Dataset

The data used in this study is from the open-source ST-EVCDP dataset [14] gathered from a publicly available mobile application that provides real-time status of charging pile availability (i.e.,

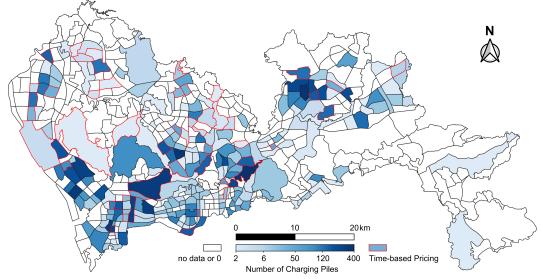


Fig. 1. Spatial distribution of the public EV charging piles in ST-EVCDP [15]

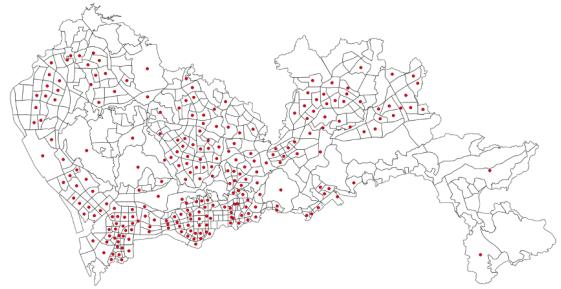


Fig. 2. Centroids of the traffic zones with at least one EV charging pile

whether they are idle or in use). The dataset covers 18,061 public charging piles in Shenzhen, China, collected over one month from June 19 to July 18, 2024, with a data collection interval of 5 minutes, resulting in 8,640 timestamps. The spatial distribution of the EV charging piles in 491 different traffic regions of the city is depicted in Fig. 1. To simplify the analysis of the problem, all charging piles in each region are assumed to be consolidated at the centroid of that region. Therefore, each region is referred to as a station from here on. Moreover, by assigning each region with at least one pile to a node, the city's charging stations network can be represented as a graph-structured data set consisting of 247 nodes. In this graph, two nodes are connected if their corresponding regions are geographically adjacent.

In addition to the geographical specifications of the regions, the ST-EVCDP also includes the occupancy and price records of all stations. The centroid nodes are shown in Fig. 2.

B. Data Preprocessing

To prepare the data for applying the proposed RL method, the raw occupancy and pricing data,

initially recorded at 5-minute intervals, are averaged to create 1-hour interval data. While demand exhibits small fluctuations over time, pricing adjustments are typically updated at intervals exceeding two hours. By aggregating the data into 1-hour intervals, we align the temporal granularity with the decision-making needs of the RL framework, reducing noise and improving the efficiency of training without losing the essential long-term patterns in demand and pricing.

In the original geographical configuration shown in Fig. 2, some regions lack charging piles, leading to a fragmented graph where certain areas are isolated. This disconnected graph structure poses challenges for GNN-based models, which rely on spatial relationships to model demand interactions across neighboring stations. A fully connected graph is crucial to ensure that spatial dependencies are accurately captured and propagated, enabling the RL agent to make informed and globally effective decisions. Without addressing these disconnections, the agent's ability to understand and optimize system-wide behavior would be limited.

To resolve this issue, regions without charging piles are merged with their nearest zone containing at least one charging pile. This merging ensures that the graph representing stations and their adjacencies becomes fully connected. The merging process is performed based on geographical proximity, maintaining realistic spatial relationships while enabling the GNN to capture spatial dependencies across the entire network effectively. This ensures the resulting graph is both meaningful and computationally feasible for modelling. The outcome of this preprocessing step and the resulting graph are illustrated in Figs. 3 and 4, respectively. These preprocessing steps ensure the spatial interdependencies among stations are accurately represented, allowing the proposed framework to balance demand across the network effectively.

