Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.057131

ARTICLE

A Location Trajectory Privacy Protection Method Based on Generative Adversarial Network and Attention Mechanism

Xirui Yang and Chen Zhang*

School of Cyber Security, Gansu University of Political Science and Law, Lanzhou, 730070, China

*Corresponding Author: Chen Zhang. Email: zc6454@gsupl.edu.cn

Received: 08 August 2024 Accepted: 30 September 2024 Published: 19 December 2024

ABSTRACT

User location trajectory refers to the sequence of geographic location information that records the user's movement or stay within a period of time and is usually used in mobile crowd sensing networks, in which the user participates in the sensing task, the process of sensing data collection faces the problem of privacy leakage. To address the privacy leakage issue of trajectory data during uploading, publishing, and sharing when users use location services on mobile smart group sensing terminal devices, this paper proposes a privacy protection method based on generative adversarial networks and attention mechanisms (BiLS-A-GAN). The method designs a generator attention model, GAttention, and a discriminator attention model, DAttention. In the generator, GAttention, combined with a bidirectional long short-term memory network, more effectively senses contextual information and captures dependencies within sequences. The discriminator uses DAttention and the long short-term memory network to distinguish the authenticity of data. Through continuous interaction between these two models, trajectory data with the statistical characteristics of the original data is generated. This non-original trajectory data can effectively reduce the probability of an attacker's identification, thereby enhancing the privacy protection of user information. Reliability assessment of the Trajectory-User Linking (TUL) task performed on the real-world semantic trajectory dataset Foursquare NYC, compared with traditional privacy-preserving algorithms that focus only on the privacy enhancement of the data, this approach, while achieving a high level of privacy protection, retains more temporal, spatial, and thematic features from the original trajectory data, to not only guarantee the user's personal privacy, but also retain the reliability of the information itself in the direction of geographic analysis and other directions, and to achieve the win-win purpose of both data utilization and privacy preservation.

KEYWORDS

Privacy protection; trajectory generation; generative adversarial networks; attention mechanism; location trajectory

1 Introduction

In the era of the Internet of Things (IoT), acquiring personal information has become increasingly effortless, with personal location data emerging as a crucial type of resource. This trend has given rise to a plethora of location-based systems, known as Location-Based Systems (LBS). These systems utilize mobile smart terminals, combined with location information, to offer various services and functionalities [1], thereby forming mobile crowd sensing networks. They find applications in numerous domains

such as social media, navigation, social services, health monitoring, and more. The core elements in the deployment of this type of system consist of two main aspects: (A) location-based services (LBSS) and (B) location data mining techniques for detecting points of interest and identifying transportation modes.

With the widespread adoption of LBSS like Facebook and Twitter, combined with location-aware technologies, LBS providers are able to collect not only people's daily travel footprints but also access large amounts of geospatially-tagged trajectory data [2]. LBS providers may share these location trajectories with third-party individuals or data analysts to perform more and more complex geographic data analysis tasks. These tasks may include recommending popular points of interest (POIs) [3], adding semantic annotations to POIs [4], analyzing people's movement patterns [5–8], and more!

From the early 21st century to the present, location-based systems have been increasingly deployed and used on the Internet of Things. Undoubtedly, they have brought significant convenience to individuals and society. For instance, trajectory big data has provided new opportunities for researching human activity patterns [9], disaster response [10], and more. However, people are also growing increasingly concerned about privacy issues stemming from the potential leakage of location information. Many people consider location information to be potentially classified as sensitive personal information. Because the leakage of information of this nature can easily jeopardize the security of the user's other information. For instance, through long-term analysis of a person's trajectory data, it becomes possible to discern sensitive information such as their workplace, residential area, social circles, and more [11]. With the help of this information, it may even be possible to deduce more intimate details such as their interests, political affiliations, economic status, and health conditions [12]. This situation can be viewed as a violation of personal freedom [13].

Track privacy means that users have the right to maintain the confidentiality of their track data and their related sensitive location information [14]. Traditional privacy-preserving methods, such as removing sensitive identifiers (e.g., name, ethnicity, etc.) from the data, have proven to be unreliable, as temporal, spatial, and thematic information contained in the track information can still provide a strong link to the user [15]; Geospatial aggregation, which aggregates personal trajectory point information into other geographic units, hides the user's location information, but it was found that this does not protect the privacy information well and reduces the effectiveness of spatial analysis to some extent [14].

To overcome these limitations, protect trajectory privacy more effectively, and enhance the analyzability and spatio-temporal analysis utility of the generated data, an innovative privacy-protecting method for generating fake trajectories, BiLS-A-GAN, is presented in this article. BiLS-A-GAN is built on a framework based on Generative Adversarial Networks (GANs). In this framework, the generator incorporates multi-layer Bidirectional Long Short-Term Memory Recurrent Neural Networks (BiLSTMs) and an improved Generator Attention Model. It encodes spatial features using an improved Generator Attention Model (GAttention) and introduces random noise in time, space, and category attributes, thereby preserving a certain level of differential privacy; The discriminator combines a dual-layer Long Short-Term Memory Recurrent Neural Network (LSTMs) with an improved Attention Model for Discriminators (DAttention) to discern sample quality more efficiently. The privacy-protecting trajectory data generated by BiLS-A-GAN retains the time, category, and spatial characteristics of the trajectory data to a certain extent. The generated trajectory is used as a substitute for the real trajectory for the sharing and publishing of trajectory data, which can effectively reduce the attacker's re-identification probability. While maintaining data privacy, the method in this

research also allows the reasonable use and research of generated data, providing a new solution in the field of trajectory data privacy protection.

2 Practical Applications and Case Studies

2.1 Related Works

Location trajectory data plays an increasingly important role in the mobile crowd-sensing network, and mobile smart terminal device applications such as location-based social networks and location-based health and fitness advice need to obtain the user's location trajectory data [16]. While location trajectory data brings convenience to people, it also causes people to think about privacy leakage security issues. Different users have different cognitive attitudes towards the identity, location, and environment [17] of people perceived by the wearable device terminals. If the perception device terminals can complete the quantitative evaluation of the perception data under the premise of minimizing the leakage of the user's location privacy, it will greatly promote the development of intelligence and informatization.

There are currently four main strategies for privacy protection of location trajectory data: (1) Geographical Hiding: A certain algorithm is used to perturb the original trajectory data in space or time to blur the original trajectory, while some spatial features still remain in the geographically hidden location [14]; Wang et al. [18] assessed eight different geographic obfuscation methods and proposed a framework for the evaluation of geographic obfuscation. (2) Trajectory Confusion: Liu et al. [19] proposed a location information protection algorithm based on bi-directional active hiding space, which realizes user location privacy protection through a virtual location privacy algorithm. (3) Differential Privacy. By introducing noise (e.g., Laplace noise) into the relevant query results, it becomes impossible for others to determine whether query data has a record. Chen et al. [20] introduced a trajectory privacy protection framework based on differential privacy in the context of mobile crowd sensing. This method offers real-time location information and can safeguard trajectory privacy even under Bayesian inference attacks. (4) Anonymization. Translating trajectory privacy into a K-anonymity problem, Zhu et al. [21] proposed a privacy-preserving framework that can construct sets of anonymized trajectories and classify the anonymized trajectories.

Although these methods have shown certain trajectory privacy protection capabilities, their focus is on confusing trajectory location information [22], and more on the study of spatial features to enhance uncertainty. However, semantic features (such as points of interest and time dimensions) have been proven to be able to infer trajectory user information with a high probability [23]. Methods used in the literature may protect trajectory data well to some extent, but little consideration has been given to the issue of balancing uncertainty and practicality in trajectory protection methods, neglecting the data quality and availability of fuzzy trajectories.

Machine learning technologies have advanced rapidly in recent years, providing new perspectives for protecting trajectories' privacy. Generative Adversarial Networks (GAN) [24] represents an unsupervised generative model framework. As shown in Fig. 1, GANs typically consist of two parts: a generator and a discriminator, which are usually neural networks. Based on the feedback from the discriminator, the generator continuously improves fake samples, aiming to generate as realistic fake samples as possible to deceive the discriminator. At the same time, the discriminator continuously optimizes its own discriminative ability, aiming at identifying the generated fake samples as soon as possible. They generate high-quality synthetic data by means of a minimal-extremely large game, which has the same distribution as the training data. GAN requires relatively little input data and is widely

applied to the production of high-quality images that appear natural, and GANs also have a wide range of applications in privacy preservation.

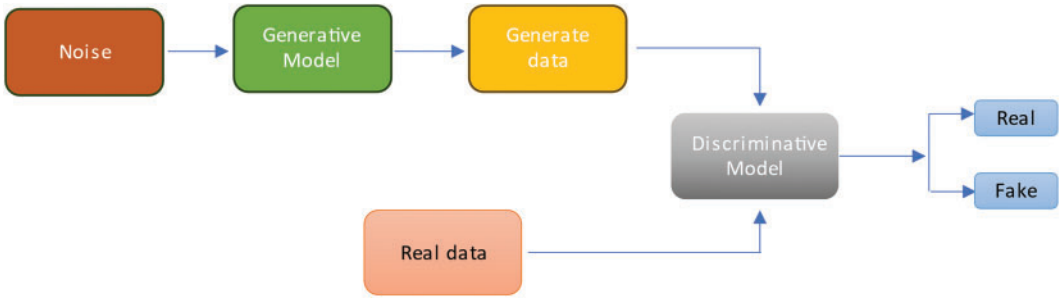

Figure 1: Diagram of GAN model

Utilizing the generated location trajectories [15] for publishing and sharing trajectory information is one of the most promising approaches to protecting user privacy. The BiLS-A-GAN proposed in this research is based on the GAN framework for synthesizing trajectory locations. During the training process, the generative model learns, thus preserving some statistical features of the original trajectories, such as POI categories, time, latitude, and longitude, in the synthesized trajectories. Based on the above features, BiLS-A-GAN can generate high-quality trajectory information that confuses the real trajectory data, which can better protect people's privacy, and at the same time retain a certain amount of geographic information of the trajectory data, striking a balance between enhancing the effectiveness of privacy protection methods and the practicality of generated data.

