# VP-NTK: EXPLORING THE BENEFITS OF VISUAL PROMPTING IN DIFFERENTIALLY PRIVATE DATA SYNTHESIS

*Chia-Yi Hsu*[1] *Jia-You Chen*[1] *Yu-Lin Tsai*[1] *Chih-Hsun Lin*[1]
*Pin-Yu Chen*[2] *Chia-Mu Yu*[1] *Chun-Ying Huang*[1]

[1]National Yang Ming Chiao Tung University, Taiwan [2]IBM Research, NY, USA

## ABSTRACT

Differentially private (DP) synthetic data has become the *de facto* standard for releasing sensitive data. However, many DP generative models suffer from the low utility of synthetic data, especially for high-resolution images. On the other hand, one of the emerging techniques in parameter efficient fine-tuning (PEFT) is visual prompting (VP), which allows well-trained existing models to be reused for the purpose of adapting to subsequent downstream tasks. In this work, we explore such a phenomenon in constructing captivating generative models with DP constraints. We show that VP in conjunction with DP-NTK, a DP generator that exploits the power of the neural tangent kernel (NTK) in training DP generative models, achieves a significant performance boost, particularly for high-resolution image datasets, with accuracy improving from 0.644±0.044 to 0.769. Lastly, we perform ablation studies on the effect of different parameters that influence the overall performance of VP-NTK. Our work demonstrates a promising step forward in improving the utility of DP synthetic data, particularly for high-resolution images.

***Index Terms*—** Differential Privacy, Visual Prompting, Data Synthesis

## 1. INTRODUCTION

Originating in the field of deep learning for natural language processing, prompt engineering has gained popularity as an innovative technique for efficiently using and adapting pre-trained language models for various downstream tasks [1]. On the other hand, the original concept of prompt engineering has blossomed and extended to other domains and data types, such as image and computer vision. In particular, visual prompting has been introduced in [2], where the method outperforms linear probing (i.e., attaching a trainable linear head to a pre-trained model) when used alongside large-scale vision models, demonstrating its effectiveness. Interestingly, model reprogramming (MR) can be seen as a generalized version of visual prompting (VP), where the well-trained models are reused in a sampling-efficient manner. Specifically, MR involves inserting an input transformation layer and an output mapping layer into a frozen, pre-trained model for fine-tuning downstream tasks. VP in [2] corresponds to MR when the input transformation consists of a trainable input perturbation and the output mapping corresponds to specified source-target label associations for label inference.

Visual prompting (VP) has been extensively studied in various applications, including image classification [2], cross-domain adaptation [3, 4], and so on. In this paper, we explore an additional advantageous facet of VP when paired with pre-trained models - its infusion with differential privacy (DP).

In particular, scaling the parameters of generative models in deep learning often leads to better performance on general tasks. However, under the goal of privacy-preserving machine learning, a strict privacy budget can lead to massive noise injection in the training process, rendering the DP generative model useless. Moreover, the demand for high-resolution images often requires that generative models be equipped with high capacity. Thus, with the above insight and the seemingly impossible tradeoff, we aim to answer the following question:

*Will incorporating VP with a well-trained generative model offer better privacy-accuracy tradeoff in DP data synthesis?*

In this paper, we provide an affirmative answer, which is empirically supported and comprehensively compared with other state-of-the-art methods. We focus on existing DP data synthesis algorithms since the improvement from applying VP could be directly observable. Our approach includes the application of VP to one of the state-of-the-art DP data synthesis methods, DP-NTK [5]. In particular, when VP is applied to DP-NTK, we observe a significant performance improvement over the original method in the downstream classification task of high-resolution images under an identical privacy budget. As a result, our results reveal advantages of VP in DP data synthesis and offer new use cases and insights for prompt engineering. We summarize our findings as follows:

- We are the first to explore the benefits of VP in DP data synthesis.
- We improve the challenge of DP data synthesis in the high-resolution image domain.
- We provide insights into how VP could reuse the well-trained generative model.

