Intelligent Layout and Optimization of EV Charging Stations: Initial Configuration via Enhanced K-Means and Subsequent Refinement through Integrated GCN

Haojie YANG, Miao LIU, Jingjing ZOU, Rongxin XU, Jiale HUANG, Peng GENG *

School of Communication and Artificial Intelligence, School of Integrated Circuits, Nanjing Institute of Technology, Nanjing 211167, China

*gengpeng@njit.edu.cn

Submitted November 10, 2024 / Accepted January 6, 2025 / Online first March 18, 2025

Abstract. *This paper proposes an optimization model for the layout of EV charging stations, aiming to ensure a wide and efficient service area to meet the increasing demand for charging. Through an in-depth study of the deployment optimization of EV charging stations, a layout algorithm based on K-Means and simulated annealing is first introduced to determine the optimal locations for new charging stations. Building on this, a layout optimization algorithm utilizing a Residual Attention Graph Convolutional Network (RAGCN) is proposed, which leverages the efficient learning capability of Graph Convolutional Networks (GCN) on graph-structured data to learn and obtain the best layout for charging stations. Finally, the effectiveness of the model is validated in Nanjing, Jiangsu Province. The results show that the optimized layout of charging stations, which added 493 new stations in high-demand areas such as business districts and corporate enterprises, significantly enhances the convenience and utilization rate of charging for EV users. Additionally, sensitivity analysis and ablation experiments based on Points of Interest (POI) data are conducted to evaluate the impact of various POI features on the layout of charging stations and to explore the contribution of different model components to classification performance.*

Keywords

Electric vehicle charging station, points of interest, fused residual network, attention mechanism, graph convolutional neural network

1. Introduction

With the rapid growth of the electric vehicle (EV) market [1], countries are gradually promoting the transition from fuel vehicles to clean energy vehicles driven by environmental protection policies. However, the popularization of EVs faces many practical challenges, among which the improvement of charging infrastructure is a key factor. At this stage, the layout and number of charging stations limit the convenience of EVs to a large extent [2], resulting in users encountering problems such as mileage anxiety and long charging waiting times in actual use.

The current development of EV charging stations still has some significant problems. Firstly, the lack of charging network coverage is an important obstacle [3–4]. In many cities, charging stations are mainly concentrated in city centers or commercial areas, while charging stations in suburbs and per-urban areas are relatively scarce, leading to drivers facing charging difficulties when traveling across regions. This uneven distribution of charging networks significantly affects the long-distance use and user experience of EVs. Secondly, the unbalanced utilization of charging equipment also causes problems for users [2]. In some hot spot areas, the peak charging demand hours can result in over-utilization of charging equipment and even the need to wait in queues, while in other areas there may be a low utilization problem. This imbalance not only affects the user experience but also increases the operating costs and weakens the economic benefits of the investment in charging facilities.

There are significant economic challenges to building EV charging stations in cities where land resources are tight and land prices are high. As charging stations require a larger space to accommodate charging piles and vehicles, this directly leads to a significant increase in the cost of constructing charging stations in areas with high land costs, such as city centers or busy business districts. In addition, the scarcity of land may also limit the size and number of charging stations, further exacerbating the conflict between supply and demand for charging facilities. High construction costs not only require operators to make a larger capital investment at the initial stage but also increase long-term operating costs, which may be passed on to consumers by raising the cost of charging services, thereby affecting the popularity of EVs and the charging experience of users.

DOI: 10.13164/re.2025.0079

Given the above problems, the rational layout of charging stations, improving the utilization rate of charging facilities and optimizing the user experience has become an important part of promoting the further development of EVs. To this end, this paper proposes an intelligent charging station layout method based on improved K-means and fused GCN, aiming at analyzing the changes in urban geographic distribution and user demand, and providing a scientific and reasonable layout plan for EV charging infrastructure to cope with the rapid growth of EV demand in the future, and to achieve optimal allocation and efficient use of charging facility resources.

The focus of this study is not on the technical details of various aspects of charging station construction but rather on the overall planning process. The layout of EV charging stations is divided into two stages: the deployment of charging stations and the optimization of their layout. In the first stage, demand points for EV charging stations are clustered based on POI demand data, which are then treated as the deployment points for electric vehicle charging stations. In the second stage, the comprehensive utilization rate of the stations is considered as a key criterion to further optimize the layout for EVs.

In recent years, scholars have conducted extensive research on the deployment and layout optimization of electric vehicle (EV) charging stations, taking into account many factors such as grid load [5], mileage, public charging station distribution and passenger distribution [6]. Meanwhile, researchers have applied several different algorithms to analyze and solve this complex problem. Based on the methods employed, these studies can be broadly classified into the following categories: evolutionary algorithmic approaches, hybrid optimization approaches, statistical approaches, and data-driven approaches. The comprehensiveness and novelty of this work are summarized in Tab. 1 through a comparative analysis of existing studies and proposed methods in the literature.

Evolutionary algorithms are widely used in charging station layout optimization, mainly including genetic algorithms and particle swarm optimization. Literature [7] proposes a multi-intelligence system approach based on genetic algorithms for optimizing the location configuration of charging stations in a given city. The method is able to integrate social network activity and mobility information to estimate the optimal configuration location, but it is highly dependent on exhaustive travel data and there are difficulties in data acquisition for practical applications. Literature [8] proposes a multi-objective genetic algorithm model aiming to minimize the total cost of the charging network and maximize the quality of service, which, although it strikes a balance between cost and quality of service, faces scalability issues when dealing with large-scale data and is computationally inefficient. Literature [9]–[10] used genetic algorithms and evolutionary algorithms to optimize the deployment of fast charging stations, respectively, and although the optimization is effective, the parameters need to be adjusted frequently in highly dynamic urban environments, which increases the computational complexity and the difficulty of practical application. The optimization algorithm MDDC proposed in the literature [11] effectively reduces the number of deployments of very fast charging stations and improves the accuracy of the deployment location by simulating the dynamic changes of electric vehicles and designing optimization rules. The evolutionary algorithm approach excels in coping with complex optimization problems but still has significant shortcomings in data dependency and computational efficiency, which limits its wide application.

Hybrid optimization approaches combine multiple algorithms or models in order to enhance optimization effectiveness and solution efficiency. Literature [12] combines the improved Huff gravity model with a two-tier optimization strategy to consider market competition factors to determine the location, size and pricing of charging stations, which demonstrates a high degree of flexibility but requires precise calibration of the gravity model parameters to ensure the accuracy of the results. Literature [13] uses a hybrid heuristic algorithm GA-BPSO for charging station optimization design, which improves the solution efficiency and optimization results, however, the computational overhead is large when facing complex dynamic demands, which affects the efficiency of practical applications. Literature [14] combines the Enhanced Heuristic Descent Gradient (EHDG) algorithm and Voronoi diagram to optimize the layout of fast charging stations for electric buses, which, despite the significant results in reducing energy consumption and operating costs, needs to be strengthened in terms of adaptability in non-uniform urban areas. Hybrid optimization methods have made some progress in improving solution quality and computational efficiency, but challenges remain in terms of data dependency and computational resource requirements, affecting their applicability in different urban environments.

