

New algorithm to ensure virtual simulation data security based on deep learning using applied innovation design

Jiao Bo ^{1,2*}, Manual Selvaraj Bexci ¹

¹ Faculty of Social Science, Art & Humanities, Lincoln University College, Malaysia

² Faculty of Art, Zhengzhou Business University, China

ABSTRACT

The virtual simulation laboratory saves the material resources and equipment investment required for simulating a real experiment environment. This paves the learners to experiment and explore in a virtual environment, reducing resource waste and cost. In addition, the virtual simulation laboratory can also realize the sharing of resources, and academic institutions can share the platform and content of the virtual laboratory to improve the efficiency of resource utilization. But the virtual simulation experiment data can be easily hacked from the network, hence making it a challenging task to study virtual simulation data security. In this paper, we research the virtual simulation data security based on deep learning through applied innovation design and proposed a new algorithm. The minimum violation sequence set in the virtual simulation data set is identified and the suppression mode of the minimum violation sequence is judged. The score table is constructed for the instances in the sequence, and the corresponding instances are selected and suppressed according to the score value. The cross-attention module of Transformer Structure is proposed to aggregate the global and local feature information between left and right graphs and obtain the long-distance dependence relationship between left and right graphs along the polar direction, which can more effectively fuse the global feature information of left and right graphs. The results show that the proposed algorithm can not only ensure the safety of trajectory data but also improve the availability of data.

Keywords: Applied innovation design, Deep learning, Virtual simulation, Data security.

OPEN ACCESS

Received: September 2, 2023


Revised: September 13, 2023

Accepted: October 4, 2023

Corresponding Author:

Jiao Bo

jiao.phdscholar@lincoln.edu.my

 **Copyright:** The Author(s). This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted distribution provided the original author and source are cited.

Publisher:

[Chaoyang University of Technology](https://www.chaoyang.edu.cn/)

ISSN: 1727-2394 (Print)

ISSN: 1727-7841 (Online)

1. INTRODUCTION

With the rapid development of information technology and the coming of the era of big data, data privacy becomes particularly important. The privacy protection technology of data release has been widely concerned by academia and industry. Differential Privacy (Yin et al., 2021; Adnan et al., 2022; Zhao et al., 2022) is a new privacy protection technology emerging in recent years, which solves the defects of the traditional privacy protection model. Its basic idea is to use stochastic algorithm to disturb the query operation results of the original data set, so as to achieve the privacy protection effect. The advantage of Differential Privacy Protection technology is that it no longer describes the background knowledge of the attacker quantitatively, but directly gives the assumption that the attacker has all the background knowledge (Tseng and Zhang, 2023). Assume that an attacker attacks a record and knows all records except this record. After analyzing the query result, the attacker can still ensure that the record is not leaked (Jia et al., 2021; Cretu et al., 2022).

With the rapid development of GPS, Wi-Fi and other positioning technologies and storage technologies, a large number of mobile users' trajectory data has been collected and stored (Zhang et al., 2023). Trajectory data contains rich temporal and spatial information, and in-depth research and analysis of trajectory data has become a research

hotspot in the field of data mining (Kesu and RamAngu, 2023; Meng et al., 2023). Trajectory data with high availability is the basis of effective trajectory data mining. Researchers can obtain valuable information through the analysis and mining of trajectory data (Hamid et al., 2021; Hu et al., 2021). However, trajectory data also contains a large amount of privacy information of mobile users. If trajectory data is not protected before release, attackers with background knowledge can infer the privacy information of users by analyzing trajectory data, such as physical condition, living habits and home address, etc., and even bring economic losses and personal safety problems to users (Dhinakaran and Joe, 2022). Therefore, how to ensure that the published trajectory data will not disclose the privacy of users and have high availability is an urgent problem to be solved (Bag et al., 2021; Birjali and Kasri, 2021).

This paper is organized as follows. In Section 2, the related works are introduced. We detailed state the proposed method in this paper in Section 3. The results and discussion are shown in Section 4. There is a conclusion in Section 5.

2. RELATED RESEARCH

When accessing published data, even if the attacker has the background knowledge obtained from other channels, the attacker cannot get any additional information about the target (Irazoqui et al., 2014), which is the basic idea of data privacy protection. Under this goal, many privacy protection models and specific methods have emerged. Traditional privacy protection techniques are based on grouping, including K-anonymity (Ren et al., 2023), L-diversity (Li et al., 2023), t-close (Soria-Comas et al., 2015) and some derivative methods (Kerestes et al., 2021). The basic idea is to anonymize and hide all records by aligning identifiers (attributes related to the background knowledge of the attacker), so that all records are divided into several equivalence classes, and thus realize a record hiding in a group of records. But they are all based on assumptions about the attacker's capabilities and background knowledge. Therefore, the above model cannot provide a guarantee of sufficient security, but it leads to many other ideas, such as privacy metrics (Yin et al., 2018). Differential privacy protection technology is proposed in (Soria-Comas et al., 2017). Its specific implementation algorithm was $f(D)$ obtained by arbitrary query f of data set D , and random algorithm M added noise x on the basis of $f(D)$, and x followed a certain distribution (e.g., Laplace distribution). It was proved that this algorithm satisfied the differential privacy definition, and finally returned $f(D) + x$ to the user.

