The Emergence of Reproducibility and Consistency in Diffusion Models

Huijie Zhang

EECS, University of Michigan

October 28, 2023

Outline

1. Introduction
2. Reproducibility for Unconditional Diffusion Models
3. Correlation between Reproducibility, Memorizability & Generalizability
4. Reproducibility beyond Unconditional Diffusion Models
5. Future Directions

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

- **Rare in deep learning:**

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

- **Rare in deep learning:**
  - Representation Learning: Proved for only linear reproducibility [13].

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

- **Rare in deep learning:**
  - Representation Learning: Proved for only linear reproducibility [13].
  - Classification problem: has reproducibility for only similar network architecture [17].

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

- **Rare in deep learning:**
  - Representation Learning: Proved for only linear reproducibility [13].
  - Classification problem: has reproducibility for only similar network architecture [17].
  - Generative model: before this work, only for Variational autoencoders (VAE) with a factorized prior distribution over the latent variables [6].

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Model reproducibility in deep learning

- **Definition:** You can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project [1].

Quick Survey: Is there any reproducibility in your research area?

- **Rare in deep learning:**
  - Representation Learning: Proved for only linear reproducibility [13].
  - Classification problem: has reproducibility for only similar network architecture [17].
  - Generative model: before this work, only for Variational autoencoders (VAE) with a factorized prior distribution over the latent variables [6].
- **Very general in diffusion model**

[1] similar to the definition uniquely identifiable encoding in [6, 7, 13, 20]

Introduction of diffusion model

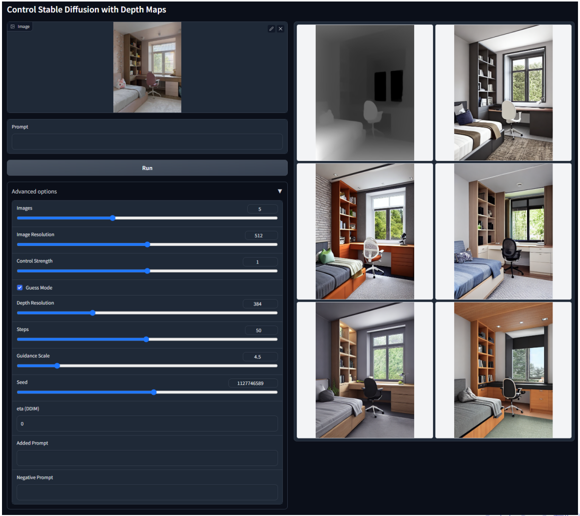
- **A powerfule generative model:**
  - **Stable Diffusion [14]**: Large text-to-image diffusion model

'An oil painting of a latent space.'

'An epic painting of Gandalf the Black summoning thunder and lightning in the mountains.'

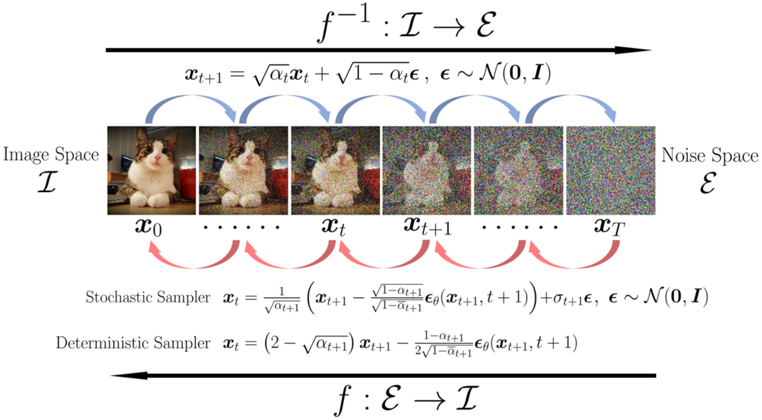
Introduction of diffusion model

- **A powerfule generative model:**
  - **ControlNet [22]**: Diffusion model with high flexiable guidance



Introduction of diffusion model

- **Definition:** A generative model $f : \mathcal{E} \mapsto \mathcal{I}$ mapping from the gaussian noise space $\mathcal{E}$ to the image manifold $\mathcal{I}$



Introduction of diffusion model

- **Training:** Only the denoiser funtion ϵ_θ requires training, following an easy pipeline:

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

Outline

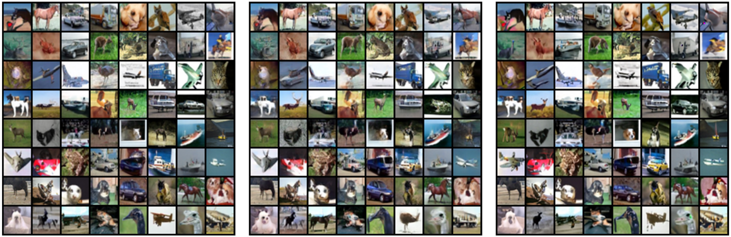
1. Introduction
2. Reproducibility for Unconditional Diffusion Models
3. Correlation between Reproducibility, Memorizability & Generalizability
4. Reproducibility beyond Unconditional Diffusion Models
5. Future Directions

