基于CNN-BiLSTM-Attention融合模型的差分隐私轨迹重构攻击

谢丽霞[1]，赵尔康[1]，杨宏宇[1,2]，刘哲理[3]，赵永新[4]
（1.中国民航大学计算机科学与技术学院，天津 300300；2.中国民航大学安全工程学院，天津 300300；
3.南开大学网络空间安全学院，天津 300350；4.天津理工大学计算机科学与工程学院，天津 300384）

摘　要：针对现有差分隐私轨迹保护机制的重构攻击方法在局部特征提取、空间信息提取以及全局依赖建模方面的不足所导致攻击性能不佳的问题，提出了一种基于CNN-BiLSTM-Attention融合模型的轨迹重构攻击方法。该方法引入卷积神经网络（CNN）捕捉轨迹数据中的空间依赖性和局部模式。通过双向长短期记忆网络（BiLSTM）建模轨迹序列中的长期时序依赖，增强轨迹序列在时间维度上的表达能力。通过注意力机制为轨迹中的每个时间步自适应分配不同的权重，捕捉轨迹中的全局信息和长时间跨度的依赖关系。实验结果表明，相较于基线方法，所提方法的欧几里得距离减少百分比平均提升5.03%，豪斯多夫距离减少百分比平均提升5.02%，轨迹凸包的杰卡德指数平均提升了2.4倍，可有效实施轨迹重构攻击。

关键词：差分隐私；轨迹重构攻击；卷积神经网络；双向长短期记忆网络；注意力机制
中图分类号：TP309.2
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025252

Trajectory reconstruction attacks on differential privacy based on a CNN-BiLSTM-Attention hybrid model

XIE Lixia[1], ZHAO Erkang[1], YANG Hongyu[1,2], LIU Zheli[3], ZHAO Yongxin[4]
1. School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China
2. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China
3. College of Cyber Science, Nankai University, Tianjin 300350, China
4. School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

Abstract: To address the poor attack performance of existing reconstruction attack methods against differential privacy trajectory protection mechanisms caused by deficiencies in local feature extraction, spatial information extraction, and global dependency modeling, a trajectory reconstruction attack method based on a CNN-BiLSTM-Attention fusion model was proposed. A convolutional neural network (CNN) was introduced to capture spatial dependencies and local patterns in trajectory data. Long term temporal dependencies in trajectory sequences were modeled using bidirectional long short-term memory (BiLSTM) network, thereby strengthening representation capability along the temporal dimension. An attention mechanism was employed to adaptively assign different weights to each time step, capturing global information and long span dependencies within trajectories. Experimental results show that, compared with baseline methods, the average percentage of Euclidean distance reduction of the proposed method is increased by 5.03%, the average improvement in the percentage reduction of Hausdorff distance is 5.02%, and the Jaccard index of the trajectory convex hull increases by an average factor of 2.4, enabling effective trajectory reconstruction attacks.

Keywords: differential privacy, trajectory reconstruction attack, convolutional neural network, bidirectional long short-term memory network, attention mechanism

收稿日期：2025–08–19；修回日期：2025–12–06
通信作者：杨宏宇，yang_hy2006@126.com
基金项目：国家自然科学基金民航联合研究基金重点项目（No.U2433205）
Foundation Item：Civil Aviation Joint Research Fund Project of the National Natural Science Foundation of China (No.U2433205)

0　引言

随着移动互联网与定位技术的广泛应用，基于位置服务[1]（LBS, location based service）所产生的时空轨迹数据呈指数级增长。这类数据在智慧城市、交通规划和商业选址等领域具有重要价值。然而，轨迹数据中蕴含的细粒度时空信息，如通勤模式与生活习惯，使其面临较高的隐私泄露风险。多起与位置信息相关的隐私泄露事件和早期研究成果，凸显了隐私保护机制在实际应用中面临的严峻挑战。例如，在2014年发生的纽约出租车行程数据泄露事件中，尽管数据已进行匿名化处理，但研究人员通过结合外部公开信息，仍成功识别出乘客和司机的个人隐私信息，这表明即使是匿名轨迹数据，其固有的时空模式仍蕴含强大的识别能力。因此，在轨迹数据公开发布前，必须采取有效的隐私保护措施，保障用户隐私的同时最大限度地保留数据可用性。

为降低轨迹数据在共享或发布过程中的隐私泄露风险，研究者提出了多种隐私保护方法。早期方法主要基于k-匿名[2]及其扩展模型，如l-多样性[3]、t-贴近性[3]等，通过对轨迹数据进行泛化处理，降低个体身份被识别的风险。然而，此类方法在实际应用中面临一定的局限性，其安全性高度依赖于攻击者所掌握的辅助信息，且缺乏严格的数学证明和理论支撑。差分隐私[4]（DP, differential privacy）作为一种具有严格数学定义和理论基础的隐私保护机制，其核心思想是在原始数据或查询结果中引入随机噪声，使任意一个个体数据的加入或删除对最终输出结果的影响在统计上可忽略不计，从而有效降低敏感信息的泄露风险。由于其在隐私保护强度上的可控性与理论完备性，DP机制已被广泛应用于轨迹数据发布场景中[5-13]。尽管现有的轨迹隐私保护机制在形式上严格满足DP定义，但由于其未能充分考虑实际地理约束条件，应用该机制后的轨迹数据常常呈现出地理或语义上的不合理性。例如，经DP机制处理后的船舶轨迹可能穿越陆地、行人轨迹可能穿越建筑物、车辆轨迹可能穿越河流等。已有研究表明，攻击者通过利用隐私保护机制引入的地理约束破坏现象，构建训练数据以引导模型捕捉扰动轨迹与真实轨迹之间的结构性差异，从而实现对原始轨迹形态的有效恢复。该类攻击被称为轨迹重构攻击，旨在揭示现有隐私保护方法在实际应用中存在的不足。

目前，在针对传统匿名化方法的攻击中，现有工作通常采用无监督学习或启发式策略恢复个体轨迹。近年来，随着深度学习技术的发展，其强大的特征学习能力也被引入轨迹重构攻击领域，并在各类场景下展现出强大的攻击潜力。然而，无论是针对传统匿名化方法还是DP机制，现有轨迹重构攻击方法仍存在以下不足：1) 在局部特征提取方面，未能有效捕捉轨迹中短时行为特征（如加速、减速、停留等），导致重构结果缺乏精细度；2) 在空间特征建模方面，轨迹数据作为在特定空间环境中展开的几何路径，现有方法缺乏有效机制建模和提取其空间特征，这导致重构轨迹在空间位置上的准确性存在较大误差；3) 在全局上下文信息的整合与利用方面，现有方法未能有效捕捉长时间跨度或大范围空间内的依赖关系，进而导致重构轨迹在整体空间形态上偏离真实轨迹。

针对上述不足，本文提出了一种基于CNN-BiLSTM-Attention融合模型的轨迹重构攻击方法，旨在从DP保护的轨迹数据中更有效地恢复原始轨迹。该方法融合了卷积神经网络（CNN, convolutional neural network）、双向长短期记忆网络（BiLSTM, bidirectional long short-term memory network）和注意力机制（Attention, attention mechanism）的优势，通过深度学习技术的协同工作，可有效提取局部空间特征并对长期时序依赖建模、聚焦全局关键信息，增强模型在复杂扰动下对原始轨迹形态的理解与还原能力，确保模型尽可能恢复原始轨迹的整体形态。

1　相关工作

随着移动感知与数据发布技术的发展，面向轨迹数据的隐私保护方法不断演进。相应地，针对发布后数据的重构攻击方法也不断涌现。轨迹重构攻击可根据所针对的隐私保护场景不同分为3类：针对传统匿名化方法的攻击、针对DP保护机制的攻击和针对其他隐私保护场景的攻击。

在传统匿名化隐私保护场景下，Montjoye等[14]提出了一种定量评估方法，揭示了轨迹数据的隐私边界。研究发现，即使是经过匿名化处理的轨迹数据，仍能通过少量的时空数据点实现高精度

的身份识别。但该方法依赖于外部信息，且在复杂环境下，隐私保护效果可能存在较大的差异。Gambs等[15]研究地理位置数据的去匿名化攻击，提出了一种基于移动马尔可夫链（MMC, mobility Markov chain）的攻击方法，通过从移动轨迹中提取兴趣点（POI, point of interest）构建MMC模型，并设计多种距离度量来量化MMC之间的相似性。但该方法对聚类算法参数敏感，且在严格评估下，其去匿名化成功率仍有提升空间。Xu等[16]聚焦于聚合统计的轨迹匿名化方法提出了一种攻击方法，利用人类移动的规律性和独特性，在不需要任何先验知识的情况下，从聚合数据中恢复个体轨迹。但该方法缺乏实现细节，使攻击机制不够清晰。此外，该方法仅在2个非公开商业数据集上进行评估，导致其结果和结论无法被验证。为解决上述问题，D'Silva等[17]重新实现该方法，并在2个开源数据集上进行评估，详细说明预处理步骤和实现过程，并且设计一系列增强措施。但该方法结果仍受数据分布差异与较高预处理开销的影响，跨数据分布的泛化性仍需进一步验证。

在DP保护机制场景下，Shao等[18]提出了iTracker框架，通过利用多条轨迹之间的相关性和位置稀疏矩阵构建，并使用2个近似算法迭代收敛至最可能的原始轨迹，实现从DP保护的轨迹数据中恢复原始轨迹。但该方法主要依赖时间和位置的稀疏结构假设，多数实验设定以几何扰动为例，语义、道路约束等外部先验信息未被显式纳入，在复杂地形或语义约束场景下可能限制其扩展性。Buchholz等[19]提出了RAoPT（reconstruction attack on protected trajectories）模型，构建基于BiLSTM的端到端重构器，将DP机制保护后的轨迹作为输入，学习并利用受保护轨迹与真实轨迹在结构上的可区分性，在公开数据集上显著缩短与原始轨迹的距离，且具备一定的数据集间迁移能力。但该方法侧重时间依赖，缺少局部空间的建模，导致细粒度恢复与空间一致性受限，进而限制其在复杂地形或道路稀疏环境下的泛化表现。

随着隐私保护技术的不断发展，其他新兴的隐私保护场景也逐渐受到关注。例如，在联邦学习隐私保护场景下，Wang等[20]提出了mGAN-AI框架，通过一个新颖的多任务判别器，在区分样本类别和真实性的同时，还能辨别客户端身份。然而，该攻击方法虽声称隐蔽，但在主动攻击模式下仍可能对共享模型性能造成影响。Hitaj等[21]提出了一种基于生成对抗网络（GAN, generative adversarial network）的攻击方法，专门针对联邦学习中的隐私保护机制。通过引入GAN，攻击者能生成与目标训练数据集相似的样本，从而泄露训练集的敏感信息。但该方法要求攻击者能够对共享模型参数有白盒访问权限，这在某些实际部署中可能难以满足。Ariyarathna等[22]提出了一种基于共享梯度的轨迹重构攻击方法，通过在攻击端训练一个梯度引导生成器，使其生成的假轨迹在目标全局模型上产生的嵌入梯度尽可能接近真实梯度，并通过欧几里得距离损失约束道路连通性和空间语义合理性。但该方法依赖于服务端能观测客户端梯度并利用路网等先验知识，当共享梯度被显著加噪或访问受限时，攻击效果会下降。在轨迹嵌入反演场景下，Han等[23]提出了一种对抗性轨迹重构攻击方法，揭示出轨迹嵌入可能带来的隐私风险。然而，该方法假设攻击者可获取目标嵌入并具备路网信息，其适用性受查询限制与先验质量的约束。

综上，现有轨迹重构攻击研究已在多种隐私保护场景下取得进展，覆盖传统匿名化、DP及联邦学习等新兴范式，并在特定设定下表现出良好的攻击性能。然而，这些方法仍存在一些局限性：1) 多数方法过度依赖特定数据分布与结构假设，导致跨场景下的泛化表现欠佳；2) 在面对噪声增强、查询受限以及无背景知识等更为严苛的攻击条件时，现有方法的鲁棒性与稳定性亟待提升。为解决上述不足，本文提出了一种基于CNN-BiLSTM-Attention的轨迹重构攻击模型。该模型利用CNN捕捉轨迹空间依赖性和局部模式的优势，结合BiLSTM建模长时序依赖的能力，并通过Attention对关键时空信息进行自适应加权，从而有效应对复杂数据分布，并在跨场景条件下表现出良好的泛化性能。

