Private Set Generation with Discriminative Information

Dingfan Chen, Raouf Kerkouche, Mario Fritz
CISPA Helmholtz Center for Information Security

https://github.com/DingfanChen/Private-Set

Motivation

- Privacy issues when deploying ML models in many sensitive domains (e.g., healthcare, financial)
- Can we release synthetic datasets for downstream tasks, while providing rigorous privacy guarantees?

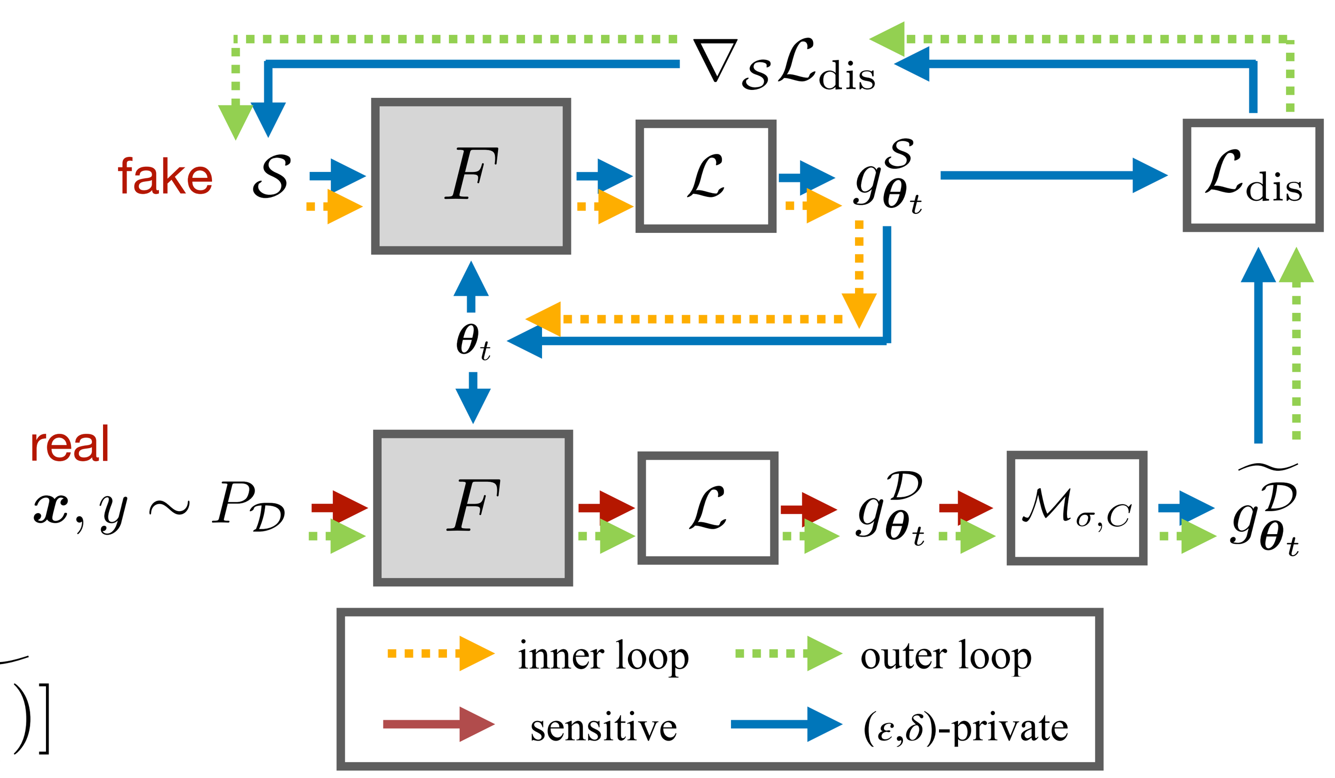
Problem: Differentially Private High-dimensional Data Generation

- **Existing approaches**
  - Aim at fitting the complete data distribution
  - Optimize deep generative models
  - Suboptimal utility: <85% for MNIST with $(\varepsilon, \delta)=(10, 10^{-5})$
- **Our approach**
  - **Generally easier:** Target at common downstream tasks (e.g., classification)
  - **Better convergence:** Directly optimize a set of representative samples
  - **Useful samples:** ~10% downstream test accuracy improvement over SOTA

