PrivImage: Differentially Private Synthetic Image Generation using Diffusion Models with Semantic-Aware Pretraining

33rd Usenix Security Symposium 2024

Presenter: Chen Gong

Kecen Li[*,1], Chen Gong[*,2], Zhixiang Li[3], Xinwen Hou[1], Tianhao Wang[2]

*Indicates Equal Contribution

[1]Chinese Academy of Sciences

[2]University of Virginia [3]University of Bristol

Privacy Leakages in Images

Training Set

Caption: Living in the light with Ann Graham Lotz

Generated Image

Prompt: Ann Graham Lotz

Carlini

It is necessary to develop safe generative models for pri image synthe

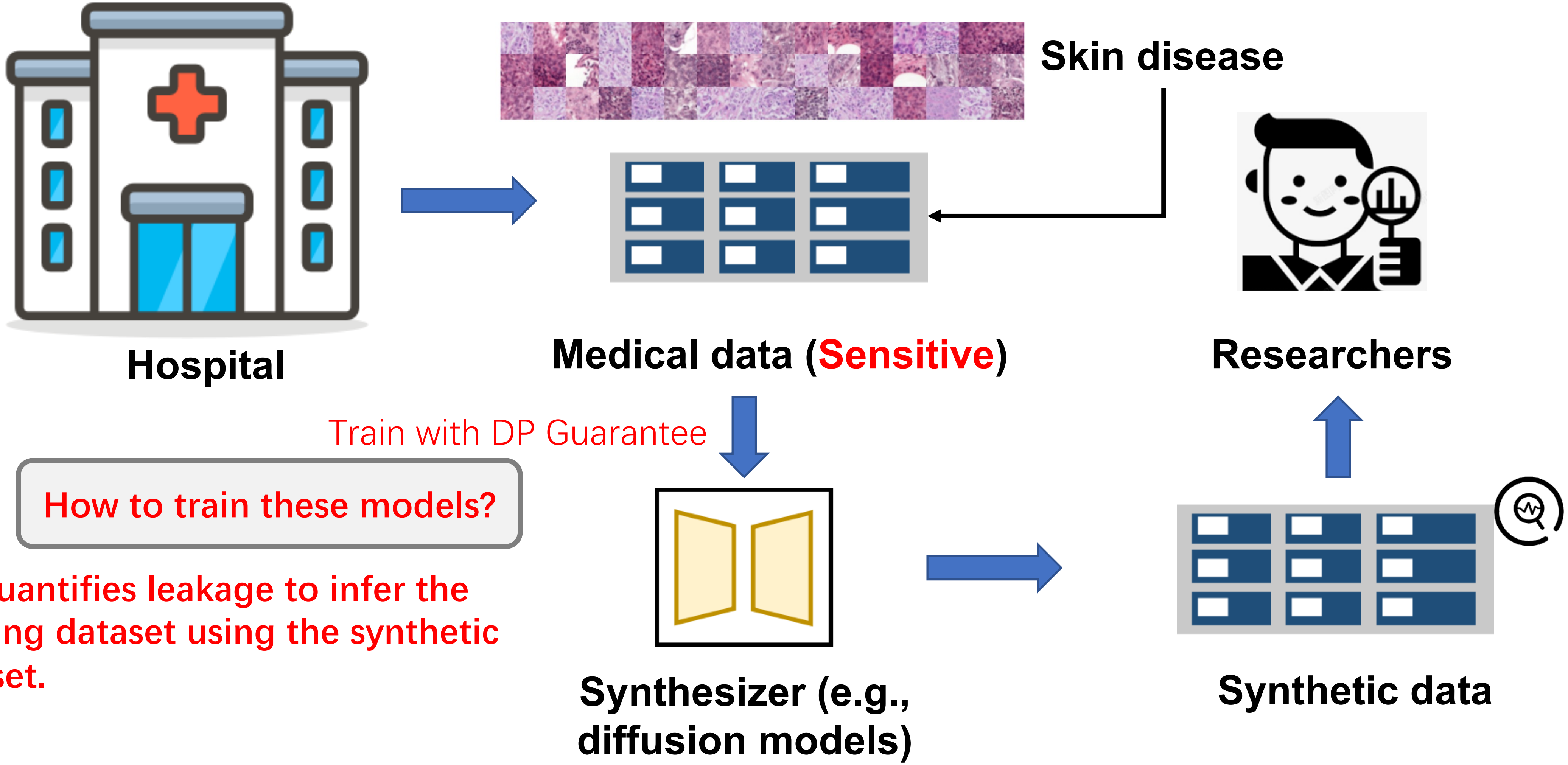
How to solve it?

Original:

Generated:

Carlini N, Hayes J, Nasr M, et al. Extracting training data from diffusion models. USENIX Security 2023.

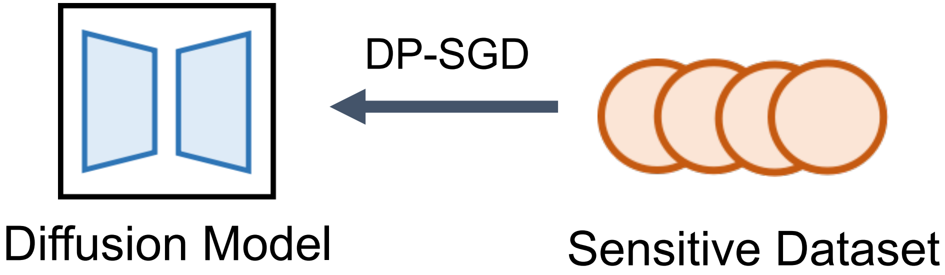
Differentially Private (DP) Image Dataset Synthesis



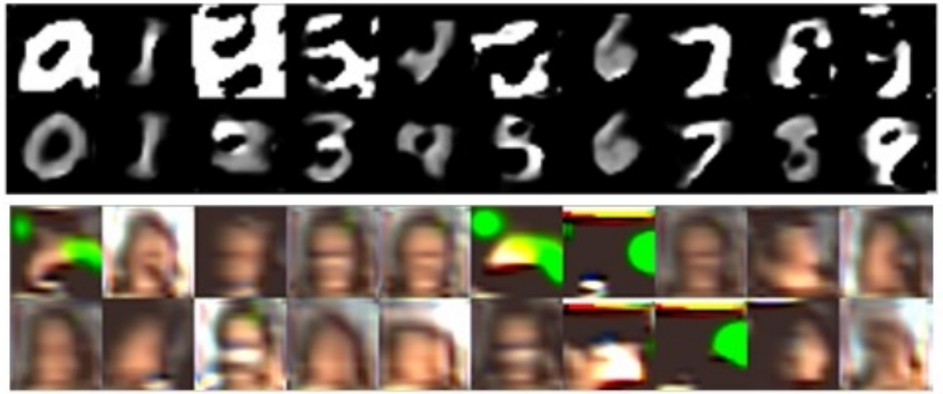
How to train these models?

