Applications of Diffusion Models

DMQA Open Seminar (2023. 11. 24)

Data Mining & Quality Analytics Lab.

박태남

발표자 소개

❖ **박태남 (Taenam Park)**
- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- M.S. Student (2023.03 ~ Present)

❖ **Research Interest**
- Deep Generative Models
- Diffusion Models

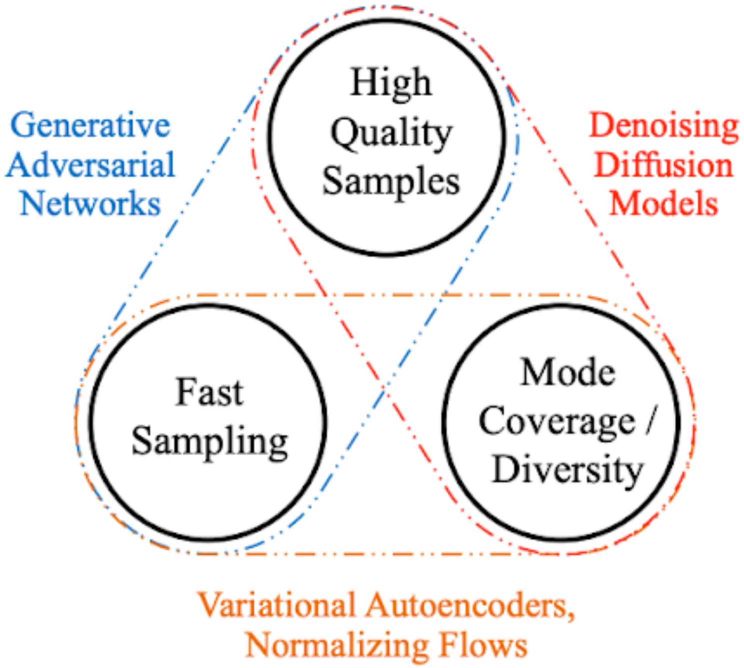
❖ **Contact**
- taenampark@korea.ac.kr

Contents

❖ Generative Models

$p(x)$

- 학습된 데이터의 표현을 학습하고 데이터 자체를 모델링
- 중요 요인들
  - ✓ High Quality Samples
  - ✓ Mode Coverage/Diversity
  - ✓ Fast Sampling



<Generative learning trilemma>

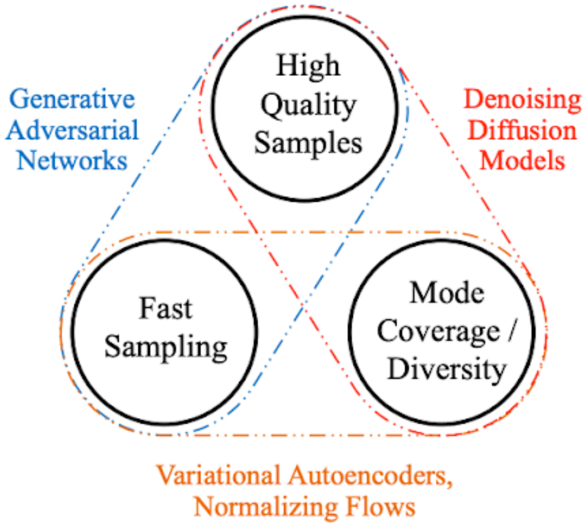
https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/

Introduction

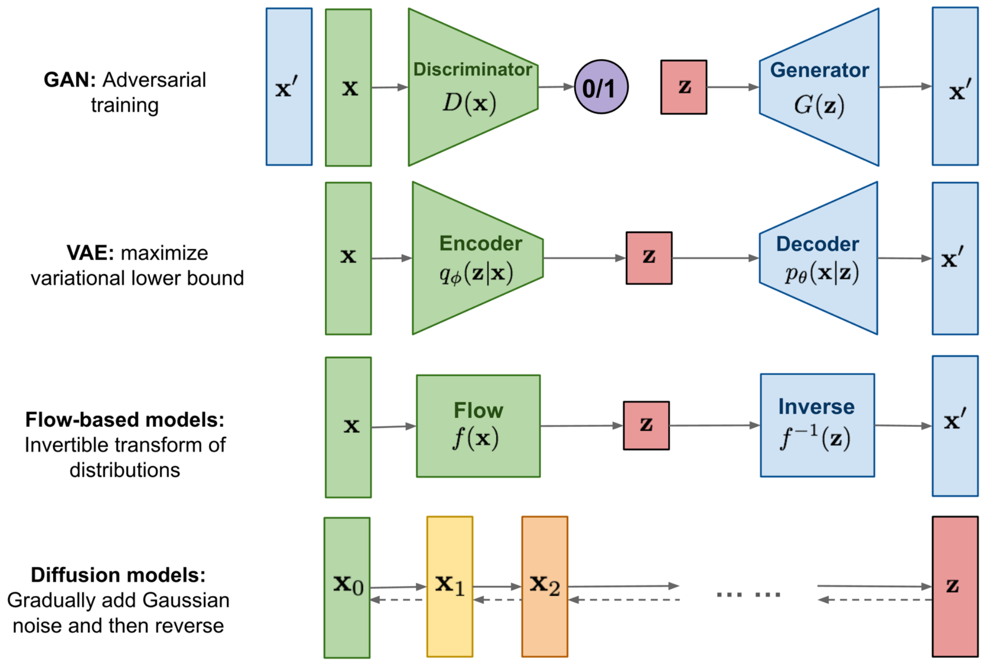
Generative Models

❖ Generative Models $p(x)$

- **Generative Adversarial Networks(GAN)**: 학습의 불안정성
- **Variational Autoencoders(VAE)**: 직접 Likelihood를 최대화하지 못함
- **Flow-based models**: 역함수가 존재하는 구조에서만 사용가능



<The Generative Model Trilemma>



<GAN, VAE, Flow-based models, Diffusion models>

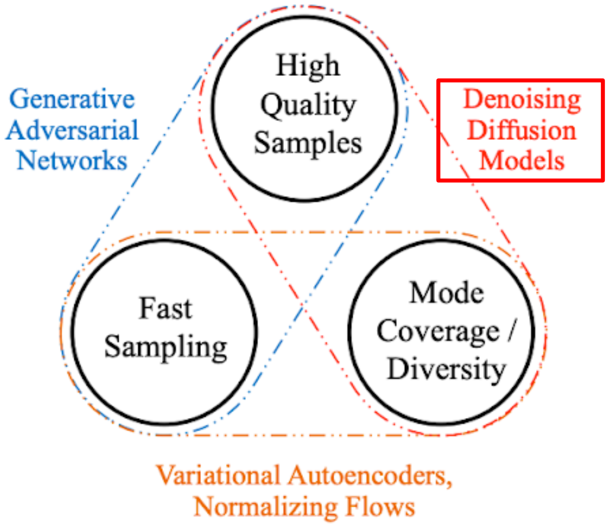
https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/

Introduction

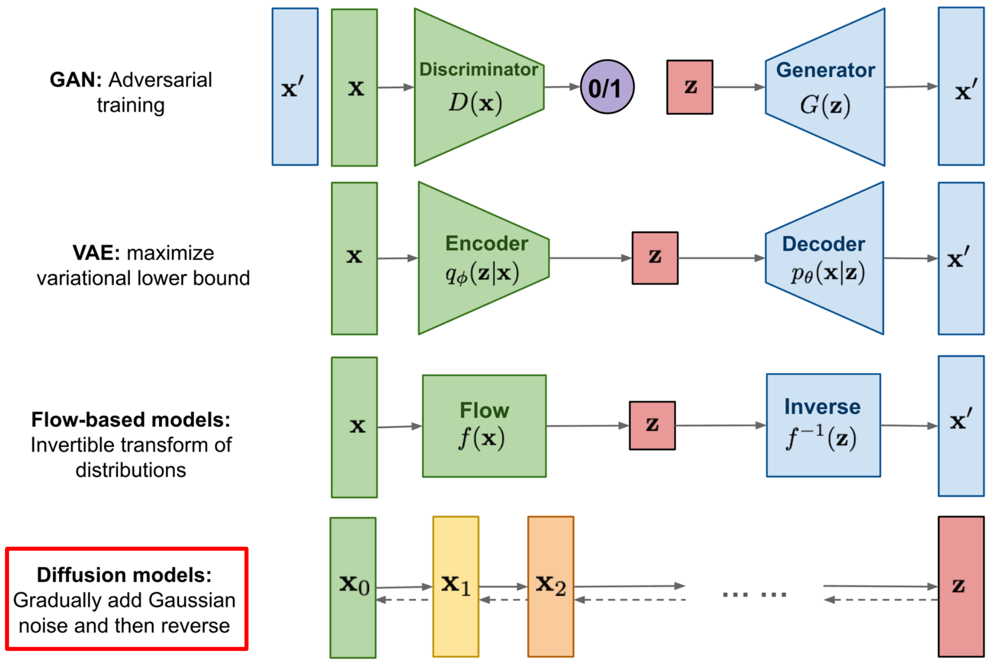
Diffusion Models

❖ Diffusion Models

- 비정형 열역학(Nonequilibrium Thermodynamics)에 기초
- 고정된 절차로 학습되며, 잠재 변수의 차원은 원본 데이터와 동일
- High Quality Samples & Mode Coverage/Diversity



<The Generative Model Trilemma>



<GAN, VAE, Flow-based models, Diffusion models>

https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/

Introduction

Diffusion Models

❖ **DDPM (Denoising Diffusion Probabilistic Models, 2020)**

- Forward Process: Data (x_0) → Noise (x_T)
- Reverse Process: Noise (x_T) → Data (x_0)

Forward Process   Data (x_0) → Noise (x_T)

$$q(\mathrm{x}_t|\mathrm{x}_{t-1}) = N(\mathrm{x}_t; \sqrt{1-\beta_t}\mathrm{x}_{t-1}, \beta_t \mathrm{I})$$

$$p_\theta(\mathrm{x}_{t-1}|\mathrm{x}_t) = N(\mathrm{x}_{t-1}; \mu_\theta(\mathrm{x}_t, t), \Sigma_\theta(\mathrm{x}_t, t))$$

Reverse Process   Noise (x_T) → Data (x_0)

Algorithm 1 Training

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

Algorithm 2 Sampling

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

<학습 및 샘플링 과정>

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.
Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015, June). Deep unsupervised learning using nonequilibrium thermodynamics. In International conference on machine learning (pp. 2256-2265). PMLR.