Statistical methods have taken an important role in charging demand forecasting and layout optimization, mainly including regression analysis. Literature [15] uses a regression analysis model to predict charging demand, which is able to better capture the trend of user behavior but is inadequate in dealing with complex non-linear relationships and is difficult to accurately reflect the dynamically changing demand. Literature [16] conducted charging demand forecasting through time series analysis methods, which performs well in short-term forecasting but has low accuracy in predicting long-term trends, limiting its application in long-term planning. Statistical methods have some advantages in dealing with basic demand forecasting and preliminary layout optimization, but they have limitations in coping with complex non-linear relationships and dynamic changes, making it difficult to meet the needs of the rapidly developing electric vehicle market.

Data-driven approaches make use of big data and machine learning techniques to optimize charging station layout and demand forecasting. Literature [17] proposed a recurrent

Reference	Algorithm Used	Network Structure	Data Dependency
[7]	Genetic Algorithm	Multi-Intelligence System	High (requires exhaustive travel data)
[8]	Multi-objective Genetic Algorithm	None	High (relies on large-scale data)
[9], [10]	Genetic and Evolutionary Algorithms	None	High
[11]	MDDC Optimization Algorithm	Dynamic Change Simulation	Moderate
[12]	Improved Huff Gravity Model + Two-Tier Strategy	Gravity Model	Moderate (requires parameter calibration)
[13]	GA-BPSO Hybrid Heuristic Algorithm	None	High
[14]	EHDG Algorithm + Voronoi Diagram	None	Moderate
[15]	Regression Analysis	None	Low
[16]	Time Series Analysis	None	Low
[17]	Recurrent Neural Network (Reinforcement Learning)	RNN-based Framework	High (requires extensive training data)
[18]	Trip Reconstruction + Two-Step Optimization	None	High (requires mobile signal data)
[19]	GPS Trajectory Optimization Model	None	High (relies on regional data)
[20]	CCSSA	Multi-objective Optimization Model	High (chaotic initialization parameters)
proposed	Simulated Annealing + K-Means + GAGCN	K-Means+Graph	Low (only POI)

Tab. 1. Summary of algorithms and their characteristics.

neural network framework based on reinforcement learning to optimize the charging station deployment with a limited budget, which improves the quality of service and reduces the user waiting time, however, the method requires a large amount of training data and computational resources, which increases the complexity of implementation. Literature [18] uses mobile phone signal data to determine the optimal location of charging stations through trip reconstruction and a two-step optimization model, although the approach is novel, the reliance on mobile phone signal data may face challenges in privacy protection and data access. Literature [19] developed an optimization model based on GPS trajectory data of electric taxis, which validated the effectiveness of charging station location selection through the case of Shenzhen, but data access may be limited in certain regions, limiting its generalisability. The multi-objective optimization model based on the Chaotic Cat Swarm Simulated Annealing Algorithm (CCSSA) proposed in the literature [20] maximizes the carbon emission reduction benefits while minimizing the construction cost and the cost of accessing the grid, and despite demonstrating innovation, the reliance on chaotic initialization parameters requires extensive tuning, which reduces its ease of use. Data-driven approaches excel in handling complex data and optimization results, but there are still some limitations in data acquisition and model complexity, which affects their wide application in different urban environments.

Despite the significant progress made by the aforementioned methods in the deployment and layout optimization of EV charging stations, several shortcomings remain. Firstly, most of the methods rely on exhaustive datasets, and their applicability is limited in areas where data are scarce or difficult to access. Second, many optimization algorithms have high computational overheads when dealing with large-scale or highly dynamic data, which affects the efficiency of practical applications. In addition, some methods lack sufficient flexibility and adaptability when facing dynamically changing urban environments or diverse demands. Finally, many advanced optimization algorithms require frequent parameter adjustments, which increases the difficulty and complexity of practical applications. To address the above shortcomings, this paper proposes a new EV charging station layout optimization algorithm that combines K-Means clustering and simulated annealing, which only requires point-of-interest (POI) data to determine the optimal location and significantly reduces the dependence on detailed travel data. Meanwhile, this paper further proposes a RAGCN model for optimizing the layout of EV charging stations. The model not only optimizes the layout process, but also improves the adaptability and scalability in diverse urban environments by simplifying the data collection process, overcoming the shortcomings of existing methods in terms of data dependency, computational efficiency and adaptability, and providing a more practical and efficient solution for EV infrastructure planning.

Our main contributions include the following:

- An EV charging station layout algorithm based on K-Means and simulated annealing algorithm is proposed, which aims to solve the local optimal problem that may result from the selection of initial clustering points in the K-Means clustering process. By introducing the simulated annealing algorithm, the algorithm can globally optimize the selection of initial clustering points, avoiding the limitations of traditional K-Means under complex data distribution.
- A GCN optimization algorithm for EV charging station layout optimization incorporating residual networks and graph attention mechanism is proposed. This algorithm improves the performance in urban charging station layout optimization by combining residual networks

and GCNs and using the graph attention mechanism. The introduction of residual networks solves the problem of gradient vanishing in deep neural network training, enabling the network to learn more complex features. The experimental results show that the algorithm outperforms the traditional layout optimization methods in several metrics and can provide strong support for the sustainable development of urban transport.

The rest of the paper is organized as follows. Section 2 introduces the research problem and framework. Section 3 proposes an electric vehicle charging station layout algorithm based on K-Means and simulated annealing algorithm. Section 4 proposes an EV charging station layout optimization algorithm based on graph convolutional neural network fusing residual network and graph attention mechanism. Section 5 conducts empirical experiments to validate the proposed method. Section 6 provides the results of POI sensitivity analysis and ablation experiments. Finally, the paper is summarized in Sec. 7.

2. Problem Formulation and Research Ideas

2.1 Description of the Problem

The aim of this paper is to address the current problem of deploying and optimizing charging infrastructure for EVs. With the rapid growth of the EV market, the demand for charging stations has increased. However, the existing charging station layout often fails to meet the charging demand during peak hours, resulting in long charging waiting times and low user satisfaction. In addition, the irrational distribution of charging stations also results in a waste of resources. In order to effectively address these challenges, this paper explores how to rationally deploy and optimize EV charging stations through scientific data analysis and optimization models to improve the service efficiency and user experience of the charging network.

2.2 Research Framework

The research framework of this paper revolves around two core components of EV charging station deployment and optimization, which are divided into the following two parts as shown in Fig. 1.

- Deployment part: in the deployment strategy, this paper uses POI data and Geographic Information System (GIS) technology to identify potential high-demand areas. Cluster analysis is used to select the optimal location of charging stations to ensure that demand hot spots can be effectively covered while taking into account cost and accessibility.

- Optimization part: in the optimization process, the study uses deep learning techniques to analyze the environmental characteristics of the charging stations, and optimizes and filters out reasonable charging station layouts by learning the POI characteristics of existing charging stations with high utilization rates.