## 2. RELATED WORKS AND BACKGROUND

**Visual Prompting (VP).** The intent of VP is to fully reuse a pre-trained model to perform a new task, without further modification of the model's weights during fine-tuning, thus stealing the computational effort.

VP through a trainable input perturbation is revisited in [2], and the authors showed competitive results on a subset of 12 image classification tasks over linear probing and full fine-tuning on pre-trained image classifiers and the CLIP model. We note that in this paper we focus exclusively on VP in the input prompt engineering setting, and leave the alternative setting of layer-wise visual prompt tuning as future work.

**Differentially Private Data Synthesis.** Since both DPSGD and PATE are common techniques to enforce DP constraints on deep learning models, there are many algorithms for DP generative models that surround both concepts. However, since the implementation of DPSGD requires gradient clipping and noise addition at each training step, this in turn leads to severe information loss in model updates. Therefore, several works are dedicated to reducing the loss, such as [6, 7, 8]. On the other hand, there are several works based on PATE such as [9] which spends the privacy budget on selecting more useful iterations so that the discriminators can learn in the right direction.

Other methods such as DP-MERF [10], DP-NTK [5] use various pre-trained perceptual features to further facilitate the DP training process by matching the perturbed data distribution with the generator. Following frameworks similar to DP-MERF, DP-Sinkhorn [11] and PEARL [12] train the generator by minimizing the Sinkhorn divergence with semi-debiased Sinkhorn loss and the feature distance, respectively. On the other hand, works such as [13, 14] focus on infusing DP with diffusion models, which are state-of-the-art generative models.

**Visual Prompting with DP.** Some recent works that combine VP and DP include Reprogrammable-FL [15] and Prom-PATE [16], where both exploit VP to improve the performance of DP classifiers.

**High-Resolution Data Synthesis.** Ensuring privacy in high-resolution data synthesis is particularly challenging due to information loss caused by DP techniques. Adapting pre-trained models for high-resolution tasks with VP may enhance performance. However, current work focuses only on exploring the benefits of VP in the construction of classifiers [17], leaving generative models unexplored.

## 3. MAIN APPROACH

In this section, we will present the main approach, VP-NTK, as a hybrid of DP generative model and visual prompting. First, we will provide a brief overview of DP-NTK, as this is our choice of DP generative model. Second, we will present how VP can be infused into DP-NTK to create our main approach, VP-NTK, with enhanced capabilities.

### 3.1. DP-NTK

Given a labeled dataset $\{(x_i, y_i)\}_{i=1}^{m} \sim P$ where $m$ represents the number of data and $P$ represents the true data distribution, DP-NTK first construct the true mean embedding of the data distribution as $\hat{\mu}_P = \frac{1}{m}\sum_{i=1}^{m} \phi(x_i) y_i^T$ where label $y_i$ is in the form of one-hot encoding and $\phi(\cdot)$ stands for feature embedding generated from the neural tangent kernel (i.e., $\phi(x) = \frac{\nabla_\theta f(x;\theta)}{||\nabla_\theta f(x;\theta)||}$ where $f(\cdot;\theta)$ and $\theta$ denote the deep network and its corresponding parameter).

Then, to preserve privacy, DP-NTK uses the Gaussian mechanism [18] to obtain a perturbed mean embedding $\tilde{\mu}_P$ by $\tilde{\mu}_P = \hat{\mu}_P + \mathcal{N}(0, \frac{4\sigma^2}{m^2} I)$ where the variance parameter $\sigma$ is controlled by the privacy parameters $\epsilon$ and $\delta$.

On the other hand, the generator $G$ generates $n$ synthetic data samples $x_i'$ with standard Gaussian noise $z_i$ and generated label $y_i'$ (i.e. $x_i' = G(z_i, y_i')$). Similarly, DP-NTK constructs the mean embedding $\hat{\mu}_Q$ of the synthetic data as $\hat{\mu}_Q = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i') y_i'^T$

Finally, DP-NTK simulates the true data distribution by minimizing the empirical maximum mean discrepancy between the two embeddings as follows $\widetilde{\text{MMD}}(P, Q) = ||\tilde{\mu}_P - \hat{\mu}_Q|_F^2$, where $||\cdot||_F$ is the Frobenius norm.