Approach

- **Target:**
  - Optimize for training downstream neural network classifier
- **Basic idea:**
  - Gradient-based **coreset generation**
  - DP stochastic gradient descent (DP-SGD)
- **Objective:**

$$\mathcal{S} = \arg\min_{\mathcal{S}} \mathbb{E}_{\boldsymbol{\theta}_0 \sim P_{\boldsymbol{\theta}_0}} \sum_{t=0}^{T-1} [\mathcal{L}_{\text{dis}}(g^{\mathcal{S}}_{\boldsymbol{\theta}_t}, \widetilde{g^{\mathcal{D}}_{\boldsymbol{\theta}_t}})]$$

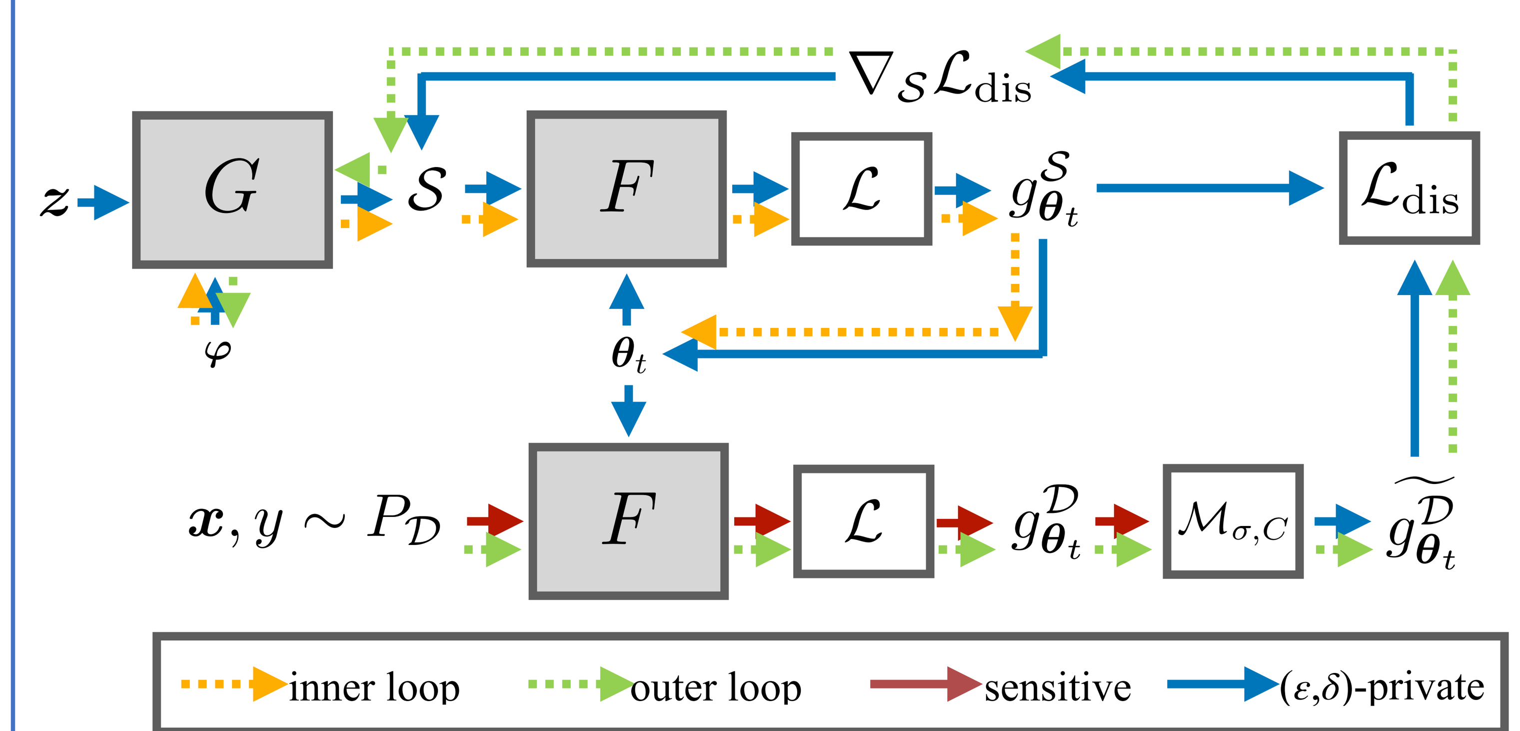