2.2 Practical Applied Research

Nowadays, smart devices integrate a variety of sensors and features, such as location sensors, health-tracking sensors, and Bluetooth, which have dramatically changed people's lifestyles. However, the widespread use of these embedded sensors after the Corona Virus Disease 2019 (COVID-19) outbreak increased the risk of user privacy leakage [25,26]. Troncoso et al. [27] proposed and analyzed a decentralized proximity-tracking system for COVID-19 propagation, that focuses on security and privacy preservation by using anonymous identifiers to tag COVID-19-positive users without the need to provide exact location information to the authorities.

Meanwhile, as the use of location data in location-based advertising (LBA) continues to grow, disclosing location information to untrusted service providers has raised significant privacy concerns [28]. Users' real-time location data can be easily collected by LBA service providers [29]. To address these challenges, Yu et al. [28] designed a system called PrivLocAd, which integrates an optimal two-stage mechanism based on geographic indistinguishability and a multilevel alternative generation mechanism to improve the balance between privacy and data utility by generating multiple proxy locations.

Additionally, the widespread use of mobile devices has increased the complexity of consumer choices, with recommendation systems collecting and analyzing user data to provide an enhanced user experience. However, this process also carries the risk of leaking sensitive user information (e.g., identity, location, religion, etc.). For this reason, Shin et al. [30] developed a novel matrix factorization algorithm based on Local Differential Privacy (LDP). In this type of recommendation system, individual users randomize their data to satisfy differential privacy requirements before sending the perturbed data to the recommender, thereby effectively protecting user privacy.

These cases show that although smart devices bring convenience to people's lives, the privacy protection issue during their data usage still needs to be carefully considered, especially the trade-off between privacy protection followed by data utility, which is one of the core challenges of trajectory privacy protection. In this study, the BiLS-A-GAN scheme is proposed aiming to strike a balance between trajectory privacy and data utility, and is planned to be deployed in a mobile smart swarm sensing network.

Human mobility trajectory data have significant application value; however, real-world mobility data can pose serious privacy concerns, making it challenging to support downstream applications. This study explores such challenges, and Fig. 2 illustrates the deployment architecture of the trajectory privacy protection method using synthetic trajectory generation on mobile devices.

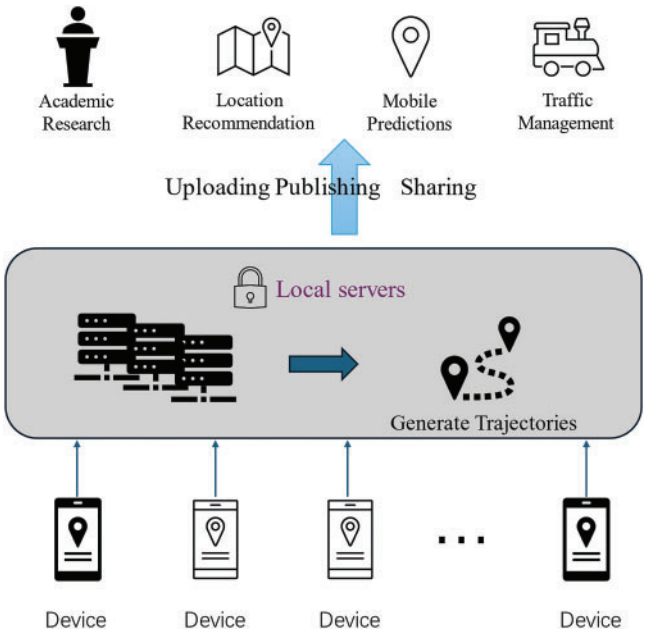

Figure 2: Deployment of the trajectory generator in mobile crowd sensing networks

Mobile devices generate trajectory data with privacy-preserving features through a local server, where the BiLS-A-GAN is integrated into the local server to generate high-quality synthetic trajectory data to protect user privacy when uploading, sharing, and publishing location information. This approach can be deployed in a variety of application scenarios. For example, in urban construction, the local server can collect, and process vehicle trajectory data and the government can provide important guidance for transportation system planning based on the travel trajectories of the generated people. In academic research, this method can provide researchers with anonymized data for pattern analysis and model training, while effectively avoiding exposing the actual locations of individuals.

All real trajectory data are stored on local servers and access is granted only with user consent, ensuring both data-sharing security and comprehensive user privacy protection. This deployment strategy not only safeguards users' personal privacy, but also meets the need for data sharing and utilization in different scenarios, achieving a balance between data utility and privacy protection.

3 BiLS-A-GAN Model

The original GAN [24] expression is:

$$\min_{G}\max_{D} V(D,G) = \min_{G}\max_{D} (E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

In (1), the discriminator D is trained to maximize the likelihood of correctly classifying training samples (i.e., maximizing $logD(x)$ and $log(1-D(G(z)))$, while the generator G is trained to minimize $log(1-D(G(z)))$, which is equivalent to maximizing D's loss. One party is fixed during the training process, the network parameters of the other party are updated, and iterations are alternated to maximize the error of the other party. Lastly, generator G is capable of estimating the distribution of sample data, which enhances the realistic nature of the generated samples. During the training process, the goals of the generator and the discriminator are contradictory, and this contradiction can be reflected in the accuracy of the discriminator's judgments. BiLS-A-GAN leverages this adversarial process through the interaction of adversarial training between the generator and the discriminator. The generator learns from the input of real trajectory data and generates synthetic trajectory data, while the discriminator receives both the real and synthetic trajectory data as input and attempts to distinguish between them. Based on the feedback from the discriminator, the generator improves the authenticity of the generated data by minimizing the discriminator's accuracy, while the discriminator enhances its ability to differentiate between real and generated data by maximizing its classification accuracy. This process is repeated multiple times until the generator can produce synthetic data that is almost indistinguishable from the real data. Through this adversarial training, more realistic trajectory data is generated, ultimately achieving a high-quality privacy protection effect.

3.1 BiLS-A-GAN Model Framework

The model in this work utilizes the idea of GAN to input the encoded data into the BiLS-A-GAN model shown in Fig. 3, and through the repeated games between the discriminator and the generator targets, it continuously distinguishes between the real data and the generated data until it generates a synthesized trajectory that is close to the "real" one. We retain more categorical, temporal, and spatial information in the synthesized trajectories, which can better ensure the quality of the user-connection (TUL) task analysis without compromising personal privacy.

3.1.1 Trajectory Generator Design

The trajectory generator, as depicted in Fig. 3, embeds trajectories by feeding the encoded real trajectories and random noise into multilayer perceptrons. When it comes to the trajectory's spatial dimension, this paper obtains a 64-dimensional vector through multi-layer perceptron embedding, and the time and category attributes are embedded into 7-dimensional, 24-dimensional, and 10-dimensional vectors, respectively.

$$u_i^s = \tau^s(\Delta lat_i, \Delta lon_i; W_{us}) \tag{2}$$

$$u_i^d = \tau^d\left(v_i^d; W_{ud}\right) \tag{3}$$

$$u_i^h = \tau^h\left(v_i^h; W_{uh}\right) \tag{4}$$

$$u_i^c = \tau^c\left(v_i^c; W_{uc}\right) \tag{5}$$

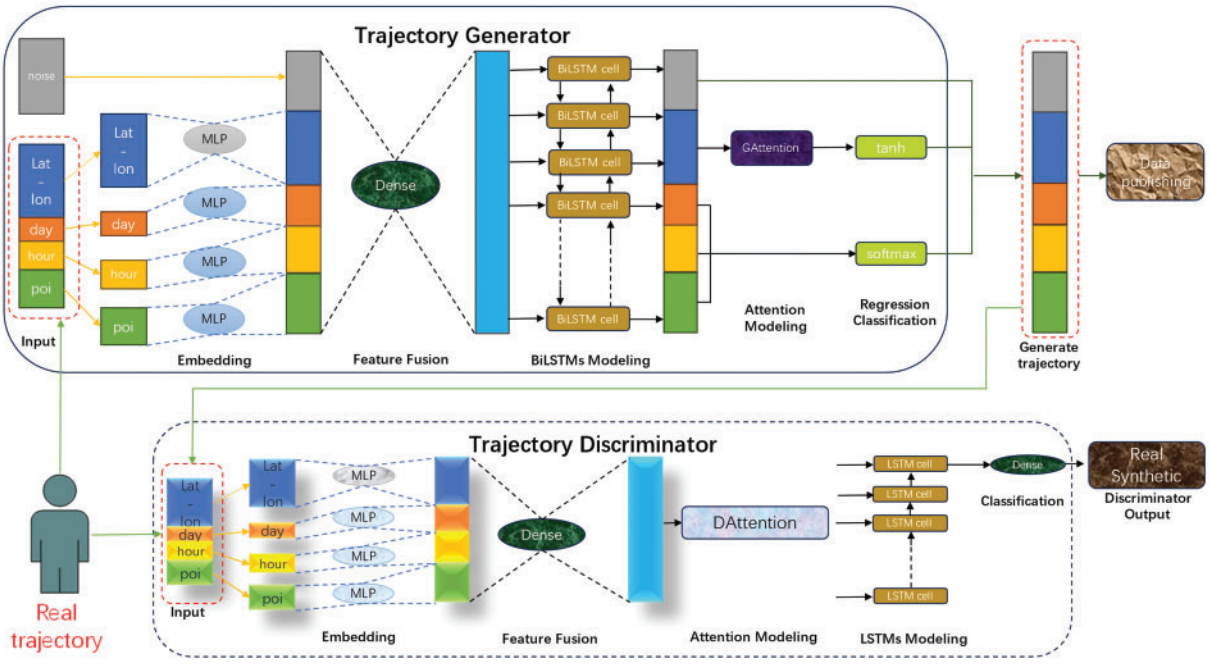

Figure 3: The BiLS-A-GAN model diagram

Sum the latitude and longitude extracted from the training and test samples and divide the value by the sum of the coordinates of the two samples to obtain the centroid coordinates to calculate the latitude and longitude deviation Δlat_i and Δlon_i of the i-th location point; v_i^d, v_i^{h} and v_i^c represent one-hot vectors of day-of-week, hourly and category attributes for the i-th trajectory point; τ^s, τ^d, τ^{h} and τ^c represent multi-layer perceptrons with activation functions for embedding spatial, weekday, hour and category attribute units, where the activation functions are all RELU (Rectified Linear Unit) and use He uniform variance Scaling initialization weights; W_{us},W_{ud}, W_{uh}, W_{uc} are the embedding weight matrices of these MLPs (Multilayer Perceptron); u_i^s, u_i^d, u_i^{h}, u_i^c are the embedding vectors of each attribute, respectively.