2 重构攻击模型

2.1 重构攻击方法形式化定义

为系统描述本文提出的轨迹重构攻击方法，本节对轨迹数据、DP保护机制和重构模型进行形式化定义。设原始轨迹数据集为$T=\{T_i\}_{i=1}^{M}$，其中M为实际轨迹数量，第i条原始轨迹表示为按时间排

序的位置点序列 $T_i=(p_{i,1},p_{i,2},\cdots,p_{i,n_i})$，$n_i$为轨迹 T_i 的有效位置点数量；每个位置点 $p_{i,j}$ 定义为 $p_{i,j}=\left(l_{\mathrm{lat},j}^{(i)},l_{\mathrm{lon},j}^{(i)}\right)$；$l_{\mathrm{lat},j}^{(i)}$和 $l_{\mathrm{lon},j}^{(i)}$分别为第 i 条轨迹在第 j 个时间步的纬度和经度。给定DP保护机制 $\mathcal{H}$ 及隐私预算 ε，原始轨迹 T_i 被随机映射为发布轨迹 $\tilde{T}_i=\mathcal{H}(T_i;\varepsilon)=\left(\tilde{p}_{i,1},\tilde{p}_{i,2},\cdots,\tilde{p}_{i,n_i}\right)$，其中 $\tilde{p}_{i,j}=\left(\tilde{l}_{\mathrm{lat},j}^{(i)},\tilde{l}_{\mathrm{lon},j}^{(i)}\right)$ 为DP保护机制处理后的发布位置点。在攻击者视角下，$\tilde{T}_i$为可观测轨迹序列。在训练阶段，假设攻击者能获得一部分发布轨迹及其对应的原始轨迹，从而构造训练样本对 $\mathcal{D}=\left\{(\tilde{T}_i,T_i)\right\}_{i=1}^{M}$。本文将轨迹重构攻击建模为一个序列到序列的回归问题：在已知发布轨迹的条件下，通过优化模型参数 θ，训练重构模型 $F_\theta:\tilde{T}_i\rightarrow\hat{T}_i$，使模型从保护后的轨迹 $\tilde{T}_i$ 中重构出的轨迹 $\hat{T}_i$ 更接近原始轨迹 T_i。

2.2　攻击流程

轨迹重构攻击流程如图1所示，可划分为3个阶段：训练数据构建、攻击模型训练和轨迹重构攻击。为训练攻击模型，攻击者需获取扰动前后的轨迹数据集。然而，在现实攻击场景中，攻击者通常无法直接获取目标系统所发布的敏感轨迹的原始版本。因此，攻击者基于公开可用的非敏感轨迹数据，并根据其背景知识选择相应的隐私保护机制与隐私预算参数进行模拟并实施，从而生成对应的扰动轨迹。通过此过程，攻击者构建出大量的原始轨迹与扰动轨迹的配对样本，用于攻击模型的训练。具体而言，将第一阶段生成的扰动轨迹作为模型输入，对应的原始轨迹作为标签用于训练攻击模型。该模型旨在学习识别并纠正扰动轨迹中不符合实际地理约束的结构性缺陷，从而构建从扰动轨迹到更具真实语义的原始轨迹形态之间的映射关系，并最小化重构轨迹与原始轨迹之间的差异。攻击模型训练完成后，进入实际应用阶段，将目标系统实际发布和已通过DP保护机制处理的轨迹数据作为模型输入，模型输出即为重构轨迹。

在该攻击过程中，攻击效果取决于保护轨迹与原始轨迹之间的结构差异。DP保护机制通过引入噪声，使保护轨迹与原始轨迹在形态和特征上表现出差异，这些不合理的结构特征为攻击者提供了轨迹重构的关键依据。攻击者通过识别和利用这些结构性差异，能够在一定程度上恢复原始轨迹，从而削弱隐私保护机制的有效性。具体而言，当隐私预算较小时，为满足隐私约束需要注入较大幅度的噪声，导致保护轨迹与原始轨迹之间的差异更为显著，从而使攻击模型能够更精确地重构轨迹。相应地，当隐私预算较大时，引入噪声的幅度降低，两类轨迹之间的差异减弱，攻击难度随之提高。

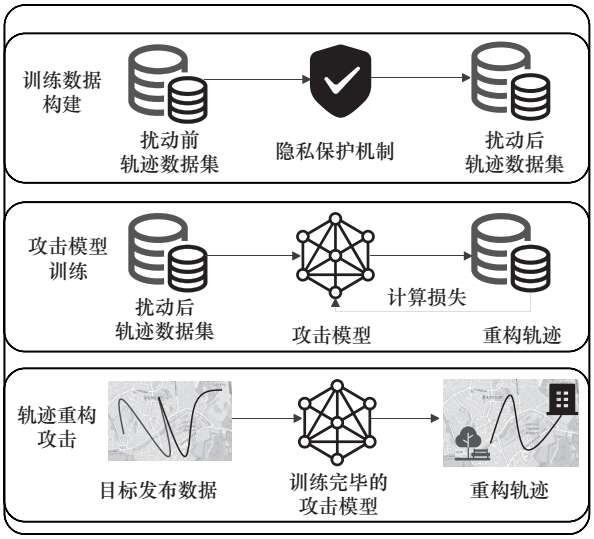


图1　轨迹重构攻击流程

2.3　模型组成

为缓解相关工作中指出的基线模型在跨场景泛化能力不足以及在严苛攻击条件下鲁棒性较弱等问题，本文设计了一种基于CNN-BiLSTM-Attention的轨迹重构模型。模型整体结构如图2所示，由输入层、CNN层、BiLSTM层、Attention层和输出层组成。该模型通过CNN、BiLSTM与Attention的组合，对发布轨迹中的关键时空结构特征进行建模，从而在不同数据集以及背景知识较弱等不利条件下，提升跨场景适应性和鲁棒性。

输入层将轨迹数据转换为深度学习模型可处理的向量形式，依次包括轨迹长度填充、one-hot编码和嵌入向量学习，最终得到嵌入表示。

在输入层基础上，为充分利用轨迹数据在时空上的局部连续性和全局周期性，CNN在短时间或短距离窗口内提取相邻轨迹点的相对位置、运动方向和局部几何形状，将原始点序列编码为片段级局部运动模式表示，为BiLSTM提供更加稳定且结构化的时序建模基础。将这些按时间排序的片段级局部特征向量作为BiLSTM的输入，在

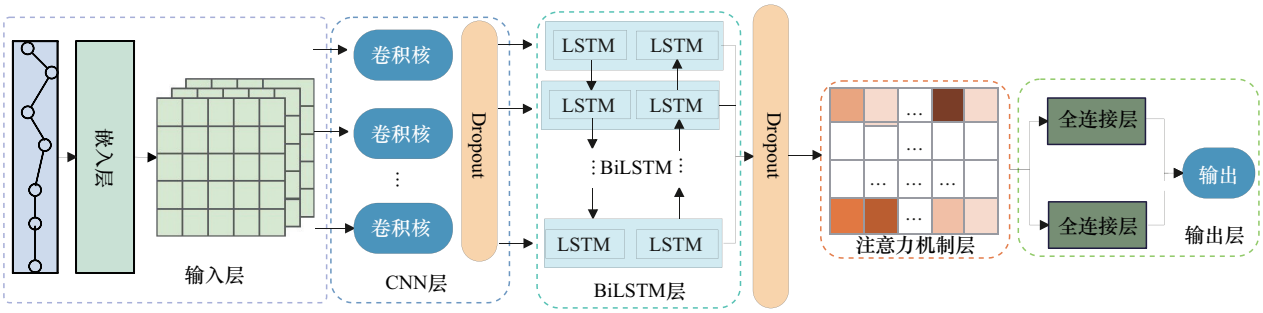


图2　基于CNN-BiLSTM-Attention的轨迹重构模型

前向和后向2个时间方向上依次更新隐藏状态，使每个时间步的表示不仅包含当前片段的信息，还显式整合其前后邻近片段的上下文信息，从而在CNN提取的局部模式上建模轨迹位置之间的长程依赖，形成对轨迹整体演化规律的高层表征。然而，对于长度较大的轨迹序列，来自早期时间步的信息在递归传播过程中容易逐渐衰减，因此在BiLSTM的隐藏状态序列上引入Attention，对各时间步与当前重构目标的相关性进行加权汇聚，建立从输出到任意时间位置的直接依赖，使模型显式聚焦于起点、终点、转折点等关键位置，减弱长链式传递导致的信息衰减，并增强对全局周期性和远距离依赖的表达能力。

输出层接收Attention层加权后的特征，采用2个独立的全连接层分别重构每个时间步的经度和纬度，并将二者拼接为地理坐标，从而生成完整的重构轨迹。上述模块的协同作用使模型能够同时获取局部空间特征、全局时序依赖与关键信息，从而识别并纠正扰动轨迹中的不合理模式，实现更精确的原始轨迹重构。

2.3.1　输入层

输入层的输入为隐私保护机制处理后的轨迹数据。首先，为使模型能够对轨迹数据进行批处理，将轨迹数据填充为统一长度N。其次，对轨迹数据进行编码，以确保数据格式适用于深度学习模型的输入。对每条轨迹的位置点，分别从其时间戳属性中提取小时和星期作为时间特征，并将经纬度坐标作为空间特征。对于时间特征，采用one-hot编码将其转换为二进制向量。对于空间特征，由于经纬度本身具有数值连续性，因此不需要额外编码。完成上述处理后，每个位置点的记录被转换为D维的特征向量，它由编码后的各类特征拼接而成。最后，各类子特征向量均通过独立的嵌入模块进行处理，得到其对应的嵌入向量。之后，将所有变换后的向量沿特征维度进行拼接，以聚合多类型特征信息，拼接后的特征向量被输入全连接层，以实现多类型特征融合。因此，对于第i条轨迹，在每个有效时间步j都得到一个融合后的特征向量。各时间步上的特征向量$\boldsymbol{F}^{(i)}_{\text{fusion},j}$共同构成该条轨迹的融合特征序列，具体过程可表示为

$$\boldsymbol{H}^{(i)}_{\text{fusion}} = [\boldsymbol{F}^{(i)}_{\text{fusion},1}, \boldsymbol{F}^{(i)}_{\text{fusion},2}, \cdots, \boldsymbol{F}^{(i)}_{\text{fusion},N}] \tag{1}$$

该融合特征序列$\boldsymbol{H}^{(i)}_{\text{fusion}}$将作为后续CNN层的输入，以实现更深层次的特征提取。由此可见，对于第i条轨迹，输入层最终得到的特征序列为

$$\boldsymbol{H}^{(0,i)}_{\text{fusion}} = (x_{i,1}, x_{i,2}, \cdots, x_{i,n_i}), \quad x_{i,j} \in R^{D} \tag{2}$$

其中，每个向量$x_{i,j}$由经纬度坐标、时间特征以及其他上下文信息经过嵌入和非线性变换拼接而成，该特征序列$\boldsymbol{H}^{(0,i)}_{\text{fusion}}$随后作为CNN层的输入，为后续轨迹时空特征建模提供统一的向量化表示。

2.3.2　CNN层

CNN层用于从输入的融合特征序列中提取轨迹数据的局部时序特征及其空间变化信息。为增强对隐私保护噪声导致局部扰动的识别能力，本文采用3层堆叠的CNN，并设置不同尺寸的卷积核以捕获不同时间尺度的局部模式与时间相关性。卷积核尺寸逐步增大可进一步建模更长范围的局部变化（包括噪声引入的非自然波动）。同时，通过对时间序列中的空间坐标变化进行卷积建模，CNN能感知局部空间结构相关的动态模式与噪声导致的轨迹偏移，并将提取到的空间相关特征传递给后续BiLSTM层。具体地，卷积核作为滑动窗口在序列上移动，对滑动窗口内数据进行卷积并经ReLU激活函数激活，得到各位置的局部特征表示。具体过程可表示为

$$\boldsymbol{H}^{(0,i)}_{\text{CNN}} = \text{ReLU}(\text{Conv}_0(\boldsymbol{H}^{(i)}_{\text{fusion}})) \tag{3}$$

$$\boldsymbol{H}_{\mathrm{CNN}}^{(1,i)} = \mathrm{ReLU}\left(\mathrm{Conv}_1\left(\boldsymbol{H}_{\mathrm{CNN}}^{(0,i)}\right)\right) \tag{4}$$

$$\boldsymbol{H}_{\mathrm{CNN}}^{(2,i)} = \mathrm{ReLU}\left(\mathrm{Conv}_2\left(\boldsymbol{H}_{\mathrm{CNN}}^{(1,i)}\right)\right) \tag{5}$$