Rethinking Private Data Generation

- **Question:** Are deep generative models the best option for this task?
- **Approach:** Deep generative models as "prior"
- **Objective:**

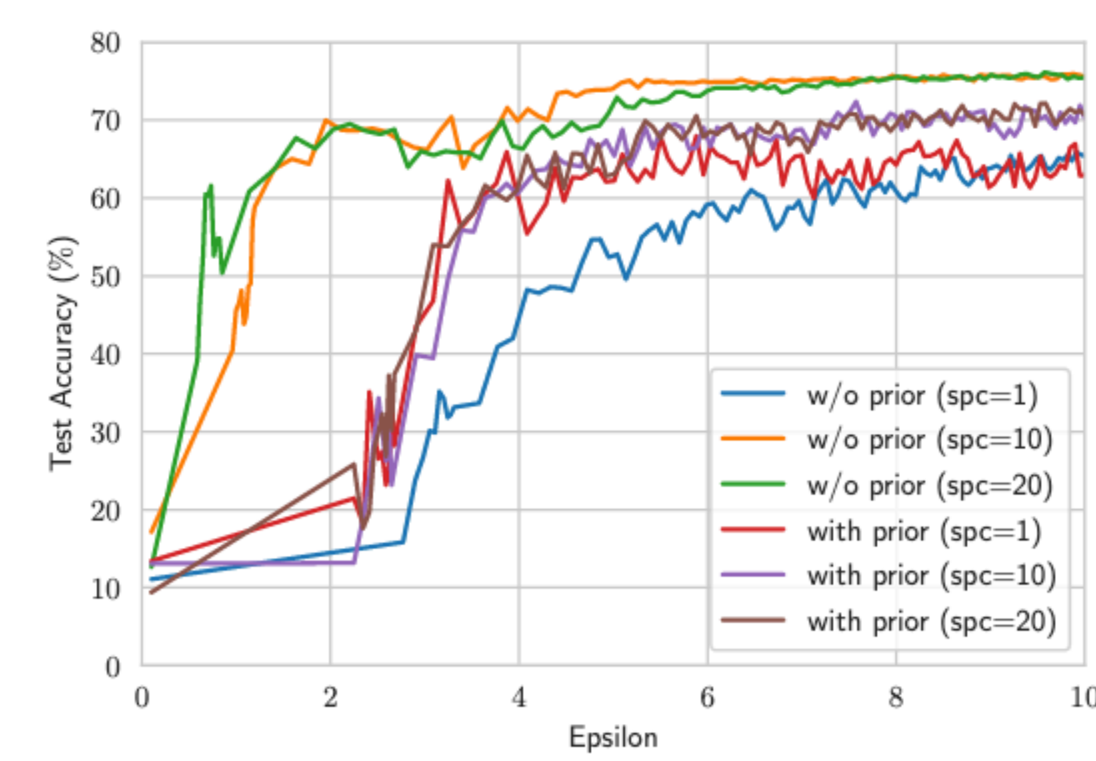
$$\min_{\boldsymbol{\varphi}} \mathbb{E}_{\boldsymbol{\theta}_0 \sim P_{\boldsymbol{\theta}_0}} \sum_{t=0}^{T-1} [\mathcal{L}_{\text{dis}}(g^{\mathcal{S}}_{\boldsymbol{\theta}_t}, \widetilde{g^{\mathcal{D}}_{\boldsymbol{\theta}_t}})]$$