Suppose the attacker already knows everyone's information except Alice's diagnosis, and the attacker wants to obtain Alice's diagnosis, so he/she issues a query request f to the medical data set shown in Table 1:

Table 1 indicates a medical data set. The attacker already knows everyone's information except Alice, so the attacker knows that the query above outputs 3 or 4. Differential

privacy algorithm M adds a noise x on the basis of the query output of 4, assuming $x = -0.7$, then returns 3.3 to the attacker. For the attacker, 3.3 makes it difficult to determine whether the output is 3 or 4, thus ensuring Alice's privacy.

Table 1. Medical data set D

Name	Age	Diagnostic results
Heshei	21	0
Shahid	22	1
Bitaf	35	1
Kokor	39	0
Tom	48	1
Alice	45	1

The core of differential privacy protection lies in the selection of parameter ϵ and the design of stochastic algorithm. In terms of random algorithm design, different types of problems can have different implementation mechanisms, the most basic protection mechanisms are Laplacian mechanism (Li et al., 2019), exponential mechanism (Gopi et al., 2022). The former is an algorithm that realizes differential privacy protection by disturbing the real output value of noise generated by Laplacian distribution, and mainly deals with some numerical data with output results. The latter mainly deals with some algorithms whose output results are non-numerical. In the interactive environment, Zhang et al. (2022) improved the traditional Laplacian mechanism, which could provide more queries under the same budget compared with the Laplacian mechanism. In a non-interactive environment, Brauwers and Frascar (2023) put forward the concept of matrix mechanism, which represented interrelated queries into a matrix, thus reducing the amount of noise added, but its efficiency and optimization effect were not ideal. Xiao et al. (2011) first implemented Haar wavelet transform on data and then adds noise, which reduced data availability and improves query accuracy. Soria-Comas et al. (2014) proposed a differential privacy protection method based on hierarchical summation and least squares to divide the query sequence into groups that met the consistency constraint, and noise was added to each group. These achievements are included in the category of matrix mechanism. Li et al. (2021) improved the above method, proposed a low-rank matrix mechanism, and adopted the method of decomposing load matrix to optimize its strategy. In terms of the selection of parameter ϵ , Jagielski et al. (2018) proposed an attack model, which provided an upper bound for the selection of parameter ϵ , but did not provide an attack algorithm. From the perspective of economics, Jacobs et al. (2022) proposed a simple economic model, which enabled users to select parameters in a principled manner. Mahawaga et al. (2020) proposed an attack algorithm of differential privacy protection technology and gave the upper bound of parameter ϵ selection. To sum up, the importance of selecting parameter ϵ is self-evident.

3. RESEARCH METHOD

Suppose the dataset D has n tuples, and the attacker knows all the background information except the sensitive information of the target. Therefore, there are a total of n possibilities for any $n-1$ tuples in dataset D to form dataset D' ($|D'| = |D| - 1$). Here, the data set D' is denoted Ψ , called the potential input set, and its size is $n = |\Psi|$.

In the scheme of Li et al. (2021), an attack model is proposed, which is based on the prior of the attacker (assuming that the average distribution is satisfied, i.e., $\rho > \frac{1}{n}$, otherwise it has no meaning) and a posterior probability deduce that the upper bound of parameter ε satisfies:

$$\varepsilon \leq \frac{\Delta f}{\Delta v} \ln\left(\frac{(1-n)\rho}{1-\rho}\right) \quad (1)$$

$$\Delta v = \max_{1 \leq i, j \leq n} |f(D'_1) - f(D'_2)| \quad (2)$$

Where ρ represents the probability that the attacker pushes the attack object in or out of the output result set. n represents the size of the potential input set. It can be seen from this inequality that when n is large, the value of the parameter ε is large.

In this section, we consider how to guess the true value of multiple attacks on the same query in a worst-case scenario (only 2 potential input sets), so as to know whether the attack object is in the data set. The data set owner obtains a result $f(D)$ based on the query request f made by the attacker against the attack object, and then returns its $f(D)$ with noise x to the attacker. In this way, the attacker performs N attacks and gets N results $f(D) + x_1, f(D) + x_2, \dots, f(D) + x_N$, and infer whether the attack object is in the data set based on N results.

Since noise x is random, it is impossible for an attacker to accurately guess the specific value of x in each query. But as long as the attacker can guess in which interval x can fall, it is enough for the attacker to make some decisions. For example, a count query, as long as the noise x falls between $[-0.5, 0.5]$, an attacker can make an accurate judgment.