After the embedding step is completed, all embedded vectors along with the added random noise are concatenated together and passed through a Dense layer with 100 units, integrating them into a 100-dimensional vector. BiLS-A-GAN uses a three-layer many-to-many BiLSTM structure with 100 units in each BiLSTM layer, uses L1 regularization to prevent overfitting, and the BiLSTMs layer returns the entire sequence of outputs, not just the output of the last time step.

$$H_{G_0} = BiLSTM\,(F; W_{lstm}) \tag{6}$$

$$H_{G_1} = BiLSTM\left(H_{G_0}; W_{lstm}\right) \tag{7}$$

$$H_{G_2} = BiLSTM\left(H_{G_1}; W_{lstm}\right) \tag{8}$$

In the given context, where F represents the fused trajectory features, f_i represents the fusion feature vector of the i-th trajectory point, and H_{G_0}, H_{G_1} and H_{G_2} respectively represent the outputs of each layer of the BiLSTMs model, with the same time step as the input. They all return complete output sequences. The matrix W_{lstm} represents the weight matrix of the BiLSTMs model. Relative to traditional single-layer LSTM structures, stacking BiLSTM layers in this manner creates a deeper

network structure, enhancing its ability to capture long-term dependencies in sequences and better extract deep relationships in time-series data. This is beneficial for modeling sequence data and feature learning because different layers of BiLSTMs can capture patterns and relationships at different levels of abstraction.

The output of the BiLSTMs layer will go to the Attention Mechanism module, where BiLS-A-GAN will perform attentional operations on spatial attributes through GAttention. In the generator, GAttention focuses on the spatial features of trajectory data, dynamically highlighting the important parts of the input sequence (location points) to more effectively capture the spatial dependencies of the trajectory data. Combined with a multi-layer bidirectional LSTM, it captures the contextual information within the trajectory data. This design structure takes sequences with specific time steps as input and generates sequences with the same time steps as output. The new trajectories generated based on this information better match the statistical characteristics of the original trajectories, enabling the generator to continuously improve during training and produce more realistic samples. The model diagram for BiLSTMs and GAttention layers in the paper is shown in Fig. 4.

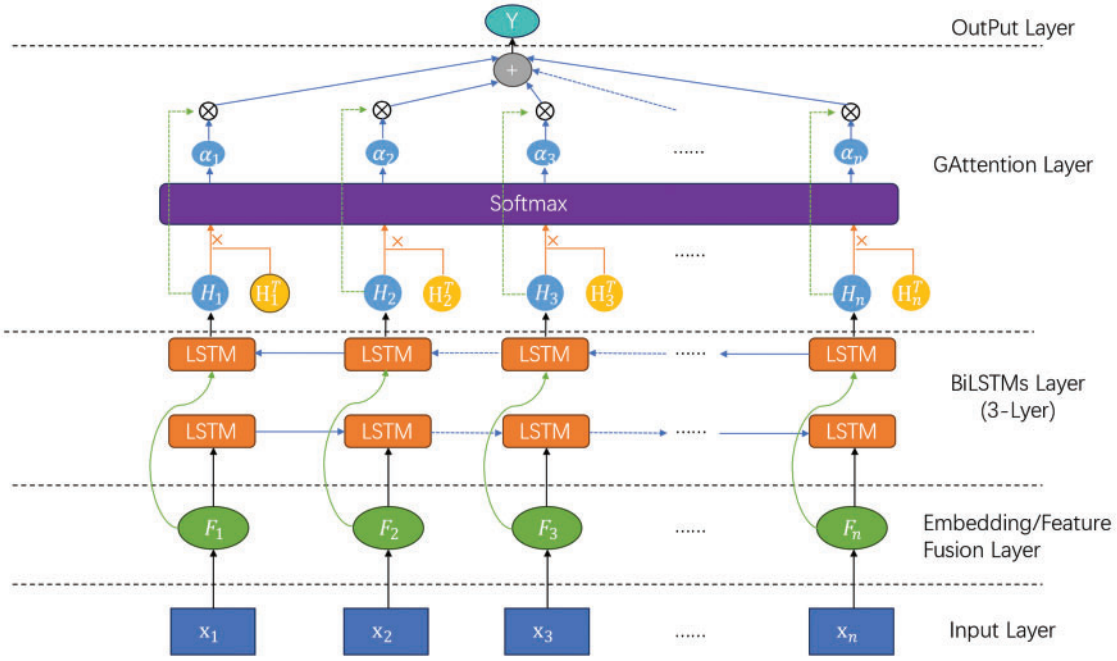

Figure 4: Model diagram of BiLSTMs and GAttention layers

Next, we will design the improved GAttention introduction model and act on the spatial attributes, the detailed algorithmic procedure of the GAttention model is shown in Fig. 5.

The vector set output by the BiLSTMs layer is expressed as H: $[h_0, h_1, h_2, h_s, \ldots, h_{maxlength-1}]$. The weight matrix of the attention layer is calculated as follows:

$$W_S = softmax\left(H_s * H_s^T\right) \tag{9}$$

W_S represents the attention weight matrix, where H_s and H_s^T represent the part of the output from the BiLSTMs model that represents spatial attributes and its transpose, respectively. Calculating their product allows for a better capture of the relationships in the spatial dimension of the input. Through a

softmax operation, it is turned into a normalized attention weight matrix, which will be used to weigh the BiLSTMs outputs.

$$C_S = W_S \cdot H_s \tag{10}$$

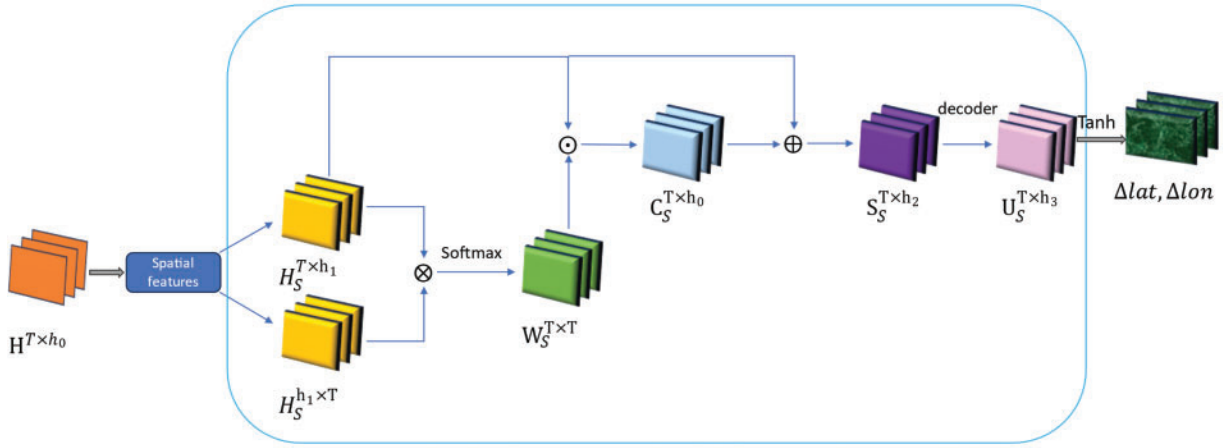

Figure 5: GAttention model algorithm

The attentional weight matrix W_S is subjected to a dot-product operation with the spatial attribute portion H_s of the BiLSTMs output in the specified dimensions to obtain the context vector C_S weighted in the spatial dimension.

$$S_S = concat\,(C_S, H_s) \tag{11}$$

The spatial attribute part H_s and the spatial context vector C_S in the original BiLSTMs output are spliced to generate an attention-enhanced feature vector S_S, which contains both the original sequence information and the spatial context information.

$$U_S = \varphi_S\,(S_S) \tag{12}$$

U_S represents the decoding output of the enhanced feature vector S_S, and φ_S represents the fully connected layer with two neurons, which acts on each time step in S_S to decode the latitude and longitude coordinates and complete the conversion of the entire sequence into a coordinate sequence.

$$(\Delta \mathrm{U}lat_i, \Delta Ulon_i) = D_T^s\,(U_S; W_{Ss}) \tag{13}$$

$\Delta \mathrm{U}lat_i, \Delta Ulon_i$ represents the latitude and longitude deviation of the i-th trajectory after decoding of the fully connected layer. W_{Ss} represents the weight matrix acting on the fully connected layer, D_T^s represents the fully connected layer whose activation function is Tanh, and W_{Ss} represents the weight matrix acting on the fully connected layer.

For the attributes like day of the week, hour, and category, the decoding output is obtained directly through a fully connected layer:

$$u_i^{'d} = D_S^d\,(H_i; W_{Td}) \tag{14}$$

$$u_i^{'h} = D_S^h\,(H_i; W_{Th}) \tag{15}$$

$$u_i^{'c} = D_S^c\,(H_i; W_{Tc}) \tag{16}$$

D_S^d, D_S^h, D_S^c represent a fully connected layer with the number of neurons. The number of neurons is the same as the vocabulary size corresponding to the different attributes of this part. Softmax is used as the activation function. H_i is enhanced by the BiLSTMs layer. After the feature vector, W_{Td}, W_{Th}, W_{Tc} represents the weight matrix, $u_i'^d$, $u_i'^h$, $u_i'^c$ represent the one-hot of week, hour and POI attributes in the i-th generated trajectory after decoding vector.

3.1.2 Trajectory Discriminator Design

From Fig. 3, we can observe that the trajectory discriminator and the trajectory generator have similar structures, but they also have the following differences:

(1) The discriminator does not require random noise as input during training.