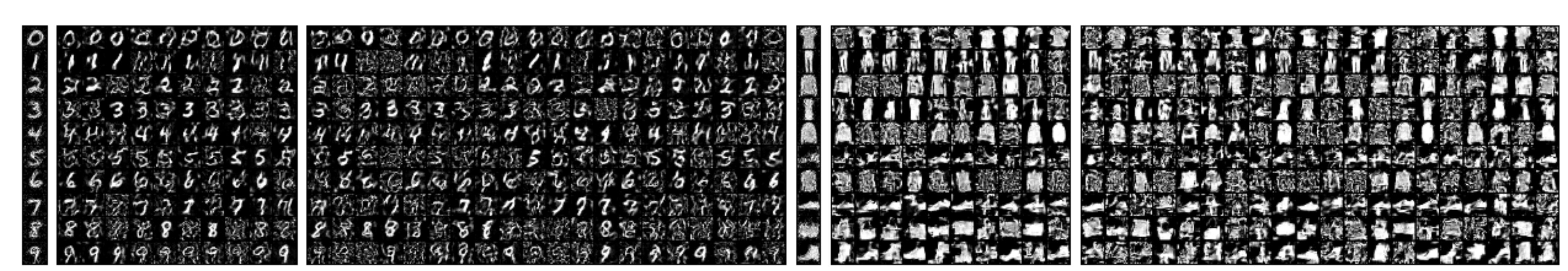
$$\text{with } \mathcal{S} = \{G(\boldsymbol{z}_i; \boldsymbol{\varphi}), y_i^{\mathcal{S}}\}_{i=1}^{M}$$



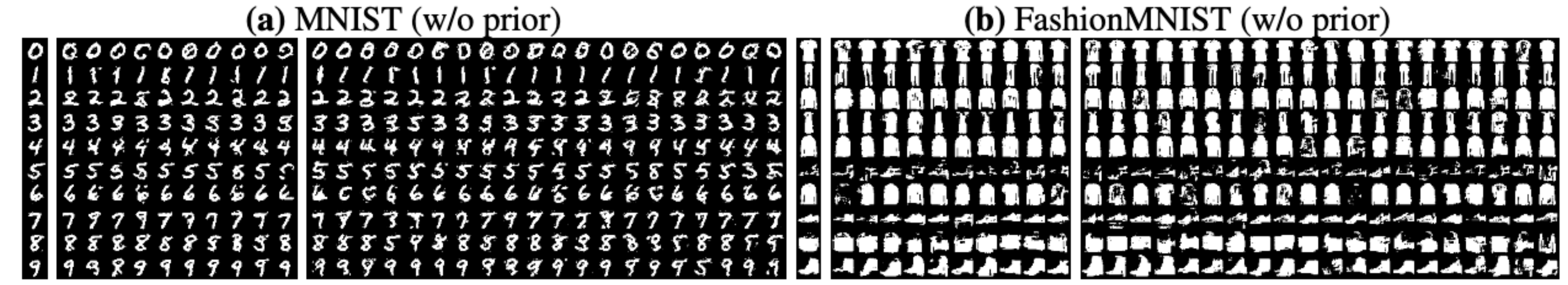
- **Findings:**
  - Deep generative models result in:
    - Better visual quality
    - Slow convergence
    - Sub-optimal downstream utility

	MNIST			FashionMNIST		
	1	10	20	1	10	20
w/o prior	81.4	**94.9**	**95.6**	**66.7**	**75.6**	**77.7**
with prior	**88.2**	92.2	90.6	63.0	70.2	70.7





(a) MNIST (w/o prior) (b) FashionMNIST (w/o prior)