Since noise x follows the *Laplace*(μ, b) distribution, an attacker can calculate the probability of x falling in a certain interval if μ and b are known. The location parameter μ has no effect on the attacker, while the scale parameter $b = \frac{\Delta f}{\varepsilon}$ directly affects the difficulty of the attack. Under a given query problem, with the increase of b , the probability of noise x falling into the fault-tolerant interval of the query becomes smaller and smaller, and the attacker's attack becomes more and more difficult. Therefore, the selection of parameter ε can reflect the above phenomenon.

In the feature extraction stage, the algorithm adopts ResNet-40 structure (Nam et al., 2023) to down-sample the left and right graphs three times. Each sub-sampling module contains 3, 4 and 6 residual modules, which can obtain the feature maps of 1/3, 1/6 and 1/12 of the original image. The

number of channels is 64, 128 and 256 respectively. Then the global average pooling is applied to the final output feature map, and finally the channel attention weight of the multi-scale feature map is calculated by two 1×1 convolution, which is used to guide the multi-scale fusion in the feature fusion stage.

The 3×3 convolution of residual modules in all scales is replaced by context attention (COA). COA can capture rich static and dynamic context information at the same time, make full use of dynamic and static context information between input keys to guide the learning of dynamic attention matrix, and enhance the representation ability of feature graphs. Because COA is computationally similar to standard 3×3 convolution, the COA module has a similar number of parameters and floating-point calculations to ResNet-40.

In COA, assuming the input 2D feature map $X \in R^{H \times W \times C}$, Key, Query, and Value are defined as $K = X, Q = X$ and $V = XW_V$, respectively. COA first uses $k \times k$ group convolution for all the neighborhood keys in the $k \times k$ grid, and the learned context Key $K^1 \in R^{H \times W \times C}$ is the static context information between the neighborhood keys. Then, the context keys Q are superimposed by 2, and the W_θ and W_δ are convolved by 2 consecutive 1×1 activation functions with ReLU and without activation functions. It gets a dynamic multi-head attention matrix:

$$A = [K^1, Q]W_\theta W_\delta \quad (3)$$

For each head in the multi-head dynamic attention matrix, A first learns the local attention matrix for each spatial position based on Query and context Key, and then multiplicities the COA matrix A and V to calculate the participation feature graph K^2 :

$$K^2 = V \otimes A \quad (4)$$

Finally, the feature mapping K^2 , which captures the dynamic feature interaction between inputs, is called dynamic context. K^2 is then fused with the static context K^1 by the attention mechanism (Yin et al., 2020) as the output of COA.

The attention-guided multi-scale fusion module aggregates the feature information of the three scales in parallel, then recalibrates the feature map at each scale, and outputs the feature map according to the original three scales (1/3, 1/6, 1/12 of the original image).

The existing research results show that, in the estimation of parallax map, the channel weights of feature map at 1/3 scale of the original image are usually fixed, and the channel weights of feature map at 1/6 and 1/12 scale of the original image are usually more specific, and the parallax distributions of different images have different influences on the channel weights of feature map. This shows the importance of recalibrating each feature channel by attention in parallel aggregation of multi-scale feature maps.

By recalibrating the channel weights of feature maps across scales, the algorithm's ability to selectively identify information features and focus on significant features can be enhanced, more comprehensive and effective features can be extracted, and matching errors can be reduced.

In this paper, the algorithm proposes three parallel aggregation modules for the feature graphs of these three scales. The aggregation modules are defined as follows:

$$\hat{F}^s = \sum_{k=1}^s f_k(F^k), s = 1, 2, \dots, S \quad (5)$$

Where S is the number of levels of feature mapping ($S = 3$ in this paper). F^k is the feature map of grade k output in the feature extraction stage. Similar to (Lee et al., 2022), f_k computes the feature mapping according to the size relationship between k and S :

$$f_k = \begin{cases} I, k = S \\ (S - k)3 \times 3, k < S \\ 1 \times 1 \text{ conv}, k > S \end{cases} \quad (6)$$

When $k = S$, I represents the identity function. When $k < S$ is used, the $(S - k)$ convolution with step 2 is used to down-sample the feature graphs to achieve the same size. When $k > S$, bilinear up-sampling is used to achieve size consistency, and then 1×1 convolution is used to align the number of channels.

After the three scales are aggregated, channel importance is recalibrated by the channel attention module, and the recalibrated feature map \hat{F}^s can be expressed as:

$$\hat{F}^s = \varphi_s(\hat{F}^s) \cdot w_s + \hat{F}^s \quad (7)$$

Where φ_s consists of two batch normalized 3×3 convolution and ReLU. w_s is the attention weight learned from the proposed attention module.

3.1 Experimental Environment and Data Set

In this paper, the synthetic data set City80K (Chen et al., 2013) is used to test the data's utility loss. City80K is a data set that simulates the movement trajectory of 80,000 pedestrians 24 h a day in a metropolis with 26 plates. It contains five sensitive attribute values, one of which is chosen as the sensitive value in the experiment (Guo et al., 2023). All comparison algorithms are implemented in MATLAB language and run on a workstation with Intel i7-5500U CPU (3.0 GHz), 8 GB memory and 7200 RPM 1 TB hard disk. The operating system is Window 11.