Samples Visualization

Q1: *Starting from the* **same noise input**, *how are the generated data samples from various diffusion models related to each other?*

Samples Visualization

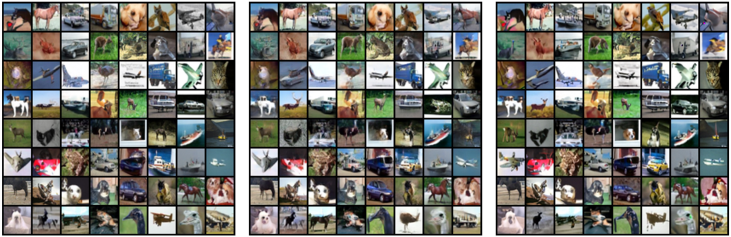
Q1: *Starting from the* **same noise input**, *how are the generated data samples from various diffusion models related to each other?*



(a) DDPMv4 (b) CT (c) U-ViT

Samples Visualization

Q1: *Starting from the* **same noise input**, *how are the generated data samples from various diffusion models related to each other?*



(a) DDPMv4 (b) CT (c) U-ViT

Training on the same dataset, sampling by a deterministic sampler.

Quantitative analysis

- **Metric**: We define *reproducibility (RP) score* to measure the this phenomenon:

$$\text{RP Score} := \mathbb{P}\left(\mathcal{M}_{\mathsf{SSCD}}(\boldsymbol{x}_1, \boldsymbol{x}_2) > 0.6\right),$$

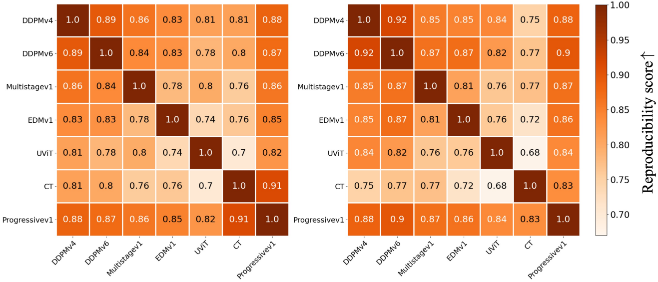
represents the *probability* of a generated sample pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ from two different diffusion models to have *self-supervised copy detection* (SSCD) similarity $\mathcal{M}_{\mathsf{SSCD}}$ larger than 0.6. We sampled 10K noise to estimate the probability. The SSCD similarity is first introduced in [12] to measure the replication between image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, which is defined as the following:

$$\mathcal{M}_{\mathsf{SSCD}}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\mathsf{SSCD}(\boldsymbol{x}_1) \cdot \mathsf{SSCD}(\boldsymbol{x}_2)}{||\mathsf{SSCD}(\boldsymbol{x}_1)||_2 \cdot ||\mathsf{SSCD}(\boldsymbol{x}_2)||_2}$$

where $\mathsf{SSCD}(\cdot)$ represents a neural descriptor for copy detection.

Quantitative analysis

- **Results**:

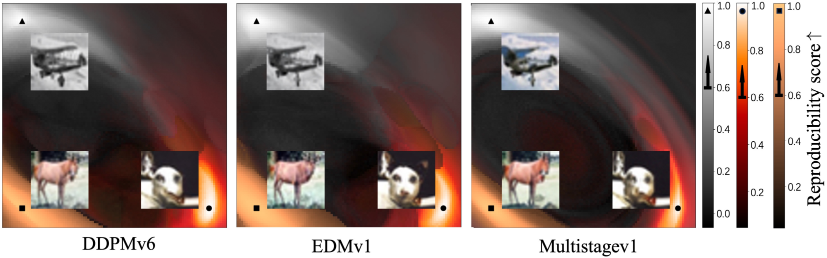


Reproducibility score MAE score

C2: *Diffusion models consistently generate* **nearly identical contents**, *irrespective of network architectures, training and sampling procedures, and perturbation kernels.*

Mapping from Noise Hyperplane to Image Manifold

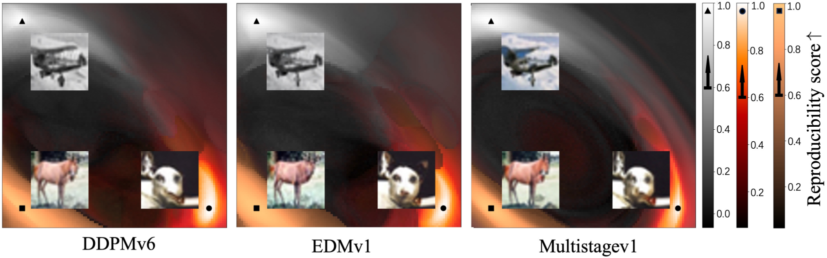
- **Results**:



