



Diffusion Model-based Mobile Traffic Generation with Open Data for Network Planning and Optimization

Haoye Chai
Department of Electronic
Engineering,
BNRist, Tsinghua University,
Beijing, China
haoyechai@mail.tsinghua.edu.cn

Tao Jiang
Research Center of 6G Mobile
Communications, School of Cyber
Science and Engineering, Huazhong
University of Science and Technology,
Wuhan, China
tao.jiang@ieee.org

Li Yu
Chinamobile Research Institute,
Beijing, China
yuliyf@chinamobile.com

ABSTRACT

With the rapid development of the Fifth Generation Mobile Communication Technology (5G) networks, network planning and optimization have become increasingly crucial. Generating high-fidelity network traffic data can preemptively estimate the network demands of mobile users, which holds potential for network operators to improve network performance. However, the data required by existing generation methods is predominantly inaccessible to the public, resulting in a lack of reproducibility for the models and high deployment costs in practice. In this article, we propose an Open data-based Diffusion model for mobile traffic generation (OpenDiff), where a multi-positive contrastive learning algorithm is designed to construct conditional information for the diffusion model using entirely publicly available satellite remote sensing images, Point of Interest (POI), and population data. The conditional information contains relevant human activities in geographical areas, which can effectively guide the generation of network traffic data. We further design an attention-based fusion mechanism to capture the implicit correlations between network traffic and human activity features, enhancing the model's controllable generation capability. We conduct evaluations on three different cities with varying scales, where experimental results verify that our proposed model outperforms existing methods by 14.36% and 13.05% in terms of generation fidelity and controllability. To further validate the effectiveness of the model, we leverage the generated traffic data to assist the operators with network planning on a real-world network optimization platform of China Mobile Communications Corporation. The source code is available online: <https://github.com/impchai/OpenDiff-diffusion-model-with-open-data>.

CCS CONCEPTS

• **Networks** → **Network simulations**; *Network management*; • **Information systems** → **Spatial-temporal systems**; **Data mining**.

KEYWORDS

Mobile traffic generation; Diffusion model; Contrastive learning; Satellite imagery

ACM Reference Format:

Haoye Chai, Tao Jiang, and Li Yu. 2024. Diffusion Model-based Mobile Traffic Generation with Open Data for Network Planning and Optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3637528.3671544>

1 INTRODUCTION

The Fifth Generation Mobile Communication Technology (5G) has shown its significant implications for industrial production and residents' daily lives, facilitating many advanced applications such as virtual reality, autonomous driving, and holography communication *etc.* As of 2023, the first batch of 5G networks has been launched, covering over a third of the global population, with the worldwide 5G Base Stations (BSs) count exceeding 4.48 million and an industry value reaching \$897.98 billion [16]. It is reported that the number of 5G networks will increase from 251 in 2022 to over 640 by 2030, serving worldwide mobile users exceeding 9 billion [1]. To ensure this fast commercialization process of the 5G networks and the massive access of mobile users, rational network planning and optimization are indispensable.

Network traffic data, as the reflection of user behaviors, provides important support for network planning and optimization. Operators can refer to the network traffic to analyze network usage, load conditions, and user demands, thereby adopting planning/optimization strategies, *e.g.*, resource scheduling, BS site selection, and energy saving. However, operators currently adopt passive and reactive strategies, *i.e.*, adjusting networks only after observing traffic data. This approach results in processing delays and fails to adapt to dynamic network conditions. Moreover, in those areas where 5G deployment is incomplete (*e.g.*, the 5G network deployment rate in Latin America is only 16% [3]), there is often a lack of relevant network traffic data for operators to conduct effective network planning. A feasible approach is to estimate the network traffic to assist the network planning and optimization. By utilizing estimated traffic, operators can grasp potential network changes in advance, so that they can refer to the estimated traffic to test the performance under different optimization strategies and choose the best solution to improve the networks (*i.e.*, counterfactual analysis [24]).



This work is licensed under a Creative Commons Attribution International 4.0 License.

KDD '24, August 25–29, 2024, Barcelona, Spain
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0490-1/24/08.
<https://doi.org/10.1145/3637528.3671544>

The traffic estimation above is essentially a **conditional** generation process, which captures the correlation between network traffic data and various factors of regions such as geography and population. Typically, it uses generative methods to synthesize network traffic with the characteristics of a specific geographic area. The methods can capture complex features of network traffic data with high diversity, which attracts attention from many researchers [27, 28, 32, 39]. Wang *et al.* [32] combined RNN and Generative Adversarial Network (GAN) to generate mobile traffic, aiming to capture the temporal dependencies of the time-series data. Xu *et al.* [35] proposed a SpectraGAN framework that transformed mobile traffic data into images and integrated a CNN with GAN to capture the spatial dependencies of traffic data. To achieve effective network optimization/planning, the generated traffic data needs to have high fidelity and the generative model should be easily deployed. However, existing studies encounter significant limitations concerning the two requirements.

First, most generative methods are heavily reliant on non-public data, which makes the model lack reproducibility. For instance, the AppShot [28] utilized many service-level data sources including Netflix, Twitter, *etc.*, which contained privacy-sensitive information. The Urban Knowledge Graph (UKG) [39] contained a vast amount of information on urban geographic locations, regional functionalities, and affiliations. Accessing such data requires collaboration and authorization from multiple official parties. Moreover, when planning/optimizing some sensitive and critical areas, *e.g.*, government facilities, substations, research and development centers, the data within these areas are difficult to directly obtain, making it challenging to replicate and utilize the generation models.

Second, the controllable generation capabilities of existing models are relatively poor. Currently, the mainstream generative methods are based on GANs. Such methods involve training both generator and discriminator simultaneously, which suffer from instability during the training process and the phenomenon of mode collapse. These issues often result in the models only generating fixed patterns of network traffic data and failing to utilize conditions to generate desired network traffic. Moreover, traditional methods typically concatenate the conditional vector with the input data before feeding into the generative network, which is not sufficient to capture the complex relations between network traffic data and contextual information, degrading the ultimate generation fidelity.

To address the challenges mentioned above, we aim to develop a generative framework that can *generate network traffic data via publicly available datasets for arbitrary geographic areas*. In this article, we propose an Open data-based Diffusion model for mobile traffic generation (OpenDiff), which utilizes satellite remote sensing imagery, POI, and population data to generate network traffic data of the target geographic areas. As publicly available data sources, satellite images, POI, and population data provide comprehensive information on geographic areas (*e.g.*, natural environments, building layouts, commercial facilities, and residential distribution), which can fully reflect the characteristics of human activities. There are numerous studies have shown a strong correlation between human activities and network traffic usage [9, 18, 34, 36]. We are thus motivated to utilize these public data to assist in generating high-fidelity network traffic data. **On the one hand**, to ensure reproducibility, we designed a multi-positive sample contrastive learning algorithm

that utilizes POI, population distribution, and geographic coordinates to extract human activity features from satellite images, where the features serve as conditions to guide the generation of network traffic. The data leveraged during the training process is entirely public, which addresses the first limitation. **On the other hand**, we propose an attention-based classifier-free guidance module to improve the controllability of our OpenDiff framework. The module feeds conditional information with a certain probability, which can explicitly incorporate the human activity features into the generation network. The attention-based fusion mechanism transforms the conditional vectors into queries, while the network traffic feature vectors are converted into keys and values. The mechanism explores the implicit correlations between network traffic and human activities by calculating their cross-attention scores, so that the diffusion model can better utilize the conditional information to enhance the controllable generation capabilities, which tackles the second limitation. We summarize the contributions as follows:

- We propose OpenDiff, a novel diffusion model framework for controllable generation of mobile traffic data with publicly available data sources. Through the designed multi-positive sample contrastive learning algorithm, our framework is capable of extracting human activity features for arbitrary geographical areas, thereby guiding the generation of network traffic. Our approach can provide high-fidelity network traffic data for operators in areas lacking of historical network traffic data, assisting in network planning and optimization.
- We design a classifier-free guidance module that can explicitly input conditional information into the diffusion model. A cross-attention mechanism is proposed to capture implicit correlations between human activities and network traffic, thereby enhancing the model's controllability.
- We conduct extensive experiments on three real-world datasets with different city scales. The results demonstrate the proposed scheme to have high fidelity and good controllability of generating mobile traffic by 14.36% and 13.05%. OpenDiff has been successfully deployed on the Jiutian Artificial Intelligence Platform of China Mobile to assist the operators with network planning, including BS deployment, energy-saving strategy optimization, and network resource allocation.