3.2 Measuring Standards

Loss rate is an important parameter to measure the practicality of trajectory data set. This paper measures the practicality loss from three aspects: instance, Modified Fractal Signature (MFS) and trajectory, as follows:

In terms of instance losses, Equation (8) is used to measure:

$$\text{Instance - Loss} = \frac{N(T) - N(T')}{N(T)} \quad (8)$$

In terms of MFS losses, Equation (9) is used to measure:

$$\text{MFS - Loss} = \frac{U(T) - U(T')}{U(T)} \quad (9)$$

Where $U(T)$ is the total number of MFS in the original data set, and $U(T')$ is the total number of MFS in the data set after algorithm processing. In this paper, MAFIA algorithm (Burdick et al., 2005) is used to calculate MFS.

In terms of trajectory loss, Equation (10) is used to measure:

$$\text{Trajectory - Loss} = \frac{P(T) - P(T')}{P(T)} \quad (10)$$

Where $P(T)$ is the number of tracks in the original data set, and $P(T')$ is the number of tracks in the data set after algorithm processing.

4. RESULTS AND DISCUSSION

In order to fully study the effectiveness of the proposed algorithm, it is compared with the KCL algorithm (Liu et al., 2021; Yu et al., 2023), DLE method (Zheng et al., 2022) and BBL method (Liu and Zhang, 2023). Figs. 1–3 and Tables 2–4 show the instance loss rate, MFS loss rate and trajectory loss rate of the two algorithms under different C values, where $L = 3$, $K = 30$, $E = 800$. As can be seen from figures, the instance loss rate, MFS loss rate and track loss rate all decrease first and then tend to be stable with the increase of C value. As C value increases, the number of minimum violation sequences decreases and tends to be stable, and the number of suppressed sequences decreases and becomes stable. Therefore, the instance loss rate, MFS loss rate and track loss decrease first and then stabilize. When the C value is small, the minimum violation sequences are mainly global suppression, and the loss rates of the two algorithms are similar. With the increase of the C value, the number of minimum violation sequences decreases and the number of local suppression sequences increases, and the number of suppressed sequences decreases and becomes stable. Compared with the two algorithms, the algorithm in this paper effectively reduces the instance loss, so its loss rate is lower. Table 2 indicates the Instance – Loss comparison from the virtual simulation data. Fig. 1 is the Instance – Loss visualization results of virtual simulation data.

As can be seen from Table 2 and Fig. 1, under different C values, instance-loss can become lower than KCL method, which means that the security of virtual simulation data can be better guaranteed. When $C = 0.4$, the curve reaches the convergence state, so in the subsequent experiment, we adopt $C = 0.4$.