Pick three initial noises $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3)$ to generate clear images $(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \boldsymbol{x}_3)$ in the image manifold $\mathcal{I}$. Second, we create a 2D noise hyperplane with $\boldsymbol{\epsilon}\left(\alpha, \beta\right) = \alpha \cdot \left(\boldsymbol{\epsilon}_2 - \boldsymbol{\epsilon}_1\right) + \beta \cdot \left(\boldsymbol{\epsilon}_3 - \boldsymbol{\epsilon}_1\right) + \boldsymbol{\epsilon}_1$. And utilize them to generate images $\boldsymbol{x}(\alpha, \beta)$, and RP Score $:= max_{k \in \{1,2,3\}}[\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_k, \boldsymbol{x}\left(\alpha, \beta\right))]$

Mapping from Noise Hyperplane to Image Manifold

- **Results**:

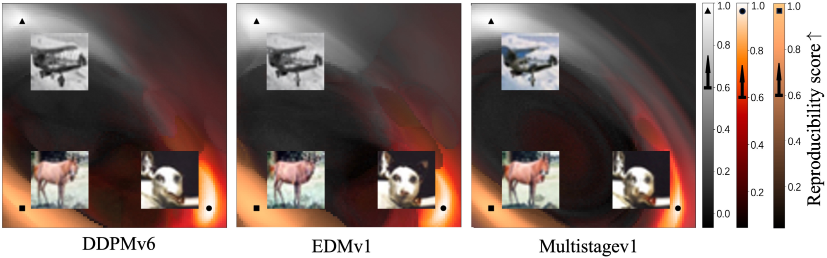


Pick three initial noises $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3)$ to generate clear images $(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \boldsymbol{x}_3)$ in the image manifold $\mathcal{I}$. Second, we create a 2D noise hyperplane with $\boldsymbol{\epsilon}(\alpha, \beta) = \alpha \cdot (\boldsymbol{\epsilon}_2 - \boldsymbol{\epsilon}_1) + \beta \cdot (\boldsymbol{\epsilon}_3 - \boldsymbol{\epsilon}_1) + \boldsymbol{\epsilon}_1$. And utilize them to generate images $\boldsymbol{x}(\alpha, \beta)$, and RP Score $:= max_{k\in\{1,2,3\}}[\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_k, \boldsymbol{x}(\alpha, \beta))]$

- **Conclusions**:
  - Similar unique encoding maps across different network architectures.

Mapping from Noise Hyperplane to Image Manifold

- **Results**:



Pick three initial noises $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3)$ to generate clear images $(\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \boldsymbol{x}_3)$ in the image manifold $\mathcal{I}$. Second, we create a 2D noise hyperplane with $\boldsymbol{\epsilon}(\alpha, \beta) = \alpha \cdot (\boldsymbol{\epsilon}_2 - \boldsymbol{\epsilon}_1) + \beta \cdot (\boldsymbol{\epsilon}_3 - \boldsymbol{\epsilon}_1) + \boldsymbol{\epsilon}_1$. And utilize them to generate images $\boldsymbol{x}(\alpha, \beta)$, and RP Score $:= max_{k \in \{1,2,3\}}[\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_k, \boldsymbol{x}(\alpha, \beta))]$

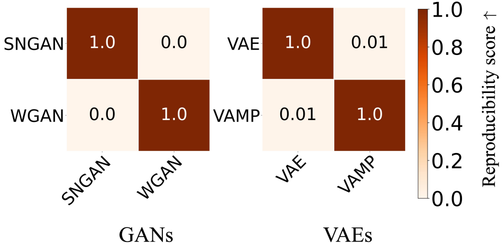
- **Conclusions**:
  - Similar unique encoding maps across different network architectures.
  - Local Lipschitzness of the unique encoding from noise to image space.

Reproducibility for other generative models

- **Question**: Dose model reproducibility appear to other generative models?

Reproducibility for other generative models

- **Question**: Dose model reproducibility appear to other generative models?

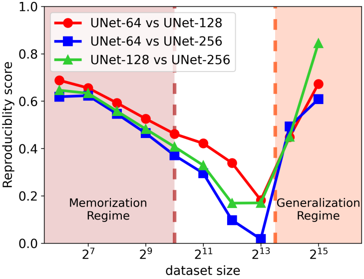


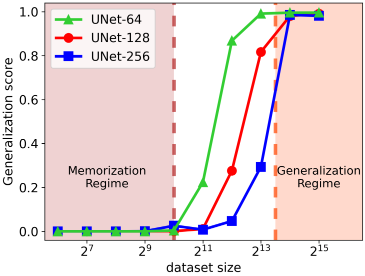
- **Conclusion**: Reproducibility doesn't appear for GAN and general VAE;

Outline

1. Introduction
2. Reproducibility for Unconditional Diffusion Models
3. Correlation between Reproducibility, Memorizability & Generalizability
4. Reproducibility beyond Unconditional Diffusion Models
5. Future Directions