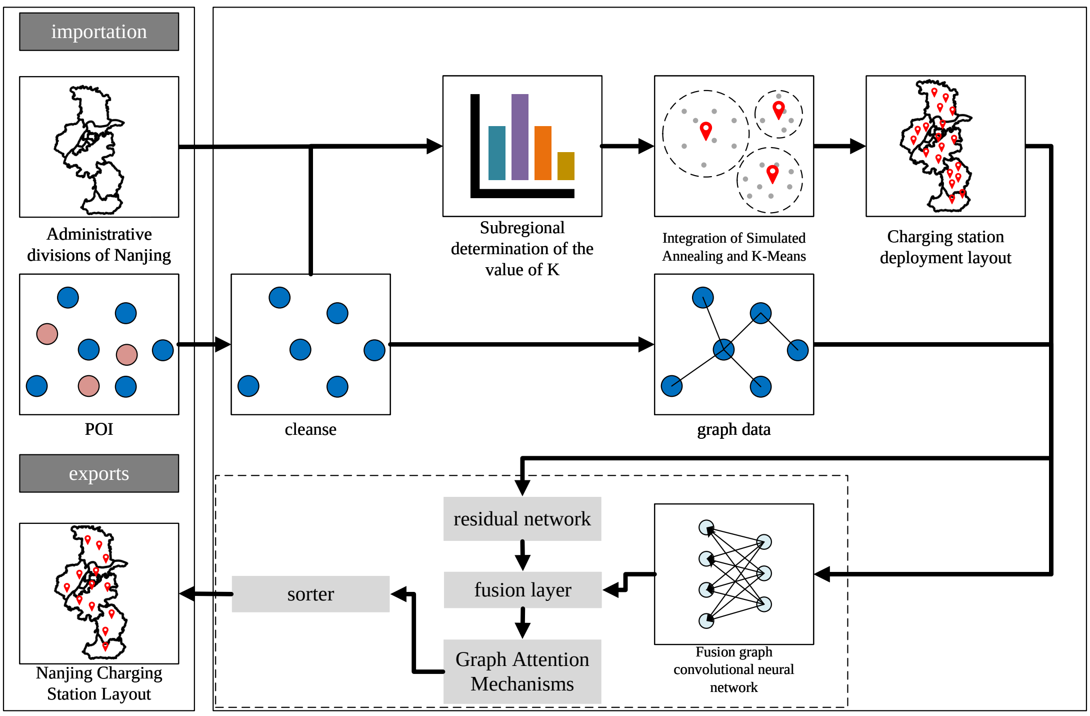

Fig. 1. Research framework.

3. EV Charging Station Deployment Algorithm

3.1 Initial Screening of POI Types

In order to facilitate data acquisition and improve the generality of the proposed method, this paper uses single-source POIs as the data source. The two districts and three centres to divide the POI data of important places in Nanjing can be divided into primary classification and secondary classification, and the specific POI division is shown in [21].

Residential areas, office areas, and commercial centres are the core areas of EV charging station construction. With long dwell time and stable charging demand, these locations can facilitate users' daily charging, alleviate anxiety, and efficiently utilise space to promote the popularity of EVs and sustainable urban development. Meanwhile, industrial centres, as the operation bases of logistics vehicles and commercial vehicles, can meet the operational needs, reduce the use of fuel vehicles and lower carbon emissions by building charging stations. Recreational centres such as parks and stadiums provide convenient charging services for residents and tourists, supporting green mobility. Reasonable deployment of charging stations can fully cover the needs of multiple scenarios and accelerate the promotion of electric vehicles.

3.2 Fusion of K-Means and Simulated Annealing Algorithms

Since the clustering center of the K-Means algorithm tends to fall into the local optimum and leads to unsatisfactory clustering results, this paper proposes a fusion of K-Means and a simulated annealing algorithm to optimize the clustering center of the K-Means algorithm to improve the clustering effect.

The model in this paper uses the silhouette coefficient to determine the value of k for each region. Silhouette Coefficient (SC) is a measure of the quality of clustering results. The value of the Silhouette Coefficient is in the range of $[-1, 1]$, the larger the value, the better the clustering effect. Initialize the clustering center $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k\}$ for each region based on the k-value of K-Means determined in the previous stage, where k is the number of clusters. Also initialize the initial temperature T_0, the minimum temperature $T_{\min}$ and the cooling rate α.

The goal of K-Means is to minimize the objective function as shown in (1):

$$J = \sum_{i=1}^{k} \sum_{\mathbf{x} \in \mathbf{C}_i} \|\mathbf{x} - \mathbf{c}_i\|^2 \tag{1}$$

where $\mathbf{x}$ is a data point, $\mathbf{c}_i$ is the center of the ith cluster, and $\mathbf{C}_i$ is the set of data points contained in the ith cluster.

The simulated annealing is used to optimize the amount of change in the objective function, and the cost difference between the new solution and the current solution is calculated $\Delta E = J_{\text{new}} - J_{\text{current}}$. From the cost difference ΔE, decide whether to accept the new solution or not according to the Metropolis criterion, and the formula is shown in (2):

$$P = \exp\left(-\frac{\Delta E}{kT}\right) \tag{2}$$

where k is Boltzmann's constant, T is the temperature, determined from $T = \alpha \cdot T$, and $\alpha \in (0, 1)$. When $\Delta E < 0$, then the new solution is accepted as the current solution. When $\Delta E > 0$, then the new solution is accepted with probability P . In this paper, the complete steps of fusing K-Means and the simulated annealing algorithm are shown in Algorithm 1.

Algorithm 1 K-Means with simulated annealing.

Input: Coordinates (longitude, latitude), number of clusters k, initial temperature T_0, cooling rate α, minimum temperature $T_{\min}$, and iterations per temperature n_{iter}
Output: Optimized cluster centroids and assigned cluster labels
Initialize K-Means with k clusters
Fit the initial model to obtain initial centroids C_{best}
Set $T \leftarrow T_0$
while $T > T_{\min}$ **do**
  for $i \leftarrow 1$ to n_{iter} **do**
    Generate perturbed centroids $C_{\text{new}} \leftarrow C_{\text{best}}$ + random noise
    Evaluate C_{new} using clustering inertia I_{new}
    if $I_{\text{new}} < I_{\text{best}}$ **or** $\exp((I_{\text{best}} - I_{\text{new}})/T) >$ rand() **then**
      $C_{\text{best}} \leftarrow C_{\text{new}}$
    end if
  end for
  Update temperature $T \leftarrow \alpha T$
end while
Assign clusters using optimized centroids C_{best}
return C_{best}, cluster labels

4. Optimisation Algorithm for EV Charging Stations

4.1 Data Processing

In this paper, the collected POI data are first categorised and classified according to pre-defined categories. These categories are shown in Tab. 2 and include, but are not limited to, restaurants and food, companies and businesses, consumer shopping, transport facilities, and financial institutions. Each POI is assigned to the corresponding category based on its primary function. After the classification was completed, these category labels were processed with solo thermal coding to facilitate subsequent data processing and analysis. Solo thermal coding is a method of converting categorical variables into a form that can be efficiently processed by a machine learning model, which creates a binary column for each category, with the value of the column for the corresponding category being 1, and the value of all other columns being 0. This coding simplifies the processing of the inputs to the model and helps to improve the efficiency and accuracy of the model training.

Feature number	Type	Feature number	Type
1	Catering	8	Tourist Attractions
2	Corporate	9	Automobile
3	Shopping	10	Business Residence
4	Transport Facilities	11	Life Service
5	Financial Institutions	12	Leisure & Entertainment
6	Hotel Accommodation	13	Healthcare
7	Science, Education and Culture	14	Sports & Fitness

Tab. 2. POI classification.

4.2 RAGCN-Based Electric Vehicle Charging Station Layout Optimisation Algorithm

4.2.1 Conventional Graph Convolutional Neural Networks

GCN is a neural network architecture specifically designed to process graph data. Unlike traditional Convolutional Neural Networks (CNNs) that mainly deal with Euclidean data, GCNs are able to perform convolution-like operations from graph data, thus overcoming the limitations of traditional deep learning algorithms when dealing with non-regular graph data.