2 PRELIMINARIES AND PROBLEM FORMULATION

2.1 Classifier-free guidance diffusion model

Denosing Diffusion Probabilistic Model (DDPM): The underlying idea of DDPM is to use two Markov chains to characterize the transition from original data to noise data [37]. The forward transition probability can be expressed as $q(x_k|x_{k-1}) = N(\sqrt{1-\beta_k}x_{k-1}, \beta_k\mathbf{I})$, where $\{\beta_k \in (0, 1), k \in (1, K)\}$ is a set of scheduled noise weight and \mathbf{I} is an identity matrix. Based on the forward transition probability, the generated noised data in step k can be calculated by $x_k = \sqrt{\alpha_k}x_0 + (1-\alpha_k)\epsilon$, where $\epsilon \sim N(0, \mathbf{I})$ is the added Gaussian noise, $\alpha_k = 1-\beta_k$ and $\hat{\alpha}_k = \prod_{k'=1}^k \alpha_{k'}$. The reversed transition probability can be expressed as $p_\theta(x_{k-1}|x_k) = N(\mu_\theta(x^k, k), \sigma_\theta(x_k, k)\mathbf{I})$, where μ_θ and σ_θ are the mean and variance, respectively. The

subscript θ indicates that the value is obtained by a well-trained neural network ϵ_θ named denoising network. By maximizing the Evidence of Lower BOund (ELBO) with log-likelihood, Ho *et al.* [11] proved that the μ_θ can be parameterized as calculated by $\mu_\theta(x_k, k) = \alpha_k^{-0.5}[x_k - \beta_k(1 - \hat{\alpha}_k)^{-0.5}\epsilon_\theta(x_k, k)]$, and σ_θ can be parameterized as $\sigma_\theta(x_k, k) = \sqrt{(1 - \hat{\alpha}_{k-1})/(1 - \hat{\alpha}_k)\beta_k}$. The denoising network can then be optimized by the following objectives:

$$\min_{\theta} L(\theta) = \min_{\theta} \mathbb{E}_{x_0 \sim q(x_0), \epsilon \sim N(0, I)} [\|\epsilon - \epsilon_\theta(x_k, k)\|_2^2], \quad (1)$$

where $\epsilon_\theta(x_k, k)$ is the estimated noise by the denoising network.

Classifier-Free Guidance (CFG): Classifier-free guidance is a type of diffusion framework used for conditional generation [12]. By controlling the weights between the condition and the unconditional generation results, it can generate high-fidelity data while maintaining good controllability. During the model training phase, the condition c is randomly set to 0 with a certain probability p . During the inference phase, the results of the conditional and unconditional results are combined based on the guidance scale, which can be represented as follows [22]:

$$\epsilon_{k+1} = (1 + \mu)\epsilon_\theta(x_k, c) - \mu\epsilon_\theta(x_k), \quad (2)$$

where μ is the guidance scale, $\epsilon_\theta(x_k, c)$ is the estimated noise when the input is x_k and the conditional information is c , whereas $\epsilon_\theta(x_k)$ is the estimated noise in the absence of conditional information. Compared to classifier guidance diffusion [8], CFG does not require pre-specifying the number of generating categories with a dedicated classifier, which reduces the complexity of the model. In addition, CFG can adjust the weights between conditional and unconditional results during the inference stage, enhancing the flexibility of the model's generation process. For network traffic data generation, traffic data in different geographical areas typically have a diverse number of patterns. Using the CFG framework for generation can effectively adapt to the variability in traffic patterns.

2.2 Contrastive learning

Contrastive Learning (CL) is a type of unsupervised learning method with the core idea of making the features of similar samples closer to each other while pushing the features of different samples farther apart [5, 33]. By creating positive sample pairs (e.g., different views of the same class data) and negative sample pairs (data from different classes) from the unlabeled data, CL utilizes neural networks to extract features from these sample pairs. Typically, CL leverages two types of loss functions to guide the neural network training process, *i.e.*, Contrastive Loss, and InfoNCE. Contrastive Loss sets a margin to determine positive and negative sample pairs, it calculates the loss value based on either Euclidean distance or dot product between one positive-sample pair and one negative-sample pair. InfoNCE draws inspiration from the concept of cross-entropy and uses the softmax function to calculate the differences between one positive-sample pair and multiple negative-sample pairs, which improves the adaptability of the CL model to large-scale datasets. The InfoNCE loss can be expressed as

$$L_{NCE} = -\mathbb{E}\left[\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{n=1}^N \exp(\text{sim}(z_i, z_n)/\tau)}\right], \quad (3)$$

where z_i and z_j form a positive sample pair, and z_i and z_n form a negative sample pair. The $\text{sim}()$ function denotes the similarity between sample pairs and the τ is the temperature parameter to control the scale of the similarity function.

2.3 Utilized public data

Satellite images. The satellite images provide detailed views of the Earth's surface, including urban areas, natural landscapes, and farmlands, *etc.* Environmental Systems Research Institute (Esri)¹ offers publicly available satellite images at various resolutions, along with corresponding latitude and longitude for each image. In our experiments, we downloaded images with a zoom level of 15 for model training.

Population density. We acquire the population density statistics from the WorldPop organization², which estimates the gridded population counts with bottom-up or top-down methods. In our evaluation, we downloaded a 100m-resolution population dataset for contrastive learning.

Point of Interests. POI are various venues that people visit during their work, leisure, or travel activities. Utilizing POI distribution data can effectively infer patterns of human activities in specific areas, and subsequently, help in estimating the usage of network traffic. In our evaluations, we manually crawled POI data for various cities from Tencent Maps services³, including 14 different categories, along with the corresponding number of POIs within each category, which can be found in detail in Appendix A.1.2.

2.4 Problem statement

We aim to use regional features as conditional information C to assist the model D in generating high-fidelity network traffic. Given the target geographic area i , our goal is to generate a time sequence $S^i = \{s_t^i\}_{t \in (0, T)}$, where s_t^i represents the aggregated mobile traffic of all the base stations within i at time t . We intend to utilize open data \mathcal{P} including satellite images, POI distributions, and population density data so that the generation model D is capable of using readily available data to produce high-fidelity traffic data. The conditional generation problem can be defined as:

Given the target area i , generating the corresponding mobile traffic sequence $S^i = \mathcal{D}(C \sim \mathcal{P})$, where $C \sim \mathcal{P}$ denote feature extraction process of conditional information C from public data \mathcal{P} .

However, it is nontrivial to realize the above goal. There are three challenges to be tackled during the generation process. i). The multiple types of public data \mathcal{P} should be effectively utilized to extract the regional features. ii). The model \mathcal{D} needs to explicitly incorporate condition information to enhance its controllability. iii). The correlation between regional features C and network traffic is required to be captured to improve generation fidelity.

3 METHOD

3.1 Framework overview

To tackle the challenges above, we propose the OpenDiff framework that incorporates three modules: multi-positive contrastive learning

¹<https://www.esri.com/en-us/home>

²<https://www.worldpop.org>

³<https://lbs.qq.com>

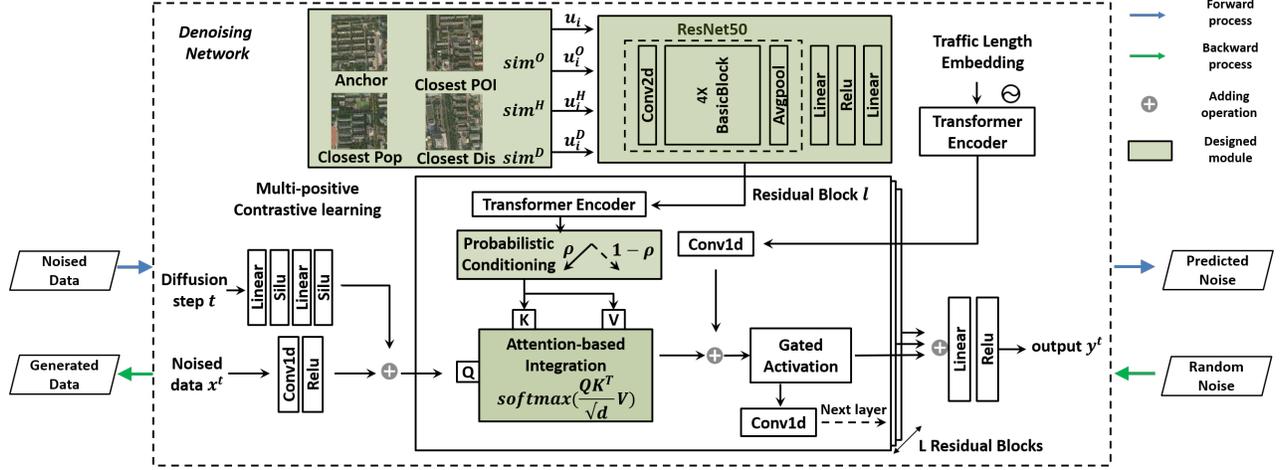


Figure 1: OpenDiff generation framework.

