Demystifying Classifier-Free Guidance: Exploring the Mechanisms and Emerging Alternatives

Open DMQA Seminar
2024.08.16

조한샘

고려대학교
KOREA UNIVERSITY

발표자 소개

- 조한샘
  - ✓ Data Mining & Quality Analytics Lab
  - ✓ 석·박통합과정 (2020.09~)
- 관심 연구 분야
  - ✓ Diffusion Models
  - ✓ Image Editing
- Contact
  - ✓ chosam95@korea.ac.kr

CONTENTs

Introduction

Diffusion Models

- Diffusion model을 사용해서 고화질 고품질의 이미지 생성 가능

Imagen 3, 2024

Stable Diffusion 3, 2024

https://deepmind.google/technologies/imagen-3/.
https://stability.ai/news/stable-diffusion-3-medium

Introduction

Classifier-Free Guidance (CFG)

- CFG는 diffusion model이 고품질 이미지를 생성하기 위한 필수적인 테크닉

CFG X

CFG O

“Two cats on the ground”

Preliminary: Diffusion Models

Diffusion Models

- **Training**: forward process를 사용해서 이미지에 노이즈 추가 후 모델에 입력

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
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

Forward Process

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$$

x_0

x_t

Model

$\epsilon_\theta(x_t, t)$

$\epsilon \sim N(0,1)$

$E_{x_0,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, t)\|^2]$

Model output

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.

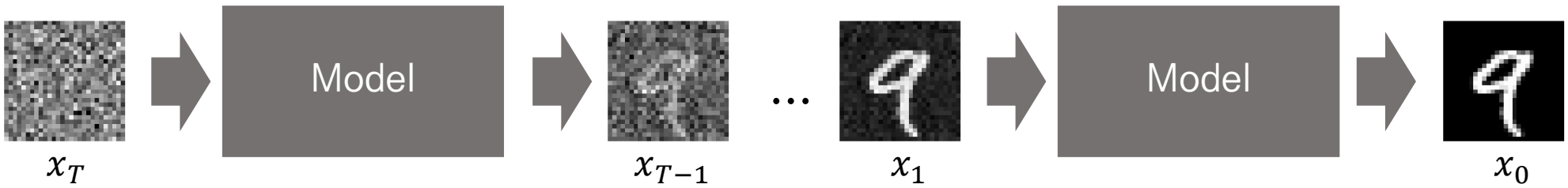
Preliminary: Diffusion Models

Diffusion Models

- **Sampling**: 각 시점 마다 노이즈 제거하며 데이터 생성

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$



Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.