GCN is executed with the input of a feature matrix, which has the shape, that represents the number of nodes in the graph, and represents the dimension of each node feature. The feature matrix contains the feature information of all the nodes in the graph, which defines the data characteristics of the graph. GCN also needs to input a neighbour matrix, which describes the connection relationship between the nodes in the graph and defines the structural characteristics of the graph. In the graph convolution process of GCN's layers, each node's features are aggregated with its neighbouring node's features to generate new features. This process of matrix can be seen as the mutual circulation of information from node to node. Each convolution operation updates the node's feature matrix so that deeper information about the graph structure can be captured and described. The layer-to-layer transfer of the neural network of the GCN is shown in (3):

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{3}$$

where $\mathbf{H}$ is the feature of each layer, $\mathbf{H}^{(l+1)}$ is the output matrix of the layer $l(l = 1, 2, \cdots, L)$, $\mathbf{H}^{(l)}$ is the input matrix of $l(l = 1, 2, \cdots, L)$, the matrix rows represent the number of nodes, and the matrix columns represent the POI node features corresponding to each node. In particular, the input layer, $\mathbf{H}$ is the feature matrix $\mathbf{X}$. σ is the nonlinear activation function, $\hat{\mathbf{A}} = \mathbf{A}+\mathbf{I}$ is the graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ adjacency matrix after adding connected edges to the graph, $\mathbf{A}$ is the adjacency matrix of the graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ reflecting the connectivity between nodes, $\mathbf{I}$ is the unit matrix. $\widehat{\mathbf{D}}$ is the degree matrix of $\hat{\mathbf{A}}$, recording the degree of each node (i.e., the number of edges connected to that node). $\mathbf{W}^{(l)}$ is the parameter matrix of the layer $l(l = 1, 2, \cdots, L)$ that can be trained and linearly transformed in the feature space.

4.2.2 RAGCN

In multilayer GCNs, node features are updated by aggregating the features of its neighbouring nodes. However, as the number of network layers increases, this aggregation process often leads to a gradual blurring of node features. This phenomenon is particularly evident in deep networks, as the loss of information at each layer accumulates and eventually leads to node features losing their original important information. In addition, as the depth of the network increases, the feature representations of different nodes tend to converge in what is known as the over-smoothing problem. This consistency of features reduces the differentiation between nodes, thus reducing the ability of the model to classify or predict different nodes.

To address the above problem, this paper proposes an improved GCN model that introduces residual networks and graph attention mechanisms. The design of residual connectivity allows information to be passed directly between multiple layers of the network, bypassing the traditional layer-by-layer aggregation process. This design effectively reduces the loss of information in the transfer process, allowing the deep network to retain more original feature information.

Meanwhile, the introduction of the graph attention mechanism brings great improvement to GCN. Unlike the traditional equal weight aggregation, the graph attention mechanism allows the model to dynamically learn which neighbour nodes are more important for the current node. In this way, the model can selectively aggregate those features that are more contributing to the current task, thus enhancing the feature differentiation among nodes and improving the expressiveness and differentiation of the model.

In this paper, the residual connection is introduced after the graph convolution layer, the traditional graph convolution operation can be expressed as shown in (3), and the output of the lth layer is updated as shown in (4) after the introduction of residual connection:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) + \mathbf{H}^{(l)} \tag{4}$$

where $\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(\mathbf{l})}$ is the input feature, and σ and $\mathbf{H}^{(l)}$ are the weight matrix and bias term of the residual connection, respectively. In this way, the introduction of residual join enables the original features $\mathbf{H}^{(l)}$ to be directly added with the output of graph convolution, which reduces the loss of information and enhances the retention of information in the deep network.

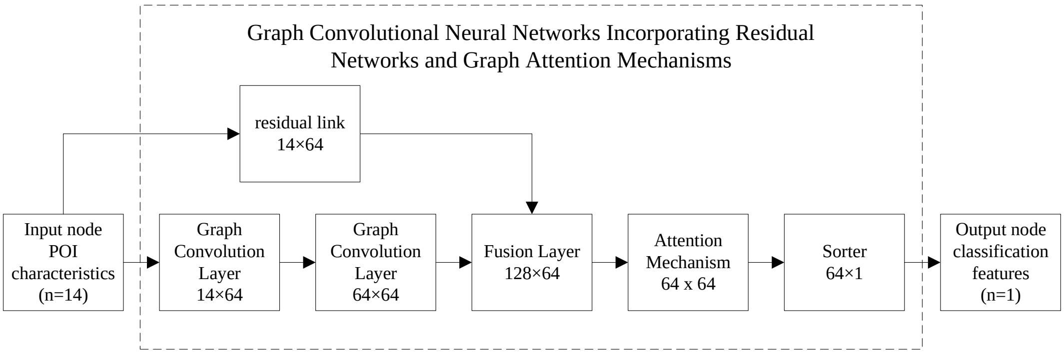

Fig. 2. RAGCN structure diagram.

To further process the features, after the residual join, the model adds the graph attention layer to enhance the generalization ability of the model, as shown in (5)–(6):

$$\alpha_{ij} = -\frac{\exp\left(\mathrm{LeakyReLU}\left(a^{\mathrm{T}}\|\mathbf{W}\mathbf{h}_i\|\mathbf{W}\mathbf{h}_j\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\mathrm{LeakyReLU}\left(a^{\mathrm{T}}\|\mathbf{W}\mathbf{h}_i\|\mathbf{W}\mathbf{h}_k\right)\right)} \tag{5}$$

$$\mathbf{h}_{ir} = \sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}\mathbf{W}\mathbf{h}_j\right) \tag{6}$$

where α_{ij} is the attention coefficient of the node i to its neighbour nodes j, $\mathcal{N}(i)$ denotes the set of neighbour nodes of the node i, $\mathbf{W}$ is the weight matrix used to transform the feature vectors, $\mathbf{h}_i$ and $\mathbf{h}_j$ are the feature vectors of node i and node j respectively, LeakyReLU is the activation function, $\mathbf{h}_{ir}$ is the updated feature vector of node i, σ is the activation function. In this paper, the structure of the graph convolutional neural network integrating residual network and graph attention mechanism is shown in Fig. 2.

5. Empirical Research

5.1 Purpose, Rationale, and Process

In this experiment, an EV charging station deployment model incorporating a simulated annealing algorithm and K-Means is constructed, and on this basis, a graph convolution neural network EV charging station layout optimization model incorporating residual network and graph attention mechanism is constructed, aiming at balancing the problem of irrational distribution of charging stations through layout optimization and then enhancing the utilization rate of EV charging stations. Firstly, POI data of important places in Nanjing were selected, and abnormal or incomplete data were deleted, and then the POI data were divided into 11 regions according to the administrative area of Nanjing, and the optimal k-value within each study sub region was determined using the contour coefficient. Next, a fusion simulated annealing algorithm and K-Means were used to find clustering centers for each study sub-region. Subsequently, the clustering center is used as a candidate charging station deployment point to find its POI features, and the candidate charging stations are classified using a graph convolution neural network that fuses residual networks and graph attention mechanisms to optimism the charging station deployment points obtained from the clustering center, and ultimately to determine the optimized layout of EV charging stations.

5.2 Research Condition

This experiment is written in PyCharm 2023 environment using Python language and executed on a computer with a 12th Gen Intel(R) Core (TM) i5-12500H processor and 32GB RAM. The point-of-interest (POI) data of Nanjing was obtained from the Gaode Map API, while the vector files of Nanjing administrative areas were obtained from the OpenStreetMap platform. All geographic data are in the WGS84 geographic coordinate system.