module, classifier-free guidance module, and adaptive conditioning module, as illustrated in Figure 1.

i). *Multi-positive contrastive learning module*: unlike traditional CL algorithms that construct a single positive sample [10, 25], we use multidimensional public data \mathcal{P} including population, POI, geographical locations to construct multiple positive samples, allowing for efficient extraction of regional features in satellite images.

ii). *Probabilistic conditioning module*: we employ the CFG diffusion model framework, which explicitly includes conditions during both training and inference processes, enhancing the controllability of the model \mathcal{D} .

iii). *Attention-based fusion module*: we introduce a mutual attention mechanism that enables the denoising network to adaptively capture the correlation between conditional information \mathcal{C} and mobile network traffic, thereby improving the generation fidelity.

3.2 Multi-positive contrastive learning module

We design a contrastive learning algorithm with multiple positive samples to extract regional features from satellite images, where three types of public data are used: POI categories and quantities, population counts, and geographical coordinates.

3.2.1 Constructing Positive samples. Unlike traditional CL algorithms that use flipping or cropping to create different views of samples, we construct the positive samples for the original satellite image u_i (referred to as anchor) based on the similarity of POI distribution, population distribution, and geographical proximity. To quantify the similarity of POI and population, we first calculate the POI and population information contained in each satellite image within its geographic coverage range. The POI information can be defined as $f_i^O = \{o_k^i\}_{k=1:K}$, where o_k^i denotes the number of type- k POI and K is the overall type of POI categories. Similarly, the population information is represented as $f_i^H = \{h_m^i\}_{m=1:M}$, where h_m^i is the population density in the sub-region m and M is the total number of sub-regions in u_i . Note that the number of sub-regions is determined by the resolution of the population data. For example, if we have a size of $2 \text{ Km}^2 \times 2 \text{ Km}^2$ satellite image and population data with a resolution of 100 m, then the number of sub-regions would

be $20 \times 20 = 400$. Then, the similarity of both POI and population can be expressed as

$$\text{sim}_{i,j}^O = \frac{f_i^O \cdot f_j^O}{\|f_i^O\| \times \|f_j^O\|}, \quad \text{sim}_{i,j}^H = \frac{f_i^H \cdot f_j^H}{\|f_i^H\| \times \|f_j^H\|}, \quad (4)$$

where $\text{sim}_{i,j}^O$ and $\text{sim}_{i,j}^H$ are the POI similarity and population similarity between image u_i and u_j , respectively. To quantify geographical proximity, we directly calculate the Euclidean distance based on the two images' center coordinates. In each training round, we choose three images that are closest to the anchor image u_i in terms of POI, population, and geographical location as positive samples, denoted as u_i^O , u_i^H , and u_i^D , while the remaining images are used as negative samples.

3.2.2 Learning regional features. Given the positive set \mathcal{P}_i with multiple positive samples $\mathcal{P}_i = \{u_i^O, u_i^H, u_i^D\}$ and the negative set \mathcal{N}_i for anchor x_i , we develop a convolutional neural network R that utilizes ResNet 50 [30] as backbone together with a multi-layer perception (MLP) layer to map the satellite images into the low dimensional representations, which can be expressed as

$$z_i = R(u_i), \quad z_p = R(u_p), \quad z_n = R(u_n), \quad p \in \mathcal{P}_i, \quad n \in \mathcal{N}_i, \quad (5)$$

where z_i , z_p , and z_n are the learned latent features of the anchor image, the three positive image samples, and the negative samples, respectively. During the training process, the CL network should minimize the distance between z_i and z_p while enlarging the difference between z_i and z_n . To achieve this goal, we extend the InfoNCE loss in Equation 3 to accommodate multiple positive samples, so that the model can capture the similarity between samples:

$$L = \sum_{i \in I} -\frac{1}{|\mathcal{P}_i|} \sum_{p \in \mathcal{P}_i} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in \mathcal{P}_i \cup \mathcal{N}_i} \exp(z_i \cdot z_a / \tau)}, \quad (6)$$

where I is the total number of training sample, $|\mathcal{P}_i|$ is its cardinality. Currently, we only consider three types of regional features: POI, population, and geographical location, *i.e.* $|\mathcal{P}_i| = 3$. According to the loss function defined above, we can train the CL network R to capture the similarity between the anchor and positive samples while distinguishing them from negative samples. Once the network

converges, we can obtain the representations z_i corresponding to the current satellite image u_i . We employ the CL framework to extract features from satellite images, using information such as POI and population density to guide the model's training. This enables extracted representation to possess certain human activity features. Then, we input the extracted features as conditions to denoising network to guide the mobile traffic generation process.

3.3 Probabilistic conditioning module

The Probabilistic Conditioning (PC) module utilizes a classifier-free guidance framework to incorporate the features z_i that are sent from the multi-positive contrastive learning module as conditional information. It simultaneously trains the network on both conditional inputs and unconditional inputs with a predefined probability ρ . Given the denoising network ϵ_θ , the latent vector of noised traffic data x_i^t for region i and conditional information z_i , the input of the denoising network during the training process can be expressed as

$$Pr\{\epsilon_\theta(x_i^t|z_i)\} = 1 - \rho, Pr\{\epsilon_\theta(x_i^t|\emptyset)\} = \rho, \quad (7)$$

where \emptyset denotes the conditional information is set to zero. During the inference process, a hyperparameter μ is adopted to control the guidance scale of the generation results, which is formulated as

$$\hat{\epsilon}_{\theta,\rho}(x_i^t, t|z_i, \mu, \rho) = (1 + \mu)\epsilon_{\theta,\rho}(x_i^t, t|z_i) - \mu\epsilon_{\theta,\rho}(x_i^t, t). \quad (8)$$

We control the proportion of satellite image features in the denoising network through the classifier-free guidance framework. This approach effectively reduces the model's reliance on conditional information, enhances the diversity of traffic data generated by the model, and mitigates the risk of model overfitting. Consequently, the two hyperparameters in the module ρ and μ have significant impacts on the generation performance. In the evaluation section, we conducted ablation experiments on these two parameters to determine the optimal values.

3.4 Attention-based integration module

After passing through the PC module, the feature vector of satellite images is fed into the denoising network with a probability ρ . Regarding the input conditional information, we design the Attention-based Integration (AI) module to capture the relations between regional features and mobile network traffic. Specifically, for the conditional information z_i and the latent vector x_i^t of noised traffic data, the adaptive module employs scaled dot-product attention, where x_i^t serves as the query and z_i serves as both the key and value, which can be expressed as

$$Q_i^t = W^Q x_i^t, K_i^t = W^K \{z_i, 0\}_\rho, V_i^t = W^V \{z_i, 0\}_\rho, \quad (9)$$

where W^Q , W^K , and W^V are weights of linear projection. The $\{z_i, 0\}_\rho$ denotes the values within the set are set to either z_i or zero vector with the probability ρ consistent with the CFG module. Afterward, the cross-attention value can be calculated by

$$A_i^t = \text{Softmax}\left(\frac{Q_i^t K_i^{t\top}}{\sqrt{d}}\right) V_i^t, \quad (10)$$

where $K_i^{t\top}$ is the transpose operation of K_i^t . By calculating the attention score, the denoising network passes the score and original

latent vectors of mobile traffic data x_i^t to a feedforward network and outputs the conditioning features, which can be represented as

$$y_i^t = \mathcal{F}(\mathcal{F}(A_i^t, x_i^t)), \quad (11)$$

where \mathcal{F} is the layer norm and feed-forward networks. The zero vector in Equation 9 is employed to align the input dimensions of the linear layer of attention mechanism, which facilitates the network design. In addition, a zero vector can signify that the input vector does not contain any conditional information, allowing the attention algorithm to differentiate between conditional and unconditional inputs. After acquiring the hidden feature y_i^t from the AI module, the denoising network will be trained with the objective in Equation 1.