Introduction

Diffusion Models

❖ DDPM (Denoising Diffusion Probabilistic Models, 2020)

- Forward Process: Data (x_0) → Noise (x_T)
- Reverse Process: Noise (x_T) → Data (x_0)

Forward Process Data (x_0) → Noise (x_T)

$q(\mathrm{x}_t|\mathrm{x}_{t-1}) = N(\mathrm{x}_t; \sqrt{1-\beta_t}\mathrm{x}_{t-1}, \beta_t \mathrm{I})$ $\alpha_t = 1 - \beta_t$

$p_\theta(\mathrm{x}_{t-1}|\mathrm{x}_t) = N(\mathrm{x}_{t-1}; \mu_\theta(\mathrm{x}_t, t), \Sigma_\theta(\mathrm{x}_t, t))$

Reverse Process Noise (x_T) → Data (x_0)

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

$q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathrm{I})$

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

$\epsilon_\theta(x_t, t)$

Algorithm 2 Sampling Inference

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **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** $\mathbf{x}_0$

<학습 및 샘플링 과정>

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840-6851.
Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015, June). Deep unsupervised learning using nonequilibrium thermodynamics. In International conference on machine learning (pp. 2256-2265). PMLR.

Introduction

Diffusion Models

❖ Conditional Diffusion Models $p(x|y)$

- **Classifier guidance**: Noisy data로 학습한 별도의 classifier의 gradients을 통해 conditional sampling
- **Classifier-free guidance(CFG)**: Diffusion Models만으로 conditional sampling 가능

w : Guidance scale

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1-\bar{\alpha}_t}\, w \nabla_{x_t} log f_\phi(y|x_t)$$

<Classifier guidance>

w : Guidance scale

$$\hat{\epsilon}(x_t, t, y) = (1+w)\epsilon_\theta(x_t, t, y) - w\epsilon_\theta(x_t, t)$$
$$= w\big(\epsilon_\theta(x_t, t, y) - \epsilon_\theta(x_t, t)\big) + \epsilon_\theta(x_t, t, y)$$

<Classifier-free guidance>

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34, 8780-8794.
Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598.

Introduction

Diffusion Models

❖ Conditional Diffusion Models $p(x|y)$

- **Classifier guidance**: Noisy data로 학습한 별도의 classifier의 gradients을 통해 conditional sampling
- **Classifier-free guidance(CFG)**: Diffusion Models만으로 conditional sampling 가능

w : Guidance scale

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1-\bar{\alpha}_t}\, w \nabla_{x_t} \log f_\phi(y|x_t)$$

<Classifier guidance>

w : Guidance scale

$$\hat{\epsilon}(x_t, t, y) = (1+w)\epsilon_\theta(x_t, t, y) - w\epsilon_\theta(x_t, t)$$
$$= w\big(\epsilon_\theta(x_t, t, y) - \epsilon_\theta(x_t, t)\big) + \epsilon_\theta(x_t, t, y)$$

<Classifier-free guidance>

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34, 8780-8794.
Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598.

Introduction

Diffusion Models

❖ Conditional Diffusion Models $p(x|y)$

- **Classifier guidance**: Noisy data로 학습한 별도의 classifier의 gradients을 통해 conditional sampling
- **Classifier-free guidance(CFG)**: Diffusion Models만으로 conditional sampling 가능

w : Guidance scale

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1-\bar{\alpha}_t}\, w \nabla_{x_t} log f_\phi(y|x_t)$$

<Classifier guidance>

w : Guidance scale

$\epsilon_\theta(x_t, t) = \epsilon_\theta(x_t, t, y = \emptyset)$

$$\begin{aligned}\hat{\epsilon}(x_t, t, y) &= (1+w)\epsilon_\theta(x_t, t, y) - w\epsilon_\theta(x_t, t) \\ &= w\big(\epsilon_\theta(x_t, t, y) - \epsilon_\theta(x_t, t)\big) + \epsilon_\theta(x_t, t, y)\end{aligned}$$

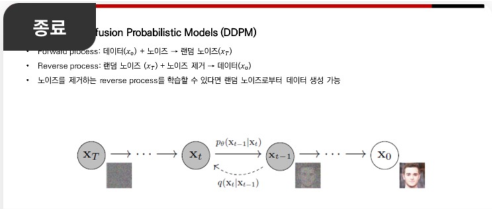
<Classifier-free guidance>

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34, 8780-8794.
Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598.

Introduction

Diffusion Models

종료



Score-based Generative Models and Diffu

발표자: 조한샘

2022년 2월 11일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

Improving Sampling Speed of Diffusion M

발표자: 조한샘

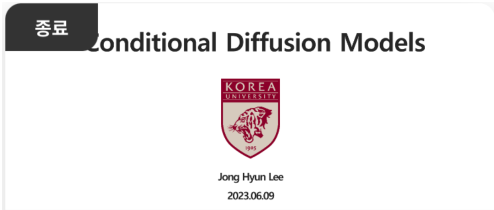
2023년 2월 10일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료



Conditional Diffusion Models

발표자: 이종현

2023년 6월 16일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

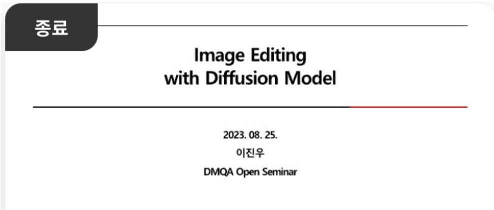


Image Editing with Diffusion Model

발표자: 이진우

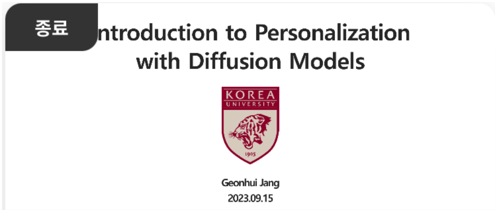
2023년 8월 25일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료



Introduction to Personalization with Diffu

발표자: 장건희

2023년 9월 15일

오후 12시 ~

온라인 비디오 시청 (YouTube)

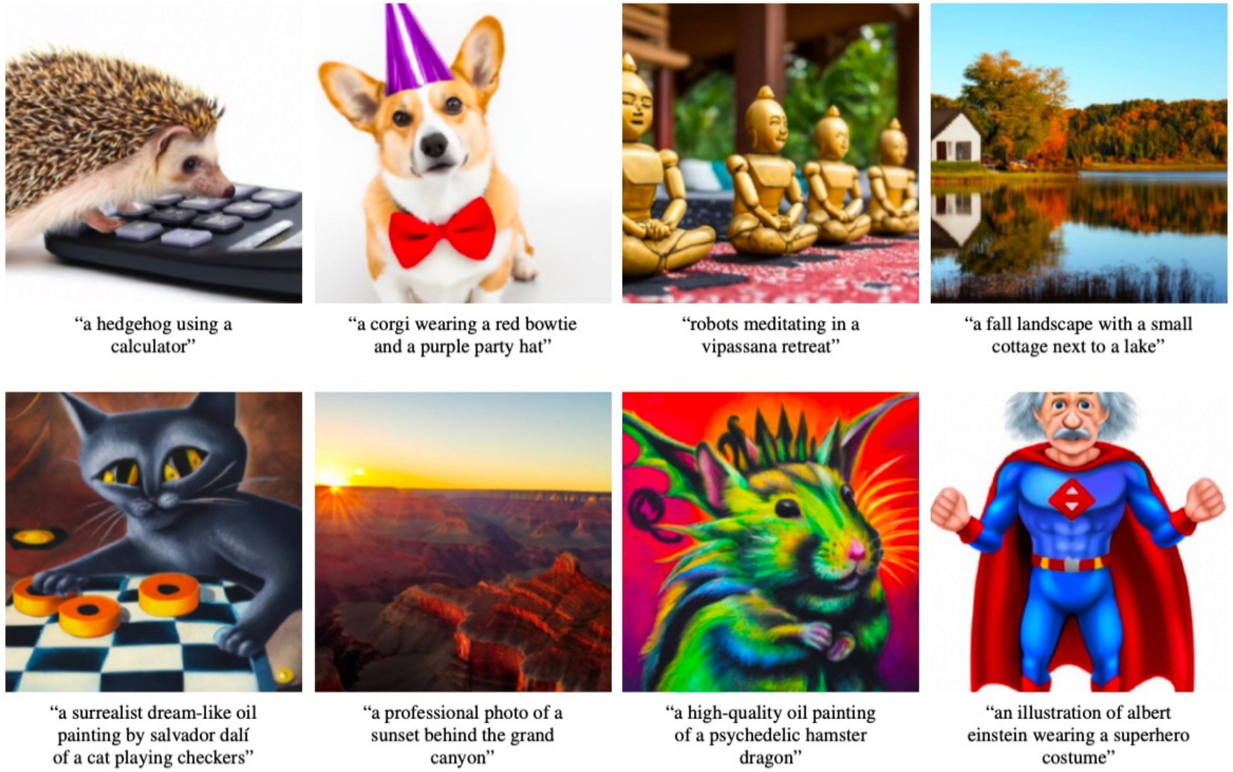
세미나 정보 보기 ⟶

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ Guided Language to Image Diffusion for Generation and Editing

- Text-conditional image synthesis



Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ GLIDE – Guidance

- Classifier guidance

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1 - \bar{\alpha}_t} w \nabla_{x_t} log f_\phi(y|x_t)$$

- Classifier-free guidance(CFG)

$$\hat{\epsilon}(x_t, t, y) = (1 + w)\epsilon_\theta(x_t, t, y) - w\epsilon_\theta(x_t, t)$$

- CLIP Guidance

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1 - \bar{\alpha}_t} \nabla_{x_t}(f(x_t) \cdot g(y))$$

Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ GLIDE – Guidance

- Classifier guidance

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1-\bar{\alpha}_t}\, w \nabla_{x_t} \log f_\phi(y|x_t)$$

- Classifier-free guidance(CFG)

$$\hat{\epsilon}(x_t, t, y) = (1+w)\epsilon_\theta(x_t, t, y) - w\epsilon_\theta(x_t, t)$$

- CLIP Guidance

Image encoder

$$\hat{\epsilon}(x_t, t) = \epsilon_\theta(x_t, t) - \sqrt{1-\bar{\alpha}_t}\nabla_{x_t}(f(x_t) \cdot g(y))$$

caption encoder

Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ Experiments

- 텍스트의 넓은 다양성을 일반화하는 능력 존재
- 고화질의 질감을 표현한 이미지 생성
- 다양한 스타일의 이미지 생성
- 여러가지 컨셉을 동시에 부여 가능

"a hedgehog using a calculator"

"a corgi wearing a red bowtie and a purple party hat"

"robots meditating in a vipassana retreat"

"a fall landscape with a small cottage next to a lake"

"a surrealist dream-like oil painting by salvador dalí of a cat playing checkers"

"a professional photo of a sunset behind the grand canyon"

"a high-quality oil painting of a psychedelic hamster dragon"

"an illustration of albert einstein wearing a superhero costume"

Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ Experiments

- 기존 SOTA 방법론들 대비 가장 현실적인 이미지 생성
- CLIP guidance보다 CFG가 더 현실적인 이미지 생성

"a green train is coming down the tracks" | "a group of skiers are preparing to ski down a mountain." | "a small kitchen with a low ceiling" | "a group of elephants walking in muddy water." | "a living area with a television and a table"

Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ Experiments - Image inpainting & Image editing

<Text-conditional image inpainting examples>

<Text-conditional SDEdit examples>

Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Text-to-Image Generation

"GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", ICML, 2022

❖ Experiments - Image inpainting & Image editing

"zebras roaming in the field"   "a girl hugging a corgi on a pedestal"

"an old car in a snowy forest"   "a man wearing a white hat"

<Text-conditional image inpainting examples>

"a corgi wearing a bow tie and a birthday hat"

"a fire in the background"

"only one cloud in the sky today"

<Text-conditional SDEdit examples>

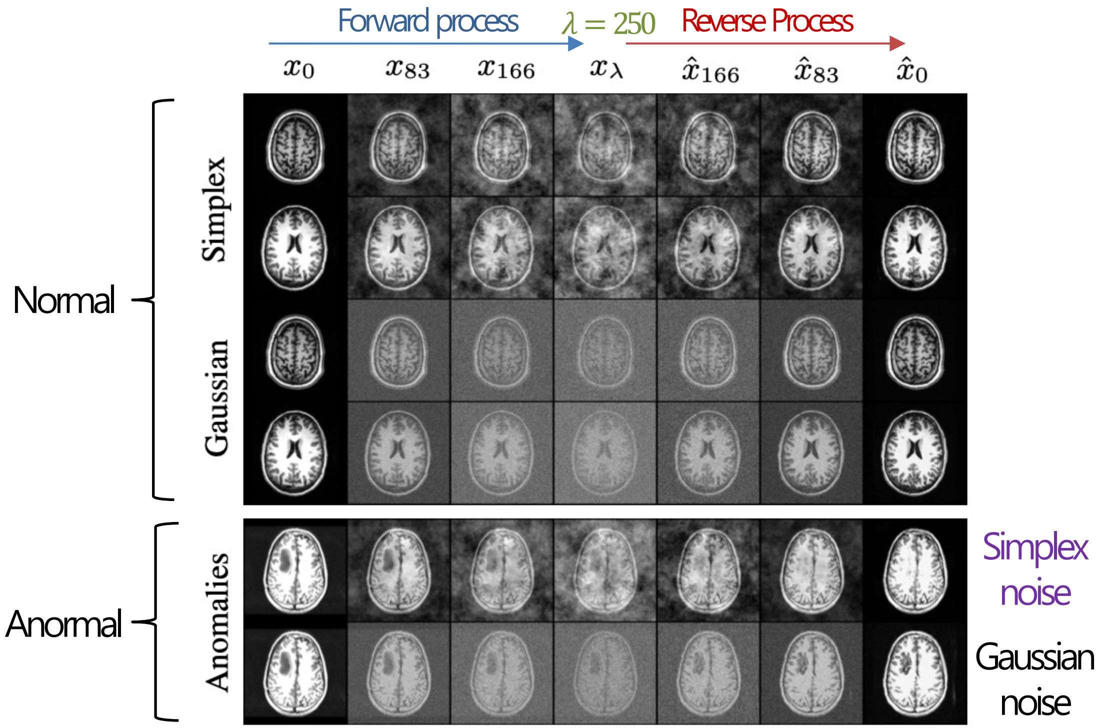
Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., Mcgrew, B., ... & Chen, M. (2022, June). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning (pp. 16784-16804). PMLR.

Diffusion Models for Anomaly Detection

"AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise", CVPR workshop, 2022

❖ AnoDDPM

- A **partial** diffusion anomaly detection
- GAN 기반 모델들에 비해 더 작은 데이터셋을 가지고도 우수한 성능 입증



https://julianwyatt.co.uk/anoddpm

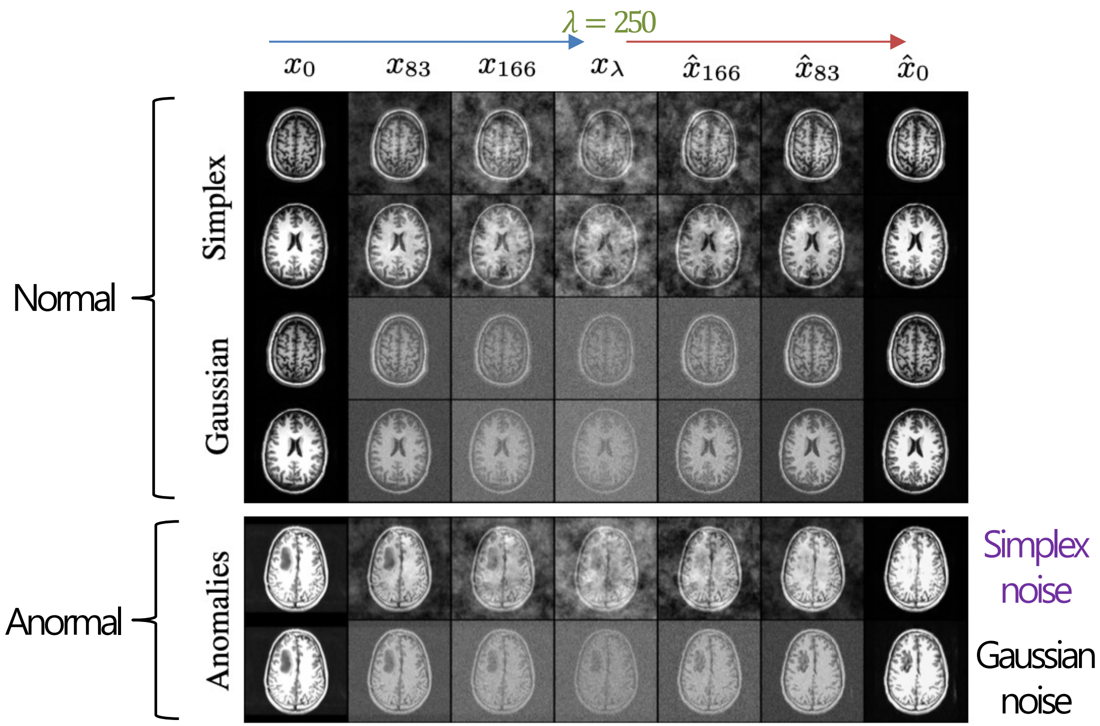
Wyatt, J., Leach, A., Schmon, S. M., & Willcocks, C. G. (2022). Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 650-656).

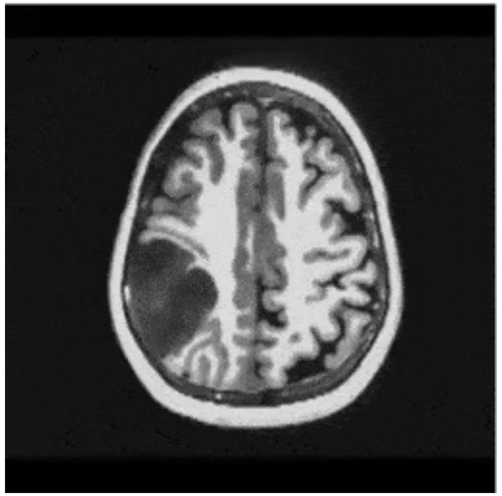
Diffusion Models for Anomaly Detection

"AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise", CVPR workshop, 2022

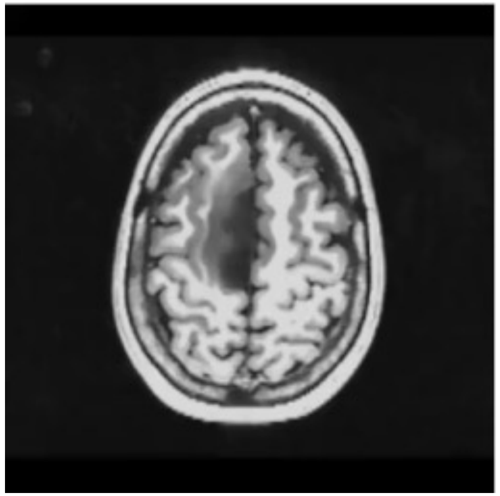
❖ AnoDDPM

- Gaussian noise을 **multi-scale simplex** noise로 대체





<Gaussian noise>



<Simplex noise>

https://julianwyatt.co.uk/anoddpm
Wyatt, J., Leach, A., Schmon, S. M., & Willcocks, C. G. (2022). Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 650-656).

Diffusion Models for Anomaly Detection

"AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise", CVPR workshop, 2022

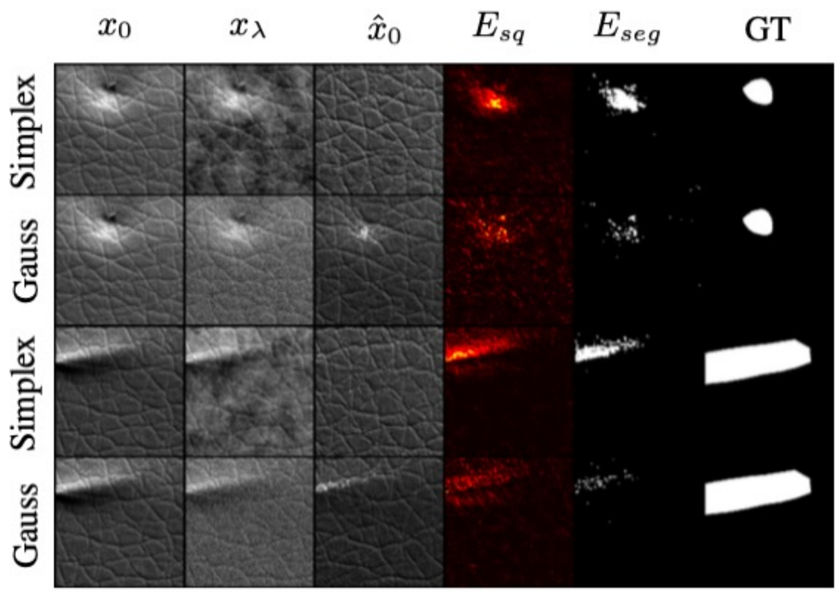
❖ Training & Inference

Algorithm 1 Training

repeat
  $x_0 \sim q(x_0)$
  $t \sim \text{Uniform}(\{1, 2, \dots, T-1, T\})$
  Randomly generate simplex seed
  $\epsilon \sim \text{Simplex}(\nu = 2^{-6}, N = 6, \gamma = 0.8)$
  Take gradient descent step on:
  $\nabla_\theta[||\epsilon - \epsilon_\theta(x_0\sqrt{\bar{\alpha}_t} + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2]$
until converged

Algorithm 2 Segmentation

$x_0 \sim A(x_0)$
$x_\lambda = x_0\sqrt{\bar{\alpha}_\lambda} + \epsilon\sqrt{1-\bar{\alpha}_\lambda}$
for $t = \lambda, \dots, 1$ **do**
  Randomly generate simplex seed
  $z \sim \text{Simplex}(2^{-6}, 6, 0.8)$ if $t > 0$ else $z = 0$
  $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)) + \tilde{\beta}_t z$
end for
$E_{sq} = (x_0 - \hat{x}_0)^2$
$E_{seg} = E_{sq} > 0.5$
return E_{seg}

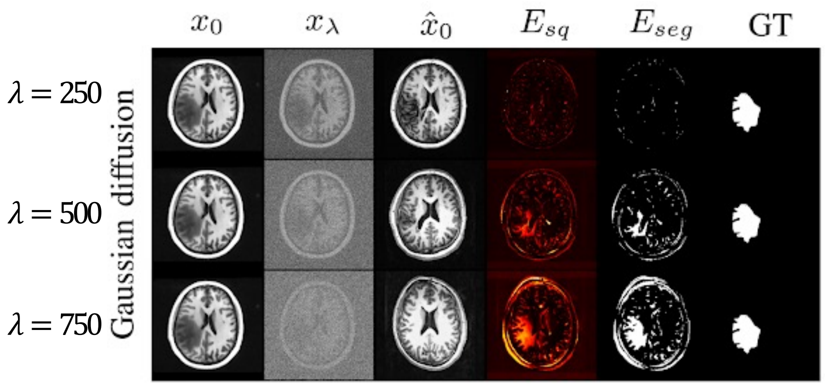


https://julianwyatt.co.uk/anoddpm
Wyatt, J., Leach, A., Schmon, S. M., & Willcocks, C. G. (2022). Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 650-656).

