[MPGD] Manifold Preserving Guided Diffusion

ICLR 2024

Yutong He [1, 2], Naoki Murata [2], Chieh-Hsin Lai [2], Yuhta Takida [2], Toshimitsu Uesaka [2], Dongjun Kim [2], Wei-Hsiang Liao [2], Yuki Mitsufuji [2, 3], J. Zico Kolter [1], Ruslan Salakhutdinov [1], Stefano Ermon [4]

[1] Carnegie Mellon University, [2] Sony AI, [3] Sony Group Corporation, [4] Stanford University

Presenter: Jia-Wei Liao (`jw@cmlab.csie.ntu.edu.tw`)

TL; DR

This paper proposes a training-free method leveraging the pre-trained diffusion models to solve the inverse problem via gradient-based optimization on sampling process.

Outline

1. Introduction
   1.1 Inverse Problem
2. Background
   2.1 Manifold Hypothesis
   2.2 Diffusion Model
   2.3 Prior Works
3. Proposed Method
   3.1 MPGD w/o Proj.
   3.2 MPGD-AE
   3.3 MPGD-Z
   3.4 MPGD-LDM
4. Experiments
5. Conclusions

Outline

1. Introduction
   1.1 Inverse Problem
2. Background
   2.1 Manifold Hypothesis
   2.2 Diffusion Model
   2.3 Prior Works
3. Proposed Method
   3.1 MPGD w/o Proj.
   3.2 MPGD-AE
   3.3 MPGD-Z
   3.4 MPGD-LDM
4. Experiments
5. Conclusions

Introduction to Inverse Problem

Inverse Problem

Given observed measurement $\boldsymbol{y} \in \mathbb{R}^m$ and forward measurement operator $\mathscr{A}(\cdot) : \mathbb{R}^m \to \mathbb{R}^n$, our goal is to reconstruct an unknown signal $\boldsymbol{x}^*$ of the form

$$\boldsymbol{y} = \mathscr{A}(\boldsymbol{x}^*) + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is additive noise.

- If $m < n$, the problem is ill-posed [1].
- If $\mathscr{A}(\cdot)$ is the linear operator ($\mathscr{A}(\boldsymbol{x}) = A\boldsymbol{x}$), then the problem has closed-form solution $\boldsymbol{x}^* = A^\dagger \boldsymbol{y} (= V\Sigma^\dagger U^\top \boldsymbol{y})$.

[1] The solution lacks existence, uniqueness, or stability.

Introduction to Inverse Problem (Conti.)

Inverse Problem (Optimization Form)

Given observed measurement $\boldsymbol{y} \in \mathbb{R}^m$ and forward measurement operator $\mathscr{A}(\cdot) : \mathbb{R}^m \to \mathbb{R}^n$, our goal is to reconstruct an unknown signal on the nature image manifold $\mathscr{M}$ via the optimization problem:

$$\hat{\boldsymbol{x}} \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathscr{M}} \left\| \mathscr{A}(\boldsymbol{x}) - \boldsymbol{y} \right\|_2^2$$

- Given the generative model $\mathscr{G}_\theta(\cdot)$ (e.g. VAE, GAN, Diffusion Model, etc.), the probelm can be reformulated as

