Synthesize Privacy-Preserving High-Resolution Images via Private Textual Intermediaries

Haoxiang Wang [1] Zinan Lin [2] Da Yu [3] Huishuai Zhang [1]

[1]Peking University [2]Microsoft Research [3]Google Research

NEURAL INFORMATION PROCESSING SYSTEMS

PEKING UNIVERSITY

Microsoft Google

Main Result

We introduce *SPTI*, a training-free yet effective pipeline to synthesize high resolution images under differential private constraints.

- *SPTI* uses text modality as intermediaries: Our pipeline uses text to generate key information for image modality under DP constraints;
- *SPTI* performs well: delivering superior performance over traditional *PE* method;
- *SPTI* is easy to extend: our method can extended to other modality like sound, etc.

Algorithm 1 SPTI: Privately Synthesize High-Resolution Images via Private Text Intermediaries

Require: Private image dataset $\mathcal{D}$, `Aug_PE`, `Aug_PE_Image_Voting`, `text_voting`
Ensure: Synthetic images $\mathcal{D}'$

```
1: if text_voting = true then
2:     Convert images D to text descriptions T via a captioning model
3:     Apply Private Evolution on text: T' = Aug_PE(T)
4: else
5:     Apply Private Evolution with image-voting: T' = Aug_PE_Image_Voting(D)
6: end if
7: Generate synthetic images D' from T' using a text-to-image diffusion model
8: return D'
```

Overview of the Synthesis via Private Textual Intermediaries (SPTI) framework

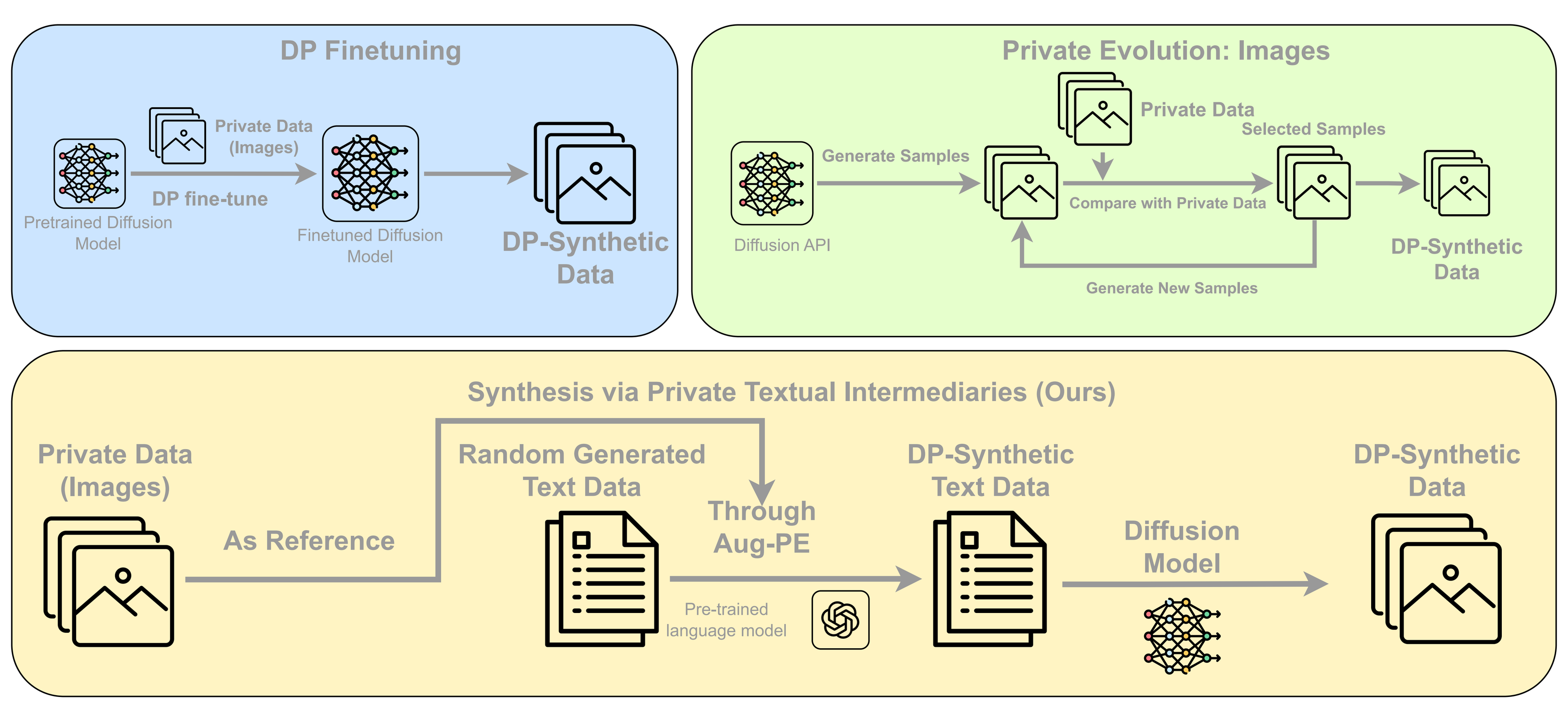

Figure 1. *Top-left: DP-finetune method. Top-right: Private Evolution (PE) on images. Bottom: SPTI.* Private image data is served as reference. A modified Augmented Private Evolution (Aug-PE) method is then applied to synthesize DP text data, which is subsequently transformed into DP synthetic images using a diffusion model API.

Privacy Analysis

SPTI ensures differential privacy by restricting all private-data access to the Private Evolution module (`Aug_PE` / `Aug_PE_Image_Voting`). This module uses private image embeddings only to guide voting over text candidates.