C. Simulation Results

The proposed RL approach is evaluated through a series of simulations conducted using the processed EV charging demand dataset. The RL model is trained in a data-driven manner, where each episode processes all training data in mini-batches. This setup allows the agent to refine its policy across multiple episodes iteratively. At the

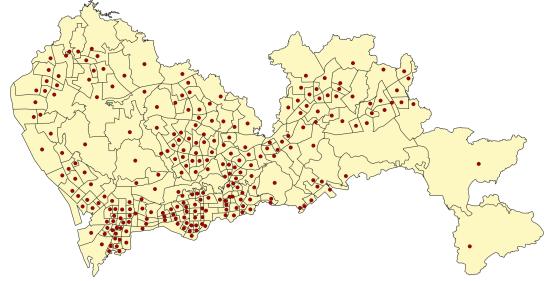


Fig. 3. Merged traffic zones map and the centroid nodes of the original stations

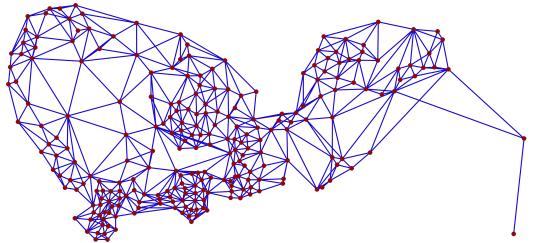


Fig. 4. Geographical adjacency graph of the EV charging stations

start of training, the agent explores various actions due to the high epsilon value ($\epsilon = 1.0$). As training progresses, epsilon decays, encouraging the agent to exploit learned policies while maintaining some level of exploration.

The action space in the proposed RL model consists of price adjustments that each charging station can apply within a predefined range. According to [8], the average price per kWh across all stations is approximately 0.99 CNY/kWh during daytime and 0.93 CNY/kWh during nighttime. The minimum price observed is 0.54 CNY/kWh, and the maximum price is 1.47 CNY/kWh. These statistics provide the context for the action space, where each action represents a relative price change within the range of $\{-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3\}$, allowing the RL agent to adjust station-specific pricing dynamically.

During training, the Q-network is updated using the Bellman equation, where the target Q-value is computed from the reward and the estimated future Q-value of the next state. In this data-driven framework, the absence of a temporal sequence of state transitions within each batch differs from the traditional interpretation of the Q-value. Instead of

reflecting long-term reward expectations over an episode, the Q-value in this setup approximates the immediate cumulative reward associated with the current state-action pair, as derived from the training data. The simulation parameters used in training the RL model are summarized in Table I.

Since the reward function defined in (1) is designed as a reciprocal function of penalties, it is not upper-bounded, as reducing penalties to near-zero results in unbounded growth of the reward, unlike typical bounded reward functions. As a result, the Q-values shown in Fig. 5 tend to rise as the agent improves its policy, indicating advancements in both system balance and load distribution. Moreover, the target network is updated every 20 episodes to stabilize the training process by providing fixed Q-value targets over multiple updates. This results in the step-like increases observed in the Q-value graph (Fig. 5). The increase in cumulative reward values, as depicted in Fig. 7, shows that the agent is successfully learning to optimize its policy.

The training loss curve, presented in Fig. 6, illustrates the learning dynamics of the Q-network over episodes. The rapid reduction in loss during the initial episodes reflects the agent's effective adaptation to the environment and convergence toward meaningful policy updates. Subsequent oscillations in the loss are attributed to target network updates, which periodically shift the optimization objective.

Additionally, the average variance of the utilization, shown in Fig. 8, highlights the agent's success in minimizing imbalances in utilization across stations while preventing overloading and underutilization across the charging network.

The penalty comparison for two different weightings of the overload penalty ($\lambda = 1$ and $\lambda = 10$) is depicted in Fig. 9. The results highlight the trade-off between variance and overload penalties as λ is adjusted. A higher λ shifts the balance of the reward function toward addressing overload penalties. Specifically, this prioritization results in higher variance penalties due to imbalances in utilization across stations. These observations demonstrate the flexibility of the proposed framework in adapting to varying operational objectives by tuning the λ parameter.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Discount factor (γ)	0.99
Learning rate	1×10^{-4}
Batch size	32
Target network update frequency	20 episodes
Epsilon decay factor	0.95
Minimum epsilon (ϵ_{\min})	0.1
Overload penalty weight λ	1
Number of episodes	100