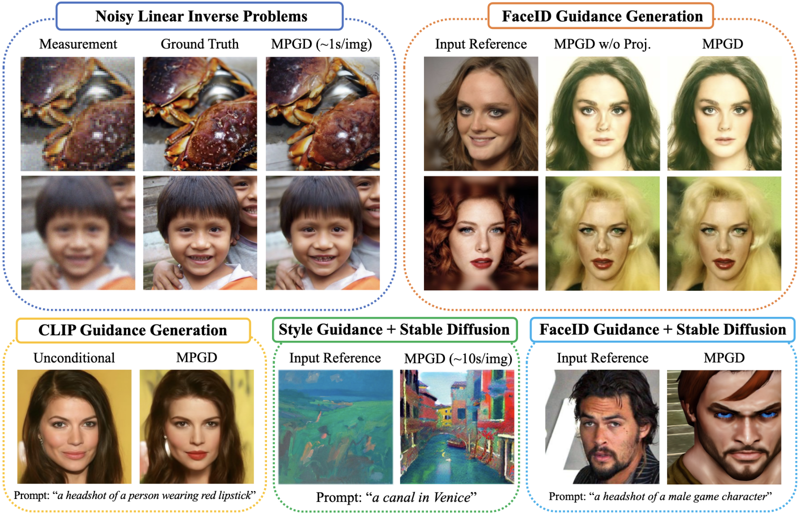
$$\hat{\boldsymbol{x}} \in \operatorname*{argmin}_{\boldsymbol{x}} \left\| \mathscr{A}(\mathscr{G}_\theta(\boldsymbol{x})) - \boldsymbol{y} \right\|_2^2 =: \operatorname*{argmin}_{\boldsymbol{x}} \mathscr{L}(\mathscr{G}_\theta(\boldsymbol{x}), \boldsymbol{y}).$$

- This slide only focuses on solving inverse problems using diffusion models.

Application of Real World Scenarios (CV Tasks)

- Image Denoising
- Image Deblurring
- Image Inpainting
- Image Colorization
- Super-Resolution
- CT reconstruction

More Applications in MPGD Paper



MPGD Achievements

Manifold Preserving Guided Diffusion (MPGD) is a framework for conditional generation using unconditionally pretrained diffusion models. It achieves

1. **Training free:** Pretrained models can be deployed without extra training.
2. **Low cost:** The method should require minimal additional computational resources and time.
3. **Generalizable:** They only require black-box access to the loss function and its gradients.
4. **High qualit:** The samples should come from a distribution that is close to the true posterior.

Outline

1. Introduction
1.1 Inverse Problem
2. Background
2.1 Manifold Hypothesis
2.2 Diffusion Model
2.3 Prior Works
3. Proposed Method
3.1 MPGD w/o Proj.
3.2 MPGD-AE
3.3 MPGD-Z
3.4 MPGD-LDM
4. Experiments
5. Conclusions

Notation

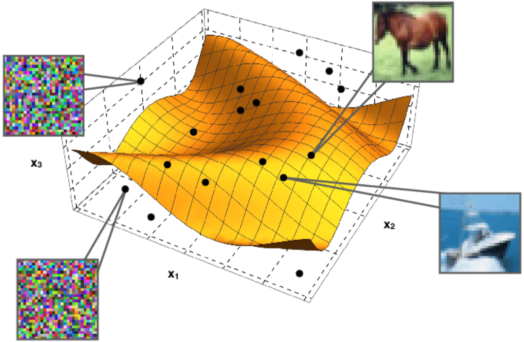
- $\mathcal{M} \subseteq \mathbb{R}^n$ be a image manifold.
- $\mathcal{Z} \subseteq \mathbb{R}^d$ be a latent space.
- $\mathcal{E} : \mathcal{M} \to \mathcal{Z}$ be a VAE encoder.
- $\mathcal{D} : \mathcal{Z} \to \mathcal{M}$ be a VAE decoder.
- ϵ_θ be the denoiser (UNet).
- $\{\bar{\alpha}_t\}_{t=1}^T$ be the parameters of diffusion process.
- $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ be the tangent space of $\boldsymbol{x}$.
- $a\mathcal{M} = \{a\boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{M}\}$.
- $d(\boldsymbol{x}, \mathcal{M}) = \inf_{\boldsymbol{x}' \in \mathcal{M}} \|\boldsymbol{x} - \boldsymbol{x}'\|_2$.
- $\mathcal{B}(\mathcal{M}; r) = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid d(\boldsymbol{x}, \mathcal{M}) < r\right\}$.
- $\mathcal{N}_{\mathcal{T}}(\boldsymbol{x})$ be the neighborhood of tangent space $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$, which is defined as

$$\mathcal{N}_{\mathcal{T}}(\boldsymbol{x}) = \left\{\boldsymbol{x}' \in \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \mid \|\boldsymbol{x}' - \boldsymbol{x}\|_2 < r\right\}.$$

Manifold Hypothesis

Manifold Hypothesis (MH)

The image distribution of interest lies on a d-dimensional manifold $\mathscr{M}$ that is embedded in a n-dimensional ambient space $\mathbb{R}^n$ such that $d \ll n$.