4 PERFORMANCE EVALUATION

We perform evaluations on real-world datasets to answer the following research question.

- **RQ1:** How does the proposed model perform in fidelity compared with existing baseline schemes?
- **RQ2:** How is the controllability of our model?
- **RQ3:** How do conditions including POI, population, and geographical adjacent, impact generation performance?

4.1 Evaluation settings

4.1.1 Mobile traffic dataset and metrics. We collect mobile traffic datasets from three Chinese cities: Beijing, Shanghai, and Nanjing. We aim to verify the accuracy of the generated data with two metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), meanwhile evaluating the fidelity of the generated data distribution with two metrics: Jensen-Shannon Divergence (JSD) and Continuous Ranked Probability Score (CRPS) [23]. Detailed descriptions of the dataset and metrics can be found in Appendix A.1.

4.1.2 Baselines. For evaluating the generation performance of our proposed scheme, we choose 6 baselines that comprise both the GAN-based schemes and diffusion model-based schemes.

transGAN The baseline [17] leveraged a transformer block-based generative structure, where multi-scale discriminators were constructed to capture the features of mobile traffic data.

tcnGAN. The baseline investigated the temporal correlation of mobile traffic sequence with Temporal Convolutional Networks (TCNs) [2]. We manually coded a GAN network with TCNs as the generator and discriminator.

KEGAN. The Knowledge-Enhanced GAN (KEGAN) [14] is a hierarchical GAN framework that utilizes a self-constructed Urban Knowledge Graph (UKG). The embeddings of UKG are concatenated with the noise vectors so that the GAN network implicitly captures the relationship between conditional information and network traffic.

CSDI. The baseline is a conditional diffusion model that is designed for time series imputation and prediction [29]. To realize the generation task, we set the condition mask in the original version to ones-matrix.

spectraGAN. The baseline [35] converted the mobile traffic generation as an image generation problem, which adopted Convolutional Neural Networks (CNNs) to extract regional features for geographical patches.

Table 1: Mobile traffic generation performance in Nanjing and Beijing datasets. Bold numbers denote the best results and underline numbers denote the second-best results. Δ denotes the improvement between the best value and the second-best value.

Datasets	Nanjing				Beijing			
	JSD	MAE	RMSE	CRPS	JSD	MAE	RMSE	CRPS
transGAN	0.7433 ± 0.00236	0.5808 ± 0.00149	0.6366 ± 0.00357	0.8205 ± 0.00110	0.5605 ± 0.00388	0.4512 ± 0.00349	0.5250 ± 0.00425	0.4575 ± 0.00765
tcnGAN	0.2509 ± 0.00276	0.2176 ± 0.00308	0.2671 ± 0.00383	0.2444 ± 0.00250	0.4225 ± 0.01039	0.2514 ± 0.00591	0.3255 ± 0.00581	0.5741 ± 0.00731
KEGAN	<u>0.2194 ± 0.00589</u>	0.3209 ± 0.00121	0.3802 ± 0.00115	0.2186 ± 0.00276	<u>0.2049 ± 0.00533</u>	0.2115 ± 0.00226	0.2731 ± 0.00339	0.4367 ± 0.00743
CSDI	0.2419 ± 0.00625	<u>0.1869 ± 0.00232</u>	<u>0.2267 ± 0.00303</u>	0.2227 ± 0.00237	0.2231 ± 0.00688	<u>0.1561 ± 0.00193</u>	<u>0.1947 ± 0.00238</u>	0.4265 ± 0.00788
spectraGAN	0.2743 ± 0.00567	0.5572 ± 0.00228	0.6835 ± 0.00299	<u>0.2133 ± 0.00191</u>	0.2670 ± 0.00382	0.4465 ± 0.00573	0.5508 ± 0.00713	<u>0.4264 ± 0.00782</u>
OpenDiff(Ours)	0.2088 ± 0.00318	0.1665 ± 0.00336	0.2036 ± 0.00446	0.2017 ± 0.00251	0.1962 ± 0.00479	0.1365 ± 0.00170	0.1759 ± 0.00210	0.3961 ± 0.00781
Δ	5.08%	12.25%	11.35%	8.38%	4.43%	14.36%	10.69%	7.65%

ADAPTIVE. ADAPTIVE [39] is a transfer generative model for mobile traffic data. The algorithm also leverages the UKG and a BS aligning scheme to transfer the knowledge from one city to another.

4.2 Mobile traffic generation fidelity (RQ1)

Table 1 shows the experimental results of mobile traffic generation. It can be seen that our proposed OpenDiff algorithm performs the best in generating traffic under different city scales (Beijing and Nanjing). Compared to the current baselines, our algorithm can improve generation performance by up to 12.25%. This demonstrates the effectiveness of the algorithm, indicating that the model can effectively capture the correlation between network traffic and geographic regions, thus improving generation accuracy. In addition, we can observe that diffusion model-based methods perform better in terms of MAE and RMSE, suggesting that diffusion models are capable of generating data that is closer to real-world situations compared to GAN-based methods. Regarding the generation results of the Shanghai dataset, please refer to Appendix A.2.

4.3 Conditional traffic generation (RQ2)

To verify the controllability of the generation capability, we set the model to generate specific traffic patterns. Specifically, we first select the indices corresponding to the traffic patterns in each dataset. For example, in the Beijing dataset, the indices closest to traffic pattern 2 are found to be 10, 21, and 23. When generating traffic data with the model, we input the conditional information (*i.e.*, the embeddings learned from satellite images) corresponding to the indices (*i.e.*, 10, 21, and 23) into the model to test its data generation performance. The results are shown in Table 2. Compared to CSDI, which does not use geographic information as conditions, both spectraGAN and the OpenDiff model exhibit better conditional generation performance. This is because both models can capture the correlation between the geographical environment and the corresponding network traffic. When specific geographical features are provided as conditional input, the models can utilize these correlations to perform relevant generation. Additionally, spectraGAN only considers information within a single patch, *i.e.*, the geographical proximity, and lacks consideration of human activity features, leading to a decrease in generation performance.

4.4 Ablation study (RQ3)

We investigate the influence of regional-related elements (*i.e.*, geographical proximity, population distribution, POI distribution) in

Table 2: Performance of conditional mobile traffic generation. We provide the model with conditional information corresponding to patterns 0, 1, and 2 for each city.

Dataset	Metric	Beijing	Shanghai	Nanjing
		JSD		
CSDI	Pattern-0	0.1881	0.4804	0.2477
	Pattern-1	0.1582	0.4866	0.2301
	Pattern-2	0.2367	0.4790	0.3111
spectraGAN	Pattern-0	0.1739	0.4582	0.1837
	Pattern-1	0.1325	0.4740	0.1755
	Pattern-2	0.1932	0.4558	0.2214
OpenDiff (Our)	Pattern-0	0.1596	0.4389	0.1758
	Pattern-1	0.1172	0.4624	0.1607
	Pattern-2	0.1833	0.4186	0.1911
Δ	Pattern-0	8.96 %	4.39 %	4.49 %
	Pattern-1	13.05 %	2.51 %	9.21 %
	Pattern-2	5.41 %	8.89 %	15.86 %

Table 3: Impact of geographical proximity, population, and POI on model performance. Ψ represents the degradation when removing corresponding elements from the model.

Dataset	Beijing							
	Metric	JSD	Ψ	MAE	Ψ	RMSE	Ψ	CRPS
No-Geo	0.1994	-10.78	0.1391	-2.04	0.1799	-1.60	0.4060	-30.59
No-Human	0.1998	-14.49	0.1378	-6.63	0.1799	-5.32	0.4059	-33.88
No-POI	0.2048	-9.29	0.1408	-2.55	0.1803	-3.72	0.4068	-32.23
OpenDiff (Our)	0.1962	-	0.1365	-	0.1759	-	0.3961	-

the condition information. Ψ represents the performance degradation when removing certain modules from the proposed OpenDiff, which can be expressed as

$$\Psi = (y_{OpenDiff} - y_i) / (y_{OpenDiff} - y_{csdi}), \quad (12)$$

where y_i is the performance values of different baselines of ablation experiments, and y_{csdi} is the values of different metrics of CSDI. The reason we designed Ψ in this way is that both CSDI and OpenDiff are methods based on the diffusion model. By comparing the performance gap between the two models, we can accurately quantify the gains from our designed denoising network or conditional information to the network performance, ensuring that these gains are not merely the result of converting a GAN model into a diffusion model.