5.3 Data Processing

To ensure the accuracy of the geographic locations of EV charging stations obtained from the clustering of single-source POI data, the collected POI data were removed from the abnormal and erroneous data and classified according to [21] and Tab. 3, and the results of the processing are shown in Fig. 3.

Different colours in the figure represent different types of POIs in Nanjing, due to too many colours in Fig. 3a, this paper only gives examples to illustrate their correspondences, for example, in Fig. 3a, the dense city centre of Nanjing, cyan represents shopping malls and blue represents leisure centres. The correspondence between the colours and POI categories in Fig. 3b is shown in Tab. 3.

Category	**Colour**	**Category**	**Colour**
Dining	Red	Business	Green
Transportation	Cyan	Financial	Magenta
Education & Culture	Orange	Tourist Spots	Purple
Residential & Business	Pink	Lifestyle Services	Gray
Health Care	Lime	Fitness & Sports	Navy
Shopping	Blue	Automotive	Brown
Accommodation	Yellow	Leisure & Entertainment	Olive

Tab. 3. Correspondence between Fig. 3b POI colours and POI categories.

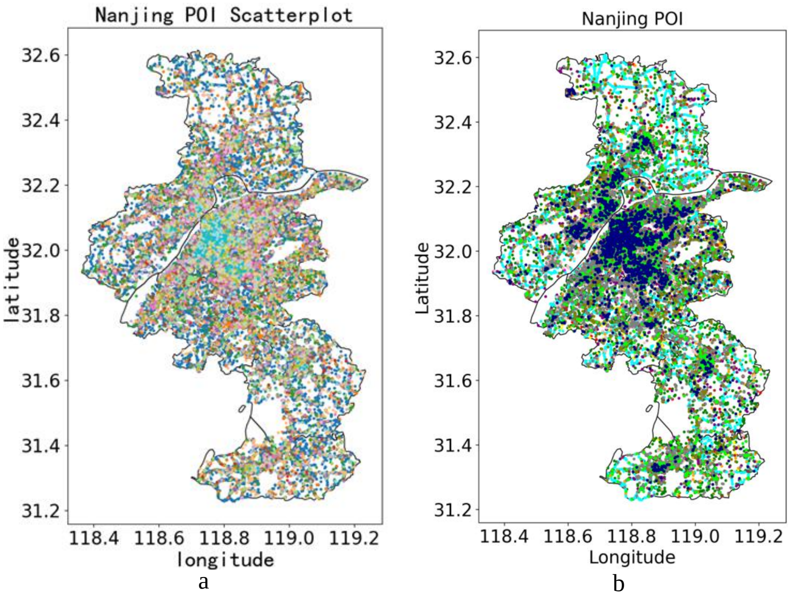

Fig. 3. Nanjing POI scatter plot.

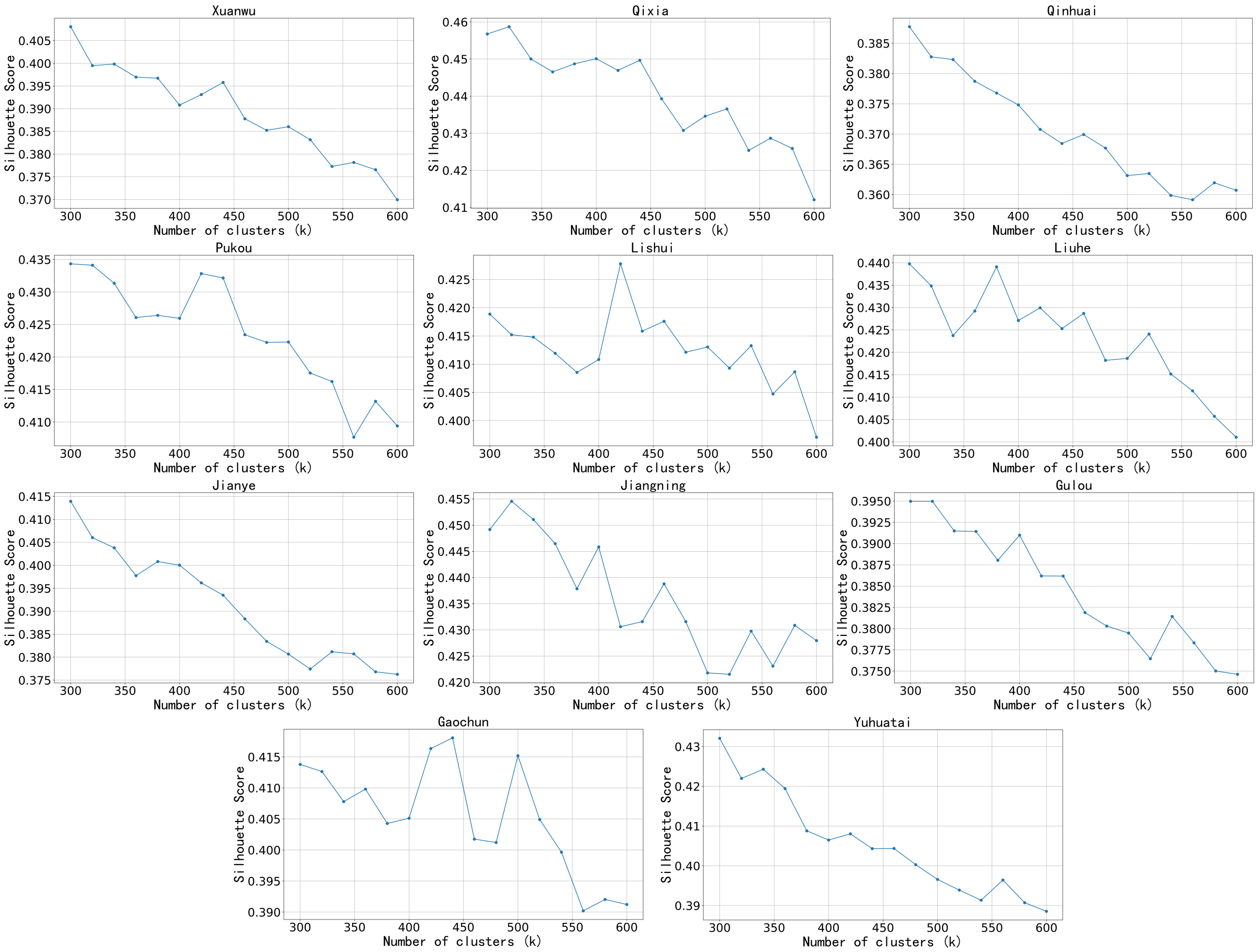

Fig. 4. Line graph of profile coefficients for each study area in Nanjing City.