其中，Conv_0、Conv_1和Conv_2分别表示不同尺寸卷积核的一维卷积操作，$\boldsymbol{H}_{\mathrm{CNN}}^{(0,i)}$、$\boldsymbol{H}_{\mathrm{CNN}}^{(1,i)}$和$\boldsymbol{H}_{\mathrm{CNN}}^{(2,i)}$分别表示第一层、第二层和第三层卷积运算并应用ReLU激活函数后的结果。

从形式上看，对于卷积核宽度为w_a的CNN层，在时间步j的输出$\boldsymbol{H}_{\mathrm{CNN},j}^{(a,i)}$仅依赖于输入序列中长度为$w_a$的局部邻域，如式(6)所示。

$$U_{w_a}(i,j) = \left\{ x_{i,t} \mid j - \left\lfloor \frac{w_a}{2} \right\rfloor \leqslant t \leqslant j + \left\lfloor \frac{w_a}{2} \right\rfloor \right\} \tag{6}$$

其中，t为局部时间窗口内的输入位置索引。卷积操作在感受野$U_{w_a}(i,j)$内共享参数，对该局部时间窗口内的特征向量进行线性变换与非线性激活，得到局部特征表示$\boldsymbol{H}_{\mathrm{CNN},j}^{(a,i)}$。因此，CNN层表示对固定长度时间窗口的局部时序特征提取。由于每个$x_{i,t}$中包含位置特征维度，卷积核在这些维度上的共享权重能够自动学习到出行方向变化、局部转弯、停留等空间形态的变化模式。

2.3.3　BiLSTM层

BiLSTM层旨在捕捉轨迹数据中的长期时序依赖关系，BiLSTM层接收CNN层提取的局部时序特征，这些特征既包含轨迹局部空间形态相关的信息，也反映轨迹在时间维度上的动态演变。通过采用BiLSTM结构，模型能同时利用过去和未来的上下文信息理解当前时间步的特征，学习轨迹数据中的时序依赖模式，帮助模型识别轨迹变化的趋势，更全面地建模轨迹序列中的长距离时序依赖。本文采用2层BiLSTM结构，以从CNN的输出中进一步提取时序特征。具体过程可表示为

$$\boldsymbol{H}_{\mathrm{BiLSTM}}^{(0,i)} = \mathrm{BiLSTM}_0\left(\boldsymbol{H}_{\mathrm{CNN}}^{(2,i)}\right) \tag{7}$$

$$\boldsymbol{H}_{\mathrm{BiLSTM}}^{(1,i)} = \mathrm{BiLSTM}_1\left(\boldsymbol{H}_{\mathrm{BiLSTM}}^{(0,i)}\right) \tag{8}$$

其中，BiLSTM_0和BiLSTM_1分别表示第一层和第二层BiLSTM。$\boldsymbol{H}_{\mathrm{BiLSTM}}^{(0,i)}$和$\boldsymbol{H}_{\mathrm{BiLSTM}}^{(1,i)}$分别表示这两层BiLSTM输出的隐藏状态序列。

由式(9)可知，第一层BiLSTM在时间步j的输出向量可以写为

$$\boldsymbol{H}_{\mathrm{BiLSTM},j}^{(0,i)} = \left[\overrightarrow{h}_j^{(0,i)} \| \overleftarrow{h}_i^{(0,i)} \right] \tag{9}$$

其中，$\overrightarrow{h}_j^{(0,i)}$和$\overleftarrow{h}_i^{(0,i)}$分别表示前向和反向LSTM的隐藏状态，符号[· ‖ ·]表示向量拼接。前向隐藏状态满足递推关系式(10)。

$$\overrightarrow{h}_j^{(0,i)} = f_{\mathrm{LSTM}}\left(\overrightarrow{h}_{j-1}^{(0,i)}, \boldsymbol{H}_{\mathrm{CNN},j}^{(2,i)}\right) \tag{10}$$

因此通过不断展开递推公式可知，$\overrightarrow{h}_j^{(0,i)}$实际上依赖于前缀$\left\{\boldsymbol{H}_{\mathrm{CNN},j}^{(2,i)}, \cdots, \boldsymbol{H}_{\mathrm{CNN},n_i}^{(2,i)}\right\}$中的所有位置。同理，反向隐藏状态$\overleftarrow{h}_i^{(0,i)}$依赖于后缀$\left\{\boldsymbol{H}_{\mathrm{CNN},j}^{(2,i)}, \cdots, \boldsymbol{H}_{\mathrm{CNN},n_i}^{(2,i)}\right\}$中的所有位置。因此，在双向结构下，每个时间步的BiLSTM输出$\boldsymbol{H}_{\mathrm{BiLSTM},i}^{(0,i)}$都综合了整条轨迹在该时间步之前和之后的时序信息，即在形式上显式建模了位置点之间的长程依赖关系。第二层BiLSTM以$\boldsymbol{H}_{\mathrm{BiLSTM},i}^{(0,i)}$为输入，重复上述处理过程，从而在第一层的全局依赖基础上进一步抽象高层时序模式。因此，在轨迹数据上，每一个位置点的高层表示都综合了整条轨迹上其他位置点的上下文信息，为后续注意力机制层和输出层提供了全局依赖建模的基础。

2.3.4　Attention层

Attention层通过更精细的方式捕捉BiLSTM层输出序列中不同位置间的复杂依赖关系，并为不同时间步动态分配权重。具体而言，CNN提取的局部空间特征与BiLSTM输出的时序特征在Attention层通过加权求和进行融合。权重分配依据每个时间步在轨迹重构过程中的重要性，动态调整每个时间步的贡献度，优先关注对轨迹恢复最为关键的部分。该加权机制使模型能根据不同任务需求自适应地聚焦关键部分，从而提高轨迹重构的精度。在低隐私预算场景下，由于隐私保护机制对数据的扰动较大，模型能更充分利用扰动产生的结构性差异提升攻击性能。在此场景下，模型更多依赖于CNN提取的局部空间特征。而在高隐私预算场景下，由于隐私保护机制对数据的扰动较小，其对数据的影响有限，模型则更多依赖于BiLSTM提取的时序特征，特别是在轨迹数据较为平稳时，BiLSTM可以捕捉轨迹序列中的长期时序依赖，进一步提升模型攻击性能。

本文采用多头自注意力机制（MHA, multi-head self-attention mechanism）增强模型对全局依赖关系的建模能力。具体而言，对于每个注意力头h，其查询$\boldsymbol{Q}_h$、键$\boldsymbol{K}_h$和值$\boldsymbol{V}_h$分别由BiLSTM层输出序列经过不同的线性投影$\boldsymbol{W}_h^{\boldsymbol{Q}}$、$\boldsymbol{W}_h^{\boldsymbol{K}}$和$\boldsymbol{W}_h^{\boldsymbol{V}}$得到。

注意力权重则通过缩放点积计算，表示为

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{11}$$

其中，$\boldsymbol{Q}$、$\boldsymbol{K}$和$\boldsymbol{V}$分别表示查询向量、键向量和值向量，d_k表示键向量的维度。这种多注意力头设计允许模型在不同的表示子空间中同时关注来自序列不同位置的信息，从多个角度捕捉和整合特征，从而更全面地理解轨迹的上下文信息。所有注意力头的输出被拼接并通过一个最终的线性变换层进行融合，形成该注意力机制层的最终输出。通过对整个序列进行动态加权，并利用MHA的多视角学习能力，模型能有效捕捉序列中各时间步之间的全局依赖关系。因此，模型在处理长时序轨迹数据时，不仅能考虑整个序列的所有信息，还能稳定并准确地捕捉轨迹的整体结构和关键转折点。具体过程可表示为

$$\boldsymbol{H}_{\text{Attention}}^{(i)} = \text{MHA}(\boldsymbol{H}_{\text{BiLSTM}}^{(i,1)},\boldsymbol{H}_{\text{BiLSTM}}^{(i,1)},\boldsymbol{H}_{\text{BiLSTM}}^{(i,1)}) \tag{12}$$

其中，$\boldsymbol{H}_{\text{BiLSTM}}^{(1,i)}$同时作为查询、键和值的输入源，MHA()表示标准的多头自注意力函数，$\boldsymbol{H}_{\text{Attention}}^{(i)}$表示经过注意力机制增强后的特征序列。

对于任意时间步j，由式(15)可知，注意力权重矩阵Attention($\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$)的第$j$行可以写为

$$\alpha_{j,t} = \frac{\exp\left(\dfrac{\boldsymbol{Q}_j\boldsymbol{K}_t^{\mathrm{T}}}{\sqrt{d_k}}\right)}{\displaystyle\sum_{u=1}^{n_i}\exp\left(\frac{\boldsymbol{Q}_j\boldsymbol{K}_u^{\mathrm{T}}}{\sqrt{d_k}}\right)},\quad t = 1,\cdots,n_i \tag{13}$$

式(13)表明，在计算第j个时间步的表示时，Attention会在整条轨迹范围内，对轨迹中每一个时间步t的特征分配一组归一化权重$\left\{a_{j,1},\cdots,a_{j,n_i}\right\}$。Attention层在时间步$j$的输出向量为

$$\boldsymbol{H}_{\text{Attention},j}^{(i)} = \sum_{t=1}^{n_i}\alpha_{j,t}\boldsymbol{H}_{\text{BiLSTM},t}^{(1,i)} \tag{14}$$

即对BiLSTM层输出序列中所有位置特征$\boldsymbol{H}_{\text{BiLSTM},i}^{(1,i)}$的加权和。权重$\alpha_{j,t}$刻画了时间步$t$对当前时间步$j$的相对重要性，权重越大，表明时间步$t$对当前时间步越重要。因此，Attention层在整条轨迹上聚合全局时序依赖的同时，显著突出对重构结果更为关键的若干位置点，从而在轨迹数据层面实现全局加权和关键位置的建模能力。

2.3.5 输出层

输出层用于生成重构轨迹中每个位置点的经度和纬度。具体而言，采用2个独立的全连接层分别对经度和纬度进行回归，最终输出每个时间步的经纬度坐标。这种设计能有效捕捉经度和纬度的独立特性，避免它们之间相互影响。具体过程表示为

$$l_{\text{lon},j}^{(i)} = \boldsymbol{W}_{\text{lon}}\boldsymbol{H}_{\text{Attention},j}^{(i)} + \boldsymbol{b}_{\text{lon}} \tag{15}$$

$$l_{\text{lat},j}^{(i)} = \boldsymbol{W}_{\text{lat}}\boldsymbol{H}_{\text{Attention},j}^{(i)} + \boldsymbol{b}_{\text{lat}} \tag{16}$$

其中，$\boldsymbol{H}_{\text{Attention},j}^{(i)}$表示Attention层输出的特征序列$\boldsymbol{H}_{\text{Attention}}^{(i)}$中第$j$个时间步的特征向量。$\boldsymbol{W}_{\text{lon}}$、$\boldsymbol{b}_{\text{lon}}$、$\boldsymbol{W}_{\text{lat}}$和$\boldsymbol{b}_{\text{lat}}$分别表示预测经度和纬度的2个独立全连接层的权重矩阵和偏置向量。最后，将2个全连接层输出的经度和纬度进行拼接，形成每个位置点的地理位置坐标，从而形成完整的重构轨迹。

2.4 损失函数

由于传统的距离度量方法（如欧几里得距离）适用于二维平面坐标系中的距离计算，但在地理位置计算中，由于地球表面为球面结构，直接采用欧几里得距离会产生计算误差。因此，本文采用基于Haversine距离[24]的平均绝对误差损失函数对模型训练结果进行量化。