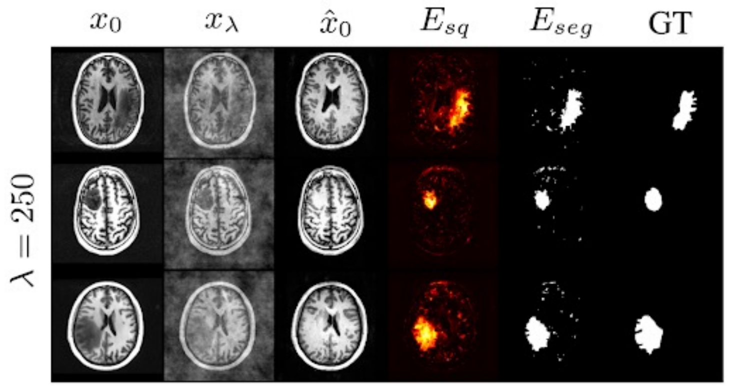
Diffusion Models for Anomaly Detection

"AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise", CVPR workshop, 2022

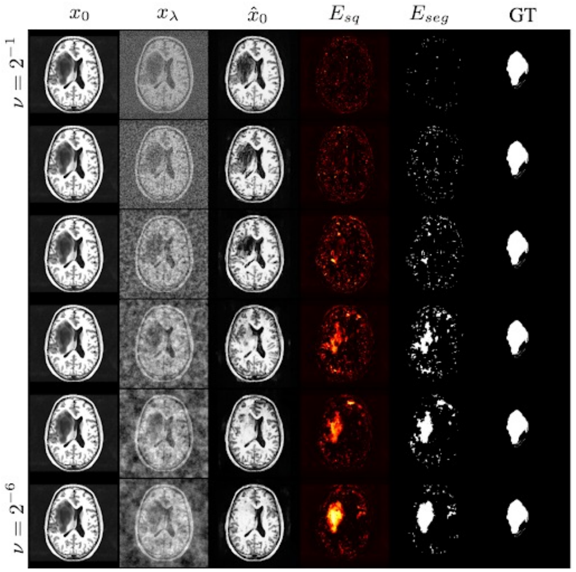
❖ Experiments



<Gaussian noise sample>



<Curated simplex noise samples>



<Simplex noise with increasing frequency>

https://julianwyatt.co.uk/anoddpm

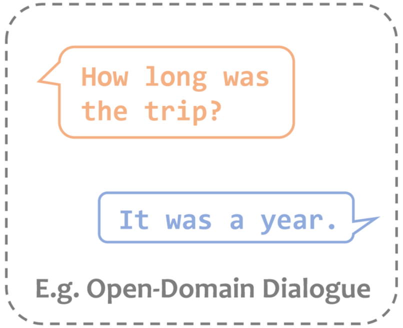
Wyatt, J., Leach, A., Schmon, S. M., & Willcocks, C. G. (2022). Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 650-656).

Diffusion Models for Natural Language Generation

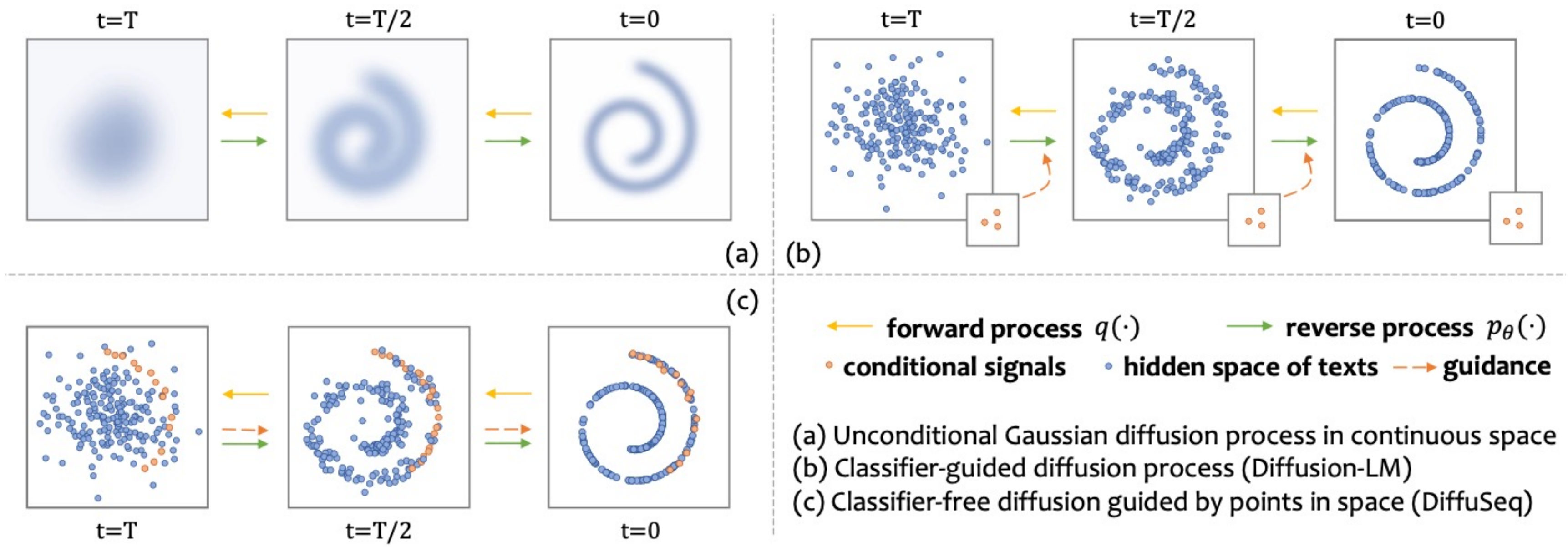
"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- Discrete한 특성을 지닌 Text에 디퓨전 모델을 적용
- 특히 Sequence-to-Sequence(Seq2Seq) Text Generation에 디퓨전 모델을 최초로 적용



<Sequence to Sequence>



<unconditional, classifier-guided, classifier-free diffusion models>

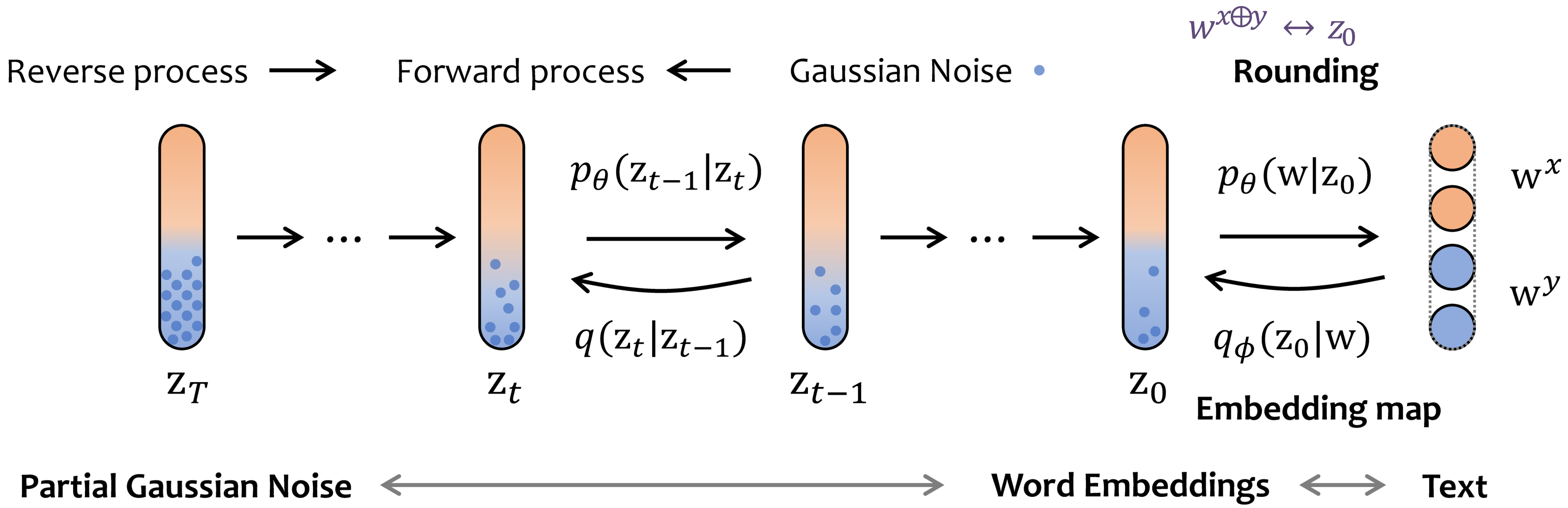
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations.*

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- **Forward process**: w^y에 대해서만 noise 부여, w^x는 un-noised 상태로 유지
- **Reverse process**: conditional signals w^x를 guidance로 활용