Specifically, during the multi-positive samples contrastive learning process, we removed specific elements, for example, "No-Human" represents the absence of population distribution data during the

Table 4: Impact of probabilistic conditioning module and attention-based integration module on model performance. Ψ represents the degradation when removing corresponding elements from the model.

Datasets		Beijing						
Metrics	JSD	Ψ	CRPS	Ψ	MAE	Ψ	RMSE	Ψ
OpenDiff (Our)	0.1962	-	0.3961	-	0.1365	-	0.1759	-
OpenDiff-NA	0.2031	-25.65%	0.4033	-23.68%	0.1407	-21.43%	0.1807	-25.53%
OpenDiff-NC	0.2054	-34.20%	0.4056	-31.25%	0.1398	-16.83%	0.1797	-20.21%
Datasets		Nanjing						
Metrics	JSD	Ψ	CRPS	Ψ	MAE	Ψ	RMSE	Ψ
OpenDiff (Our)	0.2088	-	0.2017	-	0.1665	-	0.2036	-
OpenDiff-NA	0.2187	-29.90%	0.2103	-40.95%	0.1821	-76.47%	0.2209	-74.89%
OpenDiff-NC	0.2240	-45.92%	0.2124	-50.95%	0.1793	-62.75%	0.2157	-52.38%
Datasets		Shanghai						
Metrics	JSD	Ψ	CRPS	Ψ	MAE	Ψ	RMSE	Ψ
OpenDiff (Our)	0.2321	-	0.4633	-	0.1454	-	0.1919	-
OpenDiff-NA	0.2359	-15.63%	0.4760	-55.45%	0.1479	-8.41%	0.2046	-62.56%
OpenDiff-NC	0.2396	-30.75%	0.4786	-66.81%	0.1464	-3.37%	0.2027	-53.20%

model training. All the extracted embeddings are input into the diffusion model for data generation. The experimental results shown in Table 3 indicate that removing different elements has varying impacts on the overall model performance. Among them, population density distribution has the greatest impact on the model, reaching 33.88%. This is because population density directly affects the number of users of the mobile network. Intuitively, lower population density leads to less network traffic.

We have implemented the ablation experiment of classifier-free guidance module and adaptive conditioning module in Table 4. In terms of the data distribution metrics (JSD and CRPS), the two modules can increase the model's performance by up to 66.81% and 55.45% (as indicated in the Shanghai dataset), respectively; in terms of data amplitude error metrics (MAE and RMSE), the two modules can enhance the model's generative ability by up to 76.47% and 62.75% (as indicated in the Nanjing dataset), respectively. We believe the reason for the results is that the classifier-free guidance module can effectively guide the conditional information using the guidance scale μ and conditioning probability ρ , aligning the generative network traffic data more closely with the conditional information, thereby improving the accuracy of the generated distribution. On the other hand, the adaptive conditioning module utilizes a cross-attention mechanism to adaptively capture the implicit correlations between conditional information and network traffic, which enhances the generation fidelity.

5 DEPLOYMENT APPLICATIONS

In this section, we discuss how the generated network traffic can be deployed in real network planning and optimization scenarios. Our network traffic data generation model has been deployed on the Jiutian Artificial Intelligence (AI) Platform¹. The Platform is China Mobile's self-developed AI innovation platform, providing intelligent decision-making support for mobile networks. The framework of the Jiutian platform is illustrated in Figure 2. Our model acts as a key part of the platform to generate network traffic data (the red part in the figure), by which we implement optimization and planning applications and evaluate the network performance (the

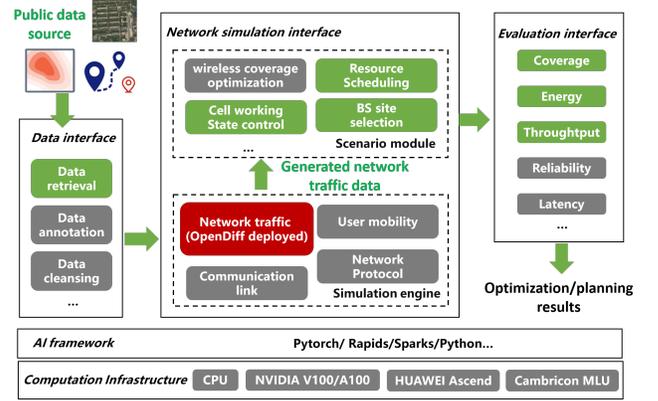


Figure 2: Framework of Jiutian AI platform.

green parts in the figure). For different application scenarios, the proposed OpenDiff framework has different usages:

i). Generating predictive network traffic data for the current geographical region. By learning the distribution of historical traffic within a specific geographical area, the model can estimate the network traffic over a certain period in the future for that region. Operators can utilize this estimated traffic data to **proactive optimize** networks, such as optimizing BS sleep strategies and scheduling network spectrum resources.

ii). Transferring the trained model to generate network traffic in any other geographical region. The model takes publicly available data for the target region, including satellite images, population density, and POI distribution. The OpenDiff then generates the network traffic data as the **potential user demands** for the target area. Operators can use the estimated network traffic demands to make network planning for the target area, such as BS deployment.

We aim to demonstrate that operators can utilize the network traffic generated by our algorithm for network planning and optimization. We validate the effectiveness of our generated traffic through three application scenarios: traffic load-based BS sleep strategy, traffic load-aware resource allocation, and traffic demand-based BS deployment.

5.1 Traffic load-based BS sleep strategy

We consider a traffic-based BS sleep strategy to save energy consumption. We set a total of m base stations and n cells in the Jiutian platform. Each base station corresponds to one BaseBand Unit (BBU), and each cell corresponds to one Remote Radio Unit (RRU). Within each cell, there is a network traffic load s_n that needs to be served (generated by the OpenDiff algorithm or derived from the real world). The RRU in each cell can choose to enter a sleep mode and offload its traffic load to another neighboring cell. The energy consumption of BSs can be divided into three parts: RRU, BBU, and air conditioner [20], which can be represented as:

$$\underbrace{k_n t_n + b_n}_{\text{RRU } P_n^r} + \underbrace{c_m}_{\text{BBU } P_m^b} + \underbrace{k_m [c_m + \sum_{\text{RRU}_n \in \text{BS}_m} (k_n t_n + b_n)]}_{\text{Air conditioner } P_{m,n}^a} + b_m, \quad (13)$$

where k_n , b_n , k_m , and k_m are the slopes and truncations of RRU n and BBU m obtained by the linear regression. The t_n denotes the network traffic on RRU, c_m is a constant representing the energy

¹<https://jiutian.10086.cn/open>

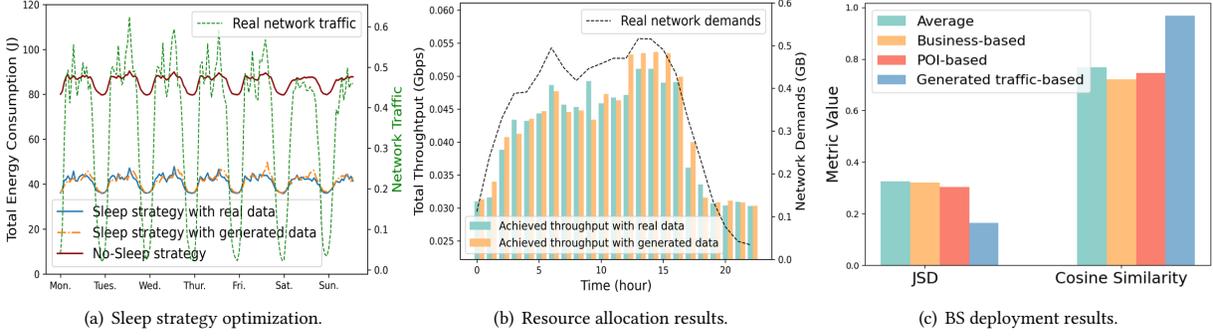


Figure 3: Network planning and optimization results via generated network traffic.

consumption of BBU. Denoting the BS sleep strategy as binary variables $\alpha_n \in \{0, 1\}$, our energy saving problem can be defined as:

$$\begin{aligned} \min E(\alpha_n) &= \sum_n \alpha_n P_n^r + \sum_m (P_m^b + \sum_n \alpha_n P_{m,n}^a), \\ \text{s.t. } \alpha_n t_n &\leq c_n, \quad \sum_n \alpha_n t_n \geq \sum_n s_n, \end{aligned} \quad (14)$$

where c_n is the maximum network traffic capacity that cell n can accommodate. The first constraint ensures that the network traffic accommodated by each active cell cannot exceed its service capacity, and the second constraint ensures that all network services should be fulfilled. We use an integer programming approach to solve the problem as shown in Figure 3(a). The sleep strategy obtained via the generated traffic data has a similar effect to that of the real traffic, and the energy consumption is significantly reduced compared to the scenario without a sleep strategy, which demonstrates the efficacy of our generative scheme. In practice, we can use the OpenDiff algorithm to preemptively generate network traffic load and proactively provide the optimal sleep strategy for BSs.