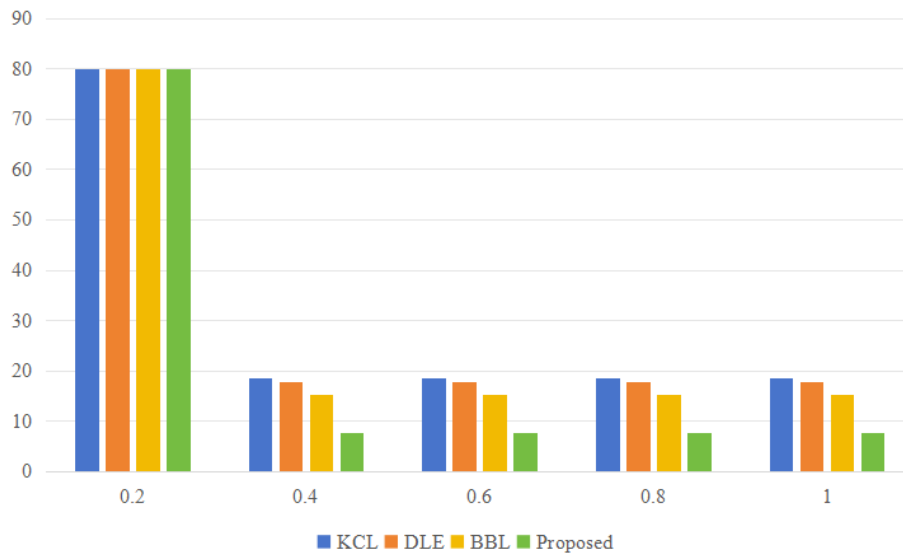


Fig. 1. Instance – Loss comparison

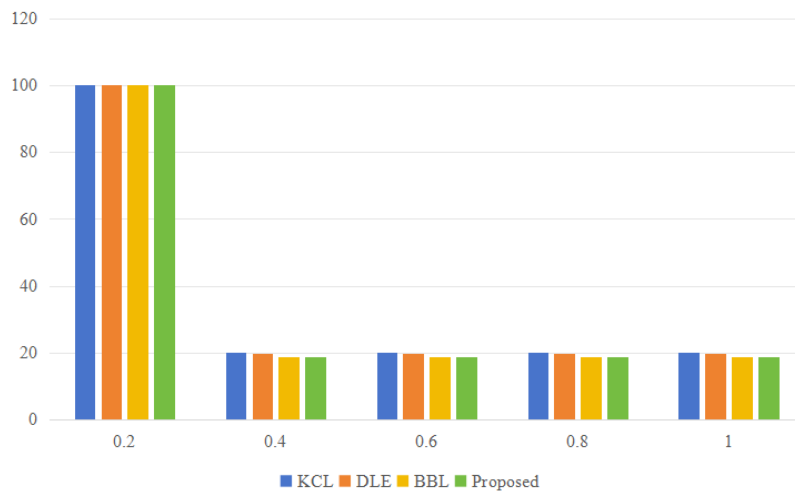


Fig. 2. MFS – Loss comparison Loss visualization results of virtual simulation data

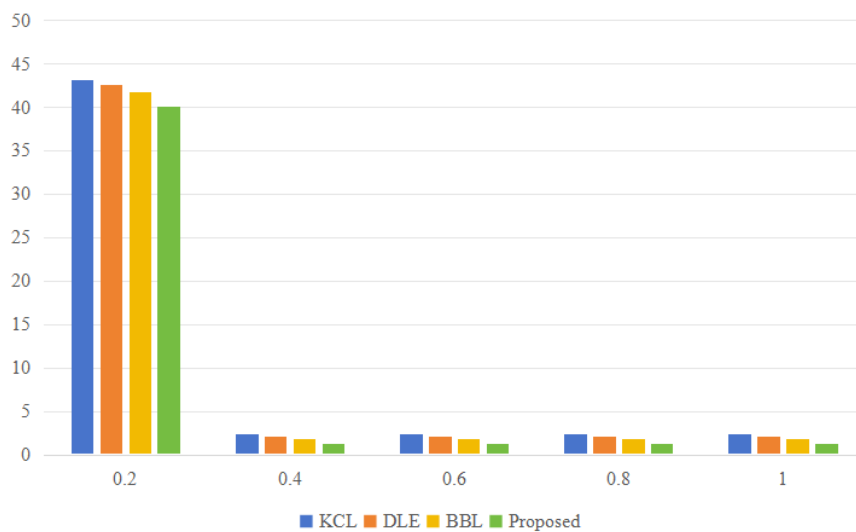


Fig. 3. Trajectory – Loss comparison

Table 2. Instance – Loss comparison (%)

C value	0.2	0.4	0.6	0.8	1.0
KCL	80.0	18.5	18.5	18.5	18.5
DLE	80.0	17.9	17.9	17.9	17.9
BBL	80.0	15.4	15.4	15.4	15.4
Proposed	80.0	7.6	7.6	7.6	7.6

Table 4. Trajectory – Loss comparison (%)

C value	0.2	0.4	0.6	0.8	1.0
KCL	43.2	2.4	2.4	2.4	2.4
DLE	42.6	2.1	2.1	2.1	2.1
BBL	41.7	1.8	1.8	1.8	1.8
Proposed	40.1	1.3	1.3	1.3	1.3

As can be seen from Tables 3 and 4 and Figs. 2 and 3, similarly, under different C values, values of MFS-loss and Trajectory-loss are also lower than that of KCL method, which signify that the security of virtual simulation data in applied innovation design can be better ensured. When C = 0.4, the curve reaches the convergence state, subsequent values do not change.

Table 3. MFS – Loss comparison (%)

C value	0.2	0.4	0.6	0.8	1.0
KCL	100.0	20.1	20.1	20.1	20.1
DLE	100.0	19.8	19.8	19.8	19.8
BBL	100.0	18.9	18.9	18.9	18.9
Proposed	100.0	18.8	18.8	18.8	18.8

Figs. 4–6 show the instance loss rate, MFS loss rate, and trajectory loss rate of the two algorithms under different K values. Where L = 3, C = 0.4, E = 800. As can be seen from Figs. 4–6, the instance loss rate, MFS loss rate and track loss rate all increase with the increase of K value. Because the increase of K value causes the increase of minimum violation sequences and global suppression sequences, the increase of suppressed sequences will correspondingly increase the data loss rate. When K value is small, the minimum violation sequence is mainly local suppression, and the loss rates of the two algorithms are similar. With the increasing of K value, the global suppression sequence and the suppressed sequence increase. Compared with the two algorithms, the proposed algorithm effectively reduces the instance loss, so its loss rate is lower.