<The diffusion process of DiffuSeq>

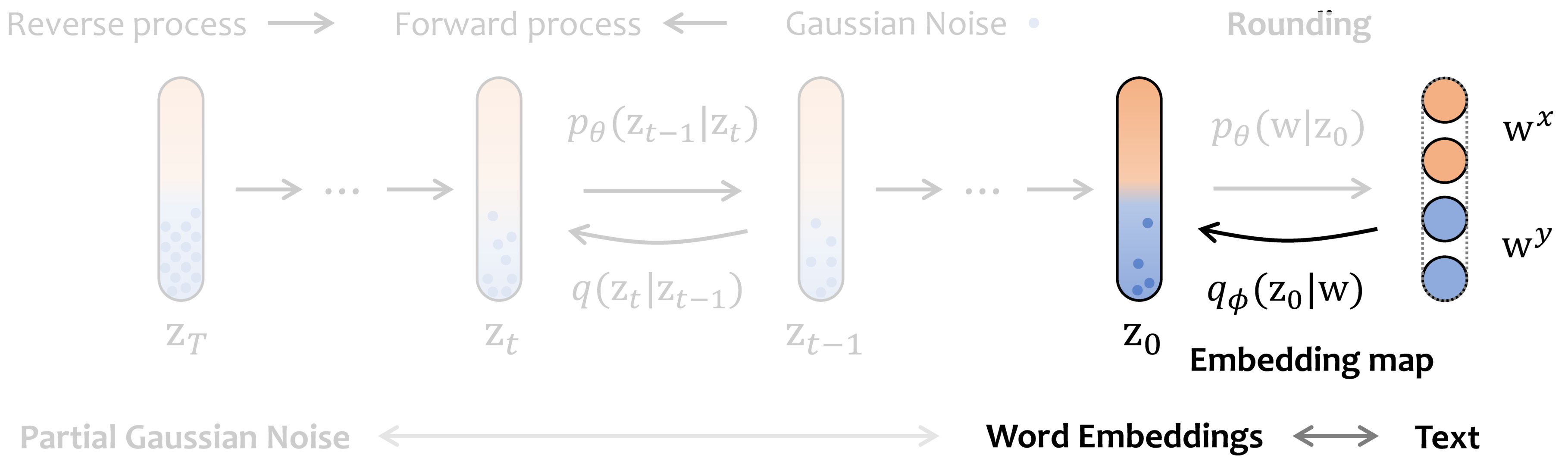
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- **Forward process**: w^y에 대해서만 noise 부여, w^x는 un-noised 상태로 유지
- **Reverse process**: conditional signals w^x를 guidance로 활용



<The diffusion process of DiffuSeq>

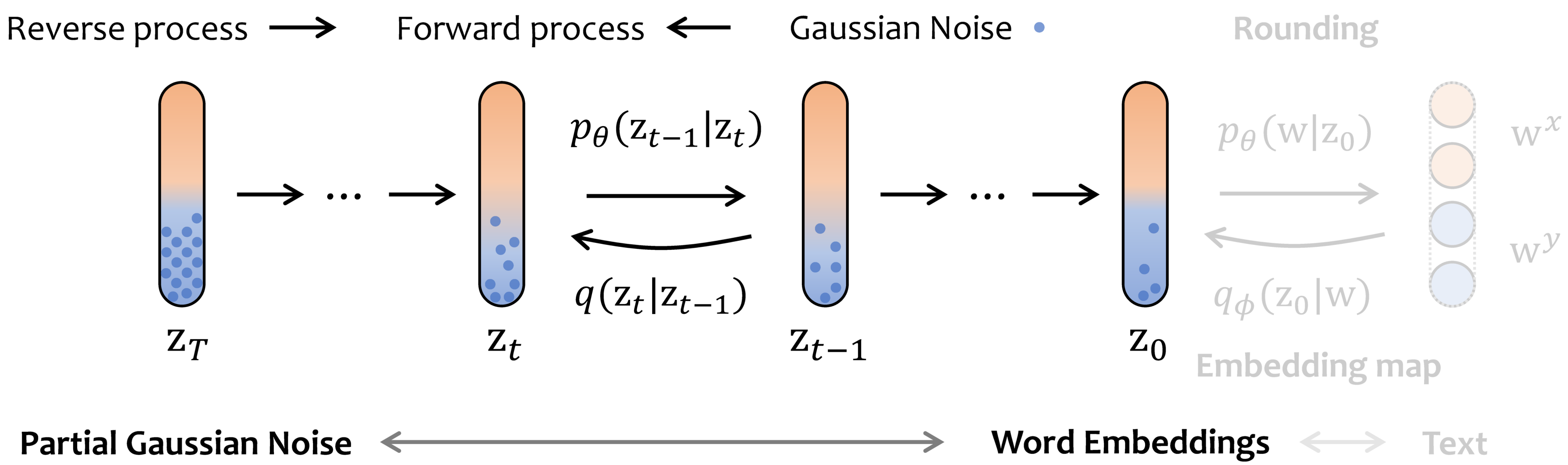
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations.*

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- **Forward process**: w^y에 대해서만 noise 부여, w^x는 un-noised 상태로 유지 → Partial Noising
- **Reverse process**: conditional signals w^x를 guidance로 활용



<The diffusion process of DiffuSeq>

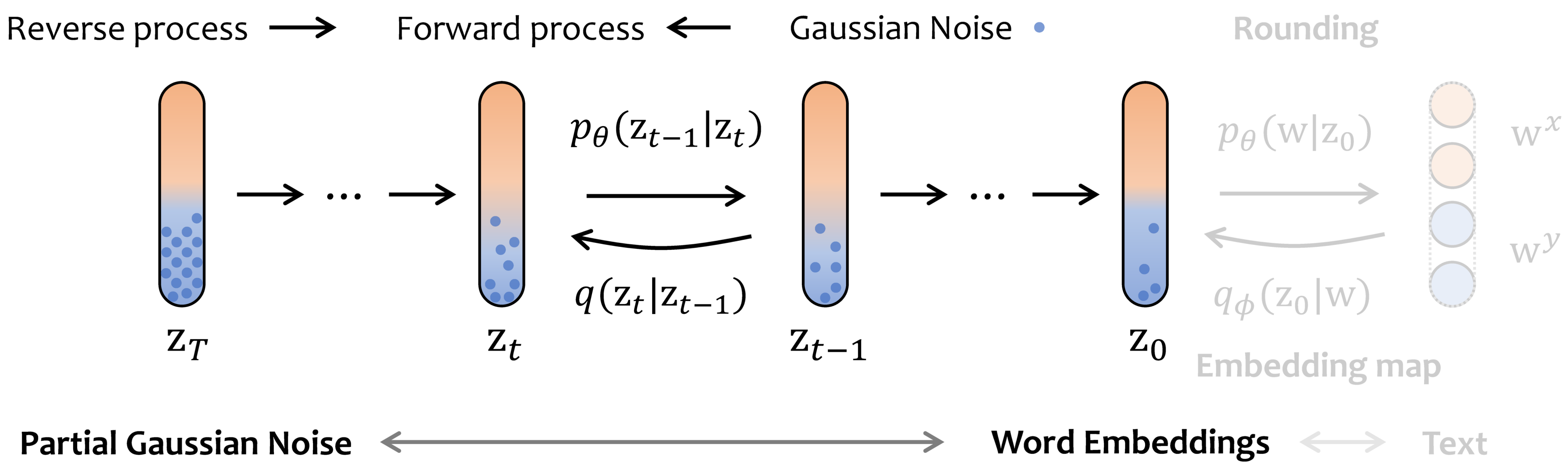
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations.*

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- **Forward process**: w^y에 대해서만 noise 부여, w^x는 un-noised 상태로 유지
- **Reverse process**: conditional signals w^x를 guidance로 활용 → Partial Denoising



<The diffusion process of DiffuSeq>

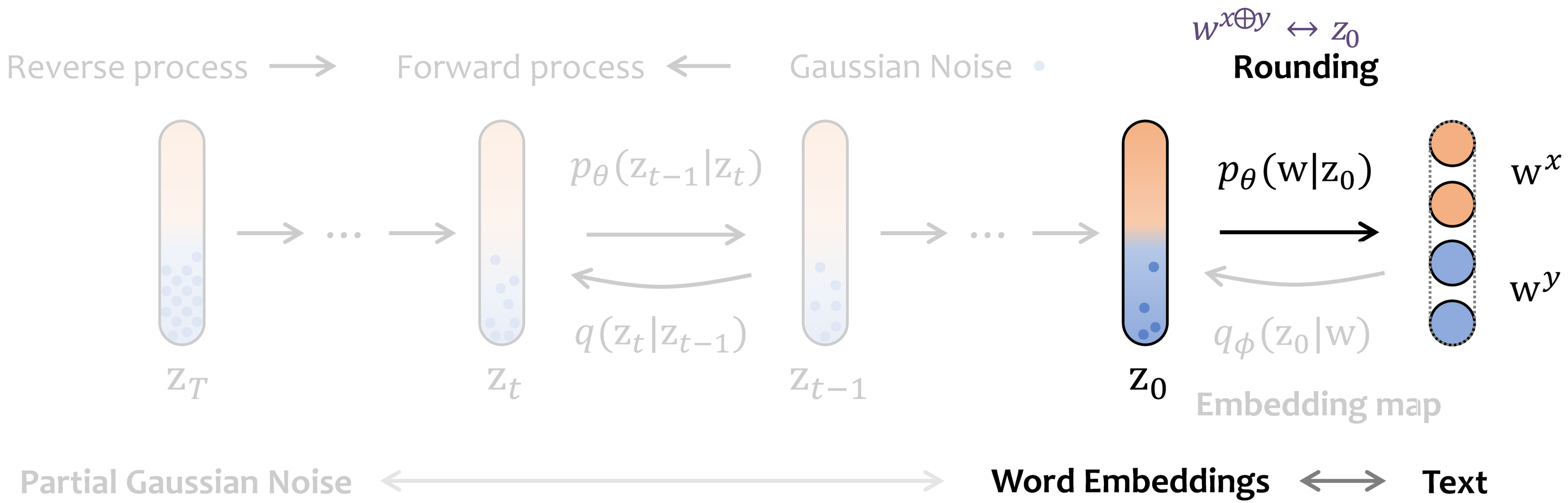
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ DiffuSeq

- **Forward process**: w^y에 대해서만 noise 부여, w^x는 un-noised 상태로 유지
- **Reverse process**: conditional signals w^x를 guidance로 활용 → Partial Denoising