What is Diffusion Model Learning?

The goal of diffusion model is to create the data from the noise [Song Yang].

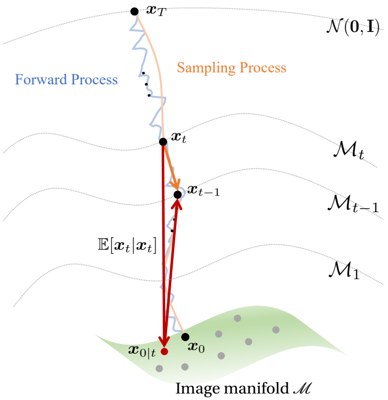
- **Forward Process:**

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t.$$

- **Sampling Process:**

$$\boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta(\boldsymbol{x}_t, t) + \sigma_t\boldsymbol{\epsilon}_t,$$

where $\boldsymbol{x}_{0|t} := \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] = \frac{1}{\sqrt{\bar{\alpha}_t}}(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{x}_t, t)).$



Concentration of Noisy Samples

Concentration of Noisy Samples

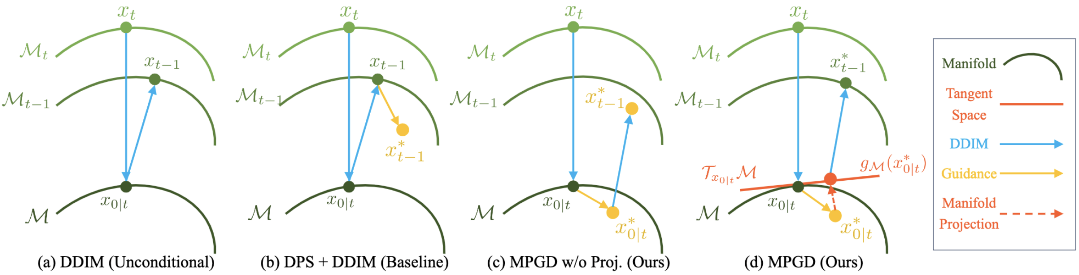
Consider the distribution of noisy data $p_t(\boldsymbol{x}_t)$ with **Linear-MH**. Then $p_t(\boldsymbol{x}_t)$ is probabilistically concentrated on the $(n-1)$-dimensional manifold $\mathscr{M}_t$. It's defined as

$$\mathscr{M}_t = \{\boldsymbol{x} \in \mathbb{R}^d \mid d(\boldsymbol{x}, \sqrt{\bar{\alpha}_t}\mathscr{M}) = \sqrt{(1-\bar{\alpha}_t)(n-d)}\}.$$

That is, for any $0 < \delta \leq 1$, there is an $0 < \epsilon = \epsilon(\delta, n, d) \leq 1$ such that

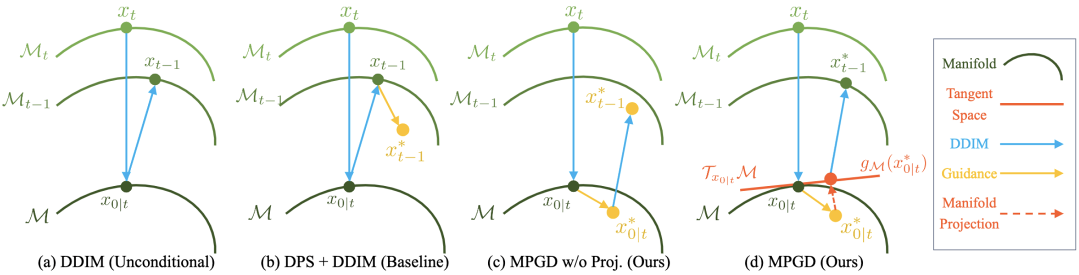
$$\mathbb{P}(\boldsymbol{x}_t \in \mathscr{B}(\mathscr{M}_t; \epsilon\sqrt{(1-\bar{\alpha}_t)(n-d)})) \geq 1 - \delta.$$