DP quantifies leakage to infer the training dataset using the synthetic dataset.

Deep Generative Models + DP-SGD[a]



Struggles in Synthesizing High-Quality Images



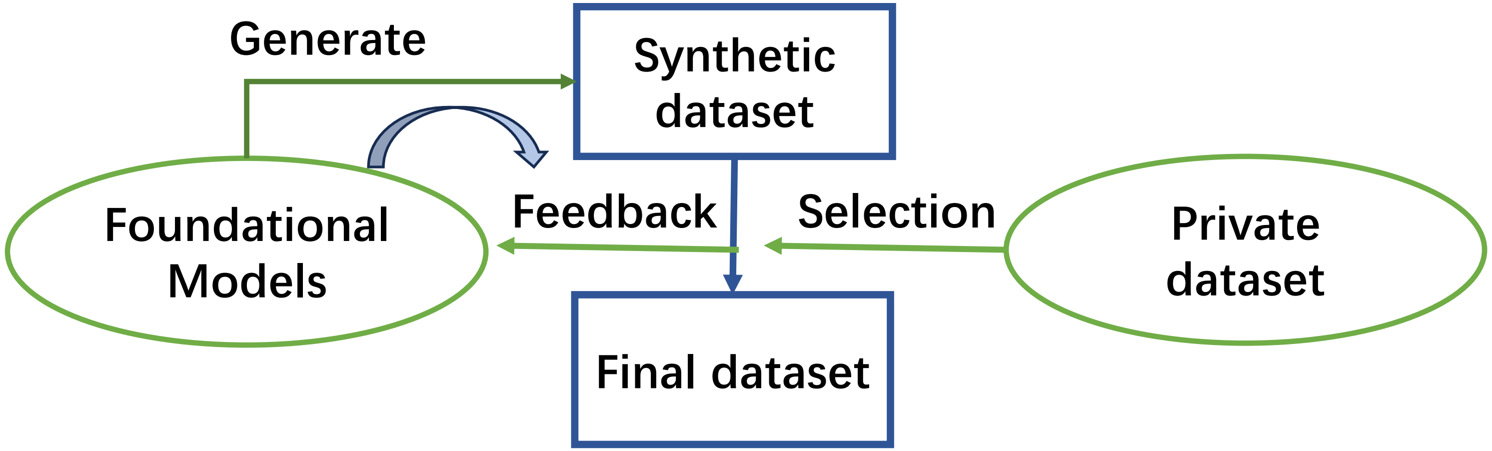
DP-MERF[b]: Differentially private mean embeddings with randomfeatures for practical privacy-preserving Data generation.

DPDM[c]: Differentially Private Diffusion Models.

[a] Abadi, Martin, et al. "Deep learning with differential privacy." CCS, 2016.
[b] Harder, Frederik, et al. "Dp-merf: Differentially private mean embeddings with randomfeatures for practical privacy-preserving data generation." AISTATS, 2021.
[c] Dockhorn, Tim, et al. "Differentially Private Diffusion Models." TMLR, 2023.

Improving DP Image Synthesis

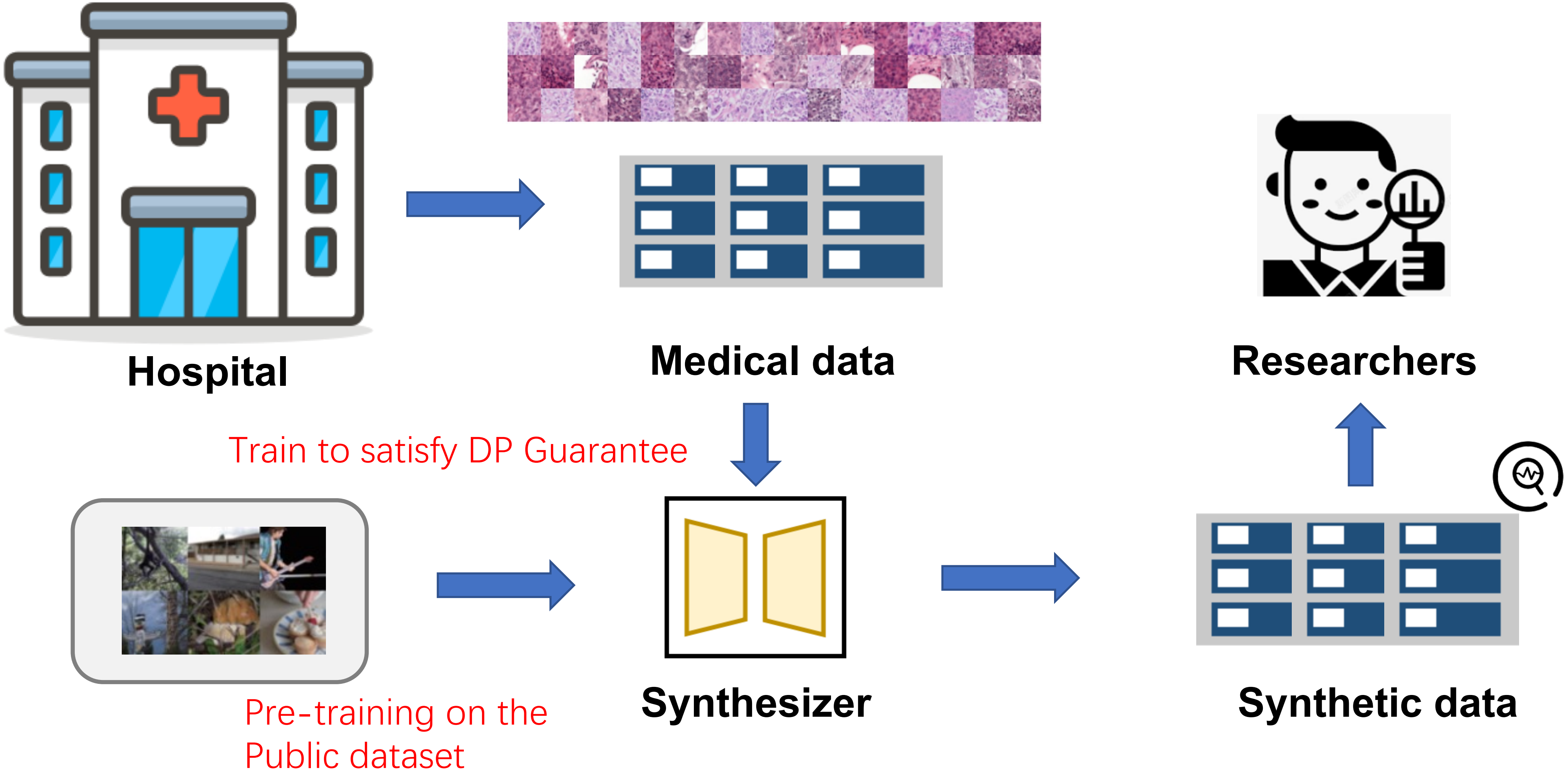
The synthesizer can already generate images that are similar to sensitive images.