本文使用的T-Drive与GeoLife均为城市短距离轨迹数据集，在此类数据集场景中，地球曲率对欧氏距离的影响通常可以忽略。然而，Haversine距离[24]作为计算球面两点间最短路径的标准方法，具有更高的理论精度与物理一致性，能够提供更接近实际地理位置的度量。为保证结果的严谨性，本文采用Haversine距离来代替欧氏距离。根据2.1节的定义，第i条原始轨迹的第j个位置点为$p_{i,j} = \left(l_{\text{lat},j}^{(i)},l_{\text{lon},j}^{(i)}\right)$，对应的模型重构位置点为$\hat{p}_{i,j} = \left(\hat{l}_{\text{lat},j}^{(i)},\hat{l}_{\text{lon},j}^{(i)}\right)$，其中$\hat{l}_{\text{lat},j}^{(i)}$和$\hat{l}_{\text{lon},j}^{(i)}$分别表示重构后轨迹位置点的纬度值和经度值。记地球半径为R（本文取R=6 371 km），则两点之间的Haversine距离定义为

$$d(p_{i,j},\hat{p}_{i,j}) = 2R\arcsin\sqrt{\sin^2\left(\frac{l_{\text{lat},j}^{(i)} - \hat{l}_{\text{lat},j}^{(i)}}{2}\right) + \cos l_{\text{lat},j}^{(i)}\cos\hat{l}_{\text{lat},j}^{(i)}\sin^2\left(\frac{l_{\text{lon},j}^{(i)} - \hat{l}_{\text{lon},j}^{(i)}}{2}\right)} \tag{17}$$

在此基础上，第i条轨迹的重构误差定义为该条轨迹所有有效位置点Haversine距离的平均值，如式(18)所示。

$$\mathrm{Loss}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} d\left(p_{i,j},\hat{p}_{i,j}\right) \tag{18}$$

其中，n_i表示轨迹T_i的有效位置点数量。对整个训练集，攻击模型参数θ通过最小化所有轨迹重构误差进行优化调整，即

$$\min_\theta \mathrm{Loss}(\theta) = \frac{1}{M}\sum_{i=1}^{M}\mathrm{Loss}_i \tag{19}$$

3　实验与结果分析

3.1　实验设置

3.1.1　实验数据集

实验评估基于2个公开的真实数据集，分别为T-Drive[25]和GeoLife数据集[26]。T-Drive数据集包含2008年2月2日至2008年2月8日的北京市出租车GPS轨迹数据，涵盖10 357辆出租车的行驶轨迹，总行驶距离为900万公里，涉及超过1 500万个位置点。GeoLife数据集由微软亚洲研究院构建，包含2007年4月至2012年8月期间182名用户的GPS轨迹数据。该轨迹数据记录了用户的家庭、工作地点以及广泛的户外活动轨迹，总行驶距离约为120万千米，持续时间为50 176小时。

3.1.2　评估指标

本文采用欧几里得距离、豪斯多夫距离和轨迹凸包的杰卡德指数作为轨迹重构攻击任务的评估指标。具体而言，欧几里得距离主要用于度量重构轨迹与原始轨迹之间的直线距离，豪斯多夫距离度量重构轨迹与原始轨迹之间的最大差异，轨迹凸包表示轨迹的空间覆盖范围，轨迹凸包的杰卡德指数反映2条轨迹的空间重叠程度。为有效展现模型的轨迹重构效果，本文在后续实验中采用距离减少百分比（DRP, distance reduction percentage）作为归一化指标，以便在不同条件下对模型性能进行比较与分析。具体计算方法为

$$\mathrm{DRP} = \frac{\mathrm{OP} - \mathrm{OR}}{|\mathrm{OP}|} \times 100\% \tag{20}$$

其中，OP表示原始轨迹与保护轨迹之间的距离，OR表示原始轨迹与重构轨迹之间的距离。

3.1.3　实验环境与参数设置

实验使用的CPU为AMD EPYC 7513 32核处理器，内存为64 GB（可用约为63 GB），GPU为NVIDIA GeForce 3090（24 GB）。本文方法使用Python3.8和TensorFlow 2.10.0（CUDA 11.2.2）深度学习框架实现。本文的实现代码已在GitHub平台公开。模型超参数设置如表1所示。

表1　　模型超参数设置

参数	值
卷积核尺寸	3，5，7
卷积核数量	64
BiLSTM隐藏层单元数（第一层）	128
BiLSTM隐藏层单元数（第二层）	64
注意力头数量	8
优化器	Adam
学习率	0.001
批次大小	512
训练轮次	500
交叉验证轮数	5
损失函数	MAE（Haversine）

3.1.4　基线攻击方法

在轨迹数据重构攻击研究中存在多种攻击方法，本文复现了iTracker[18]、RAoPT[19]、DeepSneak[22]和ARA-2[23]这4种攻击方法，并将其作为对比基线方法。iTracker的关键超参数设置如下：结构化稀疏模型参数θ设置为默认值10 000，并在[1 000, 20 000]内进行调优以获得最佳性能。近似算法中的搜索精度参数ϱ设为0.1，约束宽松系数δ设为1.5，矩阵使用独立同分布的高斯矩阵。RAoPT的关键超参数设置如下：优化器采用Adam，学习率为0.001，批次大小为512，训练轮数为500，并采用耐心值为50个epoch的早停策略，BiLSTM隐藏层单元数为100。DeepSneak的关键超参数设置如下：在数据预处理阶段，仅保留长度介于10~50个位置点的轨迹，并将轨迹统一裁剪或填充为固定长度L=20。联邦训练与梯度重构过程中采用批次大小batch_size=4。在有限查询攻击ARA-2的实现中，遵循Han等[23]的参数设置：从辅助数据中选取4条轨迹作为地标轨迹，在每个子轨迹位置上，从所有道路段中选取与目标距离坐标最接近的10条道路段并聚合为该位置的关键点。整条轨迹被均匀划分为$k \in \{1,2,4\}$段，以控制重构的关键点数量。

3.2 实验场景设计

3.2.1 威胁模型

为系统评估本文模型在不同攻击场景下的表现，设定 3 类具有不同背景知识的敌手。本文在文献[19]提出的威胁模型基础上进行扩展，设定 3 类具有不同背景知识的敌手，并在下文中分别给出其知识边界和能力假设。

敌手 1：全背景知识。假设此类敌手已知目标轨迹数据所采用的隐私保护机制和隐私预算，并能够访问与目标轨迹数据分布一致的未经隐私保护机制处理的原始轨迹数据。在这种设定下，攻击者拥有最全面的背景知识，因此具备最优攻击性能。具体来说，攻击者不仅掌握目标数据的隐私保护机制和隐私预算，并且能够利用这些数据对攻击模型进行训练，从而实现对轨迹数据的高精度恢复。需要强调的是，此类敌手旨在评估攻击模型在最优条件下的理论性能上限，以及分析不同隐私预算参数设置对模型攻击效果的影响，在实际场景中并不存在。

敌手 2：部分背景知识。假设此类敌手已知目标轨迹数据所采用的隐私保护机制和隐私预算，但无法访问与目标轨迹数据分布一致的未经隐私保护机制处理的原始轨迹数据，即无法获取轨迹数据的具体空间分布、时间分布或用户行为模式。在这种情况下，攻击者必须基于公开的隐私保护机制和数据集对攻击模型进行训练。由于公开数据集与目标发布轨迹分布不一致且属性存在差异，因此可能导致模型的攻击性能降低。因此，该设定反映模型在跨场景下的泛化能力。与敌手 1 相比，此类敌手知识有所局限，代表一种常见的现实攻击场景，尤其适用于那些隐私保护机制已知但数据分布未知的情况。

敌手 3：无背景知识。假设此类敌手既不了解目标轨迹数据所采用的隐私保护机制和隐私预算，也无法访问与目标轨迹数据分布一致的未经隐私保护机制处理的原始轨迹数据。在这种情况下，攻击者通常只能依赖公开的轨迹数据集，并通过模型学习数据中的普适轨迹结构规律，其攻击能力通常弱于具备背景知识的情形，该攻击者代表了攻击条件苛刻的情形，旨在检验模型在攻击条件苛刻下的鲁棒性。类似的无背景知识设定已在 Buchholz 等[19]的研究中被采用，并在多组真实数据集上系统评估了此类攻击者的可行性与局限性。这些普适特征可能在一定程度上帮助攻击者恢复轨迹，但在实际场景中，攻击者通常面临多个挑战，如噪声影响、个性化差异和数据分布差异过大等，这些因素会显著降低攻击效果，甚至可能导致攻击失败。此类敌手旨在验证模型在最不利条件下的表现，评估模型在缺乏背景知识情况下的攻击效果。

3.2.2 保护机制

本文研究的攻击目标是 DP 保护机制，故选取 CNoise[27]和采样距离与方向（SDD, sampling distance and direction）[27-29]这 2 种不同的保护机制。CNoise 机制作为基础的直接加噪机制，因其资源消耗量低，常被选作验证攻击模型在直接加噪扰动场景下有效性的重要基线[19,27]。SDD 机制则以指数机制为核心，在有界候选集中按距离和方向逐步对下一位置进行采样，同时结合速度上界与终点可达性等物理约束，避免采样过程中出现无界偏移与形态失真，有效平衡隐私性与可用性。SDD 机制常被用作对比基线[28-29]，反映出其在该领域的代表性。综上所述，本文选择 CNoise 与 SDD 机制评估模型在不同保护机制下的适应性与有效性。

CNoise 机制通过向轨迹每个位置点的经纬度坐标添加独立的拉普拉斯噪声实现隐私保护。给定轨迹 $T=\langle(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)\rangle$，其扰动后的坐标为

$$\tilde{x}_i = x_i + \alpha_i,\quad \tilde{y}_i = y_i + \beta_i \tag{21}$$

其中，α_i和β_i分别为第 i 个位置点的经度和纬度坐标所添加的噪声，且 α_i和β_i独立同分布于 Lap(b)。Lap(b)表示尺度参数 b>0 的拉普拉斯分布，其概率密度函数为

$$p_L(z|b) = \frac{1}{2b}\exp\left(-\frac{|z|}{b}\right) \tag{22}$$

其中，$p_{L(z|b)}$为拉普拉斯分布的概率密度函数，表示当尺度参数为 b 时，噪声变量取值为 z 的概率密度。设轨迹中相邻位置点的最大物理位移为 S。对于任意 2 条仅在单个位置点上不同的相邻轨迹，其坐标查询的全局敏感度 Δ 满足 $\Delta \leqslant 2\sqrt{2S}$。为实现 ε-DP，根据拉普拉斯机制，噪声尺度参数需满足 $b \geqslant \frac{\Delta}{\varepsilon}$，为确保在最坏情况下该不等式依然成立，设置尺度参数 b 需满足式(23)。

$$b \geqslant \frac{2\sqrt{2}\,S}{\varepsilon} \tag{23}$$

因此，隐私预算ε越小，b越大，噪声强度越高，隐私保护性越强。

SDD机制是一种相较于CNoise机制更为先进的隐私保护策略，其加噪方式并非直接向坐标点中添加噪声，而是通过迭代的和概率性的过程逐点构建隐私保护轨迹。具体而言，该机制基于指数机制，从公开的起点$\tilde{p}_0 = p_0$出发，依据已发布的扰动位置与原始轨迹的真实走向，逐步构建满足DP保护机制的轨迹。对于每个中间点i=1,2,⋯,n−1，其发布过程如下：首先基于前一步发布的扰动位置$\tilde{p}_{i-1}$和原始位置p_i计算真实位移向量$\vec{v}_i = p_i - \tilde{p}_{i-1}$，得到其距离$r_i = \|\vec{v}_i\|$和方向$\varphi_i = \arg(\vec{v}_i)$。SDD机制通过指数机制采样扰动位移，其关键是为候选距离$\rho_i \in [0,S]$和方向$\gamma_i \in [0,2\pi]$定义效用函数。

$$u_d(\rho_i) = -|\rho_i - r_i|,\quad u_a(\gamma_i) = -|\gamma_i - \varphi_i| \tag{24}$$

该效用函数衡量候选值与真实值的接近程度，值越高表示越接近。采样概率遵循式(25)和式(26)。

$$\Pr(\rho_i) \propto \exp\left(-\frac{\varepsilon|\rho_i - r_i|}{8S}\right) \tag{25}$$

$$\Pr(\gamma_i) \propto \exp\left(-\frac{\varepsilon|\gamma_i - \varphi_i|}{8\pi}\right) \tag{26}$$