### 3.2. VP-NTK

In our framework, we reuse a well-trained conditional generator and feature extractor as our source models. First, we perform a label mapping to determine how classes in the private data should correspond to classes in the conditional generator. Note that this does not introduce any plausible privacy leakage, since the mapping itself is pre-determined in the training process. Then, we acquire the features of both the private data and the images generated by the generator by feeding them through a pre-trained feature extractor that remains fixed throughout the training process. Finally, we incorporate VP by adding trainable noise to the features of $G(z, y)$ to learn the distribution of the features of the private data. Note that each class has its own trainable noise, i.e., its own visual stimulus. Finally, we connect the visual prompt feature to the original pipeline of DP-NTK to ensure that the learning process satisfies the DP guarantee. Fig 1 illustrates the workflow of VP-NTK.

We explain the details in the workflow. For label mapping, we randomly map classes of private data to the predefined conditional signals of the generator (e.g., ”dog” in the source generator corresponds to ”man” in the private data distribution). When adding noise to the features, the coefficient $\kappa$ is used to control the amount of noise. On the other hand, the original loss function of DP-NTK is to compute the empirical maximum mean discrepancy between true and synthetic

|  | $\varepsilon$ | DP-MERF | G-PATE | DP-Sinkhorn | DataLens | DP-HP | NDPDC | PEARL | VP-NTK |
|---|---|---|---|---|---|---|---|---|---|
| CelebA-Gender | $\varepsilon = 1$ | 0.594 | 0.670 | 0.543 | 0.700 | 0.656 | 0.540 | 0.634 | **0.707** |
|  | $\varepsilon = 10$ | 0.608 | 0.690 | 0.621 | 0.729 | 0.617 | 0.600 | 0.646 | **0.769** |
| CelebA-Hair | $\varepsilon = 1$ | 0.441 | 0.499 | × | 0.606 | 0.561 | 0.498 | 0.606 | **0.653** |
|  | $\varepsilon = 10$ | 0.449 | 0.622 | × | 0.622 | 0.474 | 0.462 | 0.626 | **0.641** |

**Table 1**. Classification accuracy results under $(1, 10^{-5})$-DP with image resolution as $64 \times 64$.

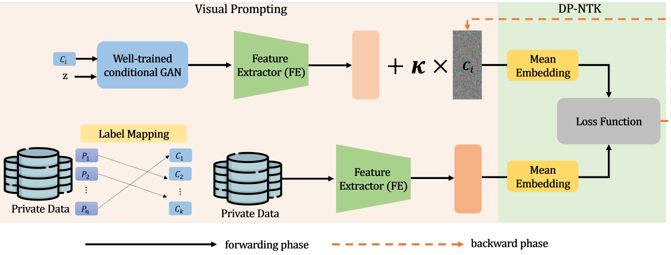

**Fig. 1**. The framework of VP-NTK.

data. Instead of relying solely on Maximum Mean Discrepancy (MMD), we also incorporate cosine similarity into our loss. Finally, we introduce the coefficient $\alpha$, which controls a penalty term on the norm of the trainable input noise to avoid over-fitting. Note that the source model remains frozen in our training process, including the well-trained conditional generator and the feature extractor itself.

**Theorem 1** *VP-NTK satisifies* $(\epsilon, \delta)$*-DP.*

*Proof*: Due to limited space, we defer the proof to web for demo.[1].

## 4. EXPERIMENTS

In this section, we conduct experiments on complicated datasets to demonstrate the performance of our proposed VP-NTK. Furthermore, we present the ablation study to show the rationale behind the hyperparameter selection.