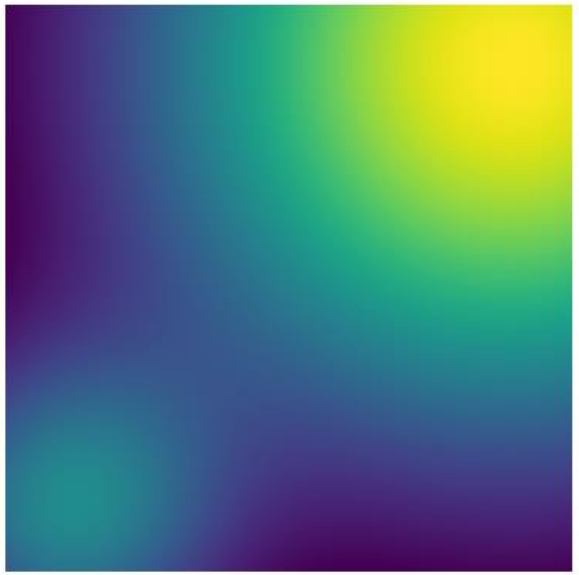
Preliminary: Diffusion Models

Diffusion Models and Score

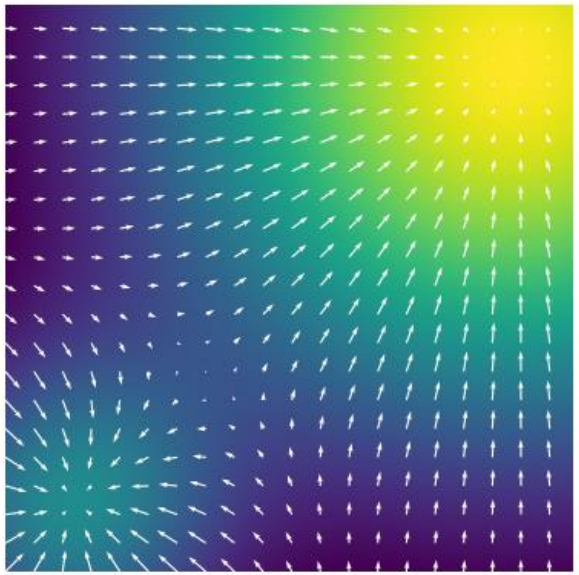
- Score: 로그 확률 밀도의 gradient
- Diffusion model의 학습 과정은 score matching으로 볼 수 있음

$$score = \nabla_x logp(x)$$

$$score = \nabla_{x_t} logp(x_t) \approx -\frac{\epsilon_\theta(x_t, t)}{\sqrt{1-\bar{\alpha}_t}}$$



Data distribution



Score

Preliminary: Diffusion Models

Unconditional to Conditional Diffusion Model

- Diffusion model의 output은 score로 치환 가능
- 특정 distribution에서 sampling하기 위해서는 특정 distribution의 score를 활용
- Unconditional diffusion model을 학습해도 classifier를 활용해서 conditional distribution에서 sampling이 가능

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Diffusion model output → Score

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$\log p(x|y) = \log p(y|x) + \log p(x) - \log p(y)$$

$$\nabla_x \log p(x|y) = \nabla_x \log p(y|x) + \nabla_x \log p\,(x)$$

Classifier gradient

Unconditional Model output

Classifier Guidance

Classifier Guidance (CG)

- Conditioning term의 scale을 키워보자
- Conditioning term의 scale을 키우는 것은 기존 distribution보다 sharpening된 distribution에서 sampling

$$\nabla_x \log p(x|y) = \underbrace{\nabla_x \log p\,(x)}_{\text{Unconditional Model output}} + \underbrace{\nabla_x \log p(y|x)}_{\text{Classifier gradient}}$$

$$\nabla_x \log \tilde{p}(x|y) = \nabla_x \log p(x) + w \cdot \nabla_x \log p(y|x)$$

Classifier Guidance

$$\nabla_x \log \tilde{p}(x|y) = \nabla_x \log p(x) + (w+1) \cdot \nabla_x \log p(y|x)$$

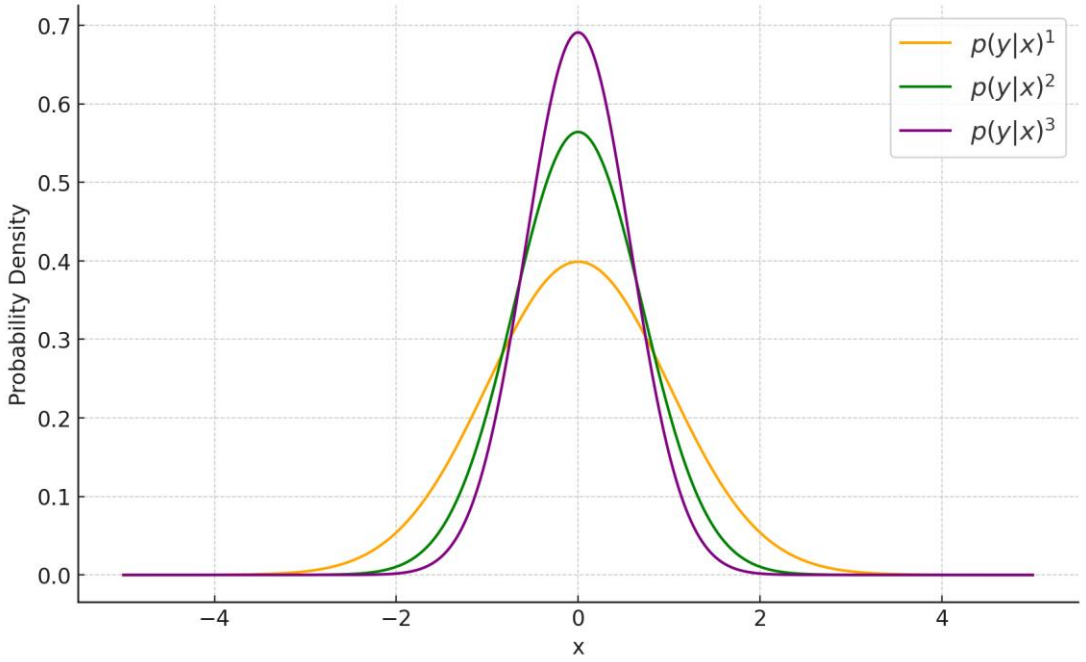
$$= \nabla_x [\log p(x) + \log p(y|x) + w \cdot \log p(y|x)]$$