其中，ρ_i表示候选距离，r_i表示真实的位移距离，Pr(ρ_i)表示第i步候选值ρ_i被选中的概率。γ_i表示待采样的扰动方向，φ_i表示真实的位移方向，Pr(γ_i)表示第i步候选方向值γ_i被选中的概率。式(25)和式(26)直观地揭示了每一步采样距离和方向的选择概率随效用函数递减的特性，效用函数值越低，指数项的衰减越剧烈，对应候选值被选中的概率按指数规律降低。这使采样结果以高概率集中在真实值附近，同时通过必要的随机性满足ε-DP。

为评估攻击模型在2种保护机制下的性能，本文在多种隐私预算ε条件下开展实验，结合现有研究与实际应用[27-29]，ε通常取[0.01,10]，该区间能够覆盖从强隐私保护场景到弱隐私保护场景的常见参数选择。此外，为评估在弱隐私保护场景条件下的潜在泄露风险，本文在CNoise机制上设置ε=100进行实验。该取值并非纯粹的理论假设，而是在现有实践和评测中被采用的弱隐私配置。2020年美国人口普查[30]的Detailed DHC-A产品在SafeTab-P算法下采用零集中差分隐私，对不同人口群体分配隐私损失预算ρ_i，折算到传统(ε,δ)-DP时其等效隐私预算处于几十这一数量级，用于保证细粒度统计的可用性。本文同样选取ε=100作为极弱隐私保护场景的代表值之一，用于观察在隐私保护约束极其宽松时轨迹重构攻击模型能够达到的性能上限。

3.3　数据预处理

为提高模型训练效率和数据质量，本文对T-Drive和GeoLife数据集进行预处理。首先，清洗轨迹数据，剔除包含缺失值（如经纬度和时间戳）和异常值的记录，以避免不完整数据对模型训练造成干扰。其次，移除重复位置点。若2个位置点在经纬度和时间戳上完全一致，则移除第二个位置点。此外，如3.2.2节所述，本文采用了SDD机制。该机制需预先定义一个最大速度阈值，以限制轨迹中用户的最高移动速度。为满足该要求，本文移除了所有用户速度超过该数据集99%分位数的位置点。具体而言，对于T-Drive数据集，移除所有速度高于90 km/h的位置点。同样，GeoLife数据集中所有速度超过100 km/h的位置点也被移除。

3.4　敌手1下的轨迹重构攻击实验

3.4.1　T-Drive数据集上的轨迹重构攻击实验

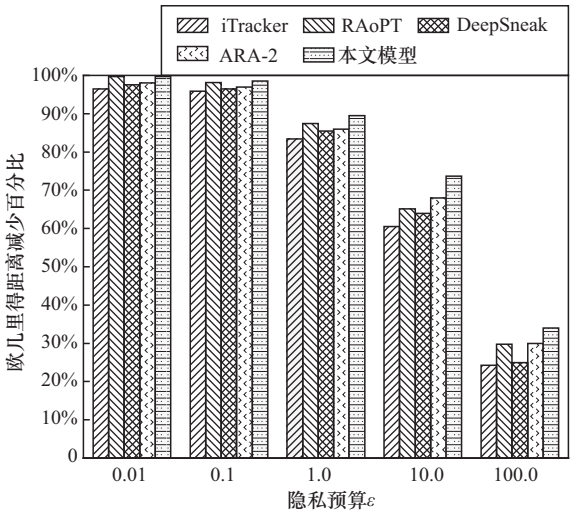
本节在T-Drive数据集上对本文模型与基线模型进行对比分析。由于该类敌手拥有全背景知识，因此，在训练攻击模型和测试攻击效果时，训练集和测试集使用的数据集、隐私保护机制和隐私预算均保持一致。具体实验设置如表2所示。

表2　T-Drive数据集上的实验设置

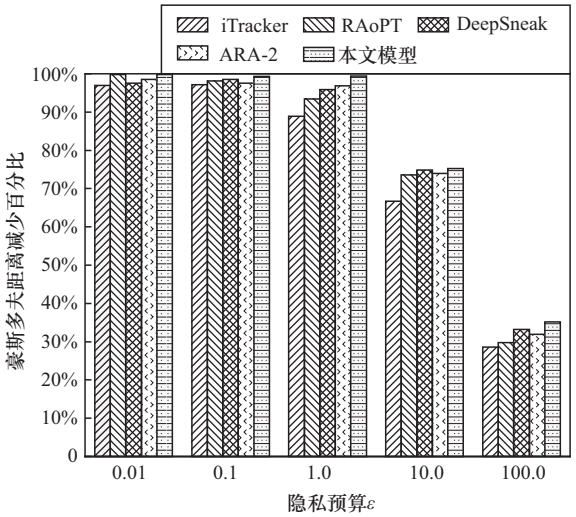
序号	隐私保护机制	隐私预算
1	CNoise	0.01
2	CNoise	0.1
3	CNoise	1.0
4	CNoise	10.0
5	CNoise	100.0
6	SDD	0.01
7	SDD	0.1
8	SDD	1.0
9	SDD	10.0

在T-Drive数据集上的实验结果如图3、图4和表3所示。由图3和图4可知，在不同隐私预算条件下，本文模型在2种距离减少百分比指标上均优

于基线模型。具体而言，在CNoise机制下，当隐私预算ε较小时，各个模型在该指标上均表现优异，表明在隐私保护强度较强时均能有效识别并利用保护轨迹与原始轨迹的结构性差异，从而实现较好的攻击效果。然而，在相同隐私预算条件下，本文模型在该指标上取得了更优的结果。随着隐私预算的增加，隐私保护强度减弱，加入的噪声减少，保护轨迹与原始轨迹之间的结构性差异减少，各个模型性能在该指标上均呈下降趋势，但本文模型的表现始终优于基线模型，展现出更高的稳定性。在SDD机制下，随着隐私预算的变化，各个模型的距离减少百分比指标基本保持稳定。这与原论文中关于SDD机制的实验结论一致，即当隐私预算在[0.01,10]内变化时，其误差基本保持不变。

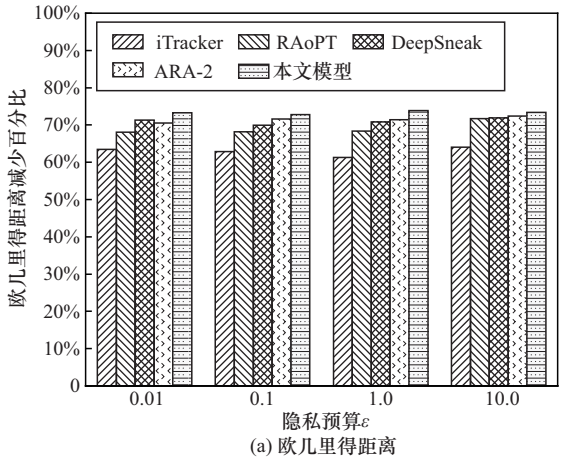


(a) 欧几里得距离

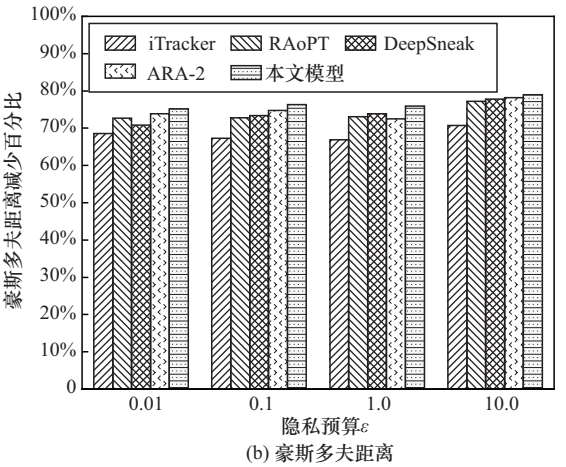


(b) 豪斯多夫距离

图3　CNoise机制下的距离减少百分比（T-Drive数据集）



(a) 欧几里得距离



(b) 豪斯多夫距离

图4　SDD机制下的距离减少百分比（T-Drive数据集）

轨迹凸包的杰卡德指数作为衡量2条轨迹覆盖区域相似度的重要指标，其数值越高，表示2条轨迹在空间范围上的重叠程度越大。为清晰区分实验效果，表3中加粗数值表示各行对比中的最优结果，本文后续表中加粗数值含义与此相同。由表3的对比结果可知，本文模型在所有实验参数设置下的杰卡德指数均高于基线模型，表明其在捕捉轨迹空间模式方面表现更优。此结果进一步验证了本文模型在局部特征提取、空间信息提取及全局信息整合方面的有效性，能更准确地恢复原始轨迹的整体形态。综上，本文模型在T-Drive数据集上的攻击效果优于基线模型。

3.4.2　GeoLife数据集上的轨迹重构攻击实验

为进一步评估本文模型的泛化能力，尤其是在不同轨迹模式下的表现。本文在GeoLife数据集上开展相关实验。GeoLife数据集包含更加多样化的运动模式，与主要由出租车轨迹组成的T-Drive数据集存在显著差异。具体实验设置如表4所示。

GeoLife数据集上的实验结果如图5、图6和表5所示。由图5和图6可知，在CNoise机制下，当隐私预算ε较小时，各类模型在距离减少百分比指标上均表现优异。随着隐私预算ε的增大，该指标呈

表3　T-Drive数据集上的杰卡德指数对比结果

序号	iTracker	RAoPT	DeepSneak	ARA-2	本文模型
1	4.78×10^{-2}	5.12×10^{-2}	5.46×10^{-2}	5.38×10^{-2}	$\mathbf{6.23\times10^{-2}}$
2	2.66×10^{-3}	3.44×10^{-3}	3.88×10^{-3}	3.75×10^{-3}	$\mathbf{4.34\times10^{-3}}$
3	3.08×10^{-2}	3.67×10^{-2}	3.59×10^{-2}	3.95×10^{-2}	$\mathbf{4.41\times10^{-2}}$
4	1.89×10^{-1}	2.66×10^{-1}	2.73×10^{-1}	2.98×10^{-1}	$\mathbf{3.58\times10^{-1}}$
5	5.05×10^{-1}	6.23×10^{-1}	6.31×10^{-1}	6.49×10^{-1}	$\mathbf{6.98\times10^{-1}}$
6	5.88×10^{-2}	7.09×10^{-2}	7.28×10^{-2}	7.38×10^{-2}	$\mathbf{7.93\times10^{-2}}$
7	5.79×10^{-2}	7.03×10^{-2}	7.10×10^{-2}	7.27×10^{-2}	$\mathbf{7.83\times10^{-2}}$
8	5.48×10^{-2}	7.13×10^{-2}	7.40×10^{-2}	7.30×10^{-2}	$\mathbf{7.76\times10^{-2}}$
9	6.96×10^{-2}	8.66×10^{-2}	8.83×10^{-2}	9.03×10^{-2}	$\mathbf{9.38\times10^{-2}}$

表4　GeoLife数据集上的实验设置

序号	隐私保护机制	隐私预算
10	CNoise	0.01
11	CNoise	0.1
12	CNoise	1.0
13	CNoise	10.0
14	CNoise	100.0
15	SDD	0.01
16	SDD	0.1
17	SDD	1.0
18	SDD	10.0

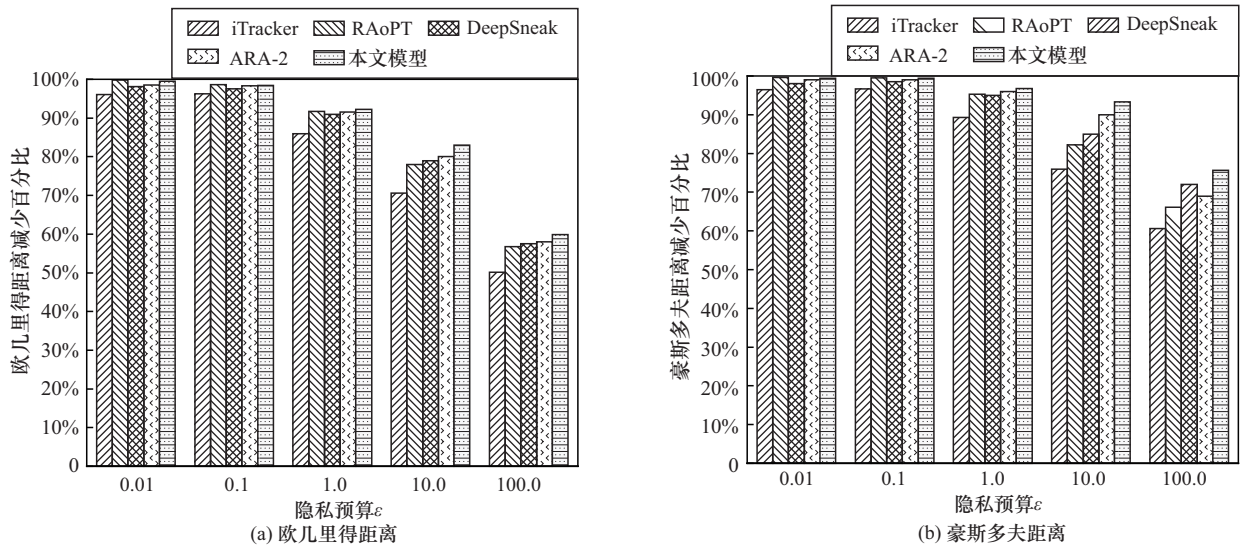


图5　CNoise机制下的距离减少百分比(GeoLife数据集)