Prior Works vs MPGD



- **DDIM:** $\boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta(\boldsymbol{x}_t, t) + \sigma_t\boldsymbol{\epsilon}_t.$
- **DPS:** $\boldsymbol{x}^*_{t-1} = \boldsymbol{x}_{t-1} - \gamma\nabla_{\boldsymbol{x}_{t-1}}\mathscr{L}(\boldsymbol{x}_{0|t}, \boldsymbol{y}).$

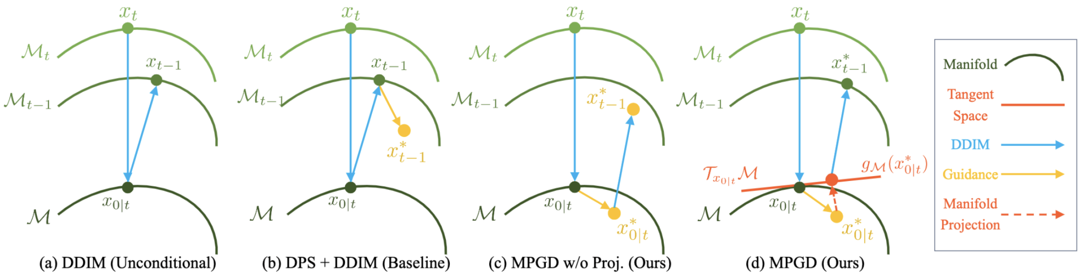
Prior Works vs MPGD



Problem of DPS

$\boldsymbol{x}^*_{t-1}$ dosen't lie on $\mathscr{M}_{t-1}$.

Prior Works vs MPGD (Conti.)



Question

1. How to project the $\boldsymbol{x}^*_{0|t}$ into tangent space $\mathscr{T}_{\boldsymbol{x}_{0|t}}\mathscr{M}$?
2. Does $\boldsymbol{x}^*_{t-1}$ lie on $\mathscr{M}_{t-1}$?

Outline

1. Introduction
   1.1 Inverse Problem
2. Background
   2.1 Manifold Hypothesis
   2.2 Diffusion Model
   2.3 Prior Works
3. Proposed Method
   3.1 MPGD w/o Proj.
   3.2 MPGD-AE
   3.3 MPGD-Z
   3.4 MPGD-LDM
4. Experiments
5. Conclusions

Overview of MPGD

4-type of MPGD

1. **MPGD w/o Proj.**: Pixel-domain optimization
2. **MPGD-AE**: Pixel-domain optimization along the tangent space via Perfect Autoencoder (i.e., $\mathscr{D} \odot \mathscr{E} = \mathrm{id}$)
3. **MPGD-Z**: Latent optimization via Perfect Autoencoder
4. **MPGD-LDM**: Latent optimization via Latent Diffusion Model (LDM)

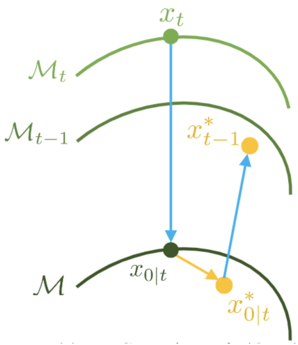
Type I: MPGD without Projection

Let the step size $\gamma_t > 0$. For $t = T, T-1, \ldots, 1$,

$$\boldsymbol{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{x}_t, t)\right)$$

$$\boldsymbol{x}^*_{0|t} = \boldsymbol{x}_{0|t} - \gamma_t \nabla_{\boldsymbol{x}_{0|t}} \mathscr{L}(\boldsymbol{x}_{0|t}, \boldsymbol{y})$$

$$\boldsymbol{x}^*_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}^*_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta(\boldsymbol{x}_t, t) + \sigma_t \boldsymbol{\epsilon}_t.$$



Type II: MPGD-AE (MPGD with Projection)