$$= \nabla_x [\log p(x)p(y|x) + \log p(y|x)^w]$$

$$\propto \nabla_x [\log p(x|y) + \log p(y|x)^w]$$

$$= \nabla_x \log p(x|y)p(y|x)^w$$

$$\tilde{p}(x|y) \propto p(x|y)p(y|x)^w$$



Classifier-Free Guidance

Classifier-Free Guidance (CFG)

- CG는 사전에 학습된 classifier가 필요
- 사전에 학습된 classifier 없이 unconditional, conditional 모델을 활용해서 guidance를 줄 수 있음
- Conditional model 학습 시 일정 확률로 condition을 제거해서 unconditional 모델로 활용

$$\nabla_x \log \tilde{p}(x|y) = \nabla_x \log p(x) + w \cdot \nabla_x \log p(y|x)$$
$$= \nabla_x \log p(x) + w \cdot (\nabla_x \log p(x|y) - \nabla_x \log p(x))$$
$$= (1-w) \cdot \nabla_x \log p(x) + w \cdot \nabla_x \log p(x|y)$$

Unconditional Model output

Conditional Model output

Perturbed Attention Guidance

Perturbed Attention Guidance

- ECCV 2024, Korea University
- 5회 인용 (2024년 8월 16일)

Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance

Donghoon Ahn[*1] Hyoungwon Cho[*1] Jaewon Min[1]
Wooseok Jang[1] Jungwoo Kim[1] SeonHwa Kim[1]
Hyun Hee Park[2] Kyong Hwan Jin[†1] Seungryong Kim[†1]

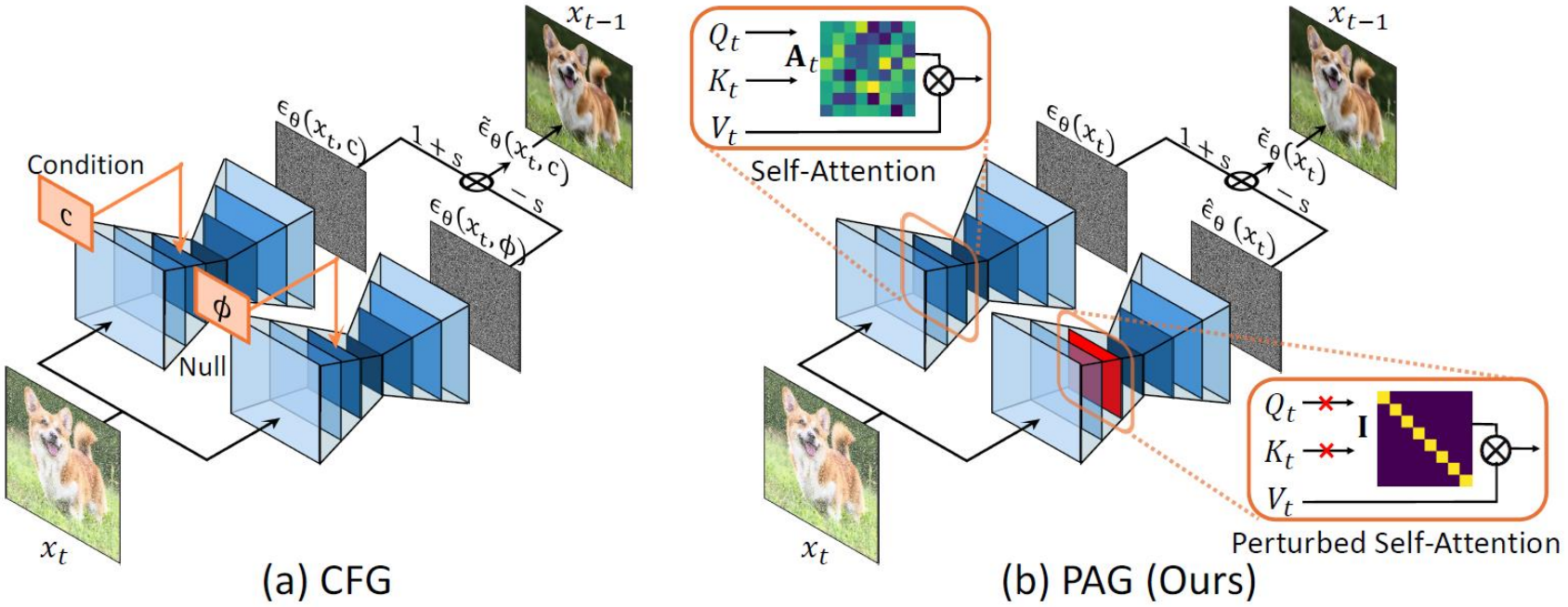
[1]Korea University [2]Samsung Electronics