<The diffusion process of DiffuSeq>

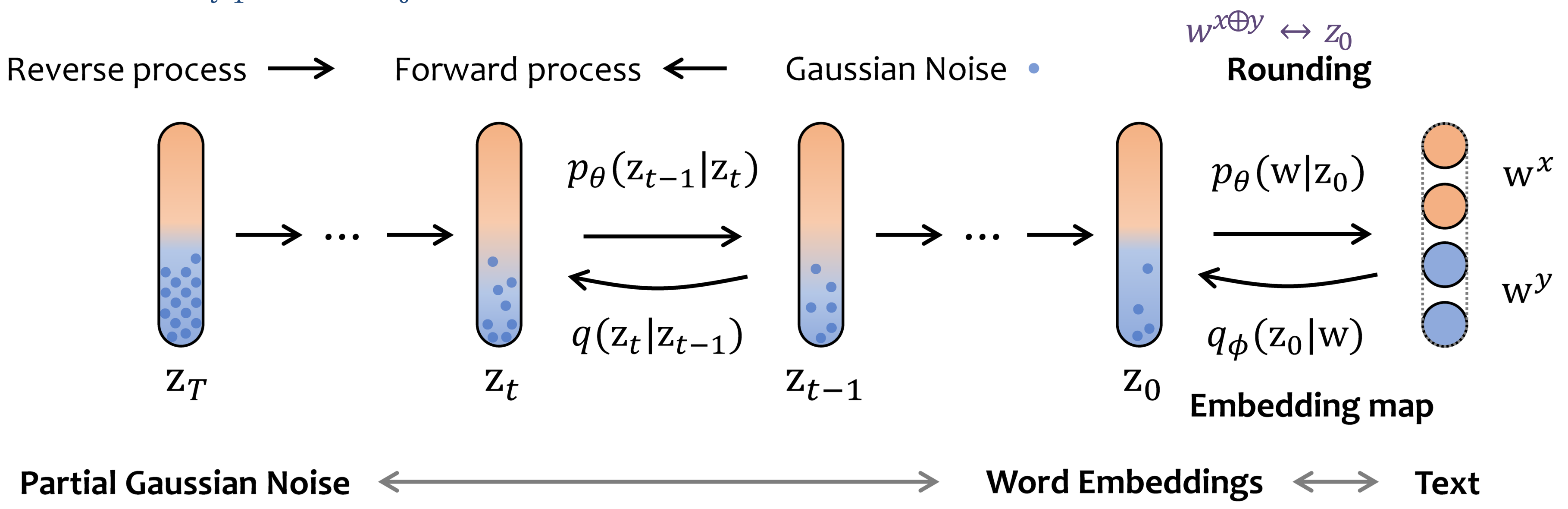
Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ Training & Inference

- **Training**: Importance sampling을 통해 효율적인 학습 진행
- **Inference**: anchoring operation을 활용
  1. z_t가 임베딩 공간상에 위치하도록
  2. z_{t-1}의 x부분을 z_0의 x 부분으로 대체



<The diffusion process of DiffuSeq>

Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ Experiments

- Sequence-to-Sequence tasks에 대해 우수한 성능을 보임
- 동일한 인풋 시퀀스가 주어졌을 때, 다양한 아웃풋을 생성하는 능력에서 우수함을 보임

Tasks	Methods	BLEU↑	R-L↑	Score↑	dist-1↑	selfB↓ / div-4↑	Len
Open Domain Dialogue	GRU-attention ◇	0.0068	0.1054	0.4128	0.8998	0.8008/0.1824	4.46
	Transformer-base ◇	**0.0189**	0.1039	0.4781	0.7493	0.3698/0.6472	19.5
	GPT2-base FT •	0.0108	**0.1508**	0.5279	0.9194	0.0182/0.9919	16.8
	GPT2-large FT •	0.0125	0.1002	**0.5293**	0.9244	0.0213/0.9938	16.8
	GPVAE-T5 •	0.0110	0.1009	0.4317	0.5625	0.3560/0.5551	20.1
	NAR-LevT ‡	0.0158	0.0550	0.4760	**0.9726**	0.7103/0.1416	4.11
	DIFFUSEQ (Ours) ‡	0.0139	0.1056	0.5131	0.9467	**0.0144/0.9971**	13.6
Question Generation	GRU-attention ◇	0.0651	0.2617	0.5222	0.7930	0.9999/0.3178	10.1
	Transformer-base ◇	0.1663	0.3441	0.6307	0.9309	0.3265/0.7720	10.3
	GPT2-base FT •	0.0741	0.2714	0.6052	0.9602	**0.1403/0.9216**	10.0
	GPT2-large FT •	0.1110	0.3215	**0.6346**	**0.9670**	0.2910/0.8062	9.96
	GPVAE-T5 •	0.1251	0.3390	0.6308	0.9381	0.3567/0.7282	11.4
	NAR-LevT ‡	0.0930	0.2893	0.5491	0.8914	0.9830/0.4776	6.93
	DIFFUSEQ (Ours) ‡	**0.1731**	**0.3665**	0.6123	0.9056	0.2789/0.8103	11.5
Text Simplification	GRU-attention ◇	0.3256	0.5602	0.7871	0.8883	0.9998/0.3313	18.9
	Transformer-base ◇	0.2693	0.4907	0.7381	0.8886	0.6924/0.5095	18.5
	GPT2-base FT •	0.3083	0.5461	0.8021	0.9439	0.5444/0.6047	16.1
	GPT2-large FT •	0.2693	0.5111	0.7882	0.9464	0.6042/0.5876	15.4
	GPVAE-T5 •	0.3392	0.5828	**0.8166**	0.9308	0.8147/0.4355	18.5
	NAR-LevT ‡	0.2052	0.4402	0.7254	**0.9715**	0.9907/0.3271	8.31
	DIFFUSEQ (Ours) ‡	**0.3622**	**0.5849**	0.8126	0.9264	**0.4642/0.6604**	17.7
Paraphrase	GRU-attention ◇	0.1894	0.5129	0.7763	0.9423	0.9958/0.3287	8.30
	Transformer-base ◇	**0.2722**	0.5748	0.8381	0.9748	0.4483/0.7345	11.2
	GPT2-base FT •	0.1980	0.5212	0.8246	0.9798	0.5480/0.6245	9.67
	GPT2-large FT •	0.2059	0.5415	0.8363	**0.9819**	0.7325/0.5020	9.53
	GPVAE-T5 •	0.2409	**0.5886**	**0.8466**	0.9688	0.5604/0.6169	9.60
	NAR-LevT ‡	0.2268	0.5795	0.8344	0.9790	0.9995/0.3329	8.85
	DIFFUSEQ (Ours) ‡	0.2413	0.5880	0.8365	0.9807	**0.2732/0.8641**	11.2

Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations.*

Diffusion Models for Natural Language Generation

"DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models", ICLR, 2023

❖ Experiments

- Sequence-to-Sequence tasks에 대해 우수한 성능을 보임
- 동일한 인풋 시퀀스가 주어졌을 때, 다양한 아웃풋을 생성하는 능력에서 우수함을 보임

***Original sentence**: How do I make friends.* ***Paraphrase reference**: How to make friends ?*

GPT2-large finetune	**GPVAE-T5**	**DIFFUSEQ**
How can I make friends?	How can I make friends?	How can I make friends better?
How can I make friends?	How do I make friends?	How can I make friends?
How can I make friends?	How can I make friends?	How do you make friends?
How can I make friends?	How can I make friends?	What is the best way to make friends?
How do I make friends and keep them?	What's the best way to make friends and make make friends?	How can I make friends and more something?

Gong, S., Li, M., Feng, J., Wu, Z., & Kong, L. (2022, September). DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Diffusion Models for Time-series Forecasting

"Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting", ICML, 2021

❖ TimeGrad

- 다변량 시계열 예측을 위한 **autoregressive** denoising diffusion model

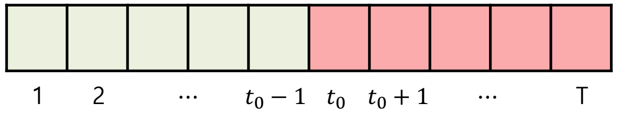
Data

$$\underbrace{x_1, x_2, \cdots, x_{t_0-1}}_{\text{Past}}, \underbrace{x_{t_0}, \cdots, x_{T-1}, x_T}_{\text{Future}} \in R^D$$

Coditional distribution

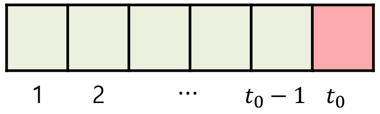
$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, \overset{\text{covariate}}{c_{1:T}}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}\right)$$



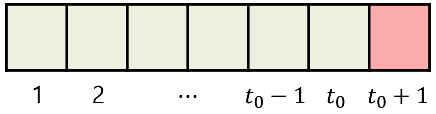
$$=$$

$$q_x\left(x^0_{t_0} \middle| x^0_{1:t_0-1}\right)$$



$$\times$$

$$q_x\left(x^0_{t_0} \middle| x^0_{1:t_0-1}\right)$$

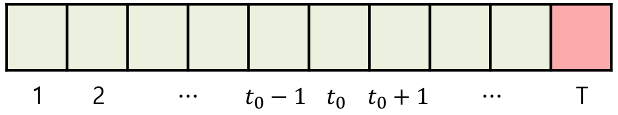


$$\times$$

$$\vdots$$

$$\times$$

$$q_x\left(x^0_T \middle| x^0_{1:T-1}\right)$$



Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Diffusion Models for Time-series Forecasting

"Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting", ICML, 2021

❖ TimeGrad

- 다변량 시계열 예측을 위한 **autoregressive** denoising diffusion model

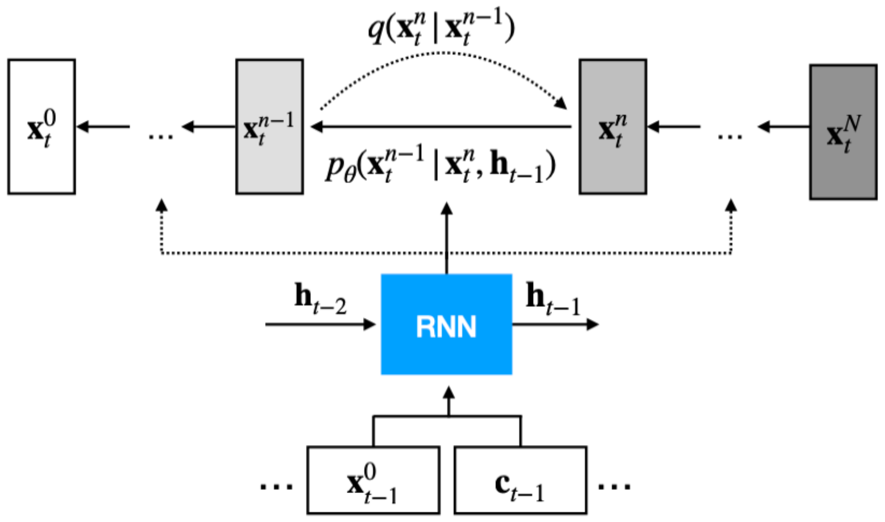
Data

$$\underbrace{x_1, x_2, \cdots, x_{t_0-1}}_{\text{Past}}, \underbrace{x_{t_0}, \cdots, x_{T-1}, x_T}_{\text{Future}} \in R^D$$

Coditional distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, \overset{\text{covariate}}{c_{1:T}}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$

$$\rightarrow \prod_{t=t_0}^{T} p_\theta\left(x^0_t \middle| h_{t-1}\right)$$



$$h_t = \mathrm{RNN}_\theta(concat(x^0_t, c_t), h_{t-1})$$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Diffusion Models for Time-series Forecasting

"Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting", ICML, 2021

❖ Training & sampling

Data

$$\underbrace{x_1, x_2, \cdots, x_{t_0-1}}_{\text{Past}}, \underbrace{x_{t_0}, \cdots, x_{T-1}, x_T}_{\text{Future}} \in R^D$$

Coditional distribution

$$q_x\left(x^0_{t_0:T} \middle| x^0_{1:t_0-1}, \underbrace{c_{1:T}}_{\text{covariate}}\right) = \prod_{t=t_0}^{T} q_x\left(x^0_t \middle| x^0_{1:t-1}, c_{1:T}\right)$$

$$\rightarrow \prod_{t=t_0}^{T} p_\theta\left(x^0_t \middle| h_{t-1}\right)$$

Algorithm 1 $\boxed{\text{Training}}$ for each time series step $t \in [t_0, T]$

Input: data $\mathbf{x}^0_t \sim q_\mathcal{X}(\mathbf{x}^0_t)$ and state $\mathbf{h}_{t-1}$
repeat
  Initialize $n \sim \text{Uniform}(1, \ldots, N)$ and $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient step on

$$\nabla_\theta \| \epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_n}\mathbf{x}^0_t + \sqrt{1-\bar{\alpha}_n}\epsilon, \mathbf{h}_{t-1}, n)\|^2$$

until converged

Algorithm 2 $\boxed{\text{Sampling}}$ $\mathbf{x}^0_t$ via annealed Langevin dynamics

Input: noise $\mathbf{x}^N_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and state $\mathbf{h}_{t-1}$
for $n = N$ **to** 1 **do**
  if $n > 1$ **then**
    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  else
    $\mathbf{z} = \mathbf{0}$
  end if
  $\mathbf{x}^{n-1}_t = \frac{1}{\sqrt{\alpha_n}}(\mathbf{x}^n_t - \frac{\beta_n}{\sqrt{1-\bar{\alpha}_n}}\epsilon_\theta(\mathbf{x}^n_t, \mathbf{h}_{t-1}, n)) + \sqrt{\Sigma_\theta}\mathbf{z}$
end for
Return: $\mathbf{x}^0_t$

Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

Diffusion Models for Time-series Forecasting

"Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting", ICML, 2021

❖ Experiments

Table 2. Test set CRPS_{sum} comparison (lower is better) of models on six real world data sets. Mean and standard error metrics for `TimeGrad` obtained by re-training and evaluating 10 times.

Method	Exchange	Solar	Electricity	Traffic	Taxi	Wikipedia
VES	**0.005**$\pm$0.000	0.9$\pm$0.003	0.88$\pm$0.0035	0.35$\pm$0.0023	-	-
VAR	**0.005**$\pm$0.000	0.83$\pm$0.006	0.039$\pm$0.0005	0.29$\pm$0.005	-	-
VAR-Lasso	0.012$\pm$0.0002	0.51$\pm$0.006	0.025$\pm$0.0002	0.15$\pm$0.002	-	3.1$\pm$0.004
GARCH	0.023$\pm$0.000	0.88$\pm$0.002	0.19$\pm$0.001	0.37$\pm$0.0016	-	-
KVAE	0.014$\pm$0.002	0.34$\pm$0.025	0.051$\pm$0.019	0.1$\pm$0.005	-	0.095$\pm$0.012
Vec-LSTM ind-scaling	0.008$\pm$0.001	0.391$\pm$0.017	0.025$\pm$0.001	0.087$\pm$0.041	0.506$\pm$0.005	0.133$\pm$0.002
Vec-LSTM lowrank-Copula	0.007$\pm$0.000	0.319$\pm$0.011	0.064$\pm$0.008	0.103$\pm$0.006	0.326$\pm$0.007	0.241$\pm$0.033
GP scaling	0.009$\pm$0.000	0.368$\pm$0.012	0.022$\pm$0.000	0.079$\pm$0.000	0.183$\pm$0.395	1.483$\pm$1.034
GP Copula	0.007$\pm$0.000	0.337$\pm$0.024	0.0245$\pm$0.002	0.078$\pm$0.002	0.208$\pm$0.183	0.086$\pm$0.004
Transformer MAF	**0.005**$\pm$0.003	0.301$\pm$0.014	0.0207$\pm$0.000	0.056$\pm$0.001	0.179$\pm$0.002	0.063$\pm$0.003
TimeGrad	0.006$\pm$0.001	**0.287**$\pm$0.02	**0.0206**$\pm$0.001	**0.044**$\pm$0.006	**0.114**$\pm$0.02	**0.0485**$\pm$0.002

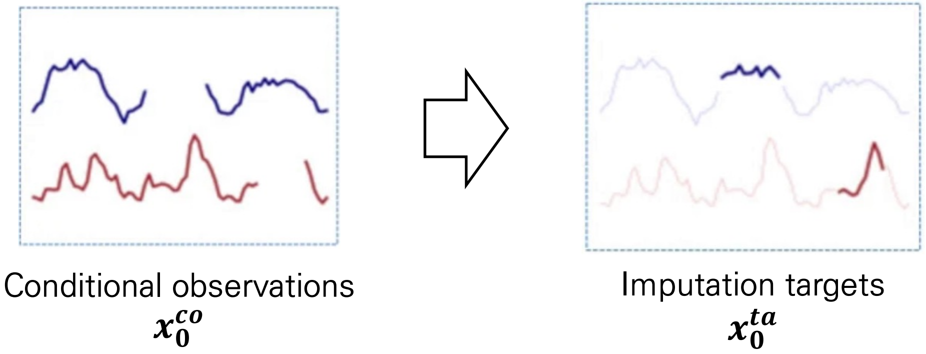
Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.

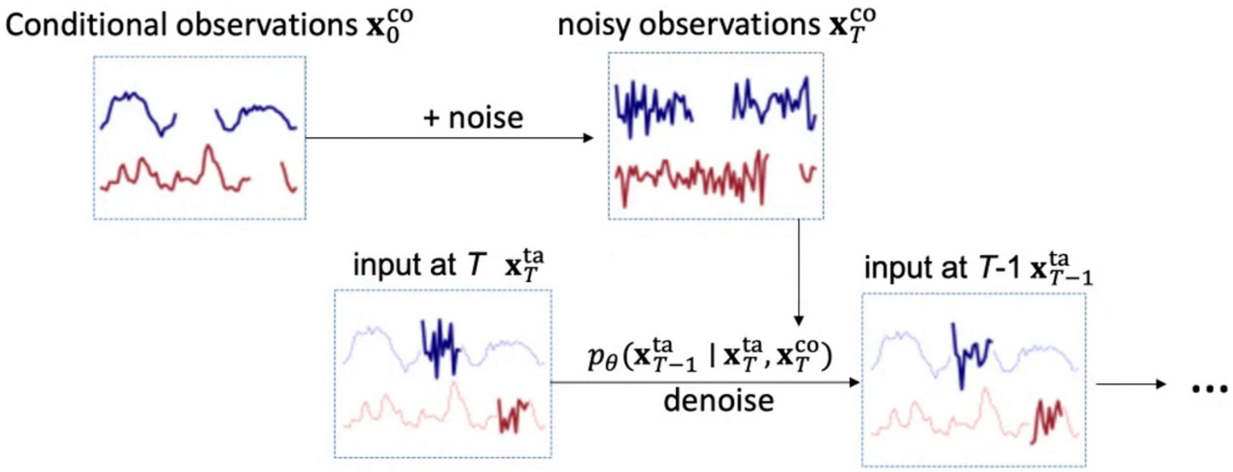
Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ CSDI

- **확률적 시계열 imputation을 위한** Conditional Score-based Diffusion Model
- Imputation task : $x_0^{co} \rightarrow x_0^{ta}$





<Imputation by score-based models>

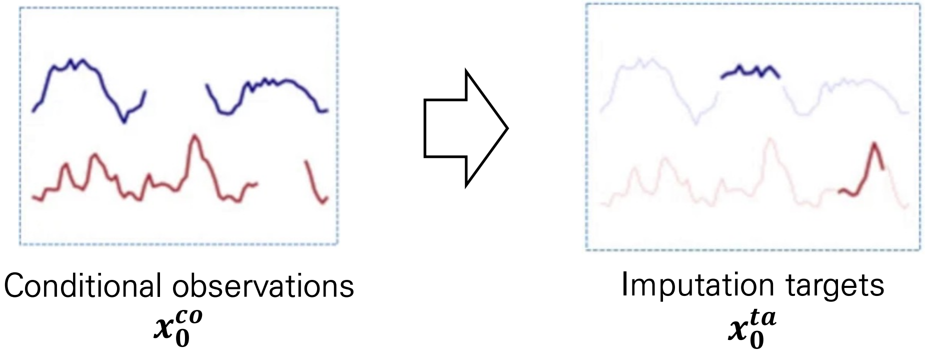
Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

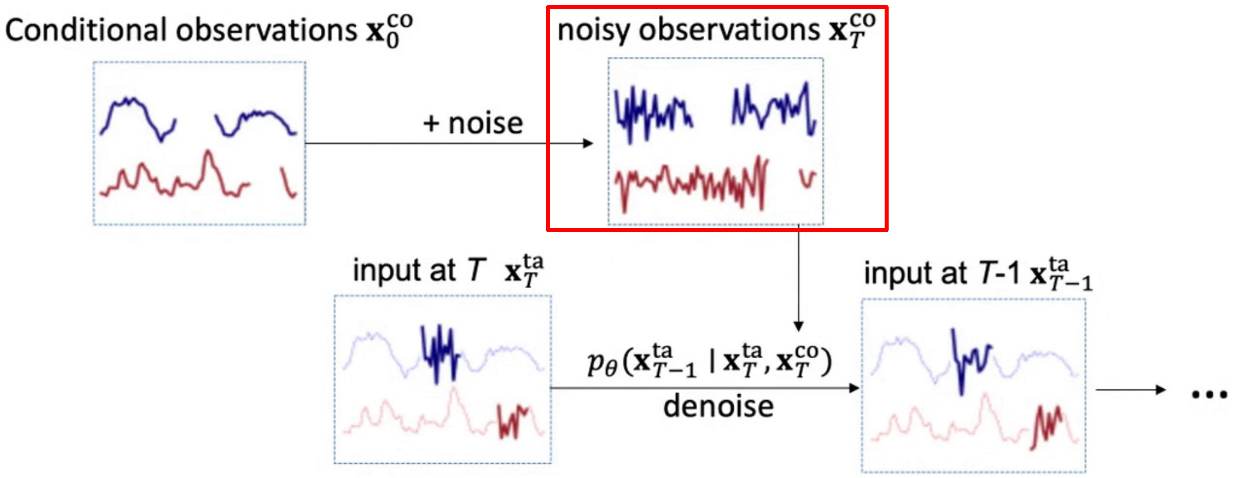
Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ CSDI

- **확률적 시계열 imputation을 위한** Conditional Score-based Diffusion Model
- Imputation task : $x_0^{co} \rightarrow x_0^{ta}$