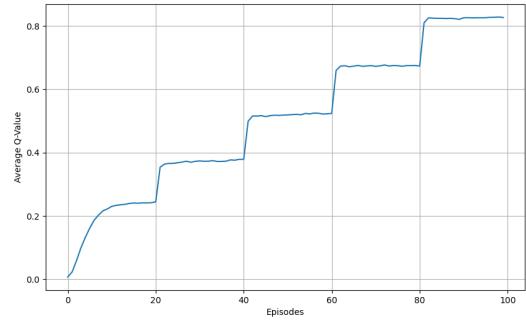


Fig. 5. Average Q-Values over episodes

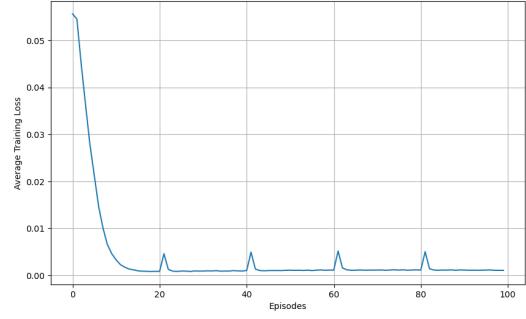


Fig. 6. Average training loss over episodes

V. CONCLUSION

This paper proposes a reinforcement learning-based framework for dynamic pricing and load balancing in EV charging networks. By leveraging a pre-trained price-adjusted graph neural network (PAG) for demand prediction and a reward function designed to minimize both utilization variance and overload penalties, the framework demonstrates its ability to optimize resource allocation and achieve equitable station utilization. The simulation results validate the efficacy of the proposed approach, with significant improvements observed

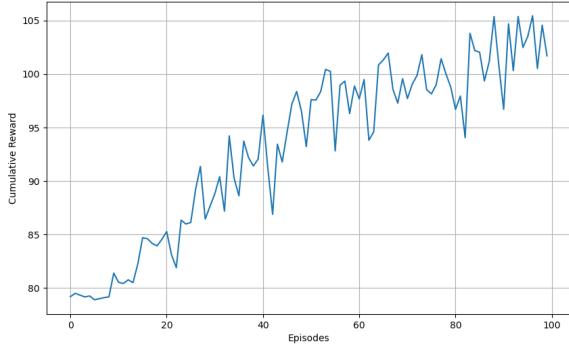


Fig. 7. Cumulative rewards over episodes

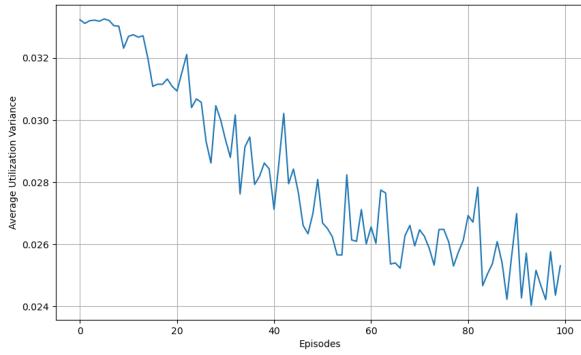


Fig. 8. Average variance of utilization over episodes

in balancing network demand and preventing capacity overloads across stations. Additionally, the flexibility of the reward function is highlighted through its adaptability to varying operational priorities by adjusting the penalty weighting factor λ . Future work will focus on extending this framework to incorporate real-time constraints, such as power resource management and user preferences, further enhancing the applicability of the proposed solution to real-world scenarios.