5.2 Traffic load-aware resource allocation

In addition to BS sleep strategy optimization, we can use the generated traffic as the prediction of the network load to optimize the scheduling of Physical Resource Blocks (PRBs). In the Jiutian platform, we define x users, y base stations, and z PRBs. Each user has a network traffic load of d_x and can associate with one BS by binary variable $\beta_{x,y}$ to request one PRB by binary variable $\gamma_{x,z}$ [13], the Signal-to-Interference-plus-Noise ratio (SINR) between users and BSs can be represented as:

$$r_{x,y}^z = \frac{\beta_{x,y} \gamma_{x,z} p_x h_{x,y}}{\sigma + \sum_{x' \neq x} \beta_{x',y} \gamma_{x',z} p_{x'} h_{x',y}}, \quad (15)$$

where p_x is the transmit power of user x and σ is the Gaussian noise, and $\beta_{x,y} \gamma_{x,z} p_x h_{x,y}$ is used to calculate the received signal power of the user x to base station y . Our goal is to maximize network channel capacity meanwhile minimizing the utilization of PRBs, with the satisfaction of users' network demands, *i.e.*, $\max \sum_x B \log(1 + r_{x,y}^z) - \tau \sum_x \gamma_{x,z}$, where B is broadband and τ is the punishment parameter. The experimental results, as shown in Figure 3(b), indicate that using generated network data closely approximates the system performance of real network data. This

Table 5: Coverage rate of different BS deployment strategies.

Time	POI-based	Business-based	Generated traffic-based
t=0	83.28%	81.29%	86.21%
t=30	81.27%	80.30%	86.13%
t=50	80.11%	78.51%	85.32%

demonstrates that the proposed OpenDiff algorithm can generate user traffic demands that closely resemble real-world scenarios, thereby guiding system resource allocation.

We currently assume that each user can access at most one PRB and do not consider the scenario where a user is assigned multiple PRBs simultaneously. In practical scenarios, operators can redefine the objective function based on actual PRB allocation strategies. By utilizing the generated network traffic, operators can proactively monitor changes in network traffic and dynamically schedule wireless resources based on the network load, which can reduce network response latency and greatly enhance user experience.

5.3 Traffic demand-based BS deployment

In real network planning, a crucial issue is how to deploy base stations in new areas. Network traffic serves as a vital reference for the deployment of BSs. However, in new areas, there is often a lack of historical traffic data, making the deployment particularly challenging. In this case, we can utilize our OpenDiff to generate synthetic mobile traffic data to guide the deployment of BSs.

We design four naive BS deployment strategies by utilizing: generated network traffic density (generated traffic-based), POI distribution density (POI-based), the distribution of commercial districts (business-based), and uniform deployment within the area (average). Our testing scenario covers a 256 km \times 256 km region in Beijing, which is further divided into 256 areas. We assume there are a total of 2000 BSs to be deployed within the overall region. The deployment method is deploying BS in proportion to the density of different data sources within each area. We evaluate the deployment performance via the JSD and cosine similarity, indicating the distribution matching degrees between different deployment strategies and the real deployment in Beijing. The results are shown in Figure 3(c), where the distribution of the generated traffic-based scheme is closest to the distribution of real BSs (with the lowest JSD

and the highest cosine similarity), demonstrating the effectiveness of using synthetic network traffic for BS deployment.

In addition, to assess the performance of different BS deployment strategies, we resort to network coverage ratio for mobile users. We assume each BS can cover an average of 3000 users within its communication range, and the number of users that can be covered is calculated based on the total number of BSs in each area. The mobility of users between different areas is also considered, which is calculated via the gravity model. In the model, the population flow between two areas is proportional to their population counts (denoted as M_i and M_j), and inversely proportional to the square of the distance between them (denoted as $D_{i,j}$), which can be expressed as $F_{i,j} = G \frac{M_i M_j}{D_{i,j}^2}$. The coverage results are listed in Table 5, it can be observed that using the generated traffic-based approach achieves the maximum coverage ratio, and it remains effective when the mobile users move between different areas. This is because our adopted OpenDiff method utilizes both POI and population data to generate network traffic for better characterization of the implicit correlation between network traffic and the environment, which aids in the optimization of BS deployment strategies.

6 RELATED WORK

Mobile Traffic Prediction and Generation. Mobile traffic synthesis is widely regarded as a time-series forecasting problem to assist network planning and optimization [38], which can be broadly classified into two categories: traffic prediction and generation. Mobile traffic prediction involves modeling future data based on the historical data's features, which utilizes sequence modeling schemes such as auto-regressive, recurrent Neural Networks (RNN), and LSTM [4, 6, 40] to capture the temporal correlations of network traffic. WANG *et al.* [31] proposed a Q-learning-based mechanism to update the LSTM network, so that the LSTM network can better predict network traffic. Li *et al.* [19] adopted a combination of GCN and attention mechanism to simultaneously capture the relationships between traffic data in both temporal and spatial domains. Whereas the prediction-based schemes fail to capture the complex distributions and show weakness in data diversity. Moreover, existing works primarily focused on the temporal correlations of network traffic, which failed to capture the correlation between contextual information and mobile traffic. Mobile traffic generation leverages generative models to directly learn the implicit distribution of data and can generate diverse data at any moment through sampling. For example, Ring *et al.* [27] utilized three GANs to generate multiple network flows. Lin *et al.* [21] proposed DoppelGanger for generating both the meta attributes and series features. Wang *et al.* [32] proposed an RNN-based GAN for generating cellular traffic, which captures the temporal dependencies of time-series data. In addition, generative methods can incorporate condition information during the model training process, which can represent the potential correlations between the environment and network traffic distribution, enhancing the fidelity of the generation performance.

Conditional Generation for Mobile Traffic. To capture the correlation between contextual and mobile traffic, existing literature proposed to extract the features of the environment and use them as conditions for the generative networks. Xu *et al.* [35] proposed a SpectraGAN framework that considered multiple city contexts

like POI, and land use as the initial embeddings of GANs. Hui *et al.* [14] constructed the contextual embeddings for each BS to generate mobile traffic within the coverage. They also generated the mobile traffic of IoT devices by feeding the features of devices [15]. Sun *et al.* [28] developed a city context-empowered generated scheme from a web service level. Zhang *et al.* [39] proposed a transferring learning paradigm based on hierarchical GANs, where a UKG aligning scheme was proposed to enhance the generation performance of mobile traffic across multiple cities.

Discussion. However, the current methods for generating network traffic predominantly rely on non-public data. For instance, Urban Knowledge Graphs (UKGs) were proposed for network data generation [14, 39], which require the construction of the complex relationships between POIs. The AppShot [28] utilized many service-level data sources including Netflix, Twitter, *etc.*, which contain privacy-sensitive information. Obtaining such data requires collaboration and authorization from multiple parties, which weakens the replicability of models. Furthermore, the mainstream generation methods are based on GAN models, which simultaneously train two networks and suffer from mode collapse during the training process. This leads to the models only generating fixed network traffic patterns regardless of conditional information. Therefore, we aim to generate network traffic within the diffusion model framework by using entirely publicly available data, to enhance the replicability of the model while ensuring its capability for conditional generation.