### 4.1. Experiment Setup

We demonstrate the utility of VP-NTK on CelebA datasets at different resolutions, including $64 \times 64$, $128 \times 128$, and the high-resolution $256 \times 256$. Derived from CelebA, we generated two additional datasets: CelebA-Gender for binary gender classification and CelebA-Hair for multiclass hair color classification (including black, blonde, and brown). Both are colorful, increasing the challenge of preserving the utility of the synthetic data. For the DP guarantee, we set $(\epsilon, \delta)$ to $(1, 10^{-5})$ and $(10, 10^{-5})$. All experiments are done on NVIDIA GeForce RTX 4090. In the following, we present the baseline, the VP-NTK settings, and how to evaluate the synthetic data.

- **Baselines.** We compare VP-NTK with state-of-the-art DP image synthesis methods such as DP-MERF [10], DataLens [19], G-PATE [9], DP-Sinkhorn [11], DP-HP [20], Nonlinear DPDC (NDPDC) [21], and PEARL [12]. All comparisons are based on their official codes.
- **VP-NTK.** Our framework includes a pre-trained GAN and a feature extractor (FE). For all experiments, we use the IC-GAN [22] and employ the ResNet18 as the FE, pre-trained on the Tiny-ImageNet dataset. In addition, there are four important hyper-parameters in our experiment which are $\kappa$, $\eta$, $\alpha$, and the choice of loss functions. $\eta$ represents the learning rate. For the standard solution of our experiments, we set $\kappa = 16$, $\eta = 10^{-2}$, $\alpha = 0.05$, and the loss function combines MMD and cosine similarity.
- **Evaluation.** All synthetic data generated by each method are used to train classifiers with the same architecture. We compare the test accuracy on the test sets of CelebA-Gender and CelebA-Hair, where high test accuracy represents higher utility.

### 4.2. Comparison with Existing DP Generative Models

We also compare VP-NTK with several SOTA DP generative models. Table 1 shows the result of our method and other reference methods on $64{\times}64$ image resolution. Here we note that while we also experiment on images with higher resolutions, such as $128{\times}128$ and $256{\times}256$, existing methods present no data for further comparison. As can be seen from Table 1, VP-NTK has outstanding performance against all GAN-based DP generative models and an accuracy gap of more than 4% with $\epsilon = 10$ on CelebA-Gender.

For other larger image resolutions, such as $128{\times}128$ and $256{\times}256$, Table 2 shows that our method also performs well on them. However, for higher resolution data, other methods require DP image generators with high capacity(i.e., large parameter counts), which will result in worsening performance than the result in Table 1, while VP-NTK benefits from the potential of visual prompting, allowing the DP generative model to handle data with increasing resolution.

### 4.3. Ablation Study

For all experiments in this section, we evaluate the performance on both the CelebA-Gender and CelebA-Hair datasets

[1] Proof Website

| Dataset | Image size | $\epsilon$ | Accuracy±Std(%) |
|---|---|---|---|
| **CelebA-Gender** | 128×128 | 1 | 79.24±0.38 |
| | 256×256 | | 82.67±1.75 |
| **CelebA-Gender** | 128×128 | 10 | 79.01±0.41 |
| | 256×256 | | 82.28±1.98 |
| **CelebA-Hair** | 128×128 | 1 | 64.14±0.17 |
| | 256×256 | | 65.48±0.30 |
| **CelebA-Hair** | 128×128 | 10 | 63.66±0.41 |
| | 256×256 | | 65.85±0.35 |

**Table 2**. Accuracy for different dataset and $\epsilon$

with a resolution of 128×128 and a privacy budget of $\varepsilon = 1$.

**The Effect of $\kappa$.** The $\kappa$ is used to control the amount of noise. The table 3 shows how $\kappa$ affects performance. If the value of $\kappa$ is too low, for example, $\kappa = 2$, it may be difficult to effectively transfer the synthetic features to the private data features. On the other hand, if $\kappa$ is too large, it can lead to overfitting. There are the same trends between CelebA-Gender and Hair that the performances are better as $\kappa$ in a reasonable range (e.g., $\kappa = 4, 8, 16$) shown in table 3.