https://ku-cvlab.github.io/Perturbed-Attention-Guidance

Perturbed Attention Guidance

Perturbed Attention Guidance

- CFG는 conditional model에서 이미지를 생성하는 경우만 활용가능
- Unconditional model에서 이미지를 생성할 때도 CFG와 같은 기능을 하는 방법론 제안



Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Perturbed Attention Guidance

Perturbed Attention Guidance

- CFG는 unconditional output을 anchor 삼아서 conditional output으로 extrapolate
- PAG는 원하지 않는 샘플을 내부적인 변형을 통해 생성 (self-attention layer를 identity matrix로)
- 원하는 샘플은 기존 diffusion model 활용

$$\nabla_x \log \tilde{p}(x|y)$$
$$= (1-w) \cdot \nabla_x \log p(x) + w \cdot \nabla_x \log p(x|y)$$

원하지 않는 sample

원하는 sample

$$\mathrm{SA}(Q_t, K_t, V_t) = \underbrace{\mathrm{Softmax}\left(\frac{Q_t K_t^T}{\sqrt{d}}\right)}_{structure} \overbrace{V_t}^{appearance} = \mathbf{A}_t V_t,$$

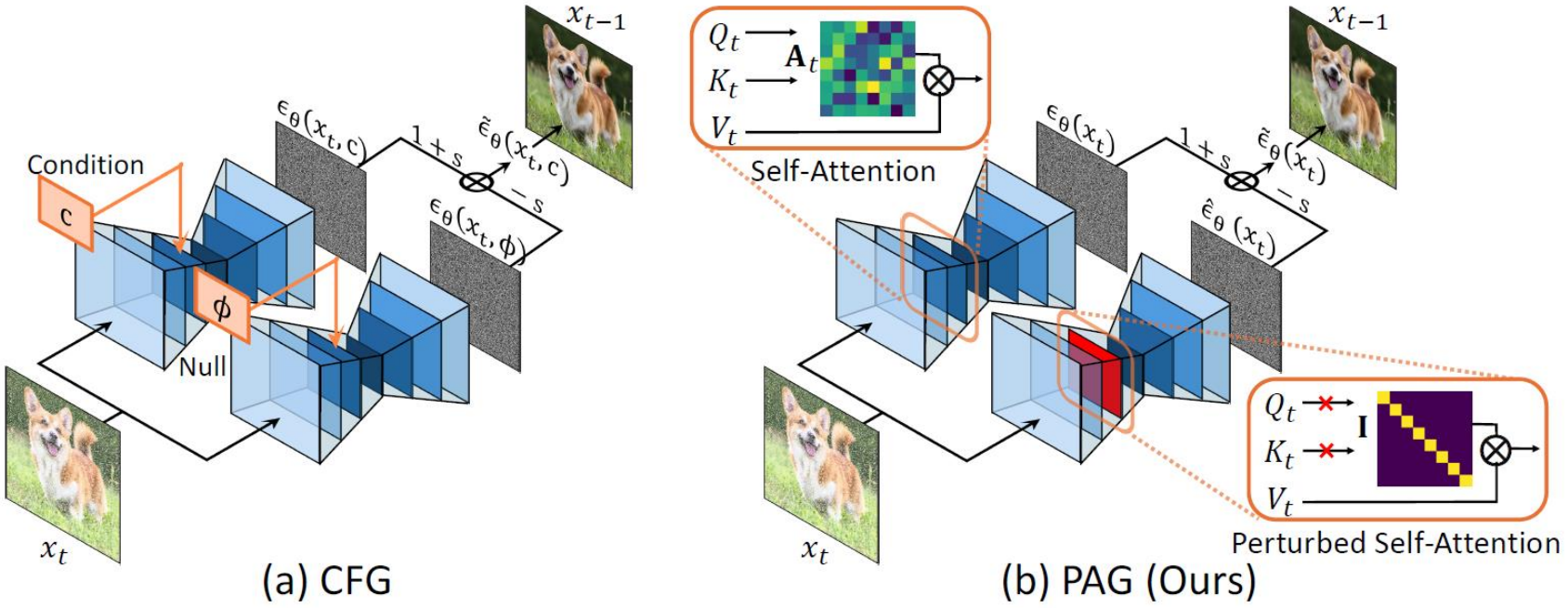
$$\mathrm{PSA}(Q_t, K_t, V_t) = \mathbf{I} V_t = V_t,$$

Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Perturbed Attention Guidance

Perturbed Attention Guidance

- CFG는 unconditional output을 anchor 삼아서 conditional output으로 extrapolate
- PAG는 원하지 않는 샘플을 내부적인 변형을 통해 생성 (self-attention layer를 identity matrix로)
- 원하는 샘플은 기존 diffusion model 활용



Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Perturbed Attention Guidance

Experiments

- PAG를 활용했을 때 생성되는 이미지 퀄리티 증가

Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Perturbed Attention Guidance

Experiments

- CFG와의 결합을 통해서 생성되는 이미지 퀄리티를 향상시킬 수도 있음

Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Perturbed Attention Guidance

Experiments

- Self attention을 변형하는 다양한 기법을 시도

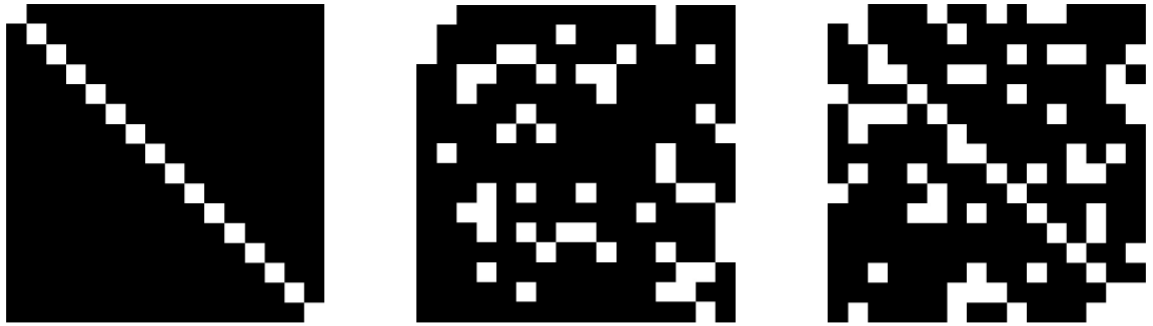


Fig. 10: Ablations study of self-attention map masking strategy. For the evaluation of FID [17], we sample 5K images from ADM [10] ImageNet [8] 256×256 unconditional model for each method. Black entries indicate the masked (set to $-\infty$) elements of the $Q_t K_t^T / \sqrt{d}$ component in Eq. 12 before the Softmax operation is applied. **(a)** Replacing attention map with identity matrix. **FID: 32.34**, **(b)** Random masking (ratio: 0.25). **FID: 40.20**, **(c)** Random masking off-diagonal entries (ratio: 0.25). **FID: 39.49**.

Ahn, D., Cho, H., Min, J., Jang, W., Kim, J., Kim, S., ... & Kim, S. (2024). Self-Rectifying Diffusion Sampling with Perturbed-Attention Guidance. arXiv preprint arXiv:2403.17377.