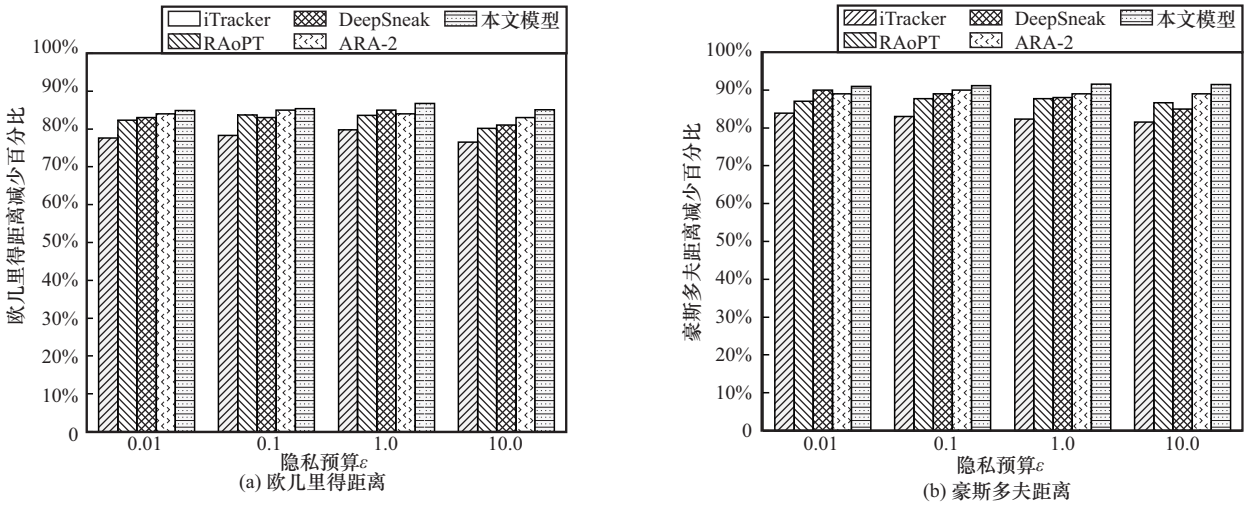


图6　SDD机制下的距离减少百分比(GeoLife数据集)

下降趋势，但本文模型在各隐私预算条件下的表现均优于基线模型，且在高隐私预算条件下优势更加显著，表明其在噪声较小时仍能有效捕捉和利用轨迹的结构性差异。此外，在 SDD 机制下，各模型性能在该指标上基本保持稳定，但本文模型的表现仍优于基线模型。

表 5　GeoLife 数据集上的杰卡德指数对比结果

序号	iTracker	RAoPT	DeepSneak	ARA-2	本文模型
10	1.08×10^{-5}	1.72×10^{-5}	1.88×10^{-5}	1.79×10^{-5}	$\mathbf{2.08\times10^{-4}}$
11	6.91×10^{-4}	8.08×10^{-4}	8.14×10^{-4}	8.30×10^{-4}	$\mathbf{8.66\times10^{-4}}$
12	9.98×10^{-4}	1.53×10^{-3}	1.73×10^{-3}	1.70×10^{-3}	$\mathbf{1.98\times10^{-3}}$
13	7.79×10^{-3}	9.42×10^{-3}	9.98×10^{-3}	1.01×10^{-2}	$\mathbf{1.17\times10^{-2}}$
14	5.04×10^{-2}	6.78×10^{-2}	6.87×10^{-2}	6.90×10^{-2}	$\mathbf{7.73\times10^{-2}}$
15	2.03×10^{-3}	2.55×10^{-3}	2.98×10^{-3}	3.27×10^{-3}	$\mathbf{4.56\times10^{-3}}$
16	2.07×10^{-3}	2.55×10^{-3}	3.14×10^{-3}	3.28×10^{-3}	$\mathbf{5.77\times10^{-3}}$
17	1.99×10^{-3}	2.56×10^{-3}	3.36×10^{-3}	3.56×10^{-3}	$\mathbf{6.68\times10^{-3}}$
18	7.08×10^{-4}	8.86×10^{-4}	9.97×10^{-4}	1.03×10^{-3}	$\mathbf{2.16\times10^{-3}}$

表 5 中轨迹凸包的杰卡德指数对比结果表明，由本文模型重构得到的轨迹在该指标上性能均高于基线模型，说明其对原始轨迹的空间形态特征重构更有效。综上所述，在 GeoLife 数据集上的实验结果验证了本文模型的泛化能力与适用性。

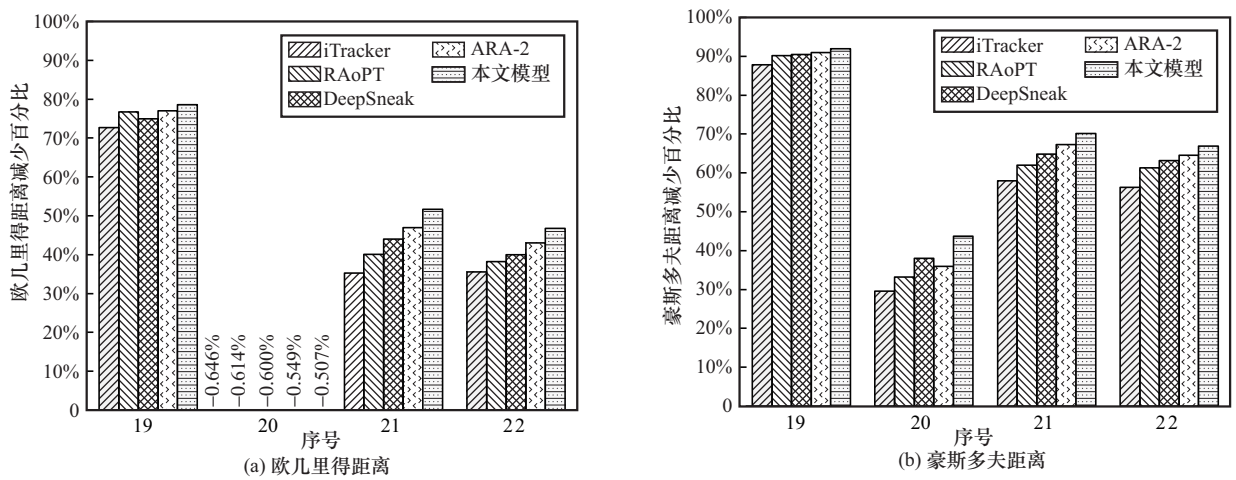
3.5　敌手 2 下的轨迹重构攻击实验

在 3.4 节中，本文在敌手 1 的背景知识假设下开展实验，验证了所提模型在不同数据集上相较于基线模型的优越性，但敌手 1 的背景知识假设过于理想化，难以适应现实场景。为评估模型在更贴近实际场景下的攻击效果，本节在敌手 2 的背景知识假设下开展实验。具体实验设置如表 6 所示。

表 6　　敌手 2 背景知识下的实验设置

序号	训练数据集	测试数据集	隐私保护机制	隐私预算
19	T-Drive	GeoLife	CNoise	1.0
20	T-Drive	GeoLife	CNoise	10.0
21	T-Drive	GeoLife	SDD	0.1
22	T-Drive	GeoLife	SDD	1.0
23	GeoLife	T-Drive	CNoise	1.0
24	GeoLife	T-Drive	CNoise	10.0
25	GeoLife	T-Drive	SDD	0.1
26	GeoLife	T-Drive	SDD	1.0

各个攻击模型在距离减少百分比指标上的对比结果如图 7 和图 8 所示。实验结果表明，本文模型在该指标上性能均优于基线模型。然而，在序号 20 的实验设置下，从 T-Drive 数据集迁移至 GeoLife 数据集时出现攻击失败的情况。而在相同的参数设置下，从 GeoLife 数据集迁移到 T-Drive 数据集时，攻击却是成功的。造成这一现象的原因可从数据分布差异与扰动信号强度两方面解释。首先，T-Drive 数据集主要由出租车轨迹构成，其轨迹类型相对单一，使模型在训练过程中学习到的轨迹结构规律覆盖面有限。而 GeoLife 数据集包含更为多样的运动出行模式，在行为模式与轨迹形态上更为复杂。因此，模型在由 T-Drive 数据集迁移到 GeoLife 数据集时更容易出现过度校正，从而导致攻击失败；相反，GeoLife 数据集的轨迹类型更丰富，其



(a) 欧几里得距离　　(b) 豪斯多夫距离

图 7　T-Drive 数据集至 GeoLife 数据集的距离减少百分比结果

中可能包含与T-Drive数据集相似的运动出行模式，因此在GeoLife数据集上训练得到的模型具有更强的泛化能力，能够迁移并适用于T-Drive数据集，从而实现攻击成功。其次，序号20对应ε=10.0，在CNoise机制下噪声强度较弱，保护轨迹与原始轨迹之间的结构性差异随之减小，模型在测试阶段缺乏明确的修正目标，更易出现修正幅度与真实偏差不匹配的情况。

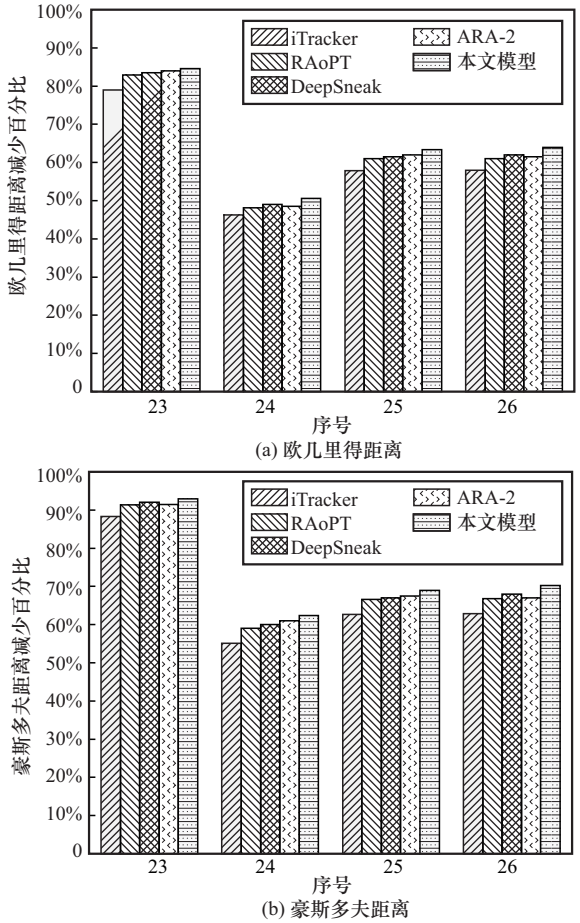


图8　GeoLife数据集至T-Drive数据集的距离减少百分比结果

各个模型在轨迹凸包的杰卡德指数指标上的对比结果如表7所示。实验结果表明，本文模型在跨数据集上的表现优于基线模型，表明本文模型重构出的轨迹在整体空间形态方面具备更高的准确性。综上所述，在敌手2背景知识假设下，本文模型在多个评价指标上均优于基线模型，验证了其在训练与测试数据分布不一致的跨域场景下的良好适应能力。在无法获取与发布轨迹具有相同分布的原始数据的现实假设下，模型能借助训练阶段所学习到的轨迹模式进行迁移，展现出更强的泛化能力与攻击稳定性。总体来看，在敌手2所代表的跨场景设定下，本文模型在大多数实验配置中优于基线模型，表明该模型在多种目标数据分布、隐私机制与隐私预算组合下具有较好的跨场景泛化能力，从而在一定程度上缓解了跨场景泛化能力不足的问题。