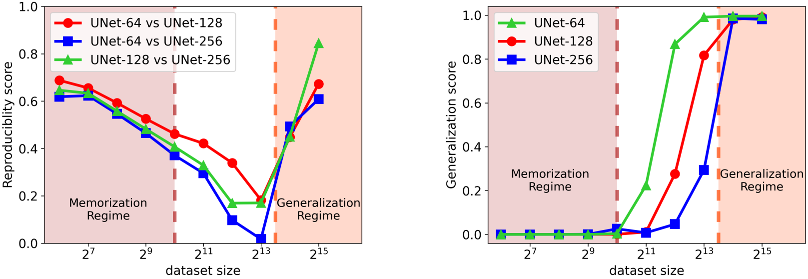
“Memorization” and “Generalization” regimes





generalization (GL) score $:= 1 - \mathbb{P}\left(\max_{i \in [N]} \left[\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}, \boldsymbol{y}_i)\right] > 0.6\right)$ between the generated sample $\boldsymbol{x}$ and all samples from training dataset $\{\boldsymbol{y}_i\}_{i=1}^{N}$.

"Memorization" and "Generalization" regimes



generalization (GL) score $:= 1 - \mathbb{P}\left(\max_{i\in[N]}\left[\mathcal{M}_{\mathsf{SSCD}}(\boldsymbol{x},\boldsymbol{y}_i)\right] > 0.6\right)$ between the generated sample $\boldsymbol{x}$ and all samples from training dataset $\{\boldsymbol{y}_i\}_{i=1}^{N}$.

C3: *The reproducibility of diffusion models manifests in two distinct training regimes, both* **strongly correlated** *with the model's generalizability.*

Theory for reproducibility in Memorization Regime

Theorem (1)

Suppose we train a diffusion model denoiser function $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ with parameter $\boldsymbol{\theta}$ on a training dataset $\{\boldsymbol{y}_i\}_{i=1}^N$ of N-samples, by minimizing the training loss

$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\epsilon}_{\boldsymbol{\theta}}; t) = \mathbb{E}_{\boldsymbol{x}_0 \sim p_{data}(\boldsymbol{x})} \mathbb{E}_{\boldsymbol{x} \sim p_t(\boldsymbol{x}|\boldsymbol{x}_0)} [||\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)||^2], \tag{1}$$

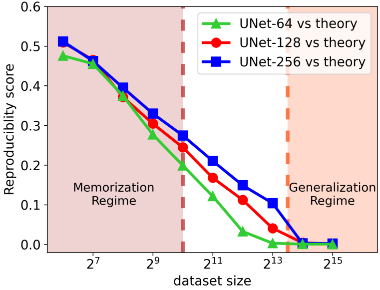
assuming data $\boldsymbol{x}_0$ follows a multi-delta distribution $p_{data}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^N \delta(\boldsymbol{x} - \boldsymbol{y}_i)$, and the perturbation kernel $p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; s_t \boldsymbol{x}_0, s_t^2 \sigma_t^2 \mathbf{I})$ with parameters s_t, σ_t. Then we show the optimal denoiser $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^*(\boldsymbol{x}; t) = \arg\min_{\boldsymbol{\epsilon}_{\boldsymbol{\theta}}} \mathcal{L}(\boldsymbol{\epsilon}_{\boldsymbol{\theta}}; t)$ is:

$$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^*(\boldsymbol{x}; t) = \frac{1}{s_t \sigma_t} \left[\boldsymbol{x} - s_t \frac{\sum_{i=1}^N \mathcal{N}(\boldsymbol{x}; s_t \boldsymbol{y}_i, s_t^2 \sigma_t^2 \mathbf{I}) \boldsymbol{y}_i}{\sum_{i=1}^N \mathcal{N}(\boldsymbol{x}; s_t \boldsymbol{y}_i, s_t^2 \sigma_t^2 \mathbf{I})} \right]. \tag{2}$$

Moreover, suppose a trained diffusion model could converge to the optimal denoiser $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^*(\boldsymbol{x}; t)$ and we use a deterministic ODE sampler to generate images using $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^*(\boldsymbol{x}; t)$, then $f : \mathcal{E} \mapsto \mathcal{I}$, which is determined by the $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^*(\boldsymbol{x}; t)$ and the ODE sampler, is an invertiable mapping and the inverse mapping f^{-1} is a unique identifiable encoding.

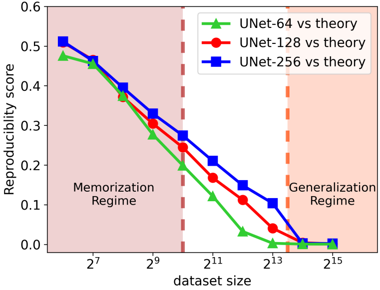
Experimental verification of Theorem (1)

- Reproducibility score between theoretical results and experimental results.



Experimental verification of Theorem (1)

- Reproducibility score between theoretical results and experimental results.



- **Conclusion**: Diffusion model could converge to the theoretical solution when model capacity is enough.

Outline

1. Introduction
2. Reproducibility for Unconditional Diffusion Models
3. Correlation between Reproducibility, Memorizability & Generalizability
4. Reproducibility beyond Unconditional Diffusion Models
5. Future Directions

More than unconditional diffusion model

C4: *Model reproducibility holds* **more generally** *across conditional diffusion models, diffusion models for inverse problems, the fine-tuning of diffusion models.*

Intro to conditional diffusion model

- Enable conditional generation, e.g. class condition, text-to-image, image-to-image translation.[14].