Voting Mechanism

Each private embedding votes for its nearest generated image, forming a histogram H. Adding Gaussian noise $\mathcal{N}(0, \sigma^2\mathbf{I})$ to H applies the Gaussian mechanism.

DP Guarantee

The histogram has L_2 sensitivity 1. With Gaussian noise of variance σ^2 applied for G iterations, *SPTI* satisfies (ε, δ)-DP whenever:

$$\Phi\left(\frac{\sqrt{G}}{2\sigma} - \frac{\varepsilon\sigma}{\sqrt{G}}\right) - e^{\varepsilon}\Phi\left(-\frac{\sqrt{G}}{2\sigma} - \frac{\varepsilon\sigma}{\sqrt{G}}\right) \leq \delta.$$

Post-Processing

All later steps—text mutation and diffusion-based image generation—use only the privatized histogram. By post-processing immunity, they add no privacy cost.

Performance of *SPTI*

		$\varepsilon=10$	$\varepsilon=5$	$\varepsilon=1$
LSUN Bedroom	*SPTI* (ours)	**25.88**	**25.87**	**26.39**
	PE image	41.72	41.08	40.36
	DP-finetune	31.28	31.34	31.76
European Art	*SPTI* (ours)	**41.42**	**42.71**	**57.64**
	PE image	76.25	74.41	76.50
	DP-finetune	61.10	61.82	63.97
Wave-ui-25k	*SPTI* (ours)	**20.16**	**22.53**	**35.18**
	PE image	39.28	48.95	50.45
	DP-finetune	49.84	52.09	62.08

Table 1. **FID** values (lower is better) across multiple datasets to compare three different DP methods: *SPTI*, PE Image, and DP-finetune.

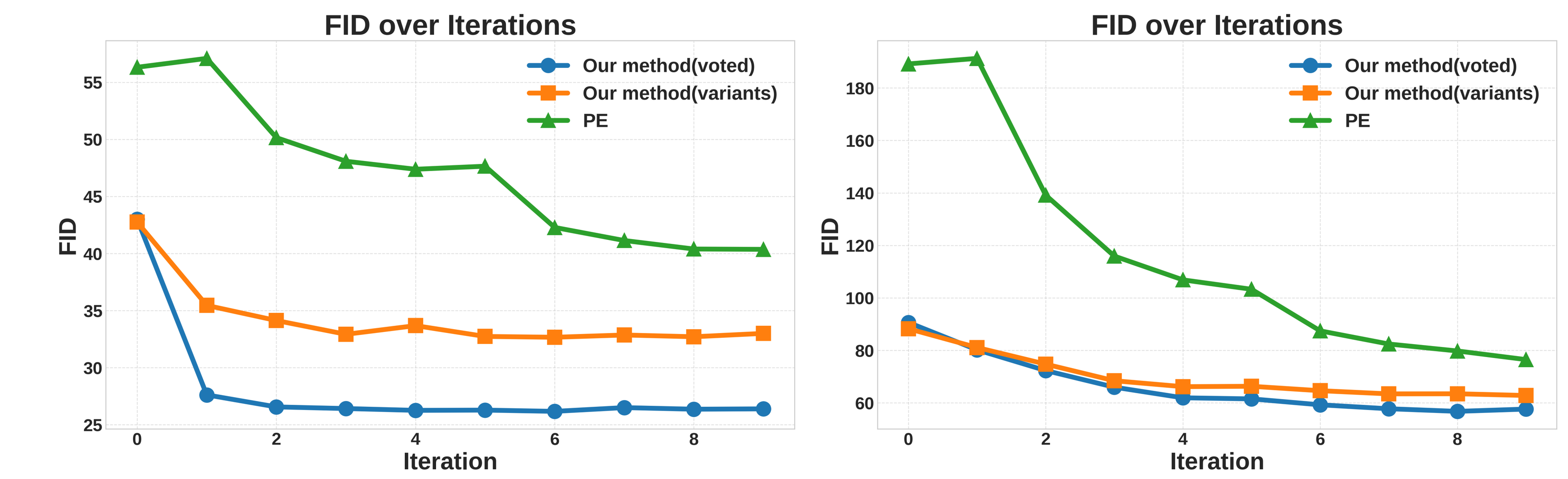

Figure 2. (**FID** lower is better). Quality of synthesized image samples in each iteration.

Ablation Study

		$\varepsilon=10$	$\varepsilon=5$	$\varepsilon=1$
LSUN bedroom	Image Voting	**27.92**	**29.50**	**38.00**
	Text Voting	39.82	37.83	41.17
Europeart	Image Voting	**46.69**	**48.45**	**64.04**
	Text Voting	74.37	74.04	74.91
Wave-ui-25k	Image Voting	**34.39**	**42.02**	**70.87**
	Text Voting	92.53	92.11	95.77

Table 2. **FID** values (lower is better) across multiple datasets to compare the *SPTI* method with Image Voting and that with Text Voting.

	SDXL-Turbo	SDXL-base-1.0	Infinity
Meta-Llama-3-8B-Instruct	26.71	25.42	30.66
qwen-plus	26.65	24.44	31.28

Table 3. **FID** (lower is better) on **LSUN Bedroom** with $\epsilon = 1.0$, tested using different LLM and diffusion APIs.

Downstream Task

We also validate the quality of generated images in downstream tasks. To be more specific, we test the classification accuracy on **CelebA** dataset using **WRN-40-4** model.

Figure 3. *Left:* samples generated by *SPTI*. *Right:* samples generated by PE method.