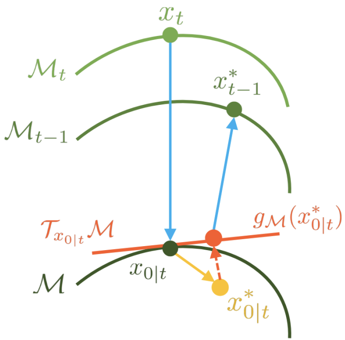
Theorem

With **Linear-MH**, if $\nabla_{\boldsymbol{x}_{0|t}} \mathcal{L}(\boldsymbol{x}_{0|t}, \boldsymbol{y}) \in \mathcal{T}_{\boldsymbol{x}_{0|t}} \mathcal{M}$. Then the marginal distribution of $\boldsymbol{x}_{t-1}$ is probabilistically concentrated on $\mathcal{M}_{t-1}$.

$$\boldsymbol{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\boldsymbol{x}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\boldsymbol{x}_t, t) \right)$$

$$\boldsymbol{x}^*_{0|t} = \boldsymbol{x}_{0|t} - \gamma_t \nabla_{\boldsymbol{x}_{0|t}} \mathcal{L}(\boldsymbol{x}_{0|t}, \boldsymbol{y})$$

$$\boldsymbol{x}^*_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \boldsymbol{x}^*_{0|t} + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \epsilon_\theta(\boldsymbol{x}_t, t) + \sigma_t \boldsymbol{\epsilon}_t.$$



Type II: MPGD-AE (MPGD with Projection) (Conti.)

Theorem (Manifold Projection with Perfect Autoencoder)

If $(\mathscr{E}, \mathscr{D})$ is a Perfect Autoencoder, then

$$\nabla_{\boldsymbol{x}_0} \mathscr{L}((\mathscr{D} \circ \mathscr{E})(\boldsymbol{x}_0), \boldsymbol{y}) = \left(\frac{\partial \mathscr{L}}{\partial \boldsymbol{x}'} \cdot \frac{\partial \mathscr{D}}{\partial \boldsymbol{z}} \cdot \frac{\partial \mathscr{E}}{\partial \boldsymbol{x}} \right) \Bigg|_{\boldsymbol{x}=\boldsymbol{x}_0, \mathbf{z}=\mathscr{E}(\boldsymbol{x}_0), \boldsymbol{x}'=\mathscr{D}(\boldsymbol{z})} \in \mathscr{T}_{\boldsymbol{x}_0} \mathscr{M}.$$

MPGD-AE

We can modify the update rules as

$$\boldsymbol{x}^*_{0|t} = \boldsymbol{x}_{0|t} - \gamma_t \nabla_{\boldsymbol{x}_{0|t}} \mathscr{L}((\mathscr{D} \circ \mathscr{E})(\boldsymbol{x}_{0|t}), \boldsymbol{y}).$$

Type III: MPGD-Z (Manipulating the Latents)

Theroem

If $(\mathscr{E}, \mathscr{D})$ is a Perfect Autoencoder, then

$$\mathscr{D}(\nabla_{\boldsymbol{z}} \mathscr{L}(\mathscr{D}(\boldsymbol{z}), \boldsymbol{y})) \in \mathscr{T}_{\boldsymbol{x}} \mathscr{M}.$$

MPGD-Z

We can modify the update rules as

$$\begin{aligned}
\boldsymbol{z}_{0|t} &= \mathscr{E}(\boldsymbol{x}_{0|t}), \\
\boldsymbol{z}^*_{0|t} &= \boldsymbol{z}_{0|t} - \gamma_t \nabla_{\boldsymbol{z}_{0|t}} \mathscr{L}(\mathscr{D}(\boldsymbol{z}_{0|t}), \boldsymbol{y}), \\
\boldsymbol{x}^*_{0|t} &= \mathscr{D}(\boldsymbol{z}_{0|t}).
\end{aligned}$$