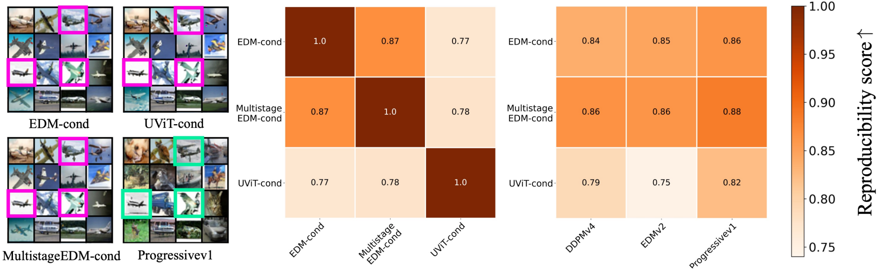
Intro to conditional diffusion model

- Enable conditional generation, e.g. class condition, text-to-image, image-to-image translation.[14].
- Change the denoiser from $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ to $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, c)$ for $c \in \mathcal{C}$. Total class $\mathcal{C}$ could be the class label, text, image, and so on.

Intro to conditional diffusion model

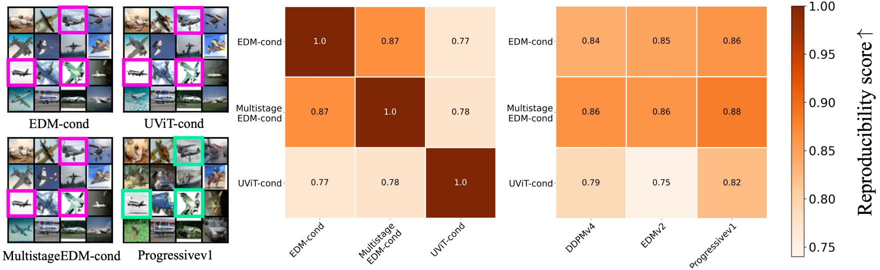
- Enable conditional generation, e.g. class condition, text-to-image, image-to-image translation.[14].
- Change the denoiser from $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ to $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, c)$ for $c \in \mathcal{C}$. Total class $\mathcal{C}$ could be the class label, text, image, and so on.
- Utilize $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, c)$ for both training and sampling.

Reproducibility of conditional diffusion model



(a) Visualization (b) Conditional diffusion models (c) Conditional and unconditional diffusion models

Reproducibility of conditional diffusion model

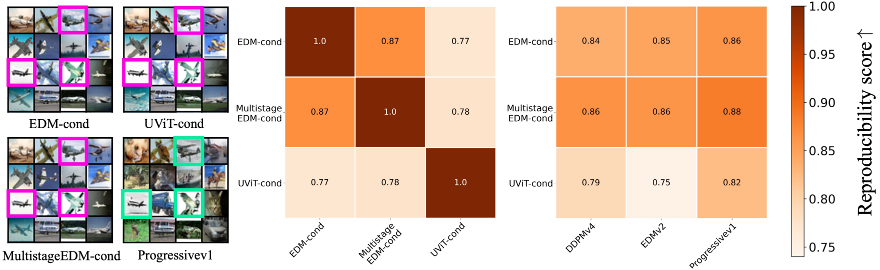


(a) Visualization (b) Conditional diffusion models (c) Conditional and unconditional diffusion models

RP$_{cond}$ Score $:= \mathbb{P}\left(\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_1^c, \boldsymbol{x}_2^c) > 0.6 \mid c \in \mathcal{C}\right)$, $(\boldsymbol{x}_1^c, \boldsymbol{x}_2^c)$ are generated by two conditional models from the same initial noise and conditioned on the class $c \in \mathcal{C}$

RP$_{between}$ Score $:= \mathbb{P}\left(\max_{c \in \mathcal{C}} [\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_1, \boldsymbol{x}_2^c)] > 0.6\right)$, for an unconditional generation $\boldsymbol{x}_1$ and conditional generation $\boldsymbol{x}_2^c$ starting from the same noise.

Reproducibility of conditional diffusion model



(a) Visualization (b) Conditional diffusion models (c) Conditional and unconditional diffusion models

RP$_{cond}$ Score $:= \mathbb{P}\left(\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_1^c, \boldsymbol{x}_2^c) > 0.6 \mid c \in \mathcal{C}\right)$, $(\boldsymbol{x}_1^c, \boldsymbol{x}_2^c)$ are generated by two conditional models from the same initial noise and conditioned on the class $c \in \mathcal{C}$

RP$_{between}$ Score $:= \mathbb{P}\left(\max_{c \in \mathcal{C}}[\mathcal{M}_{\text{SSCD}}(\boldsymbol{x}_1, \boldsymbol{x}_2^c)] > 0.6\right)$, for an unconditional generation $\boldsymbol{x}_1$ and conditional generation $\boldsymbol{x}_2^c$ starting from the same noise.