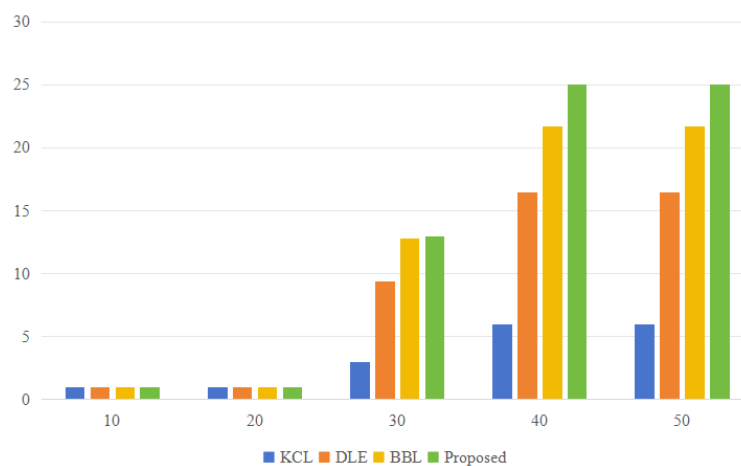


Fig. 4. Instance – Loss comparison visualization results of virtual simulation data when K value is different

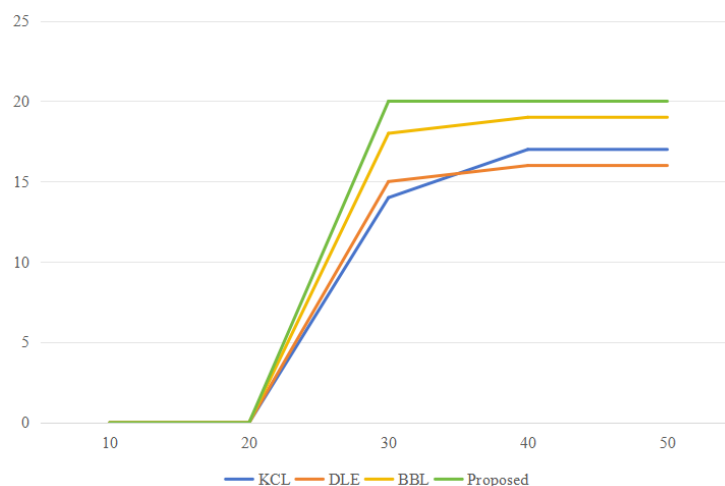


Fig. 5. MFS – Loss comparison visualization results of virtual simulation data when K value is different

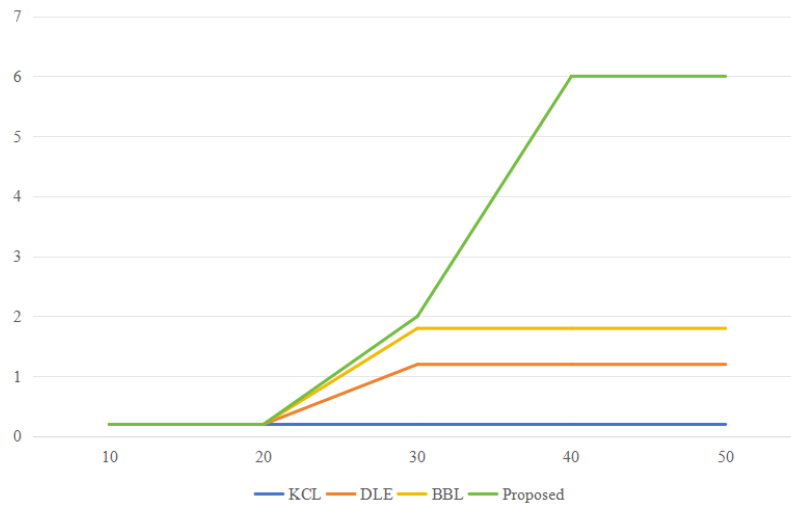


Fig. 6. Trajectory – Loss Comparison visualization results of virtual simulation data when K value is different

From Table 5 and Fig. 4, it can be seen that when $K = 10/20$, the KCL and proposed method have the same values, when K becomes bigger, the values of KCL and proposed method all become bigger too, however, their growth has been dramatic. Because the increase of K value causes the increase of minimum violation sequences and global suppression sequences, the increase of suppressed sequences will correspondingly increase the data loss rate.

Table 5. Instance – Loss comparison (%)

K value	10	20	30	40	50
KCL	1.0	1.0	3.0	6.0	6.0
DLE	1.0	1.0	9.4	16.5	16.5
BBL	1.0	1.0	12.8	21.7	21.7
Proposed	1.0	1.0	13.0	25.0	25.0

Figs. 5 and 6 and Tables 6 and 7 similarly to Fig. 4 and Table 5, the above tables and figures have the similar curve trend. From the objective analysis point of view, the increase of K value does not reduce the security guarantee of virtual simulation data. In short, the three indicators have similar trends, which shows that the method in this paper has a good effect in ensuring the security of virtual data.