Type IV: MPGD-LDM

MPGD-LDM

Given $\boldsymbol{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For $t = T, T-1, \ldots, 1$,

$$\boldsymbol{z}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t)\right),$$

$$\boldsymbol{z}^*_{0|t} = \boldsymbol{z}_{0|t} - \gamma_t \nabla_{\boldsymbol{z}_{0|t}} \mathcal{L}(\mathcal{D}(\boldsymbol{z}_{0|t}), \boldsymbol{y}),$$

$$\boldsymbol{z}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{z}^*_{0|t} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t) + \sigma_t \boldsymbol{\epsilon}_t.$$

Finally, decode the latent $\boldsymbol{z}_0$ to latent space, then we can get $\boldsymbol{x}_0 = \mathcal{D}(\boldsymbol{z}_0)$.

Summary

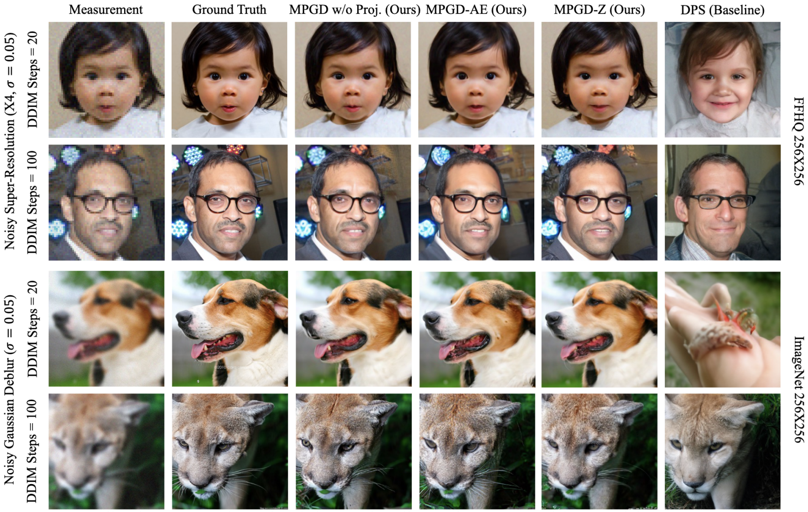
Type	By Encoder	By Decoder	Optimizing Space
MPGD w/o Proj.	No	No	Pixel Space
MPGD-AE	Every Timestep	Every Timestep	Pixel Space
MPGD-Z	Every Timestep	Every Timestep	Latent Space
MPGD-LDM	No	Last Timestep	Latent Space

Table: Comparison for 4-types of MPGD

Outline

1. Introduction
   1.1 Inverse Problem
2. Background
   2.1 Manifold Hypothesis
   2.2 Diffusion Model
   2.3 Prior Works
3. Proposed Method
   3.1 MPGD w/o Proj.
   3.2 MPGD-AE
   3.3 MPGD-Z
   3.4 MPGD-LDM
4. Experiments
5. Conclusions

Inverse Problems



Inverse Problems (Conti.)

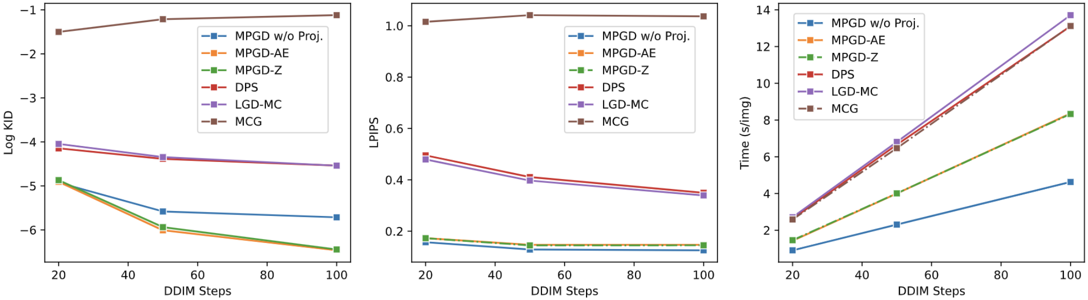
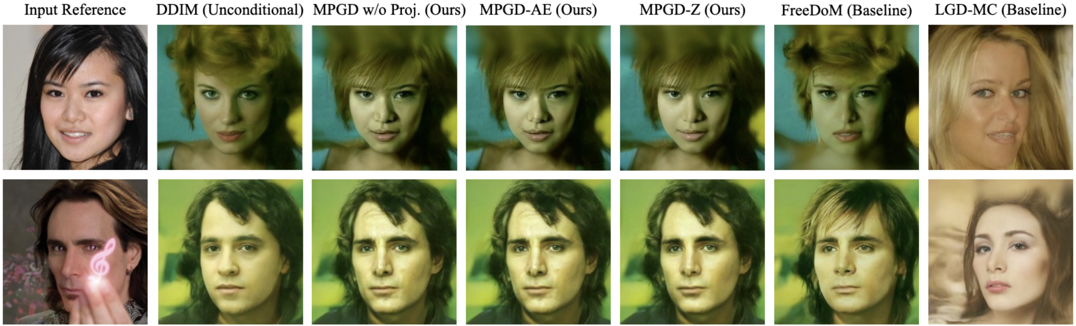