Model reproducibility of conditional models is evident and linked with unconditional counterparts.

Intro to diffusion model for inverse problem

- Inverse problem: reconstruct an unknown signal $\boldsymbol{u}$ from the measurements $\boldsymbol{z}$ of the form $\boldsymbol{z} = \mathcal{A}(\boldsymbol{u}) + \boldsymbol{\eta}$, where $\mathcal{A}$ denotes some (given) sensing operator and $\boldsymbol{\eta}$ is the noise.
- Enable conditional generation with only pre-trained unconditional denoiser $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$

Intro to diffusion model for inverse problem

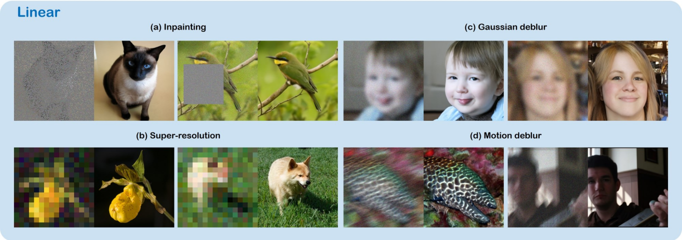
- Inverse problem: reconstruct an unknown signal $\boldsymbol{u}$ from the measurements $\boldsymbol{z}$ of the form $\boldsymbol{z} = \mathcal{A}(\boldsymbol{u}) + \boldsymbol{\eta}$, where $\mathcal{A}$ denotes some (given) sensing operator and $\boldsymbol{\eta}$ is the noise.
- Enable conditional generation with only pre-trained unconditional denoiser $\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_t, t)$
  $\boldsymbol{x}_t, \hat{\boldsymbol{x}}_0 \leftarrow \text{DeterministicSampler}(\boldsymbol{\epsilon_\theta}, \boldsymbol{x}_{t+1}, t+1,)$

Intro to diffusion model for inverse problem

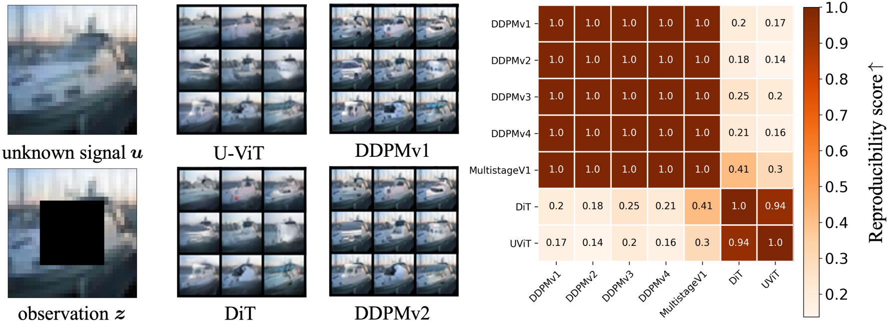
- Inverse problem: reconstruct an unknown signal $\boldsymbol{u}$ from the measurements $\boldsymbol{z}$ of the form $\boldsymbol{z} = \mathcal{A}(\boldsymbol{u}) + \boldsymbol{\eta}$, where $\mathcal{A}$ denotes some (given) sensing operator and $\boldsymbol{\eta}$ is the noise.
- Enable conditional generation with only pre-trained unconditional denoiser $\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_t, t)$
  $\boldsymbol{x}_t, \hat{\boldsymbol{x}}_0 \leftarrow \text{DeterministicSampler}(\boldsymbol{\epsilon_\theta}, \boldsymbol{x}_{t+1}, t+1,)$
  $\boldsymbol{x}_t \leftarrow \boldsymbol{x}_t - \xi_t \nabla_{\boldsymbol{x}_t} ||\boldsymbol{u} - \mathcal{A}\left(\hat{\boldsymbol{x}}_0\right)||_2^2$

Intro to diffusion model for inverse problem

- Inverse problem: reconstruct an unknown signal $\boldsymbol{u}$ from the measurements $\boldsymbol{z}$ of the form $\boldsymbol{z} = \mathcal{A}(\boldsymbol{u}) + \boldsymbol{\eta}$, where $\mathcal{A}$ denotes some (given) sensing operator and $\boldsymbol{\eta}$ is the noise.
- Enable conditional generation with only pre-trained unconditional denoiser $\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_t, t)$
  $\boldsymbol{x}_t, \hat{\boldsymbol{x}}_0 \leftarrow \text{DeterministicSampler}(\boldsymbol{\epsilon_\theta}, \boldsymbol{x}_{t+1}, t+1,)$
  $\boldsymbol{x}_t \leftarrow \boldsymbol{x}_t - \xi_t \nabla_{\boldsymbol{x}_t} ||\boldsymbol{u} - \mathcal{A}\left(\hat{\boldsymbol{x}}_0\right)||_2^2$
- Diffusion Posterior Sampling (DPS) [1]