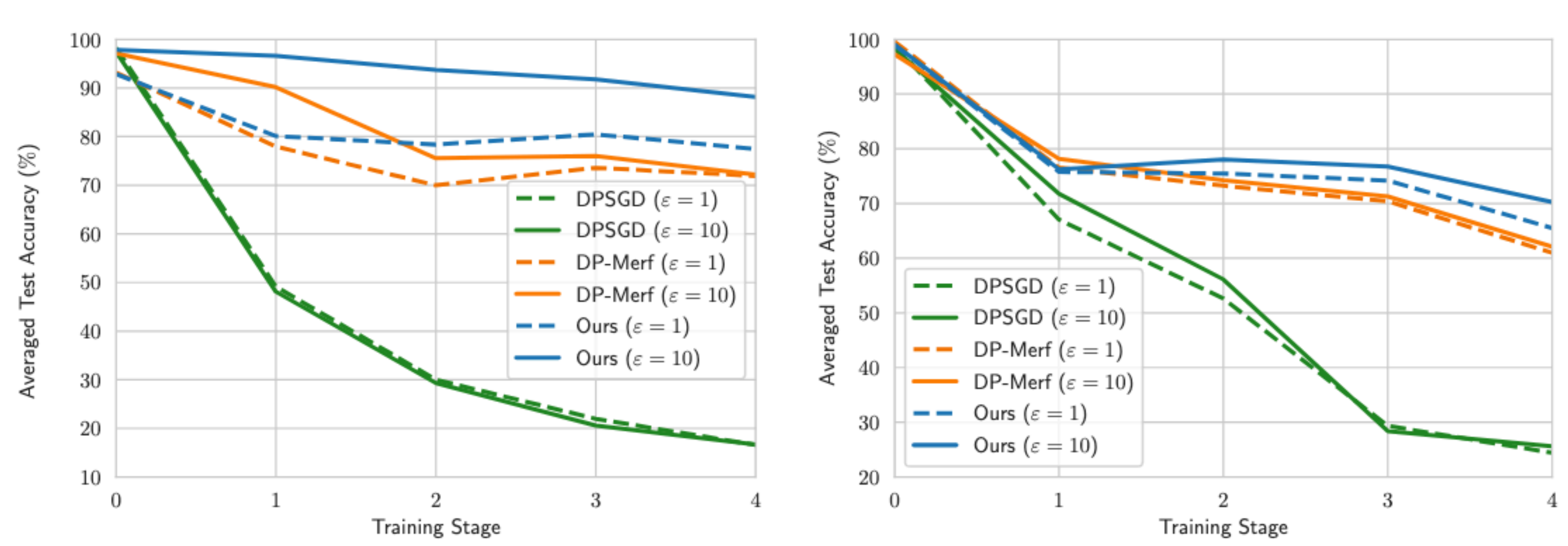


(c) MNIST (with prior) (d) FashionMNIST (with prior)

Evaluation

- **Downstream utility**
- **Generalization ability**
- **Application:** Continual learning with DP

	MNIST		FashionMNIST	
	ε=1	ε=10	ε=1	ε=10
DP-CGAN	-	52.5	-	50.2
G-PATE	58.8	80.9	58.1	69.3
DataLens	71.2	80.7	64.8	70.6
GS-WGAN	-	84.9	-	63.1
DP-Merf	72.7	85.7	61.2	72.4
DP-Sinkhorn	-	83.2	-	71.1
Ours (spc=20)	**80.9**	**95.6**	**70.2**	**77.7**



(a) SplitMNIST (b) SplitFashionMNIST

	MNIST						FashionMNIST					
	ConvNet	LeNet	AlexNet	VGG11	ResNet18	MLP	ConvNet	LeNet	AlexNet	VGG11	ResNet18	MLP
Real	99.6	99.2	99.5	99.6	99.7	98.3	93.5	88.9	91.5	93.8	94.5	86.9
DP-CGAN	50.2	52.6	52.1	54.7	51.8	54.3	50.2	52.6	52.1	54.7	51.8	54.3
GS-WGAN	84.9	83.2	80.5	87.9	89.3	74.7	54.7	62.7	55.1	57.3	58.9	65.4
DP-Merf	85.7	87.2	84.4	81.7	81.3	85.0	72.4	67.9	64.9	70.1	66.7	**73.1**
Ours (spc=10)	94.9	91.3	90.3	93.6	**94.3**	86.1	75.6	**68.0**	**66.2**	74.7	**72.1**	62.8
Ours (spc=20)	**95.6**	**93.0**	**92.3**	**94.5**	94.1	**87.1**	**77.7**	**68.0**	59.1	**76.8**	70.8	62.2