(2) After feature fusion, BiLS-A-GAN introduces the improved attention mechanism DAttention before the LSTMs layer. The DAttention attention mechanism learns the attention weights for each temporal step based on the content of the input sequence itself, without introducing additional externally coded information such as position and time. Compared to other types of attention mechanisms, such as location-based attention that introduces location coding, DAttention is simple and efficient to compute. DAttention allows for dynamic weighting of the input sequence, enabling the model to concentrate on the most important information of the sequence. The DAttention structure diagram is shown in Fig. 6, which helps the model to understand the key information of the input sequences more efficiently and thus improves the performance and representation of the model.

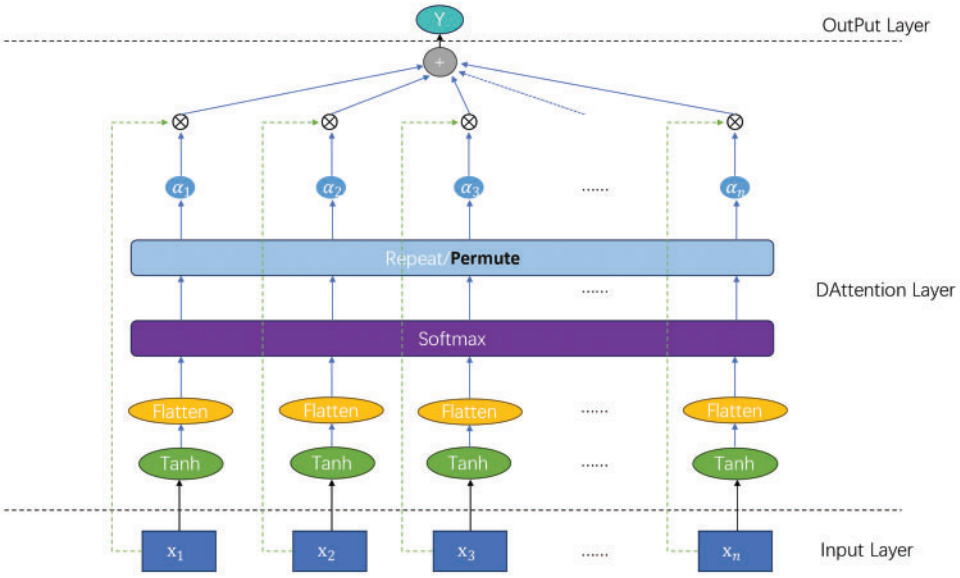

Figure 6: Model diagram of the DAttention layer

$$S_D = \varphi_D\left(Tanh\left(F_D\right)\right) \tag{17}$$

F_D represents the feature fusion of all trajectory information (i.e., $F_D = [F_0, F_1, \ldots, F_{maxlength-1}]$). After feature fusion, the sequence passes through a fully connected layer with a Tanh activation function to compute attention scores for each tensor in F_D, indicating the importance of each time step in the sequence. φ_D flattens the attention scores into 1-dimensional, and S_D represents the flattened

attention score output.

$$W_D = softmax\,(S_D) \tag{18}$$

W_D means normalizing the attention scores S_D through softmax and converting them into probability distributions representing attention weights.

$$W_D = \tau_D\,(W_D) \tag{19}$$

τ_D has the following representation: each feature vector in F_D after feature fusion is 100-dimensional, copy the normalized attention weight W_D 100 times, and align them with the length of the input sequence. Then the shape of the aligned tensor is adjusted so that the weight matrix has the same shape as the input sequence, which facilitates broadcasting the attention weight to each feature dimension.

$$Att_D = W_D * F_D \tag{20}$$

The original fused feature F_D is multiplied element-wise with the corresponding position-weighted W_D to obtain a new sequence Att_D. Each feature vector at each time step is weighted, enhancing or diminishing information from different time steps based on the attention weights. This allows for better adaptation to the task requirements.

(3) After passing through the attention layer, we use a two-layered multi-to-one LSTM to transform the input into scalar output:

$$H_{D_0} = LSTM\left(Att_D, W_{D_L}\right) \tag{21}$$

$$H_{D_1} = LSTM\left(H_{D_0}, W_{D_L}\right) \tag{22}$$

Att_D denotes the output enhanced by the attention mechanism layer, and W_{D_L} is the weight matrix of the LSTMs layer in the trajectory discriminator. Both layers of LSTMs use L1 regularization with a coefficient of 0.02, the difference is that H_{D_0} returns the complete sequence of outputs, and H_{D_1} returns only the output of the last time step, not the complete sequence.

(4) Finally, BiLS-A-GAN uses a fully connected layer with sigmoid as the activation function to binary classify the output scalars of the LSTMs layer to determine the similarity of the generated trajectories to the original trajectories.

$$out_D = \varphi_D\left(H_{D_1}, W_{D_o}\right) \tag{23}$$

φ_D represents the fully connected layer for binary classification, W_{D_o} is its weight matrix, H_{D_1} is the output of the LSTMs layer, and out_D is the final output of the discriminator.

3.2 Enhanced Optimization Loss Function

The original GAN loss is prone to shortcomings such as gradient disappearance and unstable training, in order to enhance the training quality of the BiLS-A-GAN model and better reflect the realism of the generated data, this paper invokes the TrajLoss loss metric function to train the generator, to more accurately measure the loss between the generated data and the real data.

$$TrajLoss\,(y^r, y^p, t^r, t^s) = \alpha L_{BCE}\,(y^r, y^p) + \beta L_s\,(t^r, t^s) + \gamma L_t\,(t^r, t^s) + cL_c\,(t^r, t^s) \tag{24}$$

During training, the discriminator uses the original binary cross-entropy loss L_{BCE}. This means that gradient signals are generated for both positive and negative samples that are misclassified, and

these signals can be used to adjust the weights of the neural network to minimize misclassification so that the learned weights will be more reasonable and more helpful in distinguishing between real and generated trajectories.

L_s, L_t and L_c represent the similarity losses between real and generated trajectories in terms of time, space, and category. Time and category are treated as multi-class problems in the paper, so L_t and L_c use SCE (Softmax Cross-Entropy) as the loss function. These losses are then averaged to contribute to more robust model training.

Since the trajectory data usually have different time steps or there may be no trajectory points at some time steps, the trajectory padding was previously performed to bring all the trajectory data to a uniform length, and the masking operation is performed here to deal with the trajectory data with different lengths so that the trajectory data with padding of 0 can be cropped out to reduce the effect of invalid loss on the model, and to make the loss more accurately reflected in the prediction accuracy of the trajectory segment that has trajectory segments to increase the generalization ability of the model.

α, β, γ, and c represent the weights of these losses, which can be adjusted and assigned differently depending on the context to better balance the importance of the different losses during training. This weight adjustment allows optimizing the performance of the model according to the task requirements.

4 Experiments and Analysis of Results

4.1 Data Sets and Preprocessing

This article uses the Foursquare NYC [31] check-in data set, which is the New York City check-in data extracted from the global check-in data set collected by Foursquare from April 2012 to September 2013. Here, similar to Rao et al. [32], we only extract the user ID (Identity Document), trajectory ID, location, hour, weekday, and POI category attributes from the dataset. Table 1 presents the attributes of the trajectory data in this dataset.

Table 1: Description of Foursquare trajectory data attributes in New York City

Attribute	Type	Example	Number/Range
Trajectory ID	Integer	{126, 131, 135 . . . }	3079
User ID	Integer	{371, 372, 384 . . . }	193
Latitude	Float	{40.740709, 40.744285, 40.742544 . . . }	(40.550852, 40.988332)
Longitude	Float	{−73.948376, −74.006322, −73.993963 . . . }	(−74.269644, −73.685767)
Hour	Integer	{9, 18, 22 . . . }	24
Day	String	{Monday, Tuesday, Wednesday . . . }	7
Category	String	{Arts & Entertainment, Event, Shop & Service}	10

User trajectories are collections of continuous location points with timestamps, such as $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow \ldots \rightarrow S_n (S = <X, Y, T>)$, where X represents longitude, Y represents latitude, and T represents time). However, these location point data can be challenging to use directly and often require transformation into a more easily encoded format. We consider the complexity of POI categories and temporal attributes and use one-hot to encode them as d-dimensional vectors with 1, where d refers to the number of values the attribute can take. As shown in Fig. 7, there are 24 different values for hours, so they are encoded as 24-dimensional vectors. Similarly, there are 7 different values for days

of the week, which are encoded as 7-dimensional vectors. In the Foursquare NYC dataset, if there are 10 POI categories, they would be encoded as 10-dimensional vectors.