DPSDA proposes a Private Evolution algorithm that progressively guides the pre-trained models to generate a synthetic image dataset similar to the sensitive one.

Lin, Zinan, et al. "Differentially Private Synthetic Data via Foundation Model APIs 1: Images." ICLR 2024.

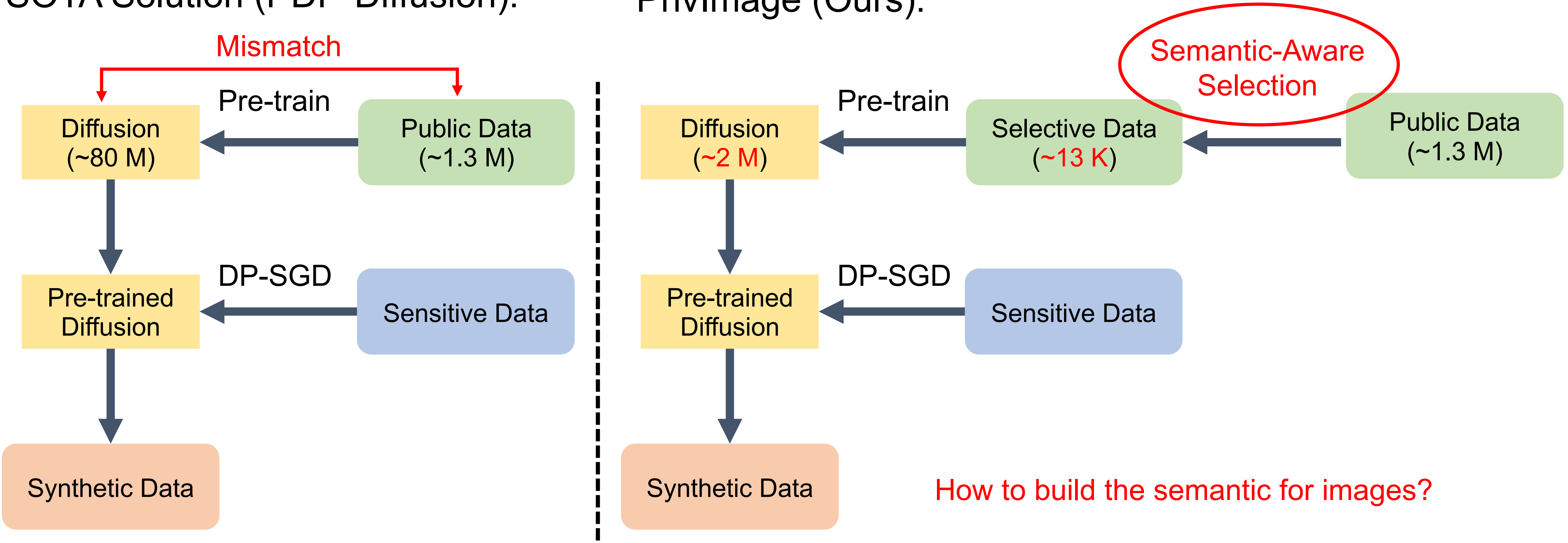
Leverage Public Dataset



Comparing Our Method With Previous Work

SOTA Solution (PDP-Diffusion):

PrivImage (Ours):



Ghalebikesabi S, Berrada L, Gowal S, et al. Differentially private diffusion models generate useful synthetic images. arXiv preprint arXiv:2302.13861, 2023.

Image Semantic

"A man on a boat holding his dog in his lap." "A man in grey shirt with camera on bench next to two dogs."

They are different at the pixel level but similar at the semantic level.