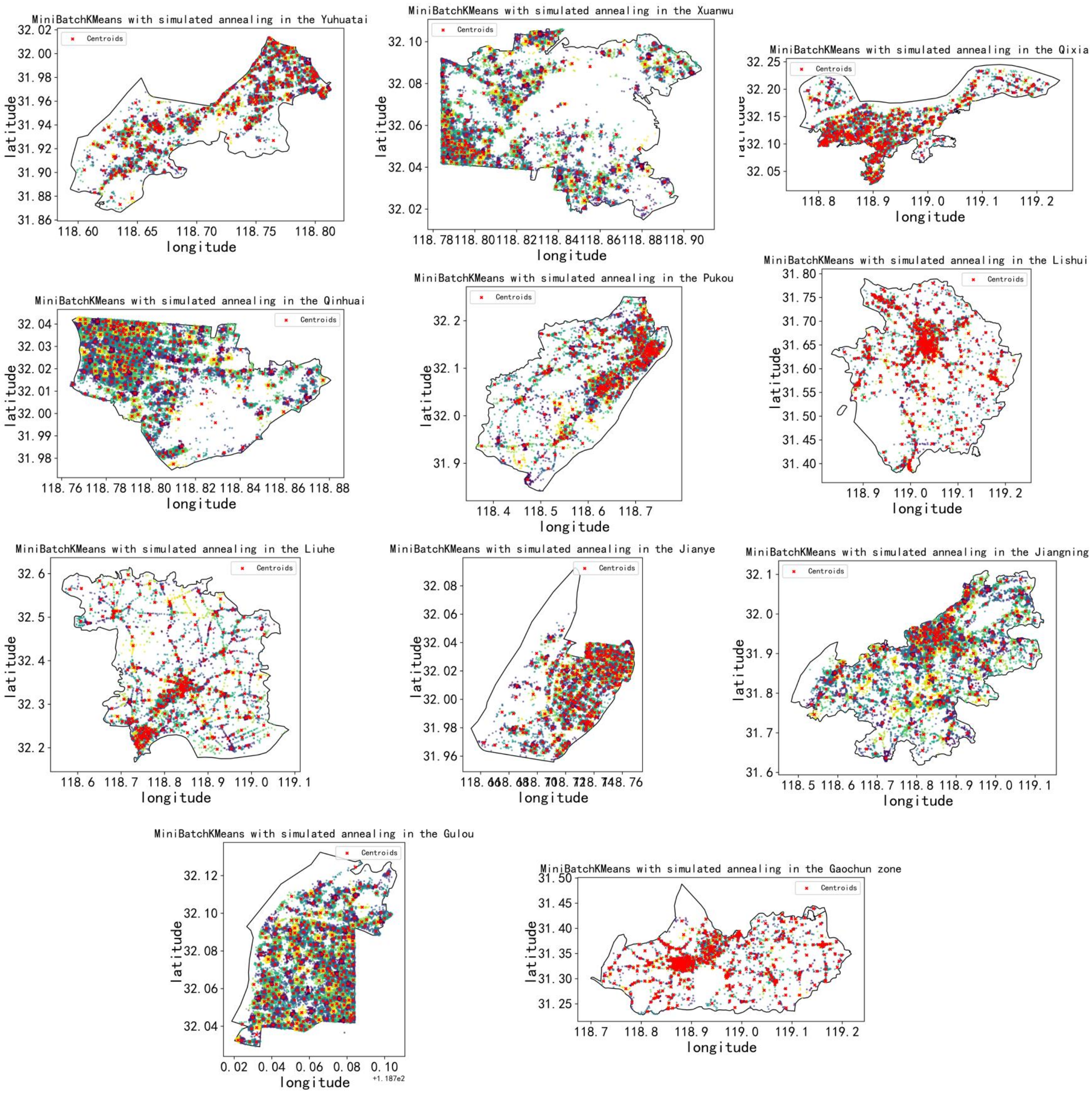

Fig. 5. Fusion of K-Means and simulated annealing EV charging station deployment model results.

5.4 Fusion of K-Means and Simulated Annealing EV Charging Station Deployment Models

In this paper, the contour coefficient is calculated to determine the optimal number of clusters (k-value) for each study area in Nanjing. The contour coefficient is an important indicator for evaluating the quality of clustering, which measures the effectiveness of clustering by considering both the tightness within clusters and the separation between clusters. The coefficient varies from –1 to 1, and the closer the value is to 1, the better the clustering is. Through repeated iterative analyses, this paper draws a line graph of the contour coefficient of each region in Nanjing, as shown in Fig. 4. According to the k-value at which the profile coefficient reaches its peak, this paper determines the final k-value for each study area in Nanjing, and the specific results are presented in Tab. 4.

Based on the k-values of each region shown in Tab. 4, this study used K-Means combined with a simulated annealing algorithm to optimize the layout of EV charging stations. The optimization results are shown in Fig. 5, where the red points mark the center position of each cluster, while the points in other colors indicate the POIs) contained in different clusters. Taking Xuanwu District in Fig. 5 as an example, it can be seen that the POI points are uniformly distributed around the cluster centers, which ensures that the distance from a POI point belonging to the same cluster to its cluster

Name of the study area	k-value
Xuanwu	300
Qinhuai	300
Jiangning	320
Pukou	300
Liuhe	300
Gulou	320
Gaochun	440
Lishui	420
Yuhuatai	300
Jianye	300
Qixia	320

Tab. 4. k-values for each of the study sub-areas in Nanjing City.

Colour	Classification
red	A real charging station exists and the model considers it reasonable.
green	The charging station is real and the model doesn't think it's reasonable.
blue	The charging station is a candidate and the model is reasonable.
grey	The candidate charging station and the model think unreasonable.

Tab. 5. Colour correspondence table.

center is the shortest distance from that point to all cluster centers. This layout effectively meets the needs of EV users seeking the shortest charging distance.

5.5 RAGCN-Based Electric Vehicle Charging Station Layout Optimization Model

Based on the EV charging station deployment model fusing K-Means clustering and simulated annealing algorithms, 3620 candidate locations of EV charging stations that meet the charging demand in the study area are identified and used as candidate nodes for analyzing the features of the area. For each node, we assign a unique identification code (ID) and use a Graph Convolutional Network (GCN) supervised learning model for label extraction and classification.

In the data preprocessing phase, the collected raw data was cleaned and organized to extract the classification labels of each node and to assess the potential of each node for charging station construction. The model was trained for 5000 iterations to obtain the training loss and classification accuracy of the nodes. The training loss of the model was reduced to about 0.11 and the classification accuracy reached about 0.95, showing the efficiency and accuracy of the model on this dataset.

In the results of node type prediction, there are 111 EV charging stations out of 160 existing EV charging stations in the test dataset with a combined utilization rate of 10% or more that are classified as reasonable EV charging stations by the model. There are 493 EV charging stations out of 3540 candidate EV charging station locations in the test dataset that are classified as reasonable EV charging stations by the model, as shown in Fig. 6 and Tab. 5.

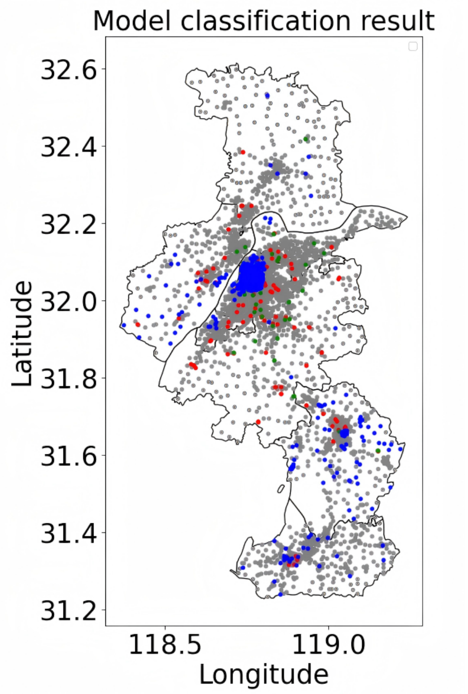

Fig. 6. Model prediction results.