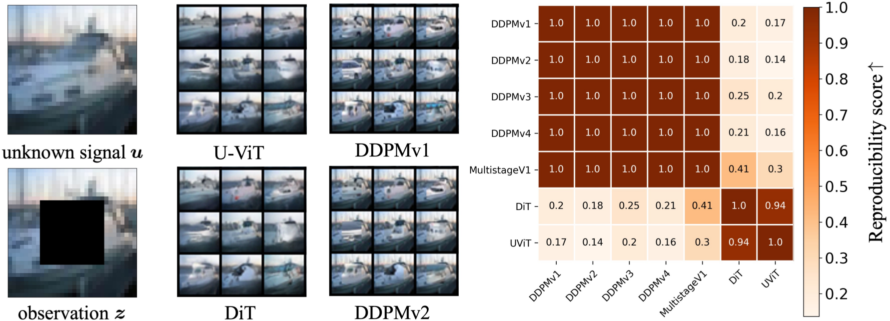
Reproducibility of diffusion model for inverse problem



(a) Visualization

(b) Reproducibility Score

Reproducibility of diffusion model for inverse problem



(a) Visualization (b) Reproducibility Score

Model reproducibility largely holds only within the same type of network architectures.

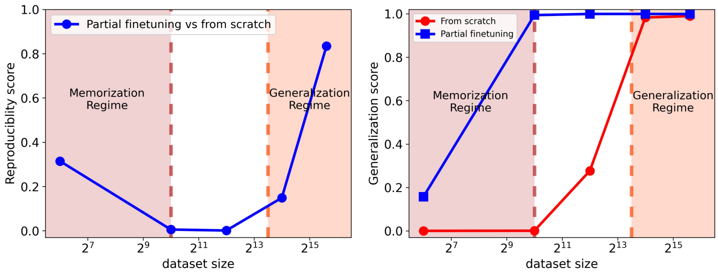
Intro to diffusion model fine-tuning

- Pre-trained large diffusion model (e.g. stable diffusion), fine-tuning only part of the diffusion model (e.g. attention layer, txt embedding.) and on few-shot images.

Intro to diffusion model fine-tuning

- Pre-trained large diffusion model (e.g. stable diffusion), fine-tuning only part of the diffusion model (e.g. attention layer, txt embedding.) and on few-shot images.
- Obtain incredible generalizability, e.g. DreamBooth [15].

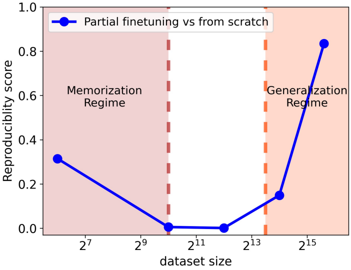
Reproducibility of diffusion model fine-tuning



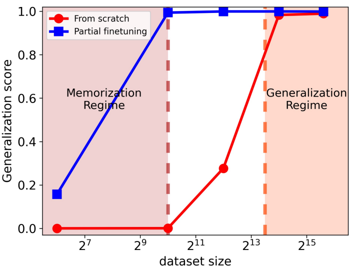
(a) Reproducibilty (b) Generalizability

Pretrained on CIFAR-100, fine-tuning on CIFAR-10. Only fine-tuning the attention layer.

Reproducibility of diffusion model fine-tuning



(a) Reproducibilty



(b) Generalizability

Pretrained on CIFAR-100, fine-tuning on CIFAR-10. Only fine-tuning the attention layer.

Partial fine-tuning reduces reproducibility but improves generalizability in "memorization regime".

Outline

1 Introduction

2 Reproducibility for Unconditional Diffusion Models

3 Correlation between Reproducibility, Memorizability & Generalizability

4 Reproducibility beyond Unconditional Diffusion Models

5 Future Directions

Future Direction

- **Practical application** Controllable data generation process through noise space.

Future Direction

- **Practical application** Controllable data generation process through noise space.
  - For text-driven image generation tasks, by leveraging the same text embeddings while varying noise inputs, potential applications including adversarial attacks [24], robust defense [23], and copyright protection [19, 18]

Future Direction

- **Practical application** Controllable data generation process through noise space.
  - For text-driven image generation tasks, by leveraging the same text embeddings while varying noise inputs, potential applications including adversarial attacks [24], robust defense [23], and copyright protection [19, 18]
  - In the context of solving inverse problems, we could select the input noise for reducing the uncertainty and variance in our signal reconstruction [5, 2, 10].

Future Direction

- **Practical application** Controllable data generation process through noise space.
  - For text-driven image generation tasks, by leveraging the same text embeddings while varying noise inputs, potential applications including adversarial attacks [24], robust defense [23], and copyright protection [19, 18]
  - In the context of solving inverse problems, we could select the input noise for reducing the uncertainty and variance in our signal reconstruction [5, 2, 10].
- **Theoretical understandings** Identifiable mapping from noise space to image space in generalization regime, this question might related to:

Future Direction

- **Practical application** Controllable data generation process through noise space.
  - For text-driven image generation tasks, by leveraging the same text embeddings while varying noise inputs, potential applications including adversarial attacks [24], robust defense [23], and copyright protection [19, 18]
  - In the context of solving inverse problems, we could select the input noise for reducing the uncertainty and variance in our signal reconstruction [5, 2, 10].
- **Theoretical understandings** Identifiable mapping from noise space to image space in generalization regime, this question might related to:
  - Analyzing the generalizability of diffusion models [21].