7 CONCLUSION

In this paper, we propose a diffusion model-based generative framework that utilizes openly available datasets including satellite images, population, and POI data. Through a multi-positive sample contrastive learning algorithm, the framework captures the implicit correlations between network traffic data and urban environments. Meanwhile, by designing an attention-based fusion mechanism, we enhance the controllable generation capability of the framework. Evaluation results demonstrate that our scheme improves fidelity and controllability by 14.06% and 13.05%, respectively. Our model can generate network traffic for arbitrary regions using publicly available data sources, showing excellent deployability and reproducibility. Our framework has been deployed on the Jiutian intelligent platform of China Mobile, where we apply the generated network data to various practical scenarios, including resource scheduling, sleep strategies, and BS deployment. The test results prove that utilizing generated network data can effectively assist operators in network planning and optimization.

ACKNOWLEDGEMENT

This research has been supported in part by the National Key Research and Development Program of China under Grant 2023YFB2904801, the National Natural Science Foundation of China under Grants 62171260 and U22B2057, and the China Postdoctoral Science Foundation under Grant 2023M742010.

REFERENCES

- [1] Telecommunication Development Industry Alliance. 2023. Global 5G/6G Industry Development Report. <https://aimg8.dlssyht.cn/u/551001/ueditor/file/276/551001/1679536723602475.pdf>.
- [2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271
- [3] Alexandra Borgeaud. 2023. 5G deployment by country in Latin America 2023.
- [4] Alisson Assis Cardoso and Flávio Henrique Teles Vieira. 2019. Generation of Synthetic Network Traffic Series Using a Transformed Autoregressive Model Based Adaptive Algorithm. *IEEE Latin America Transactions* 17, 08 (2019), 1268–1275.
- [5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML '20)*. JMLR.org, Article 149, 11 pages.
- [6] Anestis Dalgkitis, Malamati Louta, and George T. Karetos. 2018. Traffic forecasting in cellular networks using the LSTM RNN. In *Proceedings of the 22nd Pan-Hellenic Conference on Informatics (Athens, Greece) (PCI '18)*. Association for Computing Machinery, New York, NY, USA, 28–33. <https://doi.org/10.1145/3291533.3291540>
- [7] Dingsheng Deng. 2020. DBSCAN clustering algorithm based on density. In *2020 7th international forum on electrical engineering and automation (IFEAA)*. IEEE, 949–953.
- [8] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* 34 (2021), 8780–8794.
- [9] Jiahui Gong, Yu Liu, Tong Li, Haoye Chai, Xing Wang, Junlan Feng, Chao Deng, Depeng Jin, and Yong Li. 2023. Empowering Spatial Knowledge Graph for Mobile Traffic Prediction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc> (SIGSPATIAL '23)*. Association for Computing Machinery, New York, NY, USA, Article 24, 11 pages. <https://doi.org/10.1145/3589132.3625569>
- [10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9726–9735. <https://doi.org/10.1109/CVPR42600.2020.00975>
- [11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239 [cs.LG]
- [12] Jonathan Ho and Tim Salimans. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).
- [13] Wenzhen Huang, Tong Li, Yuting Cao, Zhe Lyu, Yanping Liang, Li Yu, Depeng Jin, Junge Zhang, and Yong Li. 2023. Safe-NORA: Safe Reinforcement Learning-based Mobile Network Resource Allocation for Diverse User Demands. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 885–894.
- [14] Shuodi Hui, Huandong Wang, Tong Li, Xinghao Yang, Xing Wang, Junlan Feng, Lin Zhu, Chao Deng, Pan Hui, Depeng Jin, and Yong Li. 2023. Large-Scale Urban Cellular Traffic Generation via Knowledge-Enhanced GANs with Multi-Periodic Patterns.
- [15] Shuodi Hui, Huandong Wang, Zhenhua Wang, Xinghao Yang, Zhongjin Liu, Depeng Jin, and Yong Li. 2022. Knowledge Enhanced GAN for IoT Traffic Generation.
- [16] Spherical Insights. 2023. 2024 Telecom Business Insights. <https://www.sphericalinsights.com>.
- [17] Yifan Jiang, Shiyu Chang, and Zhangyang Wang. 2021. TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up. arXiv:2102.07074 [cs.CV]
- [18] Zhao jie Yang, Jie Lin, and Yu shu Yang. 2021. Identification of network behavioral characteristics of high-expertise users in interactive innovation: The case of forum autohome. *Asia Pacific Management Review* 26, 1 (2021).
- [19] He Li, Duo Jin, Xuejiao Li, Jianbin Huang, Xiaoke Ma, Jiangtao Cui, Deshuang Huang, Shaojie Qiao, and Jaesoo Yoo. 2023. DMGF-Net: An Efficient Dynamic Multi-Graph Fusion Network for Traffic Prediction. *ACM Trans. Knowl. Discov. Data* 17, 7, Article 97 (apr 2023), 19 pages.
- [20] Tong Li, Li Yu, Yibo Ma, Tong Duan, Wenzhen Huang, Yan Zhou, Depeng Jin, Yong Li, and Tao Jiang. 2023. Carbon emissions of 5G mobile networks in China. *Nature Sustainability* 6, 12 (2023), 1620–1631.
- [21] Zinan Lin, Alankar Jain, Chen Wang, Giulia Fanti, and Vyas Sekar. 2020. Using GANs for Sharing Networked Time Series Data: Challenges, Initial Promise, and Open Questions.
- [22] Qidong Liu, Fan Yan, Xiangyu Zhao, Zhaocheng Du, Huifeng Guo, Ruiming Tang, and Feng Tian. 2023. Diffusion Augmentation for Sequential Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (<conf-loc>, <city>Birmingham</city>, <country>United Kingdom</country>, </conf-loc> (CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 1576–1586. <https://doi.org/10.1145/3583780.3615134>
- [23] James E. Matheson and Robert L. Winkler. 1976. Scoring Rules for Continuous Probability Distributions. (1976).
- [24] Navid Naderializadeh, Mark Eisen, and Alejandro Ribeiro. 2020. Wireless Power Control via Counterfactual Optimization of Graph Neural Networks. In *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 1–5.
- [25] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. 2020. Contrastive Learning for Unpaired Image-to-Image Translation. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 319–345.
- [26] Fiana Raiber and Oren Kurland. 2017. Kullback-Leibler Divergence Revisited. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval (Amsterdam, The Netherlands) (ICTIR '17)*. Association for Computing Machinery, New York, NY, USA, 117–124. <https://doi.org/10.1145/3121050.3121062>
- [27] Markus Ring, Daniel Schlör, Dieter Landes, and Andreas Hotho. 2019. Flow-based network traffic generation using Generative Adversarial Networks. *Computers and Security* 82 (2019).
- [28] Chuanhao Sun, Kai Xu, Marco Fiore, Mahesh K. Marina, Yue Wang, and Cezary Ziemlicki. 2022. AppShot: A Conditional Deep Generative Model for Synthesizing Service-Level Mobile Traffic Snapshots at City Scale. *IEEE Transactions on Network and Service Management* 19, 4 (2022).
- [29] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. 2021. CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation. arXiv:2107.03502 [cs.LG]
- [30] Dhananjay Theckedath and RR Sedamkar. 2020. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Computer Science* 1 (2020), 1–7.
- [31] Xiaojie Wang, Laisen Nie, Zhaolong Ning, Lei Guo, Guoyin Wang, Xinbo Gao, and Neeraj Kumar. 2022. Deep Learning-Based Network Traffic Prediction for Secure Backbone Networks in Internet of Vehicles. *ACM Trans. Internet Technol.* 22, 4 (2022).
- [32] Zi Wang, Jia Hu, Geyong Min, Zhiwei Zhao, and Jin Wang. 2021. Data-Augmentation-Based Cellular Traffic Prediction in Edge-Computing-Enabled Smart City. *IEEE Transactions on Industrial Informatics* 17, 6 (2021), 4179–4187. <https://doi.org/10.1109/TII.2020.3009159>
- [33] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised Graph Learning for Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (Virtual Event, Canada) (SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 726–735. <https://doi.org/10.1145/3404835.3462862>
- [34] Yanxin Xi, Tong Li, Huandong Wang, Yong Li, Sasu Tarkoma, and Pan Hui. 2022. Beyond the First Law of Geography: Learning Representations of Satellite Imagery by Leveraging Point-of-Interests. In *Proceedings of the ACM Web Conference 2022 (Virtual Event, Lyon, France) (WWW '22)*. Association for Computing Machinery, New York, NY, USA, 3308–3316. <https://doi.org/10.1145/3485447.3512149>
- [35] Kai Xu, Rajkarn Singh, Marco Fiore, Mahesh K. Marina, Hakan Bilen, Muhammad Usama, Howard Benn, and Cezary Ziemlicki. 2021. SpectraGAN: Spectrum Based Generation of City Scale Spatiotemporal Mobile Network Traffic Data.
- [36] Weiwei Yan and Yin Zhang. 2019. User behaviors and network characteristics of US research universities on an academic social networking site. *Higher Education* 78 (2019), 221–240.
- [37] Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. 2023. Spatio-temporal Diffusion Point Processes. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (<conf-loc>, <city>Long Beach</city>, <state>CA</state>, <country>USA</country>, </conf-loc> (KDD '23)*. Association for Computing Machinery, New York, NY, USA, 3173–3184. <https://doi.org/10.1145/3580305.3599511>
- [38] Junhui Zhang, Jiqiang Tang, Xu Zhang, Wen Ouyang, and Dongbin Wang. 2015. A survey of network traffic generation. In *Third International Conference on Cyberspace Technology (CCT 2015)*. 1–6. <https://doi.org/10.1049/cp.2015.0862>
- [39] Shiyuan Zhang, Tong Li, Shuodi Hui, Guangyu Li, Yanping Liang, Li Yu, Depeng Jin, and Yong Li. 2023. Deep Transfer Learning for City-Scale Cellular Traffic Generation through Urban Knowledge Graph.
- [40] Feiyue Zhu, Lixiang Liu, and Teng Lin. 2020. An LSTM-Based Traffic Prediction Algorithm with Attention Mechanism for Satellite Network. In *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition (Xiamen, China) (AIPR '20)*. Association for Computing Machinery, New York, NY, USA, 205–209. <https://doi.org/10.1145/3430199.3430208>