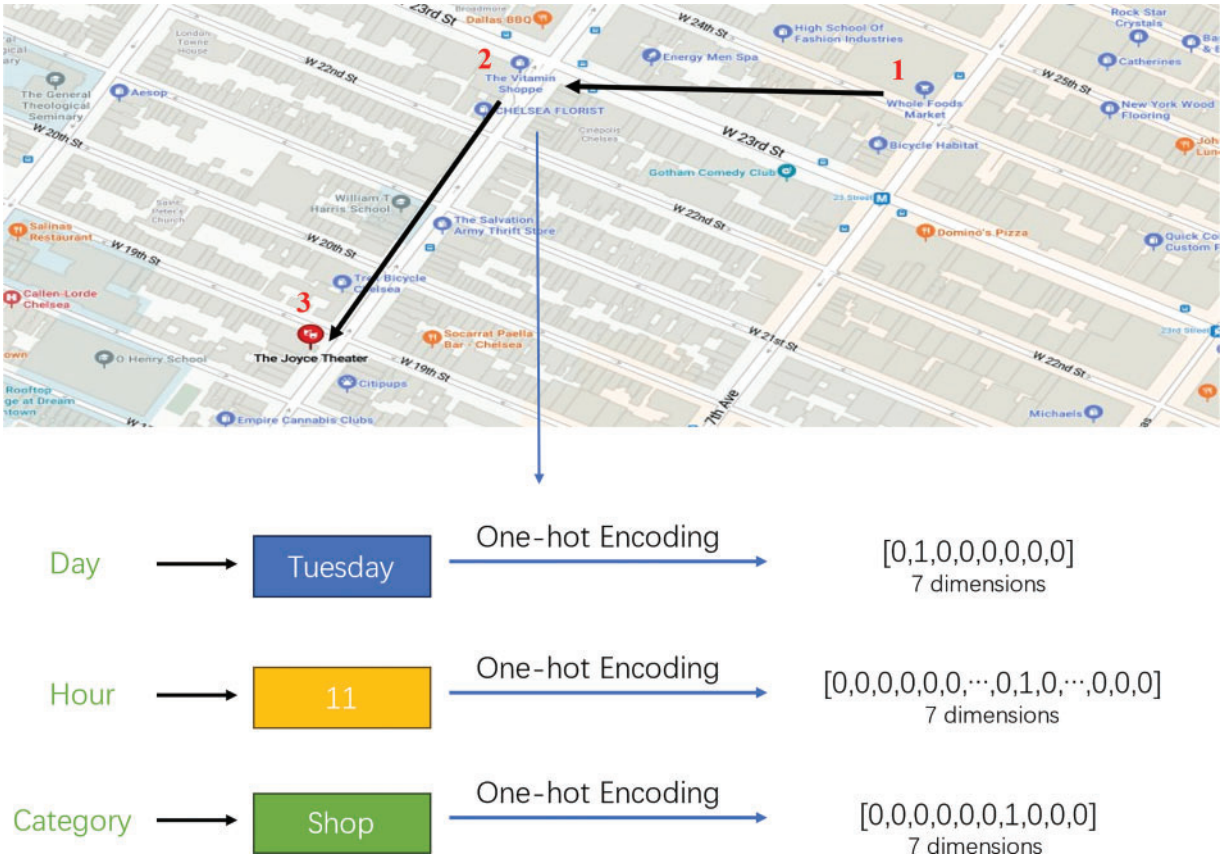

Figure 7: Trajectory encoding

In different user trajectories, the number of trajectory points is always different, the method used in this work is to fill more empty trajectory points into the trajectory data and set their values to zero to maximize the length of all the trajectories to better optimize the training.

4.2 Experimental Setup and Evaluation Indicators

4.2.1 Parameter Setting

In this work, we run experiments on the Foursquare NYC check-in dataset, where 2/3 of the data is used for experiments and 1/3 of the data is used for testing. In this experiment a total training period of 2000 rounds is set, the training parameters are saved every 100 rounds interval, and the batch size is set to 256. LS-A-GAN uses the Adam optimizer, the learning rate is set to 0.001, and clipvalue is set to 1.0, and the gradient is trimmed to prevent the gradient update from being too large, which leads to the instability of the model. All experiments were trained and tested using PyTorch on a GPU (Graphic Processing Unit) server equipped with a Tesla P100-PCIE-16G. Table 2 illustrates some parameter settings for the model.

Table 2: Parameter list

Parameters	Value	Parameters	Value
Optimizer	Adam	Sample_interval	100
Learning rate (lr)	0.001	LSTMs layers	2
Clip value	1.0	BiLSTMs layers	3
Batch_size	256	n_epochs	2000

4.2.2 Evaluation Indicators

This paper employs the advanced TUL algorithm from [23] to perform TUL tasks on our synthesized data. During this process, we utilize the following metrics for evaluation: ACC@1 (Top-1 Accuracy): Measures the accuracy of the model when predicting the most likely class; ACC@5 (Top-5 Accuracy): Measures the accuracy of the model when predicting within the top five most likely classes; Macro-P (Macro Average Precision): The average precision across all classes; Macro-R (Macro Average Recall): The average recall across all classes; Macro-F1 (Macro Average F1-Score): A metric that combines Macro-P and Macro-R to evaluate the model's balance in predicting positive and negative instances.

$$ACC@K = \frac{\mathcal{T}_K}{|\mathcal{L}_{test}|} \tag{25}$$

$\mathcal{L}_{test}$ represents the collection of trajectory data in the test set, while $\mathcal{T}_K$ represents the collection of correctly predicted classes among the top K classes with the highest confidence in trajectory prediction.

The formulas for calculating Macro-P (Macro Precision), Macro-R (Macro Recall), and Macro-F1 (Macro F1-Score) are as follows:

$$\text{Macro} - \text{P} = \frac{1}{|\mathcal{T}|} \sum_{X \epsilon \mathcal{T}} \frac{TP_X}{TP_X + FP_X} \tag{26}$$

$$\text{Macro} - \text{R} = \frac{1}{|\mathcal{T}|} \sum_{X \epsilon \mathcal{T}} \frac{TP_X}{TP_X + FN_X} \tag{27}$$

$$\text{Macro} - \text{F1} = 2 * \frac{\text{Macro} - \text{P} * \text{Macro} - \text{R}}{\text{Macro} - \text{R} + \text{Macro} - \text{P}} \tag{28}$$

The TP_X, FP_X and FN_X are the number of positive samples correctly recognized, the number of negative samples misreported, and the number of positive samples omitted, respectively, in category X.

4.3 Analysis of Experimental Results

Baselines: In this work, we select several common trajectory privacy-preserving methods, Gaussian Geomasking [14], Trajectory Synthesis for Generative Adversarial Networks [32], and Differential Privacy (DP) [33] as a baseline for experimental comparison.

Gaussian Geomasking: Its fundamental concept involves introducing Gaussian-distributed random noise to geographic location data, thereby blurring specific geographic coordinate information and making it difficult for attackers to accurately determine an individual's real location. In this research, we will add random noise that follows a Gaussian distribution, with $\mathcal{X} \sim \mathcal{N}(\mu, \sigma^2)$, to

the longitude and latitude of each trajectory point. Here, μ and σ represent the mean and standard deviation of the random noise, and in this case, we set $\mu = 0$ and $\sigma = 0.02$.

Differential Privacy (DP): $Pr[M(s) \in \mathcal{O}] \leq e^{\epsilon} \cdot Pr[M(s') \in \mathcal{O}] + \delta$. In our experiments, we employ the Laplace mechanism, denoted as $M_L(x, \epsilon) = f(x) + (Y_1, \ldots, Y_k)$, which introduces perturbations during the training process. Here, $Y_i \sim Lap(0, \frac{\Delta f}{\epsilon})$ represents the probability density function of the added random noise, where Δf is the L_1 sensitivity, and ϵ serves as the scaling factor (referred to as privacy budget in standard differential privacy). In this research, we set $\epsilon = 10$.

Trajectory Synthesis with Generative Adversarial Networks: This method involves training a generative model using the original trajectories and then using the trained model to synthesize fake trajectories that retain certain statistical characteristics of the original ones. In the experiments, we use the LSTM-TrajGAN [32] method as a baseline for comparison.

4.3.1 Analysis of the Results of Accuracy Experiments

In this work, the BiLS-A-GAN model adopts the trajectory synthesis strategy of a generative adversarial network, which incorporates the improved attention model and bi-directional long and short-term memory network to train the model to generate realistic trajectory data. The experimental results of different methods are shown in Table 3.

Table 3: Privacy protection effects of different privacy preservation methods on the TUL task (DP represents Differential Privacy)

Method	ACC@1	ACC@5	Macro-F1	Macro-P	Macro-R
LSTM-TrajGAN	0.4099	0.7098	0.3536	0.3934	0.3897
Gaussian Geomasking	0.2524	0.6484	0.2975	0.3345	0.3438
DP (Temporal)	0.1080	0.2921	0.0764	0.0860	0.0968
DP (Temporal-Spatial)	0.0068	0.0272	0.0003	0.0001	0.0057
BiLS-A-GAN	**0.3115**	**0.5920**	**0.2428**	**0.2800**	**0.2829**

The higher the TUL accuracy, the worse the effectiveness of trajectory privacy protection. Based on the results in Table 3, it can be observed that an increase in the accuracy of the model TUL is accompanied by an increase in the risk of the user's privacy being re-identified. However, the BiLS-A-GAN model in this research successfully reduced all four evaluation metrics (ACC@1, Macro-F1, Macro-P, Macro-R) to around 0.3000, and ACC@5 was also reduced to 0.5920. This shows that the model BiLS-A-GAN in this research is robustly protective not only on a single category, but also on multiple categories, and can greatly reduce the risk of trajectory privacy leakage.

After BiLS-A-GAN training is completed, we input the trajectory data and random noise from the test set into the generator and get the synthetic trajectory data after model prediction. As shown in Fig. 8, in this work, we randomly select the original trajectories with the same labels and the generated trajectories and visualize them for comparison. From the analysis of the visualization diagram, it can be seen that the user trajectories generated by the model in this research retain the spatial and temporal attributes of the original trajectories as much as possible, which provides a possibility for further geographic data analysis.

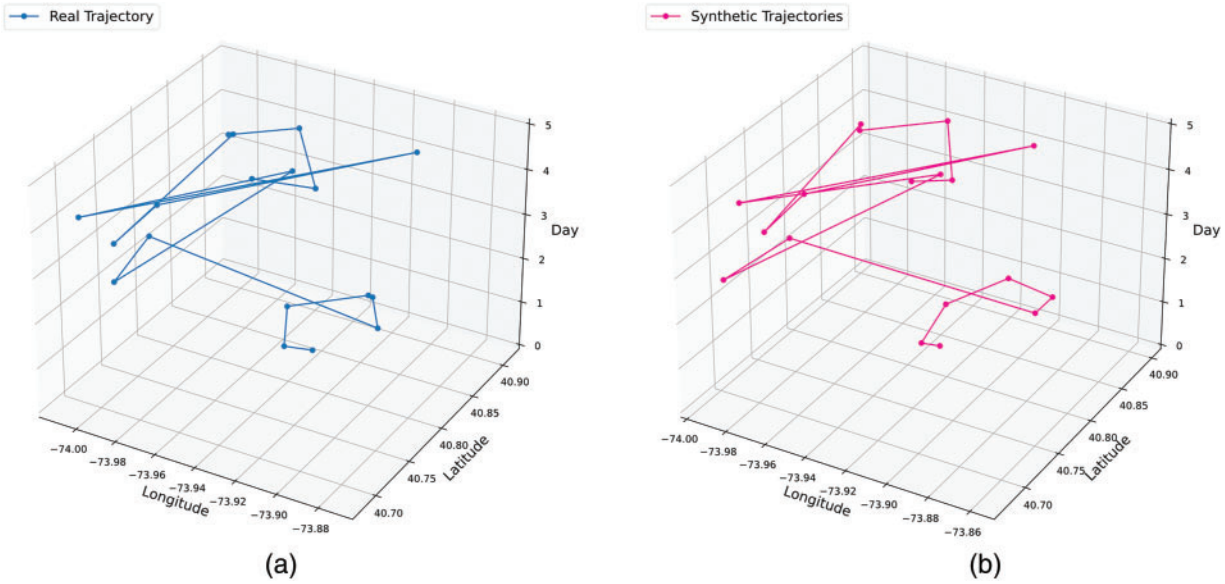

Figure 8: Visualization of real trajectories (a) and synthetic trajectories (b)

By using generated data for the upload, publication, and sharing of user trajectories, we can effectively confuse attackers while providing some level of privacy protection to prevent user re-identification. This provides fresh insights and solutions for geographic spatial artificial intelligence, particularly in addressing the analysis and processing of spatiotemporal data, spatial-temporal predictions, and related areas. Such advancements are beneficial for the application of location trajectory services within mobile crowd sensing networks.