Future Direction

- **Practical application** Controllable data generation process through noise space.
  - For text-driven image generation tasks, by leveraging the same text embeddings while varying noise inputs, potential applications including adversarial attacks [24], robust defense [23], and copyright protection [19, 18]
  - In the context of solving inverse problems, we could select the input noise for reducing the uncertainty and variance in our signal reconstruction [5, 2, 10].
- **Theoretical understandings** Identifiable mapping from noise space to image space in generalization regime, this question might related to:
  - Analyzing the generalizability of diffusion models [21].
  - Connection between diffusion models and the Schrödinger bridge [16, 3, 11, 4, 9, 8] (an optimal transport problem).

Q & A

H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye.
Diffusion posterior sampling for general noisy inverse problems.
arXiv preprint arXiv:2209.14687, 2022.

H. Chung and J. C. Ye.
Score-based diffusion models for accelerated mri.
Medical Image Analysis, 80:102479, 2022.

V. De Bortoli, J. Thornton, J. Heng, and A. Doucet.
Diffusion schrödinger bridge with applications to score-based generative modeling.
Advances in Neural Information Processing Systems, 34:17695–17709, 2021.

M. Delbracio and P. Milanfar.
Inversion by direct iteration: An alternative to denoising diffusion for image restoration.
arXiv preprint arXiv:2303.11435, 2023.

A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir.

Robust compressed sensing mri with deep generative priors.

Advances in Neural Information Processing Systems, 34:14938–14954, 2021.

I. Khemakhem, D. Kingma, R. Monti, and A. Hyvarinen.

Variational autoencoders and nonlinear ica: A unifying framework.

In *International Conference on Artificial Intelligence and Statistics*, pages 2207–2217. PMLR, 2020.

I. Khemakhem, R. Monti, D. Kingma, and A. Hyvarinen.

Ice-beem: Identifiable conditional energy-based deep models based on nonlinear ica.

Advances in Neural Information Processing Systems, 33:12768–12778, 2020.

V. Khrulkov, G. Ryzhakov, A. Chertkov, and I. Oseledets.

Understanding ddpm latent codes through optimal transport.

arXiv preprint arXiv:2202.07477, 2022.

G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar.

I^{2}sb: Image-to-image schr\odinger bridge.

arXiv preprint arXiv:2302.05872, 2023.

G. Luo, M. Blumenthal, M. Heide, and M. Uecker.
Bayesian mri reconstruction with joint uncertainty estimation using diffusion models.
Magnetic Resonance in Medicine, 90(1):295–311, 2023.

Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön.
Image restoration with mean-reverting stochastic differential equations.
arXiv preprint arXiv:2301.11699, 2023.

E. Pizzi, S. D. Roy, S. N. Ravindra, P. Goyal, and M. Douze.
A self-supervised descriptor for image copy detection.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14532–14542, 2022.

G. Roeder, L. Metz, and D. Kingma.
On linear identifiability of learned representations.
In *International Conference on Machine Learning*, pages 9030–9039. PMLR, 2021.

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer.
High-resolution image synthesis with latent diffusion models.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman.
Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023.

Y. Shi, V. De Bortoli, A. Campbell, and A. Doucet.
Diffusion schr\" odinger bridge matching.
arXiv preprint arXiv:2303.16852, 2023.

G. Somepalli, L. Fowl, A. Bansal, P. Yeh-Chiang, Y. Dar, R. Baraniuk, M. Goldblum, and T. Goldstein.
Can neural nets learn the same model twice? investigating reproducibility and double descent from the decision boundary perspective.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13699–13708, 2022.

G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein.
Diffusion art or digital forgery? investigating data replication in diffusion models.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6048–6058, 2023.

G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein.
Understanding and mitigating copying in diffusion models.
arXiv preprint arXiv:2305.20086, 2023.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole.
Score-based generative modeling through stochastic differential equations.
arXiv preprint arXiv:2011.13456, 2020.

M. Yi, J. Sun, and Z. Li.
On the generalization of diffusion model.

arXiv preprint arXiv:2305.14712, 2023.

L. Zhang, A. Rao, and M. Agrawala.
Adding conditional control to text-to-image diffusion models, 2023.

K. Zhu, J. Wang, J. Zhou, Z. Wang, H. Chen, Y. Wang, L. Yang, W. Ye, N. Z. Gong, Y. Zhang, et al.
Promptbench: Towards evaluating the robustness of large language models on adversarial prompts.
arXiv preprint arXiv:2306.04528, 2023.

A. Zou, Z. Wang, J. Z. Kolter, and M. Fredrikson.
Universal and transferable adversarial attacks on aligned language models.
arXiv preprint arXiv:2307.15043, 2023.