Autoguidance

Guiding a Diffusion Model with a Bad Version of Itself

- 6월 4일 arxiv 공개, NVIDA

Guiding a Diffusion Model with a Bad Version of Itself

Tero Karras
NVIDIA

Miika Aittala
NVIDIA

Tuomas Kynkäänniemi
Aalto University

Jaakko Lehtinen
NVIDIA, Aalto University

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

Abstract

The primary axes of interest in image-generating diffusion models are image quality, the amount of variation in the results, and how well the results align with a given condition, e.g., a class label or a text prompt. The popular classifier-free guidance approach uses an unconditional model to guide a conditional model, leading to simultaneously better prompt alignment and higher-quality images at the cost of reduced variation. These effects seem inherently entangled, and thus hard to control. We make the surprising observation that it is possible to obtain disentangled control over image quality without compromising the amount of variation by guiding generation using a smaller, less-trained version of the model itself rather than an unconditional model. This leads to significant improvements in ImageNet generation, setting record FIDs of 1.01 for 64×64 and 1.25 for 512×512, using publicly available networks. Furthermore, the method is also applicable to unconditional diffusion models, drastically improving their quality.

Karras, T., Aittala, M., Kynkäänniemi, T., Lehtinen, J., Aila, T., & Laine, S. (2024). Guiding a Diffusion Model with a Bad Version of Itself. arXiv preprint arXiv:2406.02507.

Autoguidance

Classifier Free Guidance

- Conditional model을 unconditional model보다 데이터 분포를 더 잘 학습함
- CFG는 데이터 분포를 잘 학습한 모델이 데이터 분포를 덜 학습한 모델에게 가이드 받는 기법으로 해석할 수 있음
- 해당 가이드는 high-density 영역으로 밀어주는 역할

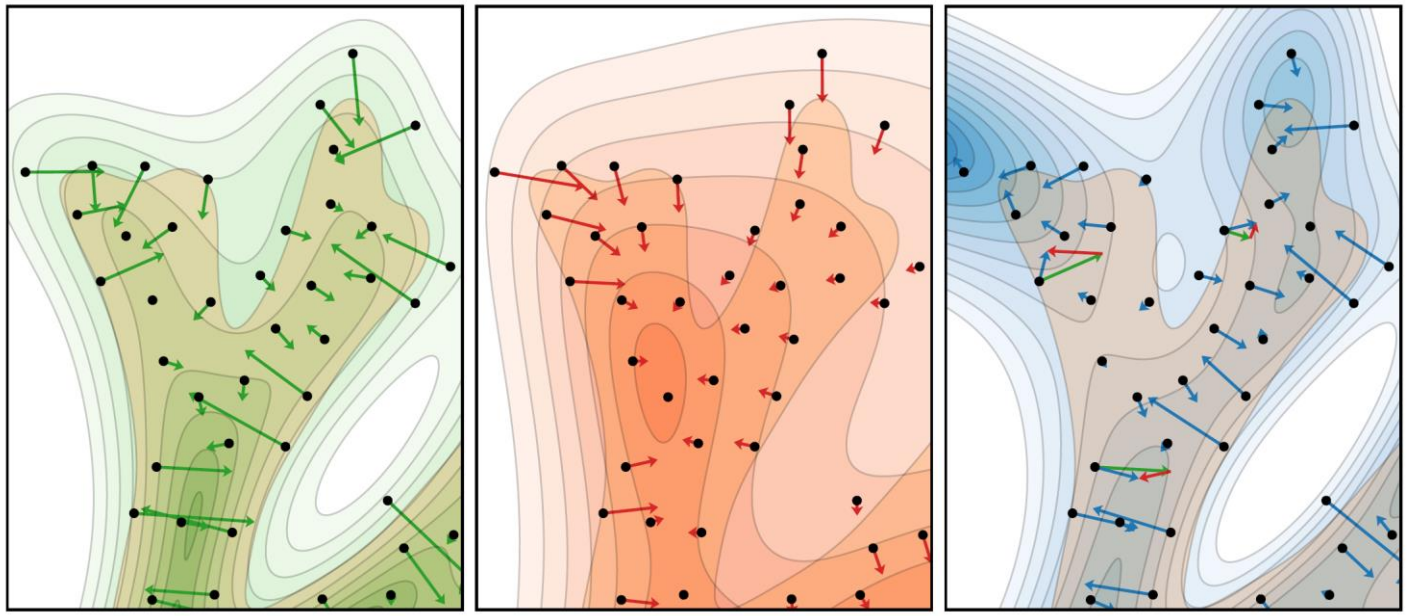
Ground Truth

Conditional

Unconditional

Ratio

$$\nabla_x \log \tilde{p}(x|y)$$
$$= \nabla_x \log p(x) + w \cdot (\nabla_x \log p(x|y) - \nabla_x \log p(x))$$
$$= \nabla_x \log p(x) + w \cdot \left(\nabla_x \log \frac{p(x|y)}{p(x)}\right)$$