It can be seen that this paper's method outperforms LSTM-TrajGAN (Fig. 9a) in all metrics. While LSTM-TrajGAN (Fig. 9a) exhibits excellent generative performance, the proposed BiLS-A-GAN (Fig. 8b) better captures human trajectory patterns, generating trajectories that more closely resemble real human trajectories. Although Gaussian geohiding achieves better privacy results on ACC@1, it is still higher than our model in other metrics. Combined with Fig. 9b, it can be seen that Gaussian geohiding perturbs the spatio-temporal characteristics of the trajectory data to a greater extent, which is also the reason for its lower ACC@1 accuracy rate. In the Differential Privacy (DP) method, we perturb the spatiotemporal features in trajectory data by adding Laplace noise. While this noise perturbation approach can provide better privacy protection, making it challenging for attackers to reidentify users' sensitive information, as shown in Fig. 10a,b, it is evident that the differential privacy method significantly disrupts the temporal and spatial attributes of trajectory data. This results in a notable reduction in the potential of generated data for further data analysis. While it achieves strong privacy protection, it sacrifices spatiotemporal similarity, which is not the desired outcome. Also, we can observe that in the case of utilizing both spatial and temporal dimensions, better privacy preservation can be achieved as compared to using only the temporal dimension. Gaussian Geomasking and LSTM-TrajGAN are more suitable for applications requiring lower levels of privacy protection, but maintaining spatio-temporal similarity is crucial for data analysis. This highlights a key limitation of both methods compared to the BiLS-A-GAN approach. On the other hand, the DP method is well-suited for high-level privacy protection applications, but this comes at the cost of data utility for specific analytical tasks. Overall, our proposed BiLS-A-GAN effectively addresses

the shortcomings of traditional methods by providing robust privacy protection while preserving the utility of the data for analysis.

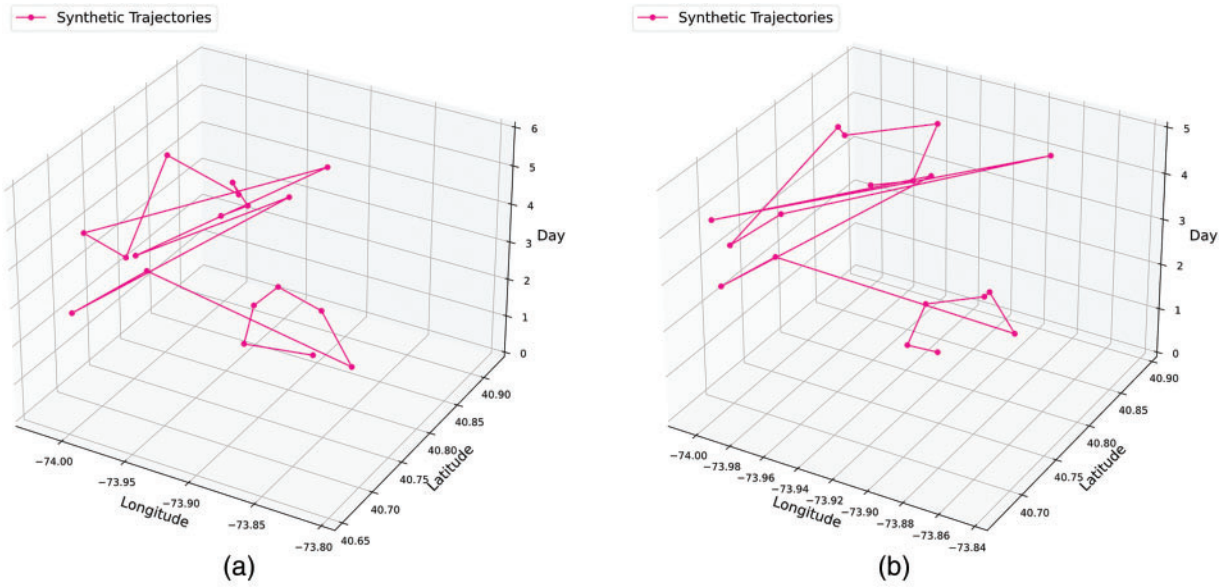

Figure 9: LSTM-TrajGAN generates trajectory maps (a), Gaussian Geomasking generates trajectory maps (b)

4.3.2 Analysis of the Results of the Data Similarity Experiment

This paper also measures the original and generated data based on the Hausdorff distance and the Jaccard coefficient. Hausdorff distance is a method of measuring the similarity between two point sets in space. It is often used to measure the spatial similarity between two trajectories. The Jaccard coefficient, also known as the Jaccard similarity coefficient or Jaccard index, is a mathematical metric used to measure the similarity between two sets and is used to compare the degree of overlap between two sets. A smaller Hausdorff distance and a larger Jaccard index indicate that the generated trajectory is more similar to the original trajectory.

In this work, we use the Hausdorff distance to measure the spatial similarity between the original trajectories and the generated trajectories. We use the Jaccard coefficient to assess the similarity of the POI categories between these two types of trajectories. The results are shown in Table 4. In terms of the Hausdorff distance metric, the BiLS-A-GAN model in this research exhibits a relatively low mean value (0.0258), while in terms of the Jaccard index, it shows the highest average (0.7649). This indicates that the trajectory data generated by the model in this research is very similar to the real data, both in terms of spatial trajectory attributes and POI (Point of Interest) categories.

In the DP method, the average Hausdorff distance reaches the highest value of 44.4751, while the average value of the Jaccard index is the lowest, only 0.3976, which shows that the Laplacian noise mechanism introduces large disturbances in spatiotemporal attributes and POI interest points. Compared with the LSTM-TrajGAN method, the BiLS-A-GAN model in this article has better performance in terms of Hausdorff distance and Jaccard index. On the Jaccard index, the two models have the same maximum and minimum values, although the variance of the model in this article is lower, but the mean value is higher than LSTM-TrajGAN, indicating that the trajectory

generated by BiLSTM-A-GAN is more similar to the real trajectory in terms of POI categories. The generated categories are basically within the original spatial domain without large deviations. Higher spatiotemporal analysis potential. In terms of Hausdorff distance, the BiLS-A-GAN model in this article has a lower mean value and shows higher similarity with the original trajectory. While Gaussian Geomasking may yield results close to those of the model proposed in this research (BiLS-A-GAN) in terms of Hausdorff distance and Jaccard index, it sacrifices more in terms of location privacy. As seen in Fig. 9b, the Gaussian Geomasking method exhibits stronger perturbation in both spatial and temporal dimensions, which provides better privacy protection but impairs the utility for further geographic analysis.

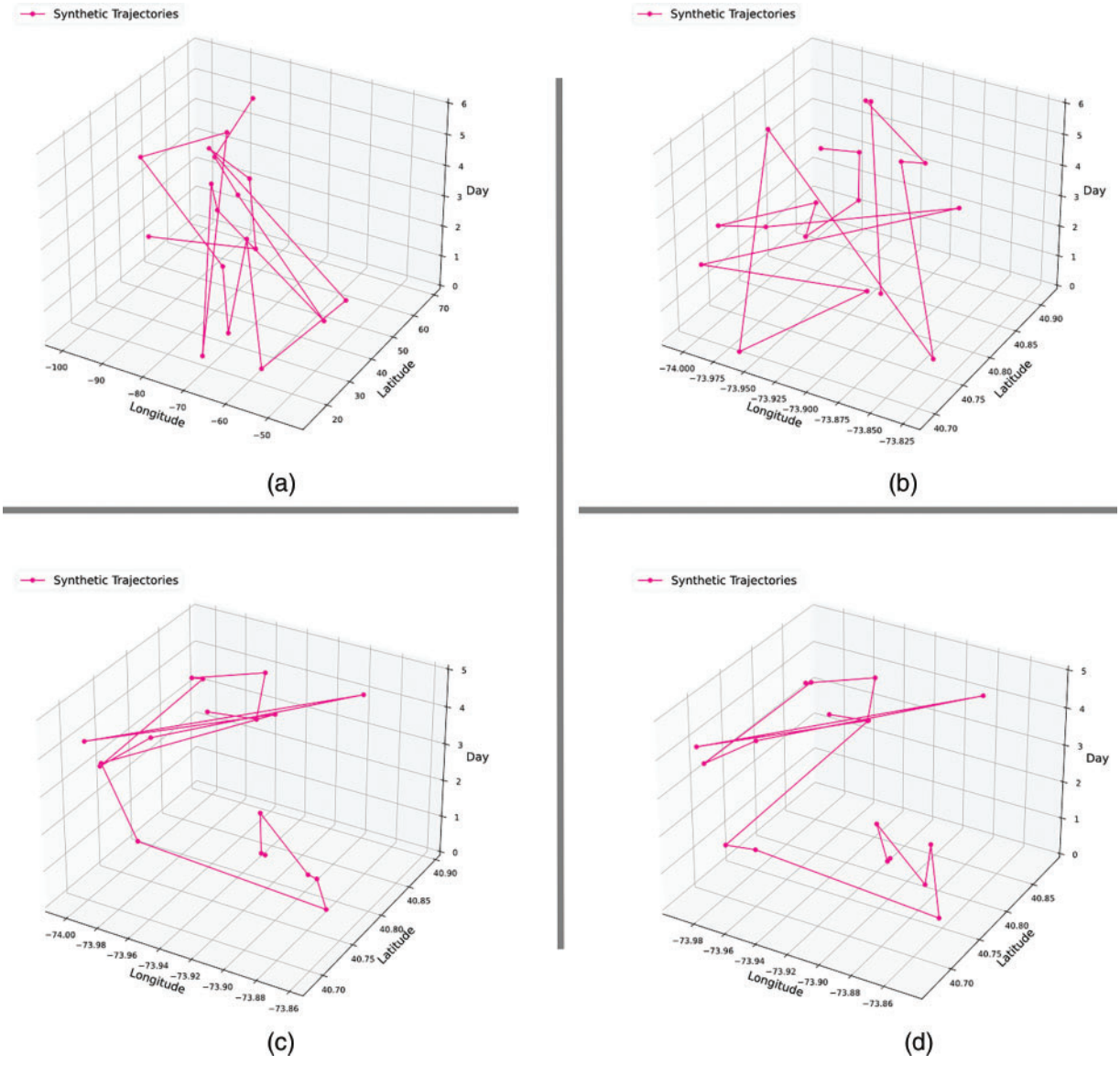

Figure 10: Trajectory generation results with different methods. Differential privacy applied to both spatial and temporal attributes (a), only applied to temporal attributes (b), only using GAttention in the generator (c), only using DAttention in the discriminator (d)