| Dataset | $\kappa$ | Accuracy±Std(%) |
|---|---|---|
| **CelebA-Gender** | 2 | 77.31±0.15 |
| | 4 | 79.23±0.31 |
| | 8 | 79.28±0.54 |
| | 16 | 79.24±0.38 |
| | 32 | 78.70±0.55 |
| **CelebA-Hair** | 2 | 53.43±0.06 |
| | 4 | 60.47±0.40 |
| | 8 | 63.05±0.48 |
| | 16 | 63.87±0.05 |
| | 32 | 64.12±1.06 |

**Table 3**. Accuracy for different $\kappa$

**The Effect of $\eta$.** Table 4 shows the result on CelebA-Gender 128×128. Similarly, Table 4 shows the result on CelebA-Hair 128×128. Obviously, for both results, using $\eta = 10^{-2}$ has the best performance. This is because scaling $\eta$ leads to an unstable result, while decreasing the learning rate leads to slower convergence and thus harder search for the optimal visual prompt.

**The Effect of $\alpha$.** We evaluate the effect of $\alpha$ on both datasets, with the results presented in Table 5. As can be seen, $\alpha$ does not have a large effect on the overall result. This is mainly due to the small magnitude of the visual cues compared to either cosine similarity or MMD loss.

**The Effect of Different Loss.** We evaluate the effect of different losses with the results shown in Table 6. As one can see from the table, MMD could provide theoretical guarantees between the true and synthetic distribution. The empirical MMD used in both VP-NTK and DP-NTK [5] could sometimes differ from the true MMD. Therefore, we include cosine similarity as an alternative metric and discovered that mixing in equal proportions will offer the best result.

| Dataset | $\eta$ | Accuracy±Std(%) |
|---|---|---|
| **CelebA-Gender** | 1e-5 | 52.81±2.26 |
| | 1e-4 | 46.77±3.26 |
| | 1e-3 | 79.22±0.75 |
| | 1e-2 | **79.24±0.38** |
| | 0.1 | 79.19±0.86 |
| | 1 | 77.03±3.02 |
| **CelebA-Hair** | 1e-5 | 29.33±0.79 |
| | 1e-4 | 33.31±1.56 |
| | 1e-3 | 63.60±0.27 |
| | 1e-2 | **63.87±0.05** |
| | 0.1 | 58.83±4.42 |
| | 1 | 58.16±3.32 |

**Table 4**. Accuracy for different $\eta$

| Dataset | $\alpha$ | Accuracy±Std(%) |
|---|---|---|
| **CelebA-Gender** | 0.01 | 78.31±0.63 |
| | 0.05 | 79.24±0.38 |
| | 0.1 | 79.44±0.37 |
| | 1 | 79.25±0.16 |
| **CelebA-Hair** | 0.01 | 63.69±1.72 |
| | 0.05 | **63.87±0.05** |
| | 0.1 | 63.41±0.46 |
| | 1 | 57.57±2.62 |

**Table 5**. Accuracy for different $\alpha$

| Dataset | loss | Accuracy±Std(%) |
|---|---|---|
| **CelebA-Hair** | MMD | 58.08±0.27 |
| | mixed | **79.24±0.38** |
| | cosine | 79.21±0.35 |
| **CelebA-Hair** | MMD | 36.65±2.12 |
| | mixed | **63.87±0.05** |
| | cosine | 63.83±0.07 |

**Table 6**. Accuracy for different loss

## 5. CONCLUSION

In this paper, we've explored the preliminary benefits of integrating visual prompting into the DP data synthesis pipeline. Specifically, when applied to one of the SOTA DP generative models, we can observe an increase in accuracies when the generative model is used for a downstream classification task. Furthermore, the integrated DP generative model could also address the challenge of generating high-resolution images. Our discovery revealed that VP is a promising method to accelerate further research in constructing DP generative models that improve the privacy-utility tradeoff.

## 6. REFERENCES

[1] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting

methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.

[2] Hyojin Bahng, Ali Jahanian, Swami Sankaranarayanan, and Phillip Isola, "Exploring visual prompts for adapting large-scale models," *arXiv preprint arXiv:2203.17274*, 2022.