表7　敌手2背景知识下的杰卡德指数对比结果

序号	iTracker	RAoPT	DeepSneak	ARA-2	本文模型
19	5.38×10^{-4}	6.62×10^{-4}	7.31×10^{-4}	7.49×10^{-4}	$\mathbf{9.63\times10^{-4}}$
20	5.05×10^{-3}	5.83×10^{-3}	6.38×10^{-3}	6.54×10^{-3}	$\mathbf{1.01\times10^{-2}}$
21	6.43×10^{-4}	8.01×10^{-4}	9.01×10^{-4}	8.75×10^{-4}	$\mathbf{3.68\times10^{-3}}$
22	7.66×10^{-4}	9.23×10^{-4}	9.87×10^{-4}	1.02×10^{-3}	$\mathbf{2.31\times10^{-3}}$
23	1.08×10^{-2}	1.66×10^{-2}	1.82×10^{-2}	1.97×10^{-2}	$\mathbf{2.68\times10^{-2}}$
24	9.65×10^{-2}	1.23×10^{-1}	1.38×10^{-1}	1.43×10^{-1}	$\mathbf{1.98\times10^{-1}}$
25	2.65×10^{-2}	3.18×10^{-2}	3.57×10^{-2}	4.14×10^{-2}	$\mathbf{6.34\times10^{-2}}$
26	2.47×10^{-2}	3.19×10^{-2}	3.69×10^{-2}	4.08×10^{-2}	$\mathbf{5.37\times10^{-2}}$

3.6　敌手3下的轨迹重构攻击实验

为更贴近现实场景，本文在敌手3的背景知识下开展实验。敌手3代表最不利的攻击情境，即训练集和测试集在数据分布、隐私保护机制以及隐私预算等方面存在差异。具体实验设置如表8所示。

各个模型在距离减少百分比指标上的对比结果如图9所示。实验结果表明，在敌手3的背景知识假设下，本文模型的表现均优于基线模型。然而，在序号27的实验设置下，各个模型均出现攻击失败的情况，其原因可从数据集与隐私保护机制两方面理解。首先，T-Drive数据集轨迹形态与行为模

表8　敌手3背景知识下的实验设置

序号	训练集	测试集	训练集保护机制	测试集保护机制	训练集隐私预算	测试集隐私预算
27	T-Drive	GeoLife	CNoise	SDD	1.0	0.1
28	GeoLife	T-Drive	SDD	CNoise	0.1	1.0
29	T-Drive	GeoLife	SDD	CNoise	1.0	0.1
30	GeoLife	T-Drive	CNoise	SDD	0.1	1.0

式更为单一，而GeoLife数据集则覆盖多类型出行与活动。因此，在从T-Drive数据集向GeoLife数据集的迁移过程中，训练端所反映的轨迹分布难以充分代表测试端的多样化行为模式，从而使攻击在少数配置下更容易受分布差异影响而失效。其次，该设置的训练端采用CNoise机制而测试端采用SDD机制，2种机制生成的发布轨迹在形态与统计特征上存在本质差异：CNoise机制对位置点进行独立加噪，扰动更接近点级随机偏移；而SDD机制基于距离和方向的采样并引入速度上界、终点可达性等约束，使发布轨迹更具几何与行为一致性，其扰动表现为更强的结构化特征。因此，当训练端与测试端的隐私保护机制不一致时，训练端数据所呈现的扰动规律可能无法覆盖测试端机制下的结构化误差形态，从而导致攻击失败。

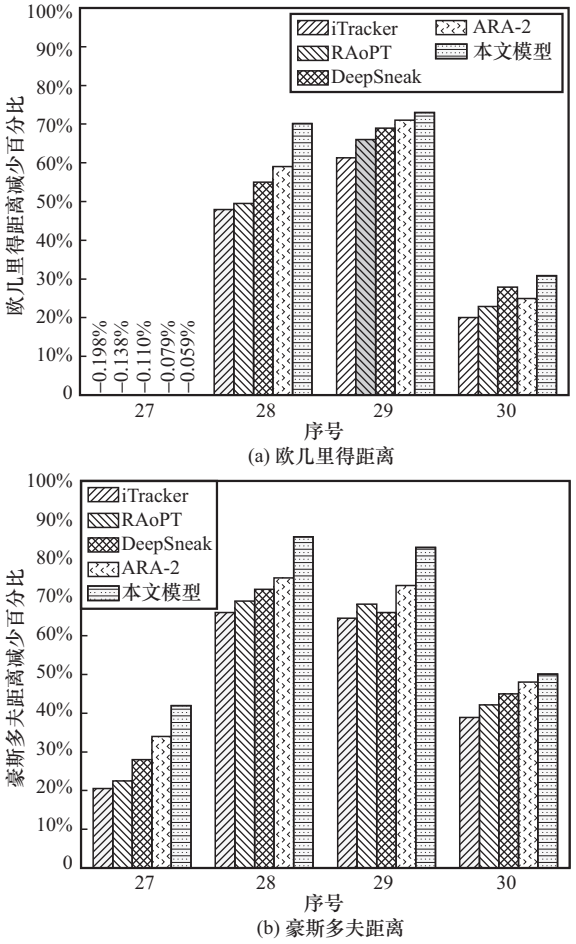


图9　敌手3背景知识下的距离减少百分比结果

各个模型在轨迹凸包的杰卡德指数指标上的对比结果如表9所示。实验结果表明，在4种实验设置下，本文模型所重构出的轨迹凸包的杰卡德指数均优于基线模型，表明该模型重构出的轨迹在活动空间范围上与原始轨迹具有更高的重叠度，能更有效地恢复出原始轨迹的空间分布特征。

表9　敌手3背景知识下的杰卡德指数对比结果

序号	iTracker	RAoPT	DeepSneak	ARA-2	本文模型
27	3.31×10^{-5}	3.92×10^{-5}	4.76×10^{-5}	5.69×10^{-5}	$\mathbf{2.98\times10^{-4}}$
28	8.99×10^{-3}	1.12×10^{-2}	1.68×10^{-2}	2.37×10^{-2}	$\mathbf{3.64\times10^{-2}}$
29	2.08×10^{-7}	2.86×10^{-7}	5.09×10^{-7}	6.38×10^{-7}	$\mathbf{1.34\times10^{-5}}$
30	8.06×10^{-3}	9.41×10^{-3}	9.89×10^{-3}	1.07×10^{-2}	$\mathbf{1.38\times10^{-2}}$

3.7　超参数实验与结果分析

3.7.1　CNN结构影响分析

本节旨在探究CNN模块中不同卷积核组成对模型性能的影响。实验从单层结构开始，逐步增加层数并扩展卷积核尺寸，评估了双层、三层和四层CNN等不同结构对模型性能的影响，结果如表10所示。由表10可知，在2个数据集上，模型的攻击性能随着CNN层数和卷积核尺寸组合的递增呈现先提升后略有下降的趋势。具体而言，当采用三层结构时，模型在各项指标上均达到最优性能。相比单层结构，这种递增尺寸的卷积核组合能有效捕获不同尺度的局部时序特征和空间相关性，增强模型对轨迹局部模式的理解，从而提升重构精度。然而，进一步增加至四层结构时，攻击性能出现轻微下降。这表明三层结构已能充分提取当前任务所需特征，过度增加网络深度或感受野可能会引入冗余信息或增加过拟合风险。因此，本文最终选择三层的CNN结构。

3.7.2　BiLSTM结构影响分析

本节旨在探究BiLSTM层数与隐藏层单元数对模型性能的影响。实验评估了单层、双层以及三层BiLSTM等不同结构对模型性能的影响，结果如表11所示。由表11可知，在2个数据集上，模型性能随着BiLSTM层数和隐藏层单元数的变化呈现先上升后趋于平稳的趋势。当采用双层BiLSTM结构时，模型在各项指标上均达到最优性能。与单层结构相比，双层结构通过增加模型深度，增强BiLSTM对轨迹序列中长距离时序依赖的捕捉能力，这对于还原轨迹的整体连贯性和全局形态具有关键作用。然而，当网络深度增加至三层时，性能未进

表10　不同CNN结构下的实验结果

数据集	CNN层数	卷积尺寸组合	欧几里得距离减少百分比	豪斯多夫距离减少百分比	杰卡德指数
T-Drive	1	3×3	70.2%	73.4%	7.18×10^{-2}
	2	3×3, 5×5	71.3%	74.5%	7.43×10^{-2}
	3	3×3, 5×5, 7×7	**73.9%**	**75.9%**	$\mathbf{7.76\times10^{-2}}$
	4	3×3, 5×5, 7×7, 9×9	73.4%	74.9%	7.65×10^{-2}
GeoLife	1	3×3	84.0%	88.3%	3.03×10^{-3}
	2	3×3, 5×5	84.9%	89.1%	4.38×10^{-3}
	3	3×3, 5×5, 7×7	**86.7%**	**91.6%**	$\mathbf{6.68\times10^{-3}}$
	4	3×3, 5×5, 7×7, 9×9	85.5%	90.7%	6.73×10^{-3}

表11　不同BiLSTM结构下的实验结果

数据集	BiLSTM层数	隐藏单元数	欧几里得距离减少百分比	豪斯多夫距离减少百分比	杰卡德指数
T-Drive	1	64	69.8%	73.9%	7.22×10^{-2}
	1	128	71.9%	75.0%	7.59×10^{-2}
	2	128，64	**73.9%**	**75.9%**	$\mathbf{7.76\times10^{-2}}$
	3	128，64，32	73.5%	75.4%	7.73×10^{-2}
GeoLife	1	64	84.3%	88.5%	3.28×10^{-3}
	1	128	85.0%	89.8%	5.03×10^{-3}
	2	128，64	**86.7%**	**91.6%**	$\mathbf{6.68\times10^{-3}}$
	3	128，64，32	86.5%	90.9%	6.60×10^{-3}

一步提升，反而有所下降。这可能源于轨迹数据的特性，过深的BiLSTM层数未能引入更多有效的高级时序特征，反而可能增加了信息冗余或导致内部状态的过度平滑，进而掩盖了关键时序信息。因此，本文最终选择双层BiLSTM结构。

3.7.3　注意力头数影响分析

本节旨在探究注意力头数对模型性能的影响。实验评估了注意力头数分别为4、8、16、32等不同设置下的模型性能，结果如表12所示。由表12可知，当注意力头数为8时，模型在各项指标上均达到了最优值。而将注意力头数进一步增加至16或32时，并未带来性能提升。这一现象表明，在当前的任务和数据集上，增加注意力头数可在一定程度上提升模型对轨迹细节的捕捉能力，从而改善性能。然而，过多的注意力头数可能导致模型计算复杂度增加，而性能提升幅度却极为有限。因此，本文最终选择8个注意力头的配置，以确保模型达到较优性能的同时，能有效控制计算开销。

表12　不同注意力头数下的实验结果

数据集	注意力头数	欧几里得距离减少百分比	豪斯多夫距离减少百分比	杰卡德指数
T-Drive	4	72.1%	73.6%	7.37×10^{-2}
	8	**73.9%**	**75.9%**	$\mathbf{7.76\times10^{-2}}$
	16	73.0%	75.1%	7.65×10^{-2}
	32	72.5%	75.0%	7.68×10^{-2}
GeoLife	4	84.9%	89.8%	5.37×10^{-2}
	8	**86.7%**	**91.6%**	$\mathbf{6.68\times10^{-3}}$
	16	85.9%	91.0%	6.65×10^{-3}
	32	86.3%	91.2%	6.60×10^{-3}

3.8　消融实验与结果分析

由于本文模型包含多个模块，为验证各模块对性能的贡献，本文设计消融实验评估Attention与CNN的有效性。具体地，构造2种变体：去除Attention的CNN-BiLSTM和去除CNN的BiLSTM-Attention，分别用于评估Attention的关键时间步加权与全局信息整合能力，以及CNN的局部时序/空间信息提取能力。