It should be noted that the maximum Jaccard index value of 1 does not imply that the interest point data of the two samples are identical. In this work, the Jaccard index measures the degree of overlap of POI category points, in a certain spatial domain, the temporal order of POI categories of the generated trajectories is perturbed to a certain extent, but still within this spatial domain, and the calculation of the Jaccard index does not involve the time series, resulting in the performance of the maximum value of one.

Table 4: Hausdorff distance and Jaccard index for different methods (DP stands for differential privacy)

Method	Hausdorff distance				Jaccard index			
	Max	min	std	mean	max	min	std	mean
LSTM-TrajGAN	0.0553	0.0075	4.0113	0.0279	1.0000	0.1666	0.0368	0.6720
Gaussian Geomasking	0.0468	0.0094	4.1265	0.0232	1.0000	0.2500	0.0265	0.7582
DP	113.4699	16.6006	192.3773	44.4751	0.8333	0.0000	0.0149	0.3976
BiLS-A-GAN	0.0588	0.0100	5.6625	0.0258	1.0000	0.1666	0.0366	0.7649

4.3.3 Ablation Study

To validate the stronger privacy protection capability of the model proposed in this study present, experiments were conducted using Bidirectional Long Short-Term Memory (BiLSTM) recurrent neural networks and attention mechanisms. These experiments were carried out to compare their impact on the experimental results of the model proposed in this work. It can be observed from Table 5 that compared with LSTM-TrajGAN, only changing the long short-term memory recurrent neural network cannot achieve good results and may even lead to worse results (BiLSTM's ACC@5 reaches 0.8042). However, the attention mechanism showed amazing results. The ACC@5 of DAttention and GAttention dropped to 0.1996 and 0.3729, respectively, and also achieved significant results in other indicators. However, the methods that solely incorporate attention mechanisms, when compared to the DP method, exhibit similar drawbacks. As shown in Fig. 10c,d, the impact of attention mechanisms on the spatiotemporal attributes is less pronounced than the DP method, but there are still noticeable differences in spatiotemporal attributes when compared to real trajectories in Fig. 8. While both of these methods excel in privacy protection, they largely sacrifice spatiotemporal similarity, thereby compromising the practical utility of the generated data.

Table 5: Privacy-preserving effect of BiLS-A-GAN and its variants on TUL tasks (DAttention indicates the addition of attention mechanism in the discriminator, GAttention indicates the addition of attention mechanism in the generator)

Method	ACC@1	ACC@5	Macro-F1	Macro-P	Macro-R
BiLSTM	0.5189	0.8042	0.4778	0.5427	0.5133
DAttention (Only Discriminator)	0.0769	0.1996	0.0538	0.0556	0.0793
GAttention (Only Generator)	0.1762	0.3729	0.1231	0.1673	0.1561
BiLS-A-GAN	**0.3115**	**0.5920**	**0.2428**	**0.2800**	**0.2829**

Overall, this paper's BiLS-A-GAN not only takes into account the privacy protection performance of generated trajectories but also strikes a good balance with spatial similarity, addressing both aspects effectively. Compared to the TUL metric of other methods, this model performs remarkably well. Although it is similar to LSTM-TrajGAN and Gaussian Geomasking in terms of Hausdorff distance and Jaccard index, this model more effectively prevents users from being re-identified. At the same time, the trajectory service can return the generated trajectory to the trajectory query, which can provide a higher level of privacy protection. It has obvious advantages in trajectory generation quality. The BiLS-A-GAN generation model not only provides high privacy protection for user trajectories, but also ensures the potential of the generated trajectory data for further data mining research, better balancing the effectiveness of privacy protection and the utility of spatio-temporal analysis.

4.4 Discussion

Synthetic trajectory data is generated by simulating the statistical properties of real data, effectively reducing the risk of exposing actual trajectory information and thereby protecting individual privacy. This study employs a dual attention mechanism (GAttention and DAttention) along with GANs to generate privacy-preserving trajectory data, aiming to balance the trade-off between privacy protection and data utility. With privacy ensured, synthetic trajectory data can be widely applied in analysis, research, and model training, particularly in scenarios involving sensitive trajectory information. Such data can be safely used for uploading, sharing, and publishing without privacy breach concerns. This enables businesses or institutions to provide valuable, de-identified synthetic trajectory data to researchers and the public, fostering innovation and scientific research. Since synthetic data do not directly involve personal information, they reduce potential ethical and legal issues during data sharing and usage, especially in regions with strict data protection laws, thus aiding compliance throughout the data handling and analysis process.

However, synthetic data may not fully retain the complexity and structural relationships of the original data, potentially missing dynamic features, such as temporal variations, especially in cases involving long-tail distributions or sparse location points. This could lead to decreased performance in certain applications, such as precise location recommendations or emergency service dispatching. Real-world trajectory data often includes complex behavior patterns and dynamic changes, like sudden events or holiday effects, which synthetic data may struggle to accurately replicate. This necessitates the design of more advanced trajectory generation models to better capture the dynamic nature of real-world trajectories and to produce more realistic data.

In addition, the quality of synthetic trajectory data is highly dependent on the design and training quality of the generated model. If the model is not adequately trained, the generated data may deviate from the actual distribution or introduce new biases. In this study, the latitude and longitude of all trajectories in the dataset were standardized, and the center of mass of the trajectory was used as the datum to calculate the deviation of each point relative to the center of mass. This approach helps the model capture spatial deviation patterns between different trajectory points more efficiently. This deviation-based approach can currently only synthesize trajectory data on urban-scale trajectory data and may not be extended to global-scale trajectory data. We will explore these limitations further in future studies.

Higher levels of privacy protection often indicate increased data perturbation or obfuscation. For example, differential privacy (DP) methods provide stronger protection for original trajectory data, but at the cost of reducing the utility of the generated data. This can lead to a decrease in practicality and usability. To protect privacy, this study aims for the synthetic data to differ from the

original data to a certain extent while retaining sufficient similarity to serve as an effective substitute for spatiotemporal modeling or analysis. By analyzing the BiLS-A-GAN approach, it is possible to choose appropriate model weights by balancing the relationship between TUL accuracy and the average Hausdorff distance. This ensures that synthetic trajectories maintain certain spatiotemporal and semantic characteristics while allowing for lower TUL accuracy when necessary, thereby balancing the effectiveness of privacy protection with the utility of the synthetic trajectories.

Although our method shows good performance in the comprehensive experiments described above, there are still some limitations of the current method. First, we analyzed the performance of generated trajectories on re-identification accuracy in the context of TUL tasks. However, a more specific evaluation of privacy leakage is yet to be explored. Second, compared to traditional trajectory obfuscation techniques, our deep learning framework demands a higher computational workload. Finally, our experiments were conducted exclusively on the Foursquare NYC check-in dataset, which may lack generalizability. We plan to address these limitations in future work by exploring broader datasets.

Trajectory data often contains sensitive information about an individual's daily routines, locations, and behaviors. Improper sharing of such data may lead to unintended privacy violations, where individuals can be tracked or profiled without their consent. In addition to the technical challenges, trajectory data privacy protection also raises significant ethical and legal concerns. The sharing and usage of such data must ensure that an individual's privacy rights are preserved in accordance with ethical standards and legal regulations such as the General Data Protection Regulation (GDPR). The high-quality trajectory data generated by the BiLS-A-GAN method effectively prevents the risk of re-identification, ensuring that the generated data cannot be linked to individuals, thereby reducing the risk of user privacy breaches. At the same time, it maintains high utility for trajectory data analysis and aligns with data privacy regulations, such as GDPR.

5 Conclusion and Future Work

In this work, we propose a new approach to fake trajectory generation, BiLS-A-GAN, to enhance trajectory privacy preservation. This method focuses on handling the temporal, spatial, and thematic features of trajectory data. It feeds the fused trajectory information to a neural network to generate high-quality fake trajectories, addressing privacy leakage issues for users when performing tasks such as uploading, publishing, and sharing trajectories. Existing trajectory privacy protection schemes tend to focus only on location privacy protection, ignoring the effectiveness of protecting the trajectory data for further geographic analysis, as can be seen from Fig. 8 and the experimental data in Tables 3–5, this paper's method provides a high level of privacy protection along with good spatio-temporal and thematic feature retention capabilities, making the generated trajectories useful for some spatio-temporal analysis. The introduction of the BiLS-A-GAN method offers new perspectives and insights in the field of trajectory privacy protection. It goes beyond simply focusing on location privacy and emphasizes the importance of preserving the effectiveness of trajectory data for further geographic analysis. This approach holds significant significance for both privacy and analysis of trajectory data.