A APPENDIX

A.1 Description of datasets

A.1.1 Mobile traffic datasets. The **Beijing** dataset covers mobile traffic data of 4000 BSs in Beijing during one week in October 2021. The **Shanghai** dataset contains the mobile traffic every 30 minutes from 5326 BSs over 150,000 users around Shanghai between August 1st and August 28th, 2019. The **Nanjing** dataset covers mobile traffic data from over 4000 BSs serving surpass 100,000 users in Nanjing between February 2nd and March 31st, 2021.

We set the time intervals of the network traffic sequence to be 1 hour. Additionally, to align with the geographical regions of satellite images, we aggregated the traffic of BSs based on the geographical coordinates corresponding to the satellite images. This allowed us to obtain the overall traffic within the geographical scope of each satellite image. Subsequently, we employed DBSCAN clustering [7], an algorithm that does not require specifying cluster centers, to cluster the traffic data and calculate the typical number of traffic patterns within the dataset, as shown in Figure 4. There are three typical traffic patterns in the Beijing and Shanghai datasets and four typical traffic patterns in the Nanjing dataset.

A.1.2 Metrics. We choose four types of metrics to investigate the performance of the algorithms.

JSD. The Jensen–Shannon Divergence (JSD) is a commonly used metric to measure the similarity between two distributions. Since the JSD calculation requires both vectors to follow a probability distribution format, which means that the sum of the elements in each vector must equal 1, we need to preprocess the generated network traffic data. For the generated network traffic data vector $\hat{S} = \{\hat{s}_t\}_{t=0:T}$, We convert it into a probability distribution form by $\hat{P} = \{\hat{s}_t / \sum_t \hat{s}_t\}_{t=0:T}$. Similarly, we convert the real network traffic data into a probability distribution form P . Then, the JSD can be formulated as:

$$J(P, \hat{P}) = \sqrt{\frac{KL(P||\hat{P}) + KL(\hat{P}||P)}{2}}, \quad (16)$$

where $KL()$ is the Kullback-Leibler divergence [26].

MAE. The Mean Absolute Error (MAE) metric evaluates the similarity of the generated network traffic data \hat{S} and the real traffic data S , which can be calculated as:

$$M(S, \hat{S}) = \frac{1}{T} \sum_{t=1}^T |s_t - \hat{s}_t|. \quad (17)$$

RMSE. The Root Mean Square Error (RMSE) measures the square root of the average squared differences between the real traffic data S and generated network traffic data \hat{S} , which can be expressed as:

$$R(S, \hat{S}) = \sqrt{\frac{1}{T} \sum_{t=1}^T (s_t - \hat{s}_t)^2}. \quad (18)$$

CRPS. The Continuous Ranked Probability Score (CRPS) [23] measures how well an estimated probability distribution aligns with an observed dataset. It serves as a metric to assess the accuracy of a generative distribution. A lower CRPS value indicates higher quality in the generated distribution. Given the cumulative distribution function (CDF) \hat{C} that can be calculated based on the probability

distribution \hat{P} , the CRPS can be formulated as:

$$CRPS(\hat{C}, S) = \int_{-\infty}^{\infty} (\hat{C}(y) - H(y - S))^2 dy, \quad (19)$$

where $H(y - S)$ is the Heaviside step function, which is 0 for $y < S$ and 1 for $y \geq S$.

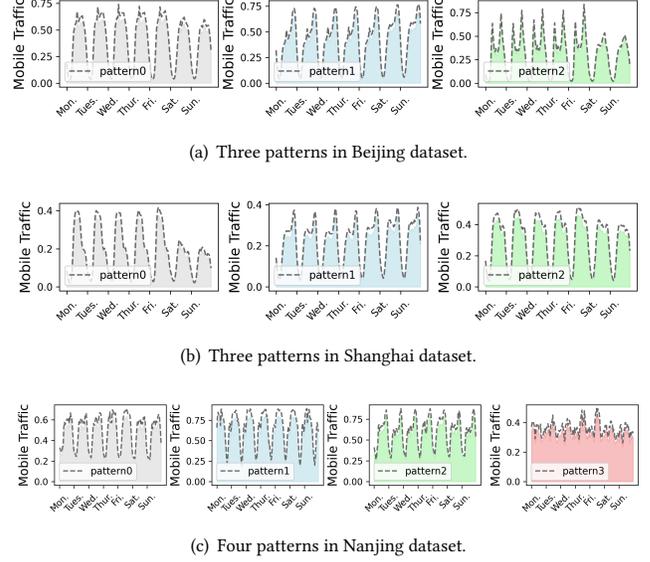


Figure 4: The traffic patterns in different datasets.

Table 6: Mobile traffic generation performance in Shanghai dataset. Bold numbers denote the best results and underline numbers denote the second-best results. Δ denotes the improvement between the best value and the second-best value. OpenDiff has good generative performance in the four metrics.

Metric	JSD	MAE	RMSE	CRPS
transGan	0.7632 ± 0.00054	0.2624 ± 0.00421	0.3390 ± 0.00313	0.8377 ± 0.00111
TCNGan	0.3258 ± 0.00152	0.2644 ± 0.00548	0.3331 ± 0.00575	<u>0.4789 ± 0.00144</u>
KEGAN	<u>0.2420 ± 0.00275</u>	0.2035 ± 0.00306	0.2537 ± 0.00253	0.4791 ± 0.01151
CSDI	0.2564 ± 0.00483	<u>0.1751 ± 0.00106</u>	<u>0.2122 ± 0.00132</u>	0.4862 ± 0.01698
spectralGan	0.3014 ± 0.00233	0.2029 ± 0.00205	0.2589 ± 0.00264	0.4873 ± 0.01796
OpenDiff (Our)	0.2321 ± 0.00479	0.1454 ± 0.00125	0.1919 ± 0.00152	0.4633 ± 0.01714
Δ	4.27	20.43	10.58	3.37

A.2 Mobile traffic generation performance of the Shanghai dataset

We also conducted the traffic data generation evaluation on the Shanghai dataset, as shown in Table 6. It can be observed that our OpenDiff algorithm also achieves the best generation performance. Similar to the results in Table 1, diffusion model-based models perform well overall in terms of MAE and RMSE, while GAN-based models perform well in terms of JSD and CRPS.