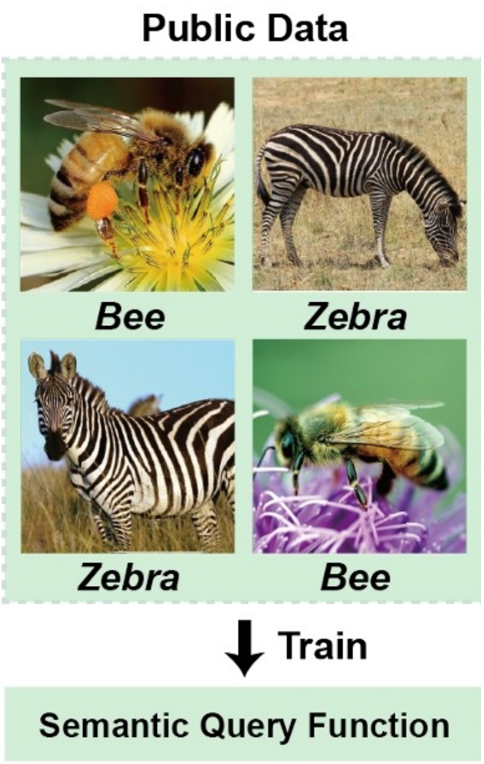
Semantics provide a high-level representation of images. Semantics capture the “meaning” of an image, which often requires higher-level processing and understanding.

If we do not have available query function, we can construct from public dataset.

Asgari Taghanaki, Saeid, et al. "Deep semantic segmentation of natural and medical images: a review." Artificial Intelligence Review 54 (2021): 137-178.

Li, Bingchen, et al. "Sed: Semantic-aware discriminator for image super-resolution." CVPR 2024.

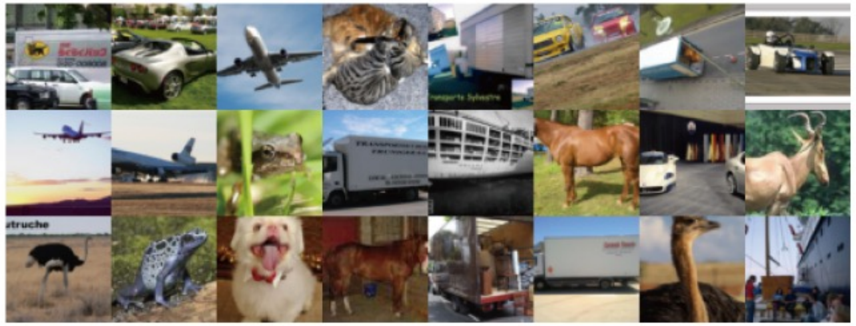
Semantic-Aware Image Selection



$$\mathcal{L}_Q := \mathbb{E}_{(x,\, s)\sim D_p}\left[\sum_{i=1}^{NS} -s_i \log Q_i(x)\right]$$

{man, dog}

0,1,0,0,1



Top-10 semantics:

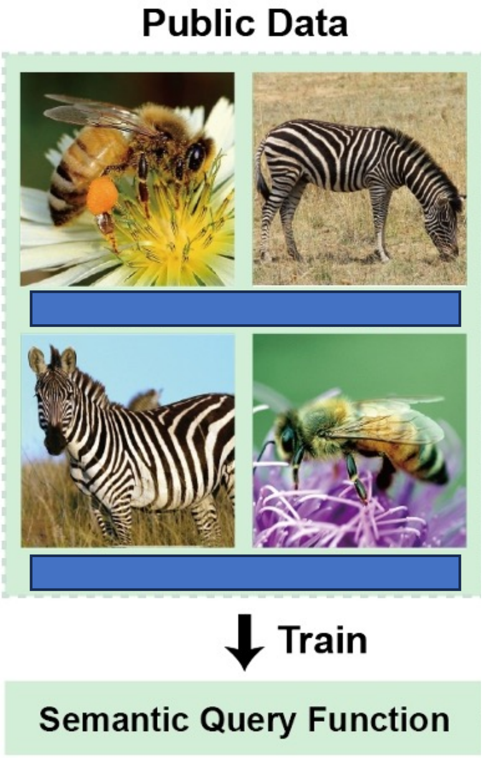
- “airline”, “sport car”, “ostrich”, “tabby”, “hartebeest”, “pekinese”, “tailed frog”, “sorrel”, “ocean liner” and “moving van”.

Step 1: Train a semantic query function. (For ImageNet, we use an 1000-categoriy image classifier.)

PrivImage: Differentially Private Synthetic Image Generation using Diffusion Models with Semantic-Aware Pretraining. Usenix Security 2024.

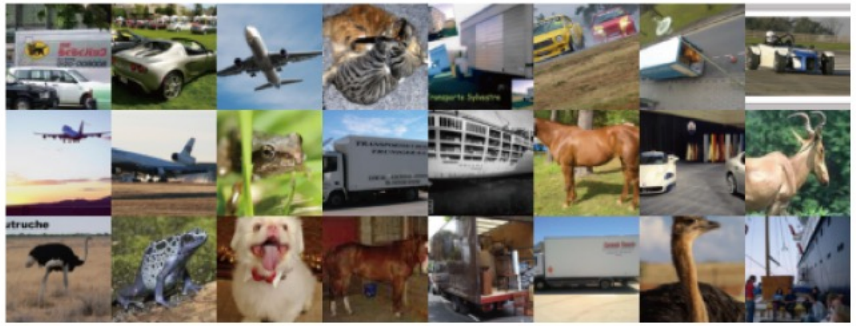
Semantic-Aware Image Selection

Contact me!!!



$$\mathcal{L}_Q := \mathbb{E}_{(x,\, s)\sim D_p}\left[\sum_{i=1}^{NS} -s_i \log Q_i\,(x)\right]$$

More than 15 co-authors helps to improve the accept possibility!!!