Table 6. MFS – Loss comparison (%)

K value	10	20	30	40	50
KCL	0	0	14.0	17.0	17.0
DLE	0	0	15.0	16.0	16.0
BBL	0	0	18.0	19.0	19.0
Proposed	0	0	20.0	20.0	20.0

Table 7. Trajectory – Loss comparison (%)

K value	10	20	30	40	50
KCL	0.2	0.2	0.2	0.2	0.2
DLE	0.2	0.2	1.2	1.2	1.2
BBL	0.2	0.2	1.8	1.8	1.8
Proposed	0.2	0.2	2.0	6.0	6.0

From the above data results, it can be seen that the method in this paper has a good security guarantee for virtual simulation data. In this case, teachers can be assured to teach without worrying about data leakage, so as to ensure the effectiveness of student learning.

5. CONCLUSION

This paper presents a new privacy protection algorithm for virtual simulation data. It adopts local suppression instead of global suppression to realize the privacy protection of trajectory data, and combines deep learning methods to reduce the trajectory loss rate, instance loss rate and MFS loss rate. Experimental results show that the data loss rate performance of the proposed algorithm is better than other privacy protection algorithms. In the future, data availability will be further improved while ensuring the efficiency of the privacy protection algorithm.

ACKNOWLEDGMENT

Thanks to data from Lincoln University College and Zhengzhou Business University. And special thanks to Dr. Surendheran for his Valuable feedback.

DECLARATION

CONFLICT OF INTERESTS

The authors would like to declare no conflict of interest in the publication of this manuscript.

CONTRIBUTIONS

Jiao Bo conceptualized, experimented and drafted the article. Manual Selvaraj Bexci co-supervised and revised

the article for submission.

REFERENCES

- Adnan, M., Kalra, S., Cresswell, J.C., Taylor, G.W., Tizhoosh, H.R. 2022. Federated learning and differential privacy for medical image analysis. *Scientific Reports*, 12, 1953.
- Bag, S., Gupta, S., Kumar, A., Sivarajah, U. 2021. An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. *Industrial Marketing Management*, 92, 178–189.
- Birjali, M., Kasri, M., Beni-Hssane, A. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.
- Brauwers, G., Frasinca, F. 2023. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35, 3279–3298.
- Burdick, D., Calimlim, M., Flannick, J., Gehrke, J., Yiu, T. 2005. MAFIA: A maximal frequent itemset algorithm. *IEEE transactions on knowledge and data engineering*, 17, 1490–1504.
- Chen, R., Fung, M., Mohammed, N., Bipin, C., Wang, K. 2013. Privacy-preserving trajectory data publishing by local suppression. *Information Sciences*, 231, 83–97.
- Cretu, A., Houssiau, F., Cully, A., Montjoye, Y. 2022. QuerySnout: Automating the discovery of attribute inference attacks against query-based systems. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 623–637.
- Dhinakaran, D., Joe Prathap, P.M. 2022. Ensuring privacy of data and mined results of data possessor in collaborative ARM. In *Pervasive Computing and Social Networking: Proceedings of ICPCSN*, 431–444.
- Gopi, S., Lee, T., Liu, D. 2022. Private convex optimization via exponential mechanism. In *Conference on Learning Theory*. PMLR, 1948–1989.
- Guo, J., Cao, W., Nie, B., Qin, Q. 2023. Unsupervised learning composite network to reduce training cost of deep learning model for colorectal cancer diagnosis. *IEEE Journal of Translational Engineering in Health and Medicine*, 11, 54–59.
- Hamid, R.A., Albahri, A.S., Alwan, J.K., Al-Qaysi, Z.T., Albahri, O.S., Zaidan, A.A., Alnoor, A., Alamoody, A.H., Zaidan, B.B. 2021. How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. *Computer Science Review*, 39, 100337.
- Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., Gong, X. 2021. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems*, 87, 101619.
- Irazoqui, G., Inci, S., Eisenbarth, T., Sunar, B. 2014. Wait a minute! A fast, cross-VM attack on AES. In *Research in Attacks, Intrusions and Defenses: 17th International Symposium, RAID 2014*, Springer International Publishing, 299–319.
- Jacobs, G., Konrad, C., Berroth, J., Huang, M. 2022. Function-oriented model-based product development. *Design Methodology for Future Products: Data Driven, Agile and Flexible*, 243–263.
- Jagielski, M., Oprea, A., Biggio, B., Liu, C., Nita-Rotaru, C., Li, B. 2018. Manipulating machine learning: Poisoning attacks and countermeasures for regression learning. In *2018 IEEE Symposium on Security and Privacy (SP)*, 19–35.
- Jia, D., Yin, B., Huang, X. 2021. Association analysis of private information in distributed social networks based on big data. *Wireless Communications and Mobile Computing*, 2021, 1–12.
- Kerestes, C., Delafield, R., Elia, J., Chong, E., Kaneshiro, B., Soon, R. 2021. It was close enough, but it wasn't close enough: A qualitative exploration of the impact of direct-to-patient telemedicine abortion on access to abortion care. *Contraception*, 104, 67–72.
- Kesu, S., Ramasangu, H. 2023. Pressure flow dynamics in cellular automata based nephron network model. *International Journal of Applied Science and Engineering*, 20, 1–12.
- Lee, H., Choi, H., Byun, M., Chang, J. 2022. Multi-scale architecture and device-aware data-random-drop based fine-tuning method for acoustic scene classification. In *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*.
- Li, H., Li, Z., Li, K., Rellermeyer, J., Chen, L., Li, K. 2021. SGD_Tucker: A novel stochastic optimization strategy for parallel sparse tucker decomposition. *IEEE Transactions on Parallel and Distributed Systems*, 32, 1828–1841.
- Li, Q., Fu, Q., Zhu, J., Sun, Y., He, H., Hu, H. 2023. Endophytic bacteria in *Ricinus communis* L.: Diversity of bacterial community, plant– Growth promoting traits of the isolates and its effect on Cu and Cd speciation in soil. *Agronomy*, 13, 333.
- Li, X., Li, H., Zhu, H., Huang, M. 2019. The optimal upper bound of the number of queries for Laplace mechanism under differential privacy. *Information Sciences*, 503, 219–237.
- Liu, C., Zhang, Y. 2023. Advances and hotspots analysis of value stream mapping using bibliometrics. *International Journal of Lean Six Sigma*, 14, 190–208.
- Liu, Z., Jiang, D., Zhang, C., Zhao, H., Zhao, Q., Zhang, B. 2021. A novel fireworks algorithm for the protein-ligand docking on the autodock. *Mobile Networks and Applications*, 26, 657–668.
- Mahawaga, P., Bertok, P., Khalil, I., Liu, D., Camtepe, S., Atiquzzaman, M. 2022. Local differential privacy for deep learning. *IEEE Internet of Things Journal*, 7, 5827–