Feature	Name	Average weights
Feature 1	Sports & Fitness	–0.1745
Feature 2	Medical & Health	–0.6869
Feature 3	Leisure & Entertainment	–0.0004
Feature 4	Life Services	–0.6412
Feature 5	Business & Residence	0.6685
Feature 6	Automobiles & Related	–0.2762
Feature 7	Tourist Attractions	0.2663
Feature 8	Science & Education	–0.7169
Feature 9	Hotels & Accommodation	–0.0095
Feature 10	Financial Institutions	–0.2717
Feature 11	Transportation & Facilities	–0.5997
Feature 12	Consumption & Shopping	–0.2733
Feature 13	Company & Enterprise	0.4083

Tab. 6. POI sensitivity analysis table.

6. Sensitivity and Ablation Analyses

6.1 Sensitivity Analysis

The graph convolutional neural network EV charging station layout optimization model, which incorporates residual networks and graph attention mechanisms, assigns different weights to different types of POI features during training, and a sensitivity analysis to assess the specific impact of these features on the EV charging station siting strategy can be used to understand which POIs have a significant impact on the EV charging station getting built. Table 6 details the average weights of each POI feature.

From the analysis in Tab. 6, it is obvious that business residential and corporate enterprises have a significant impact on the layout of EV charging stations due to their higher average weights of 0.67 and 0.40, respectively. Business residential areas usually have a high density of residential and commercial activities; therefore, EV owners are more likely to need frequent charging during their daily commute and life.

Meanwhile, corporate enterprise zones, another key layout point for EV charging stations, show that the need for EVs in the work environment in these areas cannot be ignored. Employees may be inclined to use charging facilities near their office locations or on their way to and from work, increasing the need and utility of EV charging stations in these areas.

These findings demonstrate the importance of considering the specific needs and characteristics of these critical areas when planning and constructing EV charging stations to ensure the effective layout and utilization of charging infrastructure. By targeting the construction of charging stations in these high-impact areas, not only can the coverage efficiency of the charging network be improved, but also the popularity and use of EVs can be promoted, further facilitating the electrification transition of urban transport.

6.2 RAGCN Ablation Analysis

In order to deeply explore the specific contribution of different model components to the classification performance, a series of ablation experiments are designed in this paper to analyse in detail the performance of three models, namely RAGCN, GCN+GAT and GCN+ResNet, on multiple datasets (as shown in Tab. 7). By gradually removing key modules (e.g., attention mechanism, residual network, etc.) from the models, it is possible to assess the role of each component under different dataset sizes and structural complexities.

The experimental results (as shown in Tab. 8) show that the performance of the RAGCN, GCN+GAT and GCN+ResNet models varies significantly across datasets, reflecting the effect of different structures on the feature-capturing ability. The RAGCN model achieves the best performance in most cases by combining the GCN, GAT, and residual networks, especially on Cora and CiteSeer, which are two moderately sized and complex datasets, Cora and CiteSeer. Its test accuracy and F1 score outperform the other models, thanks to RAGCN's ability to combine local features with global information, GCN performs well in aggregating local neighborhood features, and GAT's attention mechanism allows the model to dynamically focus on important neighbors, thus improving the flexibility of feature representation. Residual networks, on the other hand, further enhance the classification performance by mitigating the gradient vanishing problem of deep networks, allowing the model to integrate information at a deeper level.

On the Cora dataset, the RAGCN model performs particularly well, with a test accuracy of 0.7620 and a prediction accuracy of 0.7899. This suggests that the combination of the three models gives the RAGCN a significant advantage in dealing with complex node relationships. In contrast, the GCN+GAT model performs slightly worse, with a test accuracy of 0.7540, which may be due to the lack of residual network support in complex structures, resulting in the loss of feature information, and the GCN+ResNet model performs the worst, with a test accuracy of 0.6400, which suggests that relying on the GCN and residual network alone may not be able to fully explore the complex relationships between nodes in medium-sized datasets. Fully exploit the complex relationships between nodes.

On the CiteSeer dataset, the RAGCN model again demonstrates its advantage with a test accuracy of 0.7750, showing the effectiveness of the model combination under complex data structures. However, the GCN+GAT model, with a tested accuracy of 0.7630, although slightly lower than RAGCN, still demonstrates the potential of the attention mechanism in feature selection. The GCN+ResNet model comes close to this dataset but does not outperform RAGCN, demonstrating the limited contribution of residual networks in complex graph structures.

Dataset	Number of nodes	Number of edges	Number of classes	Feature dimensions	Dataset characteristics
Cora	2,708	5,429	7	1,433	Academic paper citation network, comprising 7 classes
CiteSeer	3,327	4,732	6	3,703	Academic paper citation network, more complex structure
Amazon	9,235	51,165	10	767	Product co-occurrence network, large scale

Tab. 7. Comparison of data sets.

Dataset	Model	Test accuracy	Prediction accuracy	Recall	F1 score
Cora	RAGCN	0.7620	0.7899	0.7620	0.7657
	GCN+GAT	0.7540	0.7745	0.7540	0.7540
	GCN+ResNet	0.6400	0.7669	0.6400	0.6363
CiteSeer	RAGCN	0.7750	0.7752	0.7750	0.7748
	GCN+GAT	0.7630	0.7661	0.7630	0.7634
	GCN+ResNet	0.7740	0.7776	0.7740	0.7738
Amazon	RAGCN	0.9503	0.9517	0.9503	0.9506
	GCN+GAT	0.9346	0.9367	0.9346	0.9329
	GCN+ResNet	0.9595	0.9599	0.9595	0.9596

Tab. 8. Results of ablation experiments.

On the Amazon dataset, the GCN+ResNet model performs exceptionally well, reaching a test accuracy of 0.9595, surpassing the RAGCN model's 0.9503. This result suggests that the residual network's strengths become more pronounced as the size of the dataset increases, mitigating the over fitting and increasing the generalization ability of the model. This suggests that the simple and efficient GCN+ResNet combination can better capture features in large-scale datasets, while the complexity of the RAGCN model may lead to performance degradation under certain conditions.

The experimental results reveal the effect of different model structures on feature capture ability and its relationship with dataset characteristics. In medium-sized and complex-structured datasets, the RAGCN model achieves the best performance by virtue of its combined advantages, while in large datasets, the GCN+ResNet model performs even better, emphasizing the importance of residual networks when dealing with large-scale data. These findings provide an important reference for future model selection and optimization.

7. Summary

This paper presents a comprehensive model for EV charging station layout optimization that incorporates K-Means clustering, a simulated annealing algorithm, and the residual network and graph attention mechanism of graph convolutional neural network (GCN). Through in-depth analysis of geographic and POI data in Nanjing, potential deployment points of charging stations are first identified using the clustering algorithm, followed by optimization of these clustered centers by the simulated annealing algorithm to obtain a more homogeneous and practical charging station layout. In addition, the introduced graph convolutional neural network further analyses the impact of each POI type on the charging station layout, which improves the prediction accuracy and operational efficiency of the model.

The experimental results show that the optimization model can significantly improve the rationality and utilization of EV charging station layout. Through sensitivity analyses, this paper also reveals the specific impact of different POI characteristics on charging station layout strategies, which provides a scientific basis for future urban EV charging infrastructure planning. Further, the internal feature-capturing ability of each model is revealed through ablation experiments, which also helps to better understand the impact and limitations of different model structures on classification accuracy under specific dataset conditions. The methodology and findings in this paper not only optimize the spatial layout of charging stations but also promote the popularity of EVs and the green development of cities. Future work will explore the application of this model to a wider geographic area, as well as the further integration of more types of data and advanced analytical techniques for more efficient and sustainable EV charging network design.