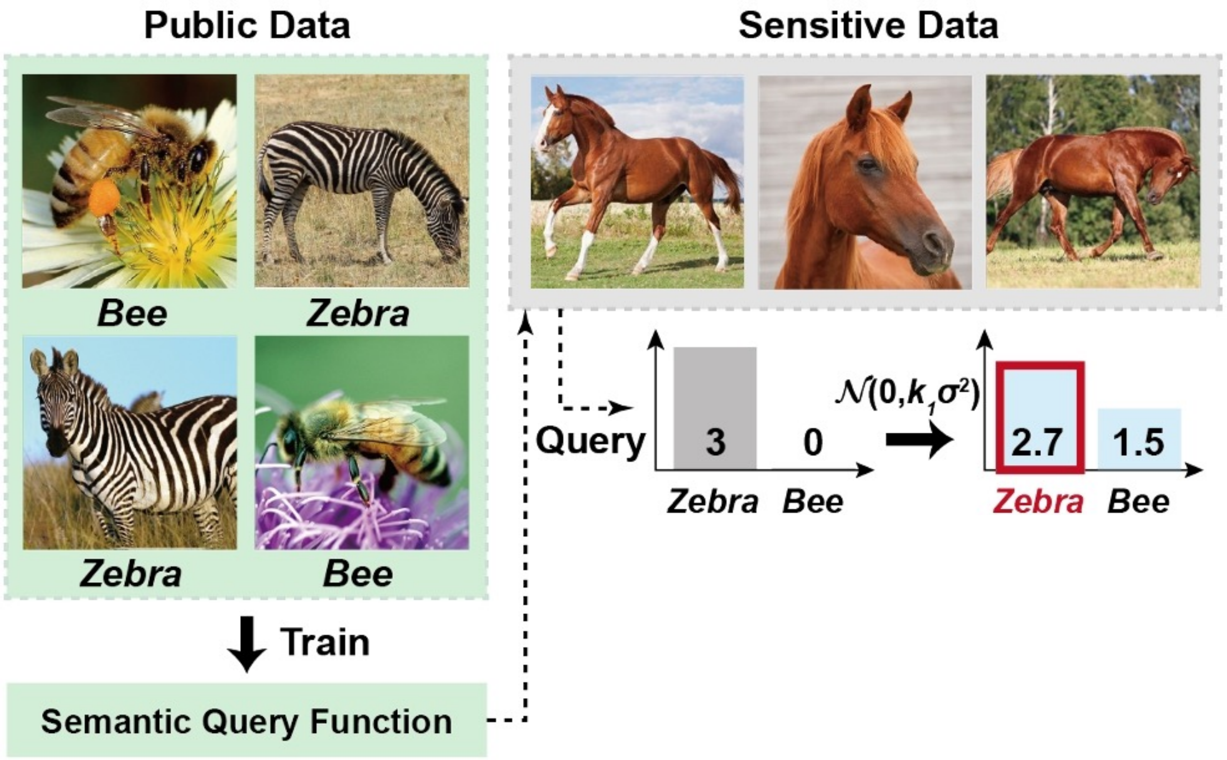


Top-10 semantics:

- “airline”, “sport car”, “ostrich”, “tabby”, “hartebeest”, “pekinese”, “tailed frog”, “sorrel”, “ocean liner” and “moving van”.

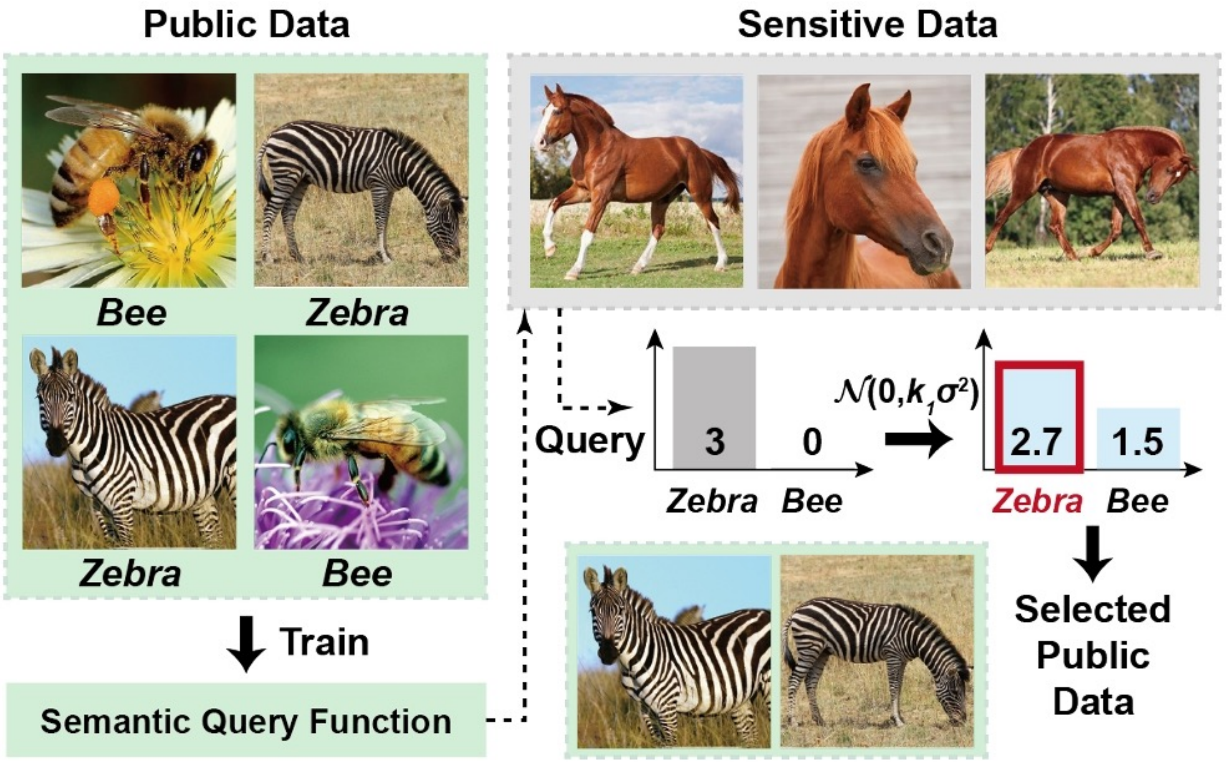
Step 1: Train a semantic query function. (For ImageNet, we use an 1000-categoriy image classifier.)

Semantic-Aware Image Selection



Step 2: Query the semantic distribution and inject Gaussian noise into the result.

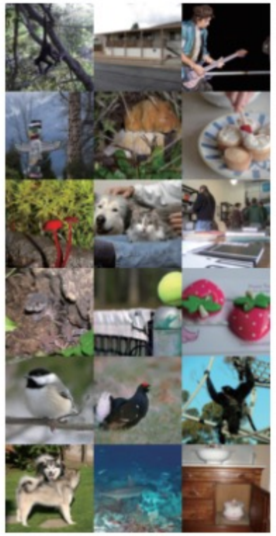
Semantic-Aware Image Selection



Step 3: Select public data which have semantics with high probability.