In our future work, we plan to conduct trajectory generation analysis on a wider range of datasets, including Chinese trajectory datasets and datasets containing trajectories with irregular clusters, as well as global-scale trajectory data. This will allow us to assess the method's ability to strike a balance between privacy protection performance and practicality in different scenarios. Additionally, we will investigate enhancements to the trajectory similarity loss metric, explore additional privacy protection

strategies for trajectory generation, and work on refining the design and effectiveness of the deep learning framework for privacy protection in trajectory data generation across diverse environments.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the Key Projects of Gansu University of Political Science and Law (GZF2023XZD18), Gansu Province 2020 Provincial Virtual Simulation First-Class Course (GZYL2020-18), 2020 School Level Education Reform Project of Gansu University of Political Science and Law (GZJG2020-B06), the 2022 Provincial Industrial Support Project (2022CYZC-57), Soft Science Special Project of Gansu Basic Research Plan under Grant (22JR11RA106), Gansu Soft Science Project (20CX4ZA074), Scientific Research Project of Colleges and Universities of Gansu Provincial Department of Education (2015A-114), and Key Research Project of Gansu University of Political Science and Law under Grant (GZF2022XZD08), Provincial Program of Innovation and Entrepreneurship Training for College Students in Gansu Province (S202311406024).

Author Contributions: Xirui Yang: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing—Original Draft. Zhang Chen: Conceptualization, Methodology, Formal Analysis, Resources, Writing—Review & Editing, Supervision, Project Administration. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] G. McKenzie, D. Romm, H. Zhang, and M. Brunila, "PrivyTo: A privacy-preserving location-sharing platform," *Trans. GIS*, vol. 26, no. 4, pp. 1703–1717, 2022. doi: 10.1111/tgis.12924.

[2] A. Boutet and M. Cunche, "Privacy protection for Wi-Fi location positioning systems," *J. Inf. Secur. Appl.*, vol. 58, 2021, Art. no. 102635. doi: 10.1016/j.jisa.2020.102635.

[3] Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz and M. Guizani, "Deep-distributed-learning-based POI recommendation under mobile-edge networks," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 303–317, 2023. doi: 10.1109/JIOT.2022.3202628.

[4] H. Wang, Y. Li, J. Lin, H. Cao, and D. Jin, "Context-aware semantic annotation of mobility records," *ACM Trans. Knowl. Discov. Data.*, vol. 16, no. 3, pp. 1–20, 2021. doi: 10.1145/3477048.

[5] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in *Proc. Sixth Int. AAAI Conf. Weblogs Soc. Med. (ICWSM)*, Dublin, Ireland, 2021, vol. 6, no. 1, pp. 58–65. doi: 10.1609/icwsm.v6i1.14278.

[6] F. Terroso-Sáenz, A. Muñoz, J. Fernández-Pedauye, and J. M. Cecilia, "Human mobility prediction with region-based flows and water consumption," *IEEE Access*, vol. 9, pp. 88651–88663, 2021. doi: 10.1109/ACCESS.2021.3090582.

[7] A. Fathalizadeh, V. Moghtadaiee, and M. Alishahi, "On the privacy protection of indoor location dataset using anonymization," *Comput. Secur.*, vol. 117, no. 15, 2022, Art. no. 102665. doi: 10.1016/j.cose.2022.102665.

[8] V. -D. Stanciu, M. van Steen, C. Dobre, and A. Peter, "Privacy-friendly statistical counting for pedestrian dynamics," *Comput. Commun.*, vol. 211, no. 12, pp. 178–192, 2023. doi: 10.1016/j.comcom.2023.09.009.

[9] Z. Wang, X. Ye, and M. -H. Tsou, "Spatial, temporal, and content analysis of Twitter for wildfire hazards," *Nat. Hazards*, vol. 83, no. 1, pp. 523–540, 2016. doi: 10.1007/s11069-016-2329-6.
[10] Q. Huang and Y. Xiao, "Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery," *ISPRS Int. J. Geo Inf.*, vol. 4, no. 3, pp. 1549–1568, 2015. doi: 10.3390/ijgi4031549.
[11] T. Ma, Q. Deng, H. Rong, and N. Al-Nabhan, "A privacy-preserving trajectory data synthesis framework based on differential privacy," *J. Inf. Secur. Appl.*, vol. 77, 2023, Art. no. 103550. doi: 10.1016/j.jisa.2023.103550.
[12] F. S. Alrayes, A. I. Abdelmoty, W. B. El-Geresy, and G. Theodorakopoulos, "Modelling perceived risks to personal privacy from location disclosure on online social networks," *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 1, pp. 150–176, 2020. doi: 10.1080/13658816.2019.1654109.
[13] J. E. Dobson and W. A. Herbert, "Geoprivacy, convenience, and the pursuit of anonymity in digital cities," in *Urban Informatics*, Singapore: Springer Singapore, 2021, pp. 567–587. doi: 10.1007/978-981-15-8983-6_32.
[14] S. Gao, J. Rao, X. Liu, Y. Kang, Q. Huang and J. App, "Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users," *J. Spatial Inf. Sci.*, vol. 2019, no. 19, pp. 105–129, 2019. doi: 10.5311/JOSIS.2019.19.510.
[15] J. Rao, S. Gao, and S. Zhu, "CATS: Conditional adversarial trajectory synthesis for privacy-preserving trajectory data publication using deep learning approaches," *Int. J. Geogr. Inf. Sci.*, vol. 37, no. 12, pp. 2538–2754, 2023. doi: 10.1080/13658816.2023.2262550.
[16] V. Shubina, A. Ometov, S. Andreev, D. Niculescu, and E. S. Lohan, "Privacy versus location accuracy in opportunistic wearable networks," in *2020 Int. Conf. Localization GNSS (ICL-GNSS)*, Tampere, Finland, 2020, pp. 1–6. doi: 10.1109/ICL-GNSS49876.2020.9115424.
[17] D. Tao, T. -Y. Wu, S. Zhu, and M. Guizani, "Privacy protection-based incentive mechanism for mobile crowdsensing," *Comput. Commun.*, vol. 156, no. 15, pp. 201–210, 2020. doi: 10.1016/j.comcom.2020.03.027.
[18] J. Wang, J. Kim, and M. -P. Kwan, "An exploratory assessment of the effectiveness of geomasking methods on privacy protection and analytical accuracy for individual-level geospatial data," *Cartogr. Geogr. Inf. Sci.*, vol. 49, no. 5, pp. 385–406, 2022. doi: 10.1080/15230406.2022.2056510.
[19] J. Liu, Z. Xu, X. Xu, and Z. Zou, "Research on user privacy protection algorithm in location service," in *Proc. 12th Int. Conf. Meas. Technol. Mechatron. Autom. (ICMTMA)*, Phuket, Thailand, 2020, pp. 953–956. doi: 10.1109/ICMTMA50254.2020.00206.
[20] X. Chen, X. Wu, X. Wang, W. Zhao, and W. Xue, "Real-location reporting based differential privacy trajectory protection for mobile crowdsensing," in *Proc. 5th Int. Confe. Big Data Comput. Commun. (BIGCOM)*, Qingdao, China, 2019, pp. 142–150. doi: 10.1109/BIGCOM.2019.00029.
[21] H. Zhu, X. Yang, B. Wang, L. Wang, and W. -C. Lee, "Private trajectory data publication for trajectory classification," in *Web Information Systems and Applications*, Cham: Springer, 2019, vol. 11817, pp. 347–360. doi. 10.1007/978-3-030-30952-7_35.
[22] B. Bostanipour and G. Theodorakopoulos, "Joint obfuscation of location and its semantic information for privacy protection," *Comput. Secur.*, vol. 107, 2021, Art. no. 102310. doi: 10.1016/j.cose.2021.102310.
[23] L. May Petry, C. Leite Da Silva, A. Esuli, C. Renso, and V. Bogorny, "MARC: A robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings," *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 7, pp. 1428–1450, 2020. doi: 10.1080/13658816.2019.1707835.
[24] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, MIT Press, 2014, vol. 2, pp. 2672–2680.
[25] Y. Bengio *et al.*, "The need for privacy with public digital contact tracing during the COVID-19 pandemic," *Lancet Digit. Health*, vol. 2, no. 7, pp. e342–e344, 2020. doi: 10.1016/S2589-7500(20)30133-3.
[26] M. A. Azad *et al.*, "A first look at privacy analysis of COVID-19 contact-tracing mobile applications," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15796–15806, 2021. doi: 10.1109/JIOT.2020.3024180.
[27] C. Troncoso *et al.*, "Decentralized privacy-preserving proximity tracing," 2020. doi: 10.48550/arXiv.2005.12273.

[28] L. Yu *et al.*, "Privacy-preserving location-based advertising via longitudinal geo-indistinguishability," *IEEE Trans. Mob. Comput.*, vol. 23, no. 8, pp. 8256–8273, 2024. doi: 10.1109/TMC.2023.3348136.

[29] P. Cheng, X. Lian, L. Chen, and S. Liu, "Maximizing the utility in location-based mobile advertising," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 776–788, 2022. doi: 10.1109/TKDE.2020.2986198.

[30] H. Shin, S. Kim, J. Shin, and X. Xiao, "Privacy enhanced matrix factorization for recommendation with local differential privacy," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 9, pp. 1770–1782, 2018. doi: 10.1109/TKDE.2018.2805356.

[31] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 45, no. 1, pp. 129–142, 2015. doi: 10.1109/TSMC.2014.2327053.

[32] J. Rao, S. Gao, Y. Kang, and Q. Huang, "LSTM-TrajGAN: A deep learning approach to trajectory privacy protection," in *11th Int. Conf. Geogr. Inf. Sci. (GIScience 2021)*, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, vol. 177, pp. 12:1–12:17. doi: 10.4230/LIPIcs.GIScience.2021.I.12.

[33] M. Selvarathnam, R. Ragel, C. C. Reyes-Aldasoro, and M. Rajarajan, "Privacy vs utility analysis when applying differential privacy on machine learning classifiers," in *19th Int. Conf. Wirel. Mob. Comput. Netw. Commun. (WiMob)*, Montreal, QC, Canada, 2023, pp. 306–311. doi: 10.1109/WiMob58348.2023.10187829.