消融实验在30组参数设置下进行。由于篇幅限制，本文选取6个代表性实例，这些实例涉及3种不同背景知识敌手下的参数设置，以评估模型在各种攻击环境下的性能。基于T-Drive和GeoLife数据集的消融实验结果如表13所示。表13中的BiLSTM一栏展示了RAoPT模型的实验结果，该模型在本文实验中作为基线模型。通过比较本文模型与CNN-BiLSTM模型的实验结果可知，本文模型的攻击效果优于CNN-BiLSTM模型。表明Attention模块能够通过自适应调整各时间步的权重，使模型有效聚焦于轨迹中的关键时间步，并整合轨迹的全局信息，从而提升攻击效果。通过比较本文模型与BiLSTM-Attention模型在T-Drive和GeoLife数据集上的实验结果可知，本文模型的攻击效果优于BiLSTM-Attention模型，表明CNN通过提取轨迹数据中的局部时序特征和空间信息，为重构过程提供了更多有效信息，这些信息有助于模型高效理解轨迹的空间分布，从而提升攻击效果。

4 结束语

针对现有DP保护机制的重构攻击方法在局部特征提取、空间信息提取以及全局依赖建模方面的不足，本文提出了一种基于CNN-BiLSTM-Attention融合模型的轨迹重构攻击方法，有效解决现有方法存在的不足。实验结果表明，本文模型在多个场景下的攻击效果均优于现有攻击模型，具有良好的鲁棒性与泛化性。尽管本文研究取得一定进展，但仍存在局限：一方面，对DP轨迹场景下攻击鲁棒性边界的理论分析仍不充分；另一方面，模型依赖现有DP保护机制的公开实现与特性，当保护机制、噪声注入策略或地理约束发生变化时，攻击效果可能下降。

从攻击结果看，尽管DP保护机制提供理论上的隐私安全保证，实际应用中仍可能因保护轨迹与原始轨迹之间的结构性差异而被攻击者利用。因此，为进一步提升隐私保护的有效性，保护机制除改进噪声注入外，还应确保发布轨迹符合真实地理约束，并综合考虑噪声与路网约束，在保证轨迹遵循真实路网结构的同时降低结构性可辨识差异。在此基础上，本文进一步给出了2种隐私增强思路：1) 在机制层面，可采用可行域内随机化采样的方

表13 消融实验结果

序号	评估指标	BiLSTM	CNN-BiLSTM	BiLSTM-Attention	CNN-BiLSTM-Attention
7	欧几里得距离减少百分比	68.2%	69.8%	70.3%	**72.8%**
	豪斯多夫距离减少百分比	72.8%	73.5%	73.2%	**76.3%**
	轨迹凸包杰卡德指数	7.03×10^{-2}	7.36×10^{-2}	7.58×10^{-2}	$\mathbf{8.13\times10^{-2}}$
13	欧几里得距离减少百分比	77.7%	79.0%	78.6%	**82.3%**
	豪斯多夫距离减少百分比	82.2%	84.3%	85.3%	**92.1%**
	轨迹凸包杰卡德指数	9.42×10^{-3}	9.88×10^{-3}	1.06×10^{-2}	$\mathbf{1.17\times10^{-2}}$
19	欧几里得距离减少百分比	76.7%	77.3%	77.1%	**78.6%**
	豪斯多夫距离减少百分比	90.2%	90.9%	91.0%	**91.9%**
	轨迹凸包杰卡德指数	6.62×10^{-2}	7.07×10^{-2}	8.29×10^{-2}	$\mathbf{9.63\times10^{-4}}$
24	欧几里得距离减少百分比	82.9%	83.4%	83.0%	**84.6%**
	豪斯多夫距离减少百分比	91.4%	91.8%	92.0%	**92.9%**
	轨迹凸包杰卡德指数	1.23×10^{-1}	1.33×10^{-1}	1.40×10^{-1}	$\mathbf{1.98\times10^{-1}}$
28	欧几里得距离减少百分比	49.5%	55.4%	59.0%	**70.1%**
	豪斯多夫距离减少百分比	69.0%	73.8%	72.9%	**85.6%**
	轨迹凸包杰卡德指数	1.12×10^{-2}	1.89×10^{-2}	2.03×10^{-2}	$\mathbf{3.64\times10^{-2}}$
30	欧几里得距离减少百分比	22.9%	25.9%	26.6%	**30.8%**
	豪斯多夫距离减少百分比	42.1%	44.8%	47.2%	**50.1%**
	轨迹凸包杰卡德指数	9.41×10^{-3}	9.97×10^{-3}	1.07×10^{-2}	$\mathbf{1.38\times10^{-2}}$

式生成发布点，即先由差分隐私保护机制产生候选扰动点，再在道路网络与运动学约束定义的可行域内进行随机采样或重采样，以避免将轨迹简单投影到最近道路导致新的确定性偏差；2) 在发布策略层面，可通过时空采样与空间分辨率下调降低细粒度连续信息密度，并结合差分隐私预算的累计核算与发布约束，对同一用户在多次发布中的隐私损失进行累积管理，限制发布次数、时长或精度，以降低攻击者获得稳定训练信号的可能性。

参考文献：

[1] 门红蕾, 曹利, 郑国莉, 等. 车联网基于稀疏用户环境的LBS隐私保护方案[J]. 计算机应用研究, 2024, 41(9): 2831-2838.
MEN H L, CAO L, ZHENG G L, et al. LBS privacy protection scheme based on sparse user environment of VANET[J]. Application Research of Computers, 2024, 41(9): 2831-2838.

[2] WU S, WANG X L, WANG S, et al. K-anonymity for crowdsourcing database[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(9): 2207-2221.

[3] SEI Y C, OKUMURA H, TAKENOUCHI T, et al. Anonymization of sensitive quasi-identifiers for l-diversity and t-closeness[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 16(4): 580-593.

[4] DWORK C. Differential privacy[C]//International Colloquium on Automata, Languages and Programming. Berlin: Springer, 2006: 1-12.

[5] JIN F M, HUA W, FRANCIA M, et al. A survey and experimental study on privacy-preserving trajectory data publishing[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 5577-5596.

[6] 冯登国, 张敏, 叶宇桐. 基于差分隐私模型的位置轨迹发布技术研究[J]. 电子与信息学报, 2020, 42(1): 74-88.
FENG D G, ZHANG M, YE Y T. Research on location trajectory publishing technology based on differential privacy model[J]. Journal of Electronics & Information Technology, 2020, 42(1): 74-88.

[7] ZHAO Y X, WANG C D. Protecting privacy and enhancing utility: a novel approach for personalized trajectory data publishing using noisy prefix tree[J]. Computers & Security, 2024, 144: 103922.

[8] NIU Y T, ZHANG J, TANG Z G, et al. Privacy-preserving spatiotemporal trajectory generalization publishing scheme with differential privacy[J]. Computers & Security, 2025, 156: 104514.

[9] ZHANG Z M, XU X L, XIAO F. LGAN-DP: a novel differential private publication mechanism of trajectory data[J]. Future Generation Computer Systems, 2023, 141: 692-703.

[10] GAO W, ZHOU S W. Privacy-preserving for dynamic real-time published data streams based on local differential privacy[J]. IEEE Internet of Things Journal, 2024, 11(8): 13551-13562.

[11] SHANTHI P, VIDIVELLI S, PADMAKUMARI P. Privacy-preserving cloud-based secure digital locker with differential privacy-based deep learning technique[J]. Multimedia Tools and Applications, 2024, 83(34): 81299-81324.

[12] YAO L, CHEN Z Y, HU H B, et al. Privacy preservation for trajectory publication based on differential privacy[J]. ACM Transactions on Intelligent Systems and Technology, 2022, 13(3): 1-21.

[13] 吴万青, 赵永新, 王巧, 等. 一种满足差分隐私的轨迹数据安全存储和发布方法[J]. 计算机研究与发展, 2021, 58(11): 2430-2443.
WU W Q, ZHAO Y X, WANG Q, et al. Safe storage and release method of trajectory data satisfying differential privacy[J]. Journal of Computer Research and Development, 2021, 58(11): 2430-2443.

[14] MONTJOYE Y A D, HIDALGO C A, VERLEYSEN M, et al. Unique in the crowd: the privacy bounds of human mobility[J]. Scientific Reports, 2013, 3: 1376.

[15] GAMBS S, KILLIJIAN M O, CORTEZ M N D P. De-anonymization attack on geolocated data[J]. Journal of Computer and System Sciences, 2014, 80(8): 1597-1614.

[16] XU F L, TU Z, LI Y, et al. Trajectory recovery from ash: user privacy is NOT preserved in aggregated mobility data[C]//Proceedings of the 26th International Conference on World Wide Web. Piscataway: IEEE Press, 2017: 1241-1250.

[17] D'SILVA N, SHAHI T, HUSVEG Ø T D, et al. Demystifying trajectory recovery from ash: an open-source evaluation and enhancement[C]//Proceedings of the 2024 17th International Conference on Security of Information and Networks (SIN). Piscataway: IEEE Press, 2024: 1-8.

[18] SHAO M L, LI J X, YAN Q B, et al. Structured sparsity model based trajectory tracking using private location data release[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(6): 2983-2995.

[19] BUCHHOLZ E, ABUADBBA A, WANG S, et al. Reconstruction attack on differential private trajectory protection mechanisms[C]//Proceedings of the 38th Annual Computer Security Applications Conference. New York: ACM Press, 2022: 279-292.

[20] WANG Z B, SONG M K, ZHANG Z F, et al. Beyond inferring class representatives: user-level privacy leakage from federated learning[C]//Proceedings of the IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 2512-2520.

[21] HITAJ B, ATENIESE G, PEREZ-CRUZ F. Deep models under the GAN: information leakage from collaborative deep learning[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 603-618.

[22] ARIYARATHNA T, MOHOMMADY M, PAIK H Y, et al. DeepSneak: user GPS trajectory reconstruction from federated route recommendation models[J]. ACM Transactions on Intelligent Systems and Technology, 2025, 16(1): 1-22.

[23] HAN H C, YANG S Y, DING J X, et al. Adversarial reconstruction of trajectories: privacy risks and attack models in trajectory embedding[C]//Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2024: 259-269.

[24] MARIA E, BUDIMAN E, HAVILUDDIN, et al. Measure distance locating nearest public facilities using Haversine and Euclidean methods[J]. Journal of Physics: Conference Series, 2020, 1450(1): 012080.

[25] YUAN J, ZHENG Y, ZHANG C Y, et al. T-drive: driving directions

based on taxi trajectories[C]//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems.New York: ACM Press, 2010: 99-108.

[26] ZHENG Y, XIE X, MA W Y. GeoLife: a collaborative social networking service among user, location and trajectory[J]. IEEE Data Engineering Bulletin, 2010, 33(2): 32-39.

[27] JIANG K F, SHAO D X, BRESSAN S, et al. Publishing trajectories with differential privacy guarantees[C]//Proceedings of the 25th International Conference on Scientific and Statistical Database Management. New York: ACM Press, 2013: 1-12.

[28] CHEN S, FU A M, SHEN J, et al. RNN-DP: a new differential privacy scheme base on recurrent neural network for dynamic trajectory privacy protection[J]. Journal of Network and Computer Applications, 2020, 168: 102736.

[29] NIU X, HUANG H Y, LI Y T. A real-time data collection mechanism with trajectory privacy in mobile crowd-sensing[J]. IEEE Communications Letters, 2020, 24(10): 2114-2118.

[30] HANEY S, BERGHEL S, CARLSON B, et al. SafeTab-P: disclosure avoidance for the 2020 census detailed demographic and housing characteristics file A (Detailed DHC-A) [J]. arXiv Preprint, arXiv: 2505.01472, 2025.

[作者简介]

谢丽霞（1974– ），女，重庆人，中国民航大学教授，主要研究方向为网络信息安全、网络安全态势分析与评估、软件漏洞分析。

赵尔康（2001– ），男，河南许昌人，中国民航大学硕士生，主要研究方向为网络与信息安全。

杨宏宇（1969– ），男，吉林长春人，博士，中国民航大学教授、博士生导师，主要研究方向为网络与系统安全、软件安全检测、网络安全态势感知。

刘哲理（1978– ），男，山东潍坊人，博士，南开大学教授、博士生导师，主要研究方向为基于密码学的隐私保护、密文数据库、差分隐私。

赵永新（1995– ），男，河南濮阳人，天津理工大学博士生，主要研究方向为网络信息安全、隐私计算、深度学习。