Investigated Datasets

Examples of Datasets

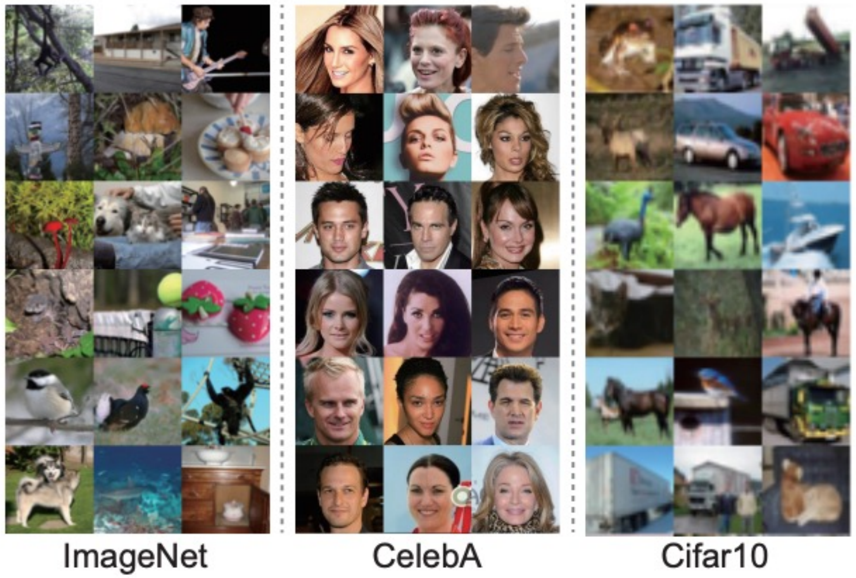


ImageNet

Tramèr, Florian, Gautam Kamath, and Nicholas Carlini. "Position: Considerations for Differentially Private Learning with Large-Scale Public Pretraining." Forty-first International Conference on Machine Learning. 2022.

Investigated Datasets

Examples of Datasets

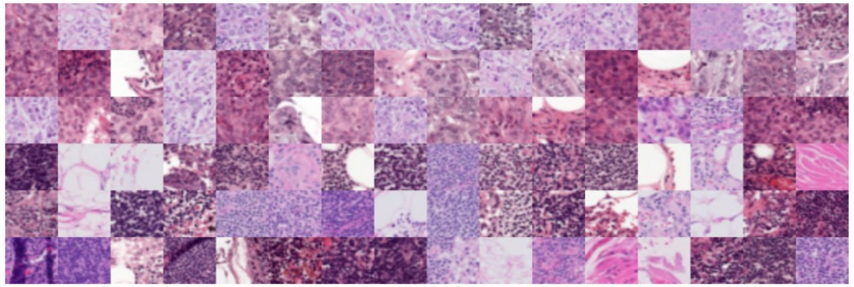


Tramèr, Florian, Gautam Kamath, and Nicholas Carlini. "Position: Considerations for Differentially Private Learning with Large-Scale Public Pretraining." Forty-first International Conference on Machine Learning. 2022.

Investigated Datasets

Examples of Datasets

ImageNet CelebA Cifar10



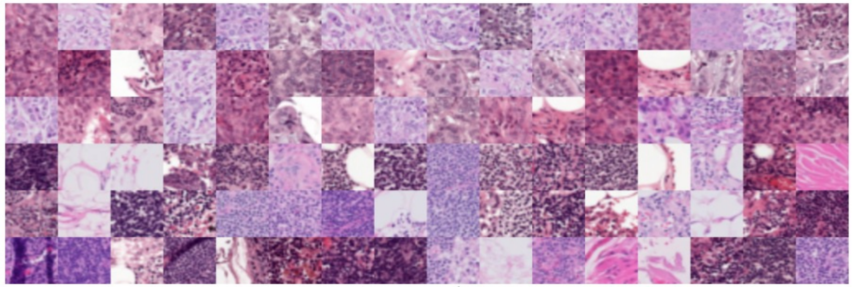
Skin Disease

Tramèr, Florian, Gautam Kamath, and Nicholas Carlini. “Position: Considerations for Differentially Private Learning with Large-Scale Public Pretraining.” Forty-first International Conference on Machine Learning. 2022.

Investigated Datasets

Examples of Datasets

ImageNet CelebA Cifar10



Skin Disease

We should not use ImageNet as pre-training dataset and Cifar10 as the sensitive dataset.

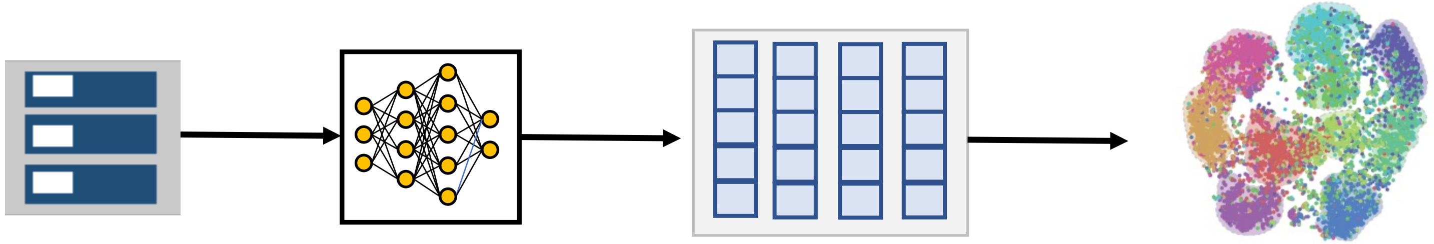
Tramèr, Florian

Gautam Kamath

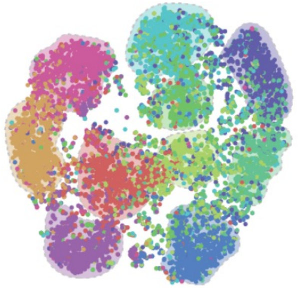
Nicholas Carlini

Tramèr, Florian, Gautam Kamath, and Nicholas Carlini. “Position: Considerations for Differentially Private Learning with Large-Scale Public Pretraining.” Forty-first International Conference on Machine Learning. 2022.