Figure: Quantitative results.

FaceID Guidance



FaceID Guidance (Conti.)

Method	KID↓	FaceID↓	Time↓
DDIM	0.0442	1.3914	3.41s
FreeDoM	0.0452	0.5690	10.65s
LGD-MC	0.0448	0.6783	14.64s
MPGD	0.0473	**0.5163**	**5.82s**
MPGD-AE	0.0467	0.5309	7.78s
MPGD-Z	**0.0445**	0.5791	6.93s

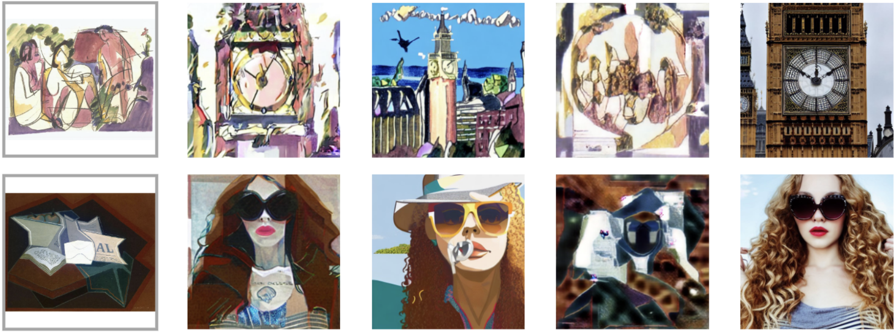
Figure: Quantitative results.

Style Guidance

Prompt | Input Reference | MPGD-LDM (Ours) | FreeDoM (Baseline) | LGD-MC (Baseline) | DDIM(Unconditional)

"a Big Ben clock towering over the city of London"

"a girl with long curly blonde hair and sunglasses"



Style Guidance (Conti.)

Method	Style↓	CLIP↑	Time↓	VRAM↓
DDIM	761.0	31.61	13.89s	10.80 GB
FreeDoM	498.8	30.14	26.50s	17.30 GB
LGD-MC	404.0	21.16	37.43s	31.65 GB
MPGD-LDM	441.0	26.61	**19.83s**	**15.53 GB**

Figure: Quantitative results.

Outline

1. Introduction
   1.1 Inverse Problem
2. Background
   2.1 Manifold Hypothesis
   2.2 Diffusion Model
   2.3 Prior Works
3. Proposed Method
   3.1 MPGD w/o Proj.
   3.2 MPGD-AE
   3.3 MPGD-Z
   3.4 MPGD-LDM
4. Experiments
5. Conclusions

Take Away

1. **Enhanced Efficiency and Quality:** MPGD prioritizes high-quality output with reduced computational and memory costs through manifold preserving guidance.
2. **Utilization of Autoencoders:** MPGD leverages pretrained autoencoders to address challenges in guided generation effectively.
3. **Optimization Strategies:** MPGD employs optimization strategies to improve sampling process efficiency.

Thank you!