Acknowledgments

This work was supported by the Jiangsu Science and Technology Think Tank Program (Youth) Project (JSKX24085), and the Jiangsu Provincial College Students Innovation and Entrepreneurship Training Plan Project (202411276037Z).

References

[1] ZHOU, Y., WANG, M., HAO, H., et al. Plug-in electric vehicle market penetration and incentives: A global review. *Mitigation and Adaptation Strategies for Global Change*, 2015, vol. 20, no. 5, p. 777–795. DOI: 10.1007/s11027-014-9611-2

[2] BURRA, L. T., AL-KHASAWNEH, M. B., CIRILLO, C. Impact of charging infrastructure on electric vehicle adoption: A synthetic population approach. *Travel Behaviour and Society*, 2024, vol. 37, p. 1–12. DOI: 10.1016/j.tbs.2024.100834

[3] SOMMER, S., VANCE, C. Do more chargers mean more electric cars? *Environmental Research Letters*, 2021, vol. 16, no. 6, p. 1–10. DOI: 10.1088/1748-9326/ac05f0

[4] GREENE, D. L., KONTOU, E., BORLAUG, B., et al. Public charging infrastructure for plug-in electric vehicles: What is it worth? *Transportation Research Part D: Transport and Environment*, 2020, vol. 78, p. 1–15. DOI: 10.1016/j.trd.2019.11.011

[5] MA, H., WANG, H., CHENG, S., et al. Research on location selection and capacity planning of urban electric vehicle charging station. In *Proceedings of the 2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP)*. Xi'an (China), 2019, p. 331–335. DOI: 10.1109/APAP47170.2019.9225059

[6] GONG, D., TANG, M., BUCHMEISTER, B., et al. Solving location problem for electric vehicle charging stations: A sharing charging model. *IEEE Access*, 2019, vol. 7, p. 138391–138402. DOI: 10.1109/ACCESS.2019.2943079

[7] JORDAN, J., PALANCA, J., DEL VAL, E., et al. Localization of charging stations for electric vehicles using genetic algorithms. *Neurocomputing*, 2021, vol. 452, p. 416–423. DOI: 10.1016/j.neucom.2019.11.122

[8] LAZARI, V., CHASSIAKOS, A. Multi-objective optimization of electric vehicle charging station deployment using genetic algorithms. *Applied Sciences*, 2023, vol. 13, no. 8, p. 1–18. DOI: 10.3390/app13084867

[9] GAN, X., ZHANG, H., HANG, G., et al. Fast-charging station deployment considering elastic demand. *IEEE Transactions on Transportation Electrification*, 2020, vol. 6, no. 1, p. 158–169. DOI: 10.1109/TTE.2020.2964141

[10] NICCOLAI, A., BETTINI, L., ZICH, R. Optimization of electric vehicles charging station deployment by means of evolutionary algorithms. *International Journal of Intelligent Systems*, 2021, vol. 36, no. 9, p. 5359–5383. DOI: 10.1002/int.22515

[11] ZHAO, Y., GUO, Y., GUO, Q., et al. Deployment of the electric vehicle charging station considering existing competitors. *IEEE Transactions on Smart Grid*, 2020, vol. 11, no. 5, p. 4236–4248. DOI: 10.1109/TSG.2020.2991232

[12] ZHONG, P., XU, A., KANG, Y., et al. An optimal deployment scheme for extremely fast charging stations. *Peer-to-Peer Networking and Applications*, 2022, vol. 15, no. 3, p. 1486–1504. DOI: 10.1007/s12083-022-01306-7

[13] HU, D., ZHANG, J., ZHANG, Q. Optimization design of electric vehicle charging stations based on the forecasting data with service balance consideration. *Applied Soft Computing*, 2019, vol. 75, p. 215–226. DOI: 10.1016/j.asoc.2018.07.063

[14] OTHMAN, A. M., GABBAR, H. A., PINO, F., et al. Optimal electrical fast charging stations by enhanced descent gradient and Voronoi diagram. *Computers & Electrical Engineering*, 2020, vol. 83, p. 1–15. DOI: 10.1016/j.compeleceng.2020.106574

[15] XU, M., MENG, Q. Optimal deployment of charging stations considering path deviation and nonlinear elastic demand. *Transportation Research Part B: Methodological*, 2020, vol. 135, p. 120–142. DOI: 10.1016/j.trb.2020.03.001

[16] ANTHOPOULOS, L., KOLOVOU, P. A multi-criteria decision process for EV charging stations deployment: Findings from Greece. *Energies*, 2021, vol. 14, no. 17, p. 1–15. DOI: 10.3390/en14175441

[17] ZHAO, Z., LEE, C. K. M., REN, J., et al. Optimal EV fast charging station deployment based on a reinforcement learning framework. *IEEE Transactions on Intelligent Transportation Systems*, 2023, vol. 24, no. 8, p. 8053–8065. DOI: 10.1109/TITS.2023.3265517

[18] JIA, J., LIU, C., WAN, T. Planning of the charging station for electric vehicles utilizing cellular signaling data. *Sustainability*, 2019, vol. 11, no. 3, p. 1–18. DOI: 10.3390/su11030643

[19] HU, D., HUANG, L., LIU, C., et al. Data-driven optimization for electric vehicle charging station locating and sizing with charging satisfaction consideration in urban areas. *IET Renewable Power Generation*, 2022, vol. 16, no. 12, p. 2630–2643. DOI: 10.1049/rpg2.12382

[20] GENG, Q., SUN, D., JIA, Y. Planning of electric vehicle charging facilities on highways based on chaos cat swarm simulated annealing algorithm. *Tehnicki Vjesnik - Technical Gazette*, 2023, vol. 30, no. 5, p. 1602–1611. DOI: 10.17559/TV-20230320000459

[21] AMAP. Amap Open Platform API Download. *Amap Open Platform* [Online]. Available: https://lbs.amap.com/api/webservice/download, Accessed: Dec. 23, 2024

About the Authors . . .

Haojie YANG was born in Jiangsu Province, China, in May 2004. He is an undergraduate majoring in Electronic Information Engineering at Nanjing Engineering Institute. During his studies, he has participated in his teachers' scientific research projects. His research interests include intelligence algorithms.

Jingjing ZOU was born in Jiangsu Province, China, in July 2004. She is an undergraduate majoring in Information Engineering at Nanjing Institute of Technology. During her studies, she has participated in her teachers' scientific research projects. Her research interests include reinforcement learning and deep learning.

Rongxin XU was born in Jiangsu Province, China, in October 2004. She is an undergraduate majoring in Information Engineering at Nanjing Institute of Technology. During her studies, she has participated in her teachers' scientific research projects. Her research interests include reinforcement learning and deep learning.

Jiale HUANG was born in Jiangsu Province, China, in October 2004. She is an undergraduate majoring in Information Engineering at Nanjing Institute of Technology. During her studies, she has participated in her teachers' scientific research projects. Her research interests include reinforcement learning and deep learning.

Peng GENG (corresponding author) was born in Hubei Province, China, in 1979. He received his Master's degree in Communication and Information System in 2007 from Guilin University of Electronic Science and Technology, Guilin, Guangxi, China. Now he is the Associate Professor of Nanjing Institute of Technology. His research interests include complex systems and deep learning.