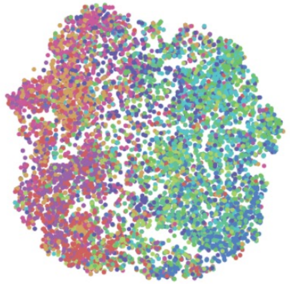
RQ1. How effective is PrivImage for synthesizing useful images?



Synthetic dataset　Classifier　Embedding　Tsne Distribution



Ours



PDP-Diffusion

A simple way to present the utility of synthetic dataset.

method.

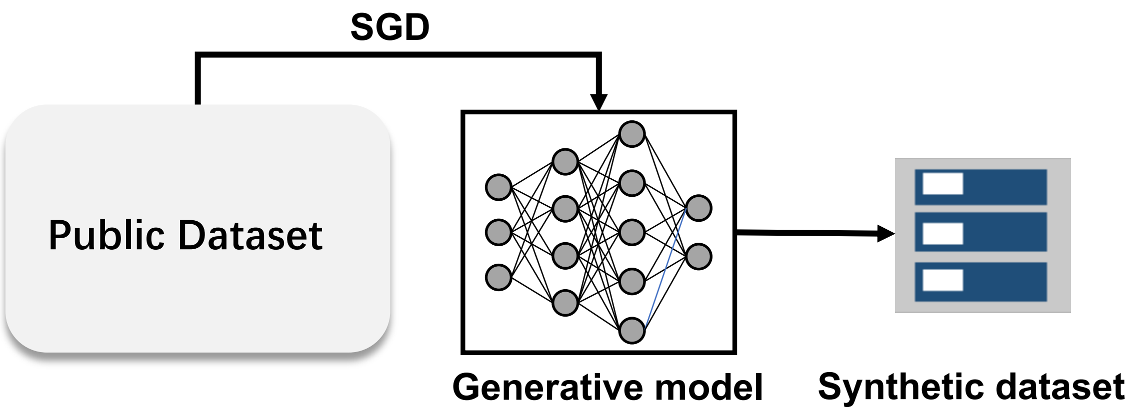
RQ1. How effective is PrivImage for synthesizing useful images?

Ours

PDP-Diffusion

Answers to RQ1: On average, the FID of the synthetic dataset is **6.8%** lower, and the CA of the downstream classification task is **13.2%** higher, compared to the state-of-the-art method.

RQ2. Is query selection useful?



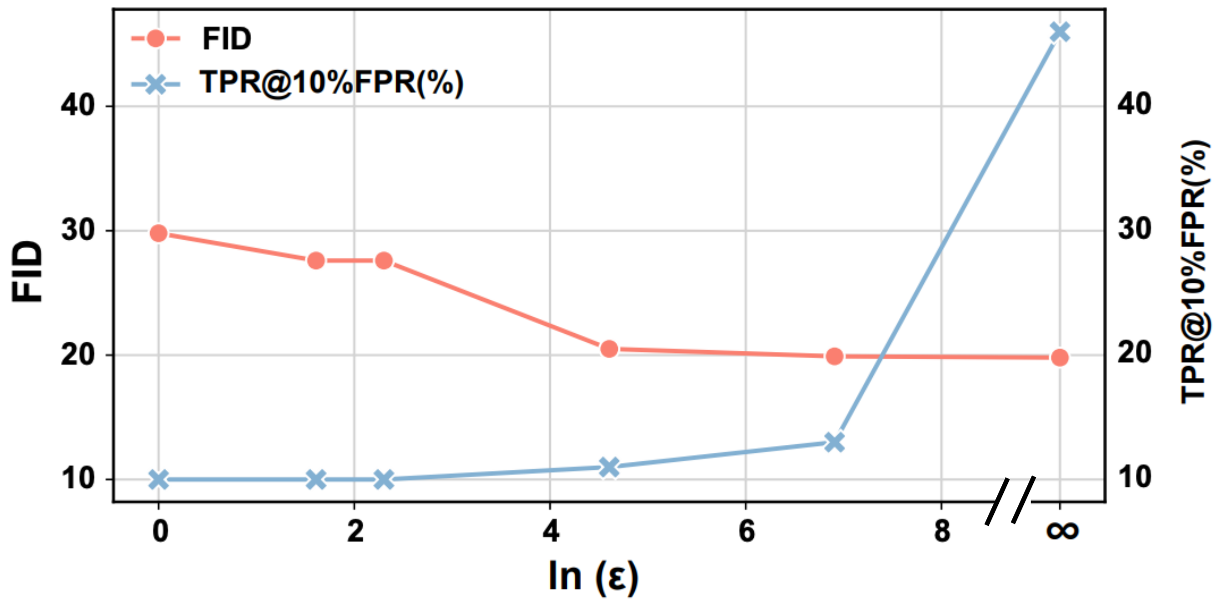
Only involved pre-training

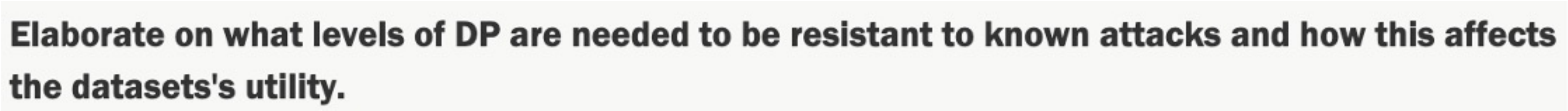
Method	CIFAR-10	CelebA32	CelebA64
PrivImage	**26.2**	**158.3**	**204.0**
PDP-Diffusion	47.3	210.8	270.3

Lower FID means the synthetic images are more similar to the sensitive images.