REFERENCES

- [1] Z. Liu, F. Wen, and G. Ledwich, “Optimal planning of electric-vehicle charging stations in distribution systems,” *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 102–110, 2012.
- [2] Z. Ye, Y. Gao, and N. Yu, “Learning to operate an electric vehicle charging station considering vehicle-grid integration,” *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3038–3048, 2022.
- [3] H. J. Feng, L. C. Xi, Y. Z. Jun, Y. X. Ling, and H. Jun, “Review of electric vehicle charging demand forecasting based on multi-source data,” in *Proceeding IEEE Sustainable Power and Energy Conference (iSPEC)*, 2020, pp. 139–146.
- [4] S. Ai, A. Chakravorty, and C. Rong, “Household EV charging demand prediction using machine and ensemble learning,” in *Proceeding 2018 IEEE International Conference on Energy Internet (ICEI)*, 2018, pp. 163–168.
- [5] S. Su, Y. Li, Q. Chen, M. Xia, K. Yamashita, and J. Jurasz, “Operating status prediction model at EV charging stations with fusing spatiotemporal graph convolutional network,” *IEEE Transactions on Transportation Electrification*, vol. 9, no. 1, pp. 114–129, 2022.
- [6] Z. Hu, K. Zhan, H. Zhang, and Y. Song, “Pricing mechanisms design for guiding electric vehicle charging to fill load valley,” *Applied Energy*, vol. 178, pp. 155–163, 2016.
- [7] A. K. Kalakanti and S. Rao, “Dynamic pricing for electric vehicle charging,” *arXiv preprint arXiv:2408.14169*, 2024.
- [8] H. Kuang, X. Zhang, H. Qu, L. You, R. Zhu, and J. Li, “Unraveling the effect of electricity price on electric vehicle charging behavior: A case study in shenzhen, china,” *Sustainable Cities and Society*, vol. 115, p. 105836, 2024.
- [9] Y. Yang, H.-G. Yeh, and R. Nguyen, “A robust model predictive control-based scheduling approach for electric vehicle charging with photovoltaic systems,” *IEEE Systems Journal*, vol. 17, no. 1, pp. 111–121, 2023.
- [10] S. Sanami, H. Mosalli, Y. Yang, H.-G. Yeh, and A. G. Aghdam, “Demand forecasting for electric vehicle charging stations using multivariate time-series analysis,” *to appear in Proceedings of American Control Conference (ACC)*, 2025.
- [11] M. Majidpour, C. Qiu, P. Chu, H. R. Pota, and R. Gadh, “Forecasting the EV charging load based on customer profile or station measurement?” *Applied energy*, vol. 163, pp. 134–141, 2016.
- [12] Y. Lu, Y. Li, D. Xie, E. Wei, X. Bao, H. Chen, and X. Zhong, “The application of improved random forest algorithm on the prediction of electric vehicle charging load,” *Energies*, vol. 11, no. 11, p. 3207, 2018.
- [13] K. Lu, W. Sun, C. Ma, S. Yang, Z. Zhu, P. Zhao, X. Zhao, and N. Xu, “Load forecast method of electric vehicle charging station using SVR based on GA-PSO,” in *IOP Conference Series: Earth and Environmental Science*, vol. 69, no. 1. IOP Publishing, 2017, p. 012196.
- [14] “Github - intelligentsystemslab/st-evcdp: A real-world dataset for ev-related research, e.g., spatiotemporal prediction and urban energy management.” [Online]. Available: <https://github.com/IntelligentSystemsLab/ST-EVCDP>
- [15] H. Qu, H. Kuang, Q. Wang, J. Li, and L. You, “A physics-informed and attention-based graph learning approach for regional electric vehicle charging demand prediction,” *IEEE Transactions on Intelligent Transportation Systems*, 2024.

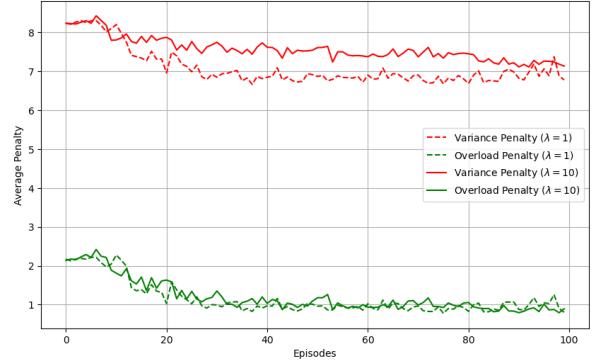


Fig. 9. Comparison of penalty components for $\lambda = 1$ and $\lambda = 10$.