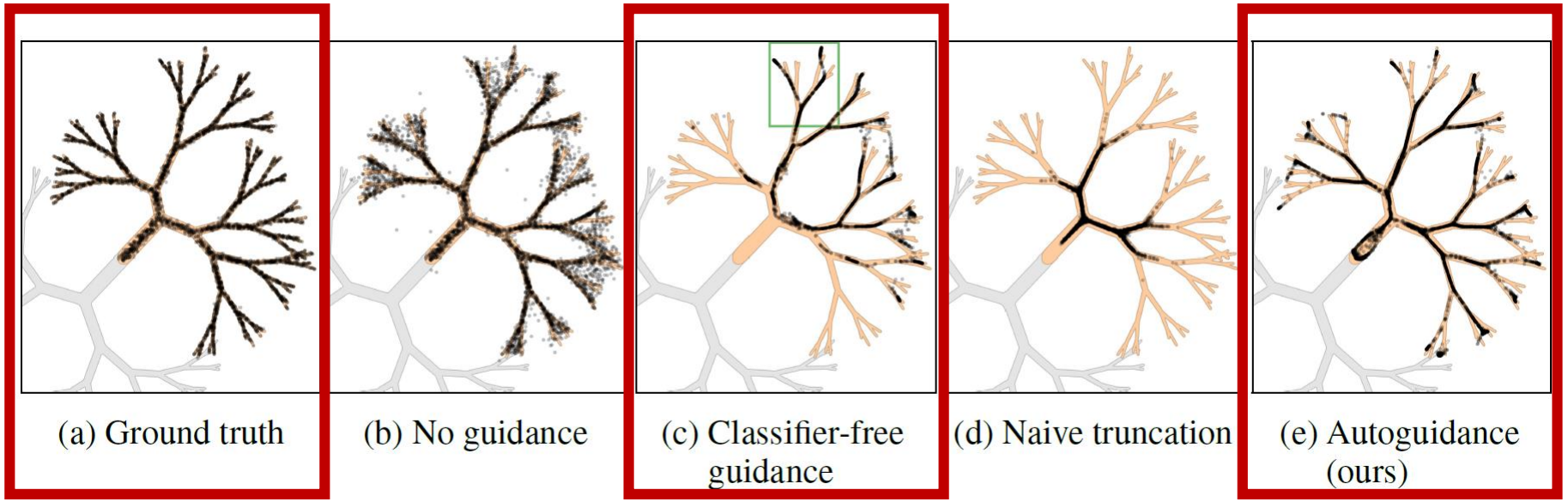
(a) $p_1(\mathbf{x}|\mathbf{c}; \sigma_{\text{mid}})$ (b) $p_0(\mathbf{x}; \sigma_{\text{mid}})$ (c) Ratio p_1/p_0

Karras, T., Aittala, M., Kynkäänniemi, T., Lehtinen, J., Aila, T., & Laine, S. (2024). Guiding a Diffusion Model with a Bad Version of Itself. arXiv preprint arXiv:2406.02507.

Autoguidance

Autoguidance

- Autoguidance는 unconditional model(UM) 대신 conditional model(CM)에 degradation을 추가한 모델을 활용 (e.g. training iteration 감소, 모델 사이즈 감소)
- UM은 CM과 서로 다른 task를 풂 → 서로 다른 영역에서 에러 발생 → 잘못된 영역에서 guidance를 줄 수 있음
- CM의 열화버전은 CM과 동일한 영역에서 에러가 발생 → 필요한 영역에서만 guidance



Karras, T., Aittala, M., Kynkäänniemi, T., Lehtinen, J., Aila, T., & Laine, S. (2024). Guiding a Diffusion Model with a Bad Version of Itself. arXiv preprint arXiv:2406.02507.

Conclusion

Accelerating Diffusion Models

- Classifier Guidance
  - Conditioning term의 scale을 키우기
- Classifier-Free Guidance
  - Classifier 없이 guidance 적용
- Perturbed Attention Guidance
  - Unconditional model 상황에서 적용할 수 있는 guidance 기법 제안
- Autoguidance
  - Degradation이 된 모델로 부터 guidance 시그널을 받는 guidance 기법 제안