[3] Paarth Neekhara, Shehzeen Hussain, Jinglong Du, Shlomo Dubnov, Farinaz Koushanfar, and Julian McAuley, "Cross-modal adversarial reprogramming," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2427–2435.

[4] Yun-Yun Tsai, Pin-Yu Chen, and Tsung-Yi Ho, "Transfer learning without knowing: Reprogramming black-box machine learning models with scarce data and limited resources," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9614–9624.

[5] Yilin Yang, Kamil Adamczewski, Danica J. Sutherland, Xiaoxiao Li, and Mijung Park, "Differentially private neural tangent kernels for privacy-preserving data generation," arXiv: 2303.01687, 2023.

[6] H. B. McMahan and G. Andrew, "A general approach to adding differential privacy to iterative training procedures," *NeurIPS Workshop on Privacy Preserving Machine Learning (PPML)*, 2018.

[7] O. Thakkar, G. Andrew, and H. B. McMahan, "Differentially private learning with adaptive clipping," *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[8] D. Yu, Huishuai Zhang, Wei Chen, and T. Liu, "Do not let privacy overbill utility: Gradient embedding perturbation for private learning," *International Conference on Learning Representations (ICLR)*, 2021.

[9] Yunhui Long, Boxin Wang, Zhuolin Yang, Bhavya Kailkhura, Aston Zhang, Carl A. Gunter, and Bo Li, "G-pate: Scalable differentially private data generator via private aggregation of teacher discriminators," *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[10] Frederik Harder, Kamil Adamczewski, and Mijung Park, "Dp-merf: Differentially private mean embeddings with random features for practical privacy-preserving data generation," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

[11] Tianshi Cao, Alex Bie, Arash Vahdat, Sanja Fidler, and Karsten Kreis, "Don't generate me: Training differentially private generative models with sinkhorn divergence," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[12] Seng Pei Liew, Tsubasa Takahashi, and Michihiko Ueno, "Pearl: Data synthesis via private embeddings and adversarial reconstruction learning," in *International Conference on Learning Representations (ICLR)*, 2022.

[13] Tim Dockhorn, Tianshi Cao, Arash Vahdat, and Karsten Kreis, "Differentially private diffusion models," *arXiv preprint arXiv:2210.09929*, 2022.

[14] Haichen Wang, Shuchao Pang, Zhigang Lu, Yihang Rao, Yongbin Zhou, and Minhui Xue, "dp-promise: Differentially private diffusion probabilistic models for image synthesis," USENIX, 2024.

[15] Huzaifa Arif, Alex Gittens, and Pin-Yu Chen, "Reprogrammable-fl: Improving utility-privacy trade-off in federated learning via model reprogramming," in *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 2023, pp. 197–209.

[16] Yizhe Li, Yu-Lin Tsai, Xuebin Ren, Chia-Mu Yu, and Pin-Yu Chen, "Exploring the benefits of visual prompting in differential privacy," *arXiv preprint arXiv:2303.12247*, 2023.

[17] Aochuan Chen, Yuguang Yao, Pin-Yu Chen, Yihua Zhang, and Sijia Liu, "Understanding and improving visual prompting: A label-mapping perspective," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 19133–19143.

[18] Cynthia Dwork and Aaron Roth, "The algorithmic foundations of differential privacy.," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.

[19] Boxin Wang, Fan Wu, Yunhui Long, Luka Rimanic, Ce Zhang, and Bo Li, "Datalens: Scalable privacy preserving training via gradient compression and aggregation," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 2146–2168.

[20] Margarita Vinaroz, Mohammad-Amin Charusaie, Frederik Harder, Kamil Adamczewski, and Mi Jung Park, "Hermite polynomial features for private data generation," in *International Conference on Machine Learning (ICML)*, 2022.

[21] Tianhang Zheng and Baochun Li, "Differentially private dataset condensation," 2023.

[22] Arantxa Casanova, Marlène Careil, Jakob Verbeek, Michal Drozdzal, and Adriana Romero-Soriano, "Instance-conditioned gan," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.