- 5842.
- Meng, X., Wang, X., Yin, S., Li, H. 2023. Few-shot image classification algorithm based on attention mechanism and weight fusion. *Journal of Engineering and Applied Science*, 70.
- Nam, H., Lee, C. 2023. Random image frequency aggregation dropout in image classification for deep convolutional neural networks. *Computer Vision and Image Understanding*, 232, 103684.
- Ren, W., Ghazinour, K., Lian, X. 2023. kt-safety: Graph release via k-anonymity and t-closeness. *IEEE Transactions on Knowledge and Data Engineering*, 35, 9102–9113.
- Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., Martínez, S. 2014. Enhancing data utility in differential privacy via microaggregation-based k-anonymity. *The VLDB Journal*, 23, 771–794.
- Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., Martínez, S. 2015. t-closeness through microaggregation: Strict privacy with enhanced utility preservation. *IEEE Transactions on Knowledge and Data Engineering*, 27, 3098–3110.
- Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., Megías, D. 2017. Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security*, 12, 1418–1429.
- Tseng, C., Zhang, S. 2023. Heuristics for parallel machine scheduling with GoS eligibility constraints. *International Journal of Applied Science and Engineering*, 20, 1–12.
- Xiao, X., Wang, G., Gehrke, J. 2011. Differential privacy via wavelet transforms. *IEEE Transactions on Knowledge and Data Engineering*, 23, 1200–1214.
- Yin, C., Xi, J., Sun, R., Wang, J. 2018. Location privacy protection based on differential privacy strategy for big data in industrial internet of things. *IEEE Transactions on Industrial Informatics*, 14, 3628–3636.
- Yin, S., Li, H., Liu, D., Karim, S. 2020. Active contour modal based on density-oriented BIRCH clustering method for medical image segmentation. *Multimedia Tools and Applications*, 79, 31049–31068.
- Yin, S., Li, H., Laghari, A., Karim, S., Jumani, K. 2021. A bagging strategy-based kernel extreme learning machine for complex network intrusion detection. *EAI Endorsed Transactions on Scalable Information Systems*. 21, e8.
- Yu, Q., Yang, F., Xiao, Z., Gong, S., Sun, L., Chen, C. 2023. Trajectory personalization privacy preservation method based on multi-sensitivity attribute generalization and local suppression. *Intelligent Data Analysis*, 27, 935–957.
- Zhang, D., Shafiq, M., Wang, L., Srivastava, G., Yin, S. 2023. Privacy-preserving remote sensing images recognition based on limited visual cryptography. *CAAI Transactions on Intelligence Technology*, 2023, 1–12.
- Zhang, K., Tian, J., Xiao, H., Zhao, Y., Zhao, W., Chen, J. 2022. A numerical splitting and adaptive privacy budget-allocation-based LDP mechanism for privacy preservation in blockchain-powered IoT. *IEEE Internet of Things Journal*, 10, 6733–6741.
- Zhao, Y., Chen, J. 2022. A survey on differential privacy for unstructured data content. *ACM Computing Surveys (CSUR)*, 54, 1–28.
- Zheng, S., Ren, S., Wang, J., Wang, C., Wang, Y. 2022. Design of network big data anti attack system for carbon emission measurement based on deep learning. *International Conference on Machine Learning for Cyber Security*, 279–293.