Answers to RQ2: Before fine-tuning, Privimage produces synthetic images with a data distribution maligned with the sensitive data. As a result, Privimage delivers enhanced DP image synthesis.

Defend Against Membership Inferences





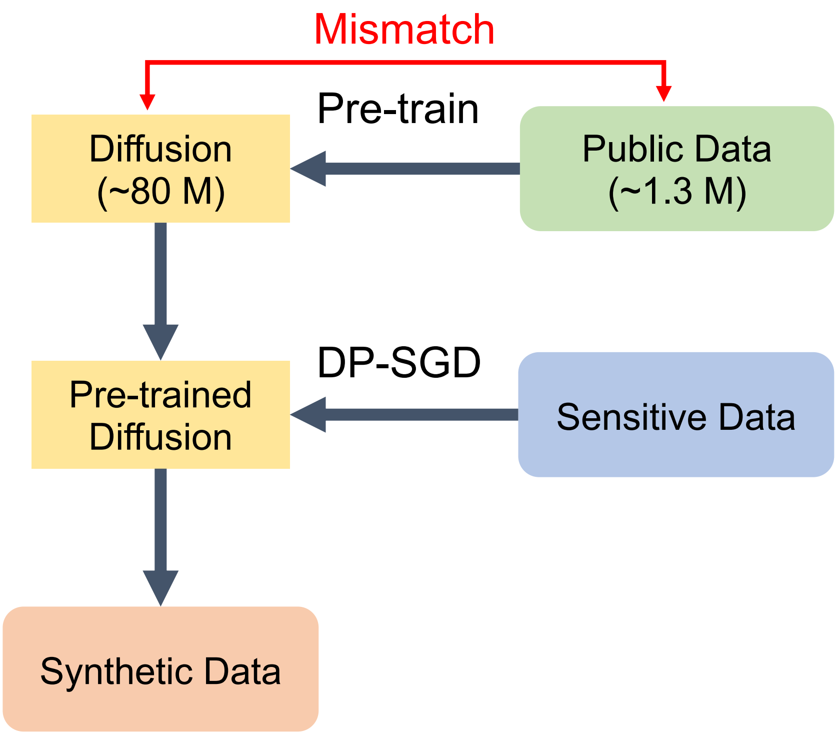
As ε increases, FID score drops and TPR@10%FPR rises.

If I want to more citations, refine my open-source repository.

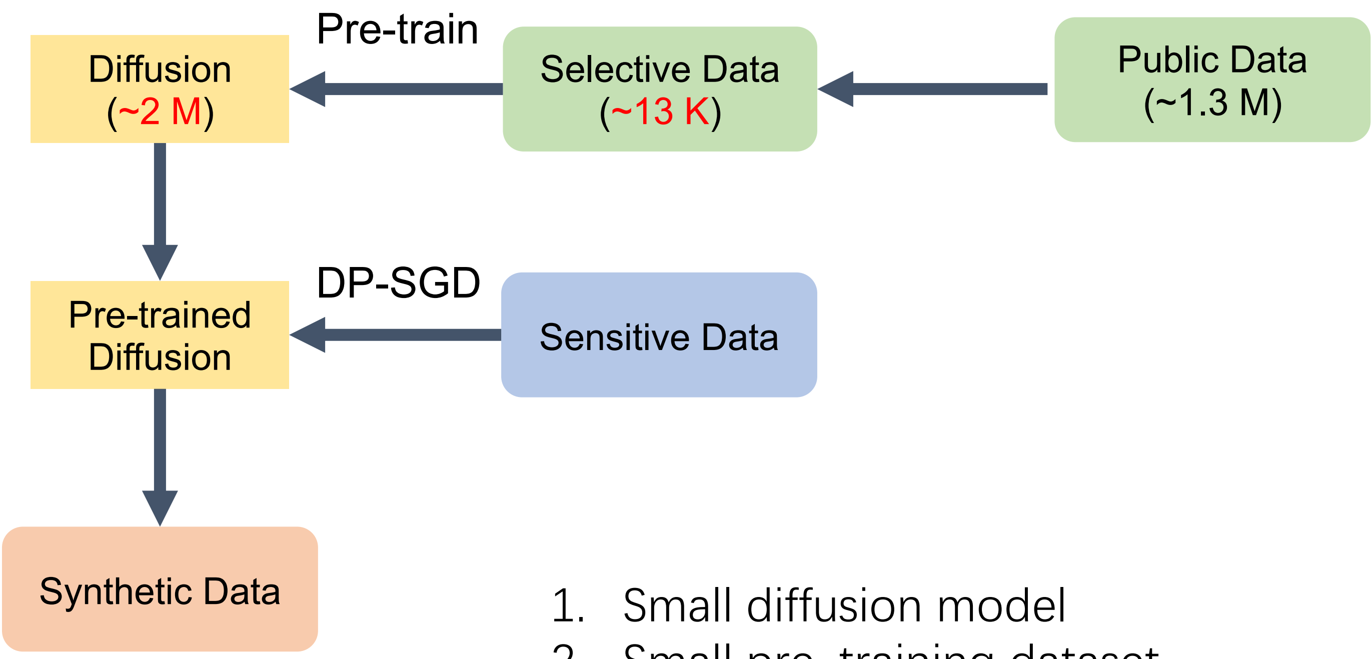
Tomoya Matsumoto, Takayuki Miura, and Naoto Yanai. Membership inference attacks against diffusion models. In 2023 IEEE Security and Privacy Workshops (SPW), pages 77–83. (The simplest method for diffusion models to replicate)

Summary

SOTA Solution (PDP-Diffusion):



PrivImage (Ours):



1. Small diffusion model
2. Small pre-training dataset

1. Large diffusion models: suffers more from DP-SGD.
2. Large pre-training dataset: huge computational cost.

Ghalebikesabi S, Berrada L, Gowal S, et al. Differentially private diffusion models generate useful synthetic images. arXiv preprint arXiv:2302.13861, 2023.

This paper proposed a DP image synthesis method using diffusion models with semantic-aware pretraining.

Hope it inspires!

Questions are welcome 😊!

chengong@virginia.edu

Paper

Artifact