<Imputation by score-based models>

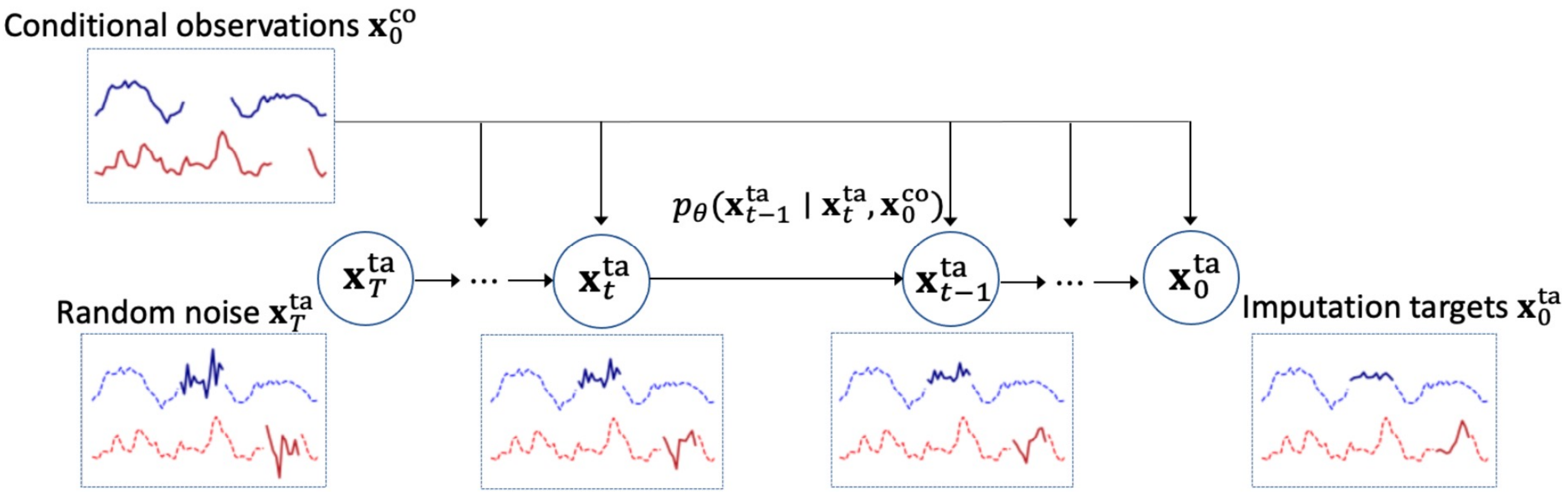
Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ CSDI

- **관측치**를 reverse process의 **conditional input**으로 활용



Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ Training

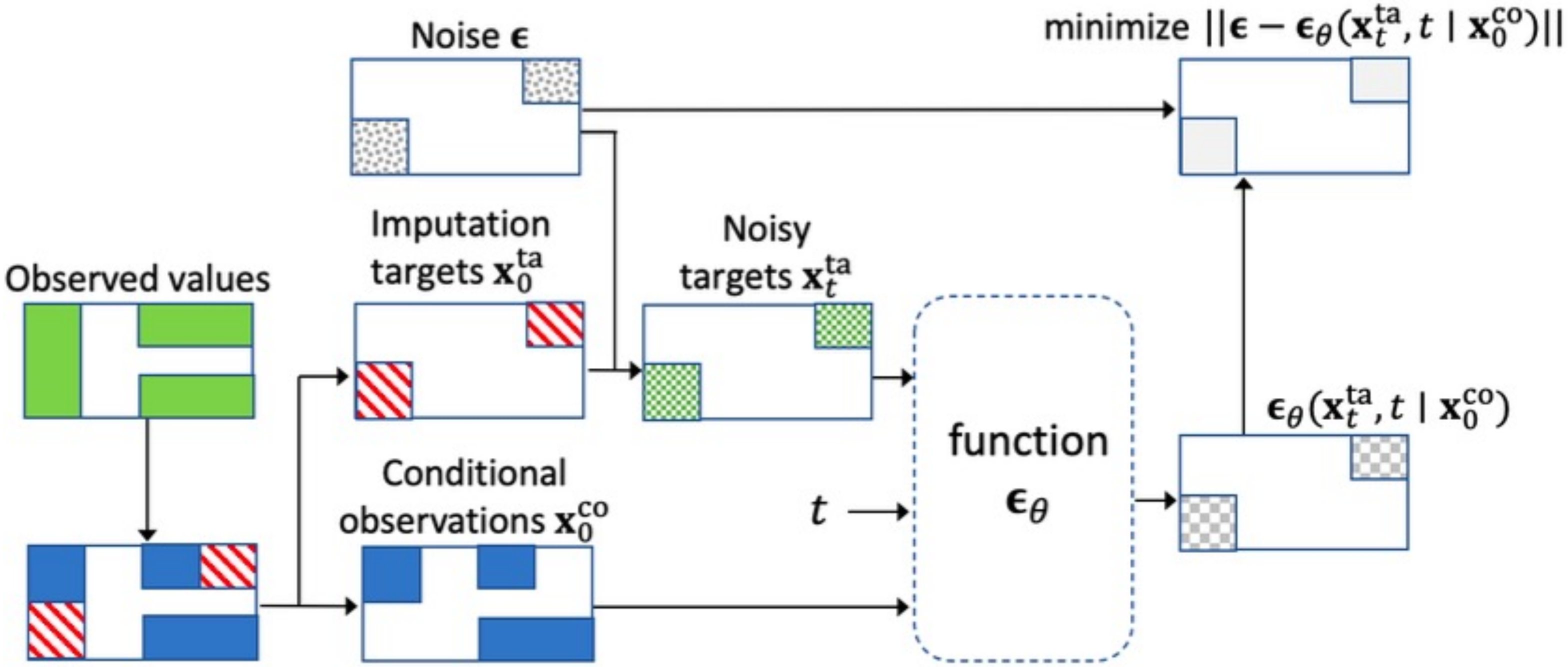
- Masked language modeling에서 착안하여 **self-supervised training method** 적용

(1) Random strategy

(2) Historical strategy

(3) Mix strategy

(4) Test pattern strategy



<The self-supervised training procedure of CSDI>

Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Diffusion Models for Time-series Imputation

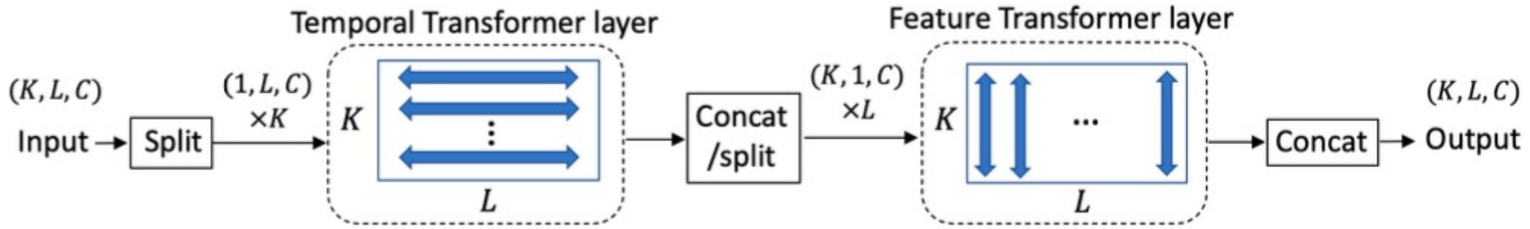
"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ x_0^{co} & x_0^{ta} at training and sampling

	imputation targets $\mathbf{x}_0^{\mathrm{ta}}$	conditional observations $\mathbf{x}_0^{\mathrm{co}}$
sampling (imputation)	all missing values	all observed values
training	a subset of the observed values (sampled by a target choice strategy)	the remaining observed values

❖ 2D attention mechanism

K : features
L : Length
C : channels



Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ Experiments

Table 2: Comparing CRPS for probabilistic imputation baselines and CSDI (lower is better). We report the mean and the standard error of CRPS for five trials.

	healthcare			air quality
	10% missing	50% missing	90% missing	
Multitask GP [31]	0.489(0.005)	0.581(0.003)	0.942(0.010)	0.301(0.003)
GP-VAE [10]	0.574(0.003)	0.774(0.004)	0.998(0.001)	0.397(0.009)
V-RIN [32]	0.808(0.008)	0.831(0.005)	0.922(0.003)	0.526(0.025)
unconditional	0.360(0.007)	0.458(0.008)	0.671(0.007)	0.135(0.001)
CSDI (proposed)	**0.238(0.001)**	**0.330(0.002)**	**0.522(0.002)**	**0.108(0.001)**

Table 3: Comparing MAE for deterministic imputation methods and CSDI. We report the mean and the standard error for five trials. The asterisks mean the results of the method are cited from the original paper.

	healthcare			air quality
	10% missing	50% missing	90% missing	
V-RIN [32]	0.271(0.001)	0.365(0.002)	0.606(0.006)	25.4(0.62)
BRITS [7]	0.284(0.001)	0.368(0.002)	0.517(0.002)	14.11(0.26)
BRITS [7] (*)	0.278	–	–	11.56
GLIMA [21] (*)	0.265	–	–	10.54
RDIS [20]	0.319(0.002)	0.419(0.002)	0.631(0.002)	22.11(0.35)
unconditional	0.326(0.008)	0.417(0.010)	0.625(0.010)	12.13(0.07)
CSDI (proposed)	**0.217(0.001)**	**0.301(0.002)**	**0.481(0.003)**	**9.60(0.04)**

Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Diffusion Models for Time-series Imputation

"CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation", NeurIPS, 2021

❖ Experiments

Table 4: Comparing the state-of-the-art interpolation methods with CSDI for the healthcare dataset. We report the mean and the standard error of CRPS for five trials.

	10% missing	50% missing	90% missing
Latent ODE [35]	0.700(0.002)	0.676(0.003)	0.761(0.010)
mTANs [22]	0.526(0.004)	0.567(0.003)	0.689(0.015)
CSDI (proposed)	**0.380**(**0.002**)	**0.418**(**0.001**)	**0.556**(**0.003**)

Table 5: Comparing probabilistic forecasting methods with CSDI. We report the mean and the standard error of CRPS-sum for three trials. The baseline results are cited from the original paper. 'TransMAF' is the abbreviation for 'Transformer MAF'.

	solar	electricity	traffic	taxi	wiki
GP-copula [34]	0.337(0.024)	0.024(0.002)	0.078(0.002)	0.208(0.183)	0.086(0.004)
TransMAF [36]	0.301(0.014)	0.021(0.000)	0.056(0.001)	0.179(0.002)	0.063(0.003)
TLAE [37]	**0.124**(**0.033**)	0.040(0.002)	0.069(0.001)	0.130(0.006)	0.241(0.001)
TimeGrad [25]	0.287(0.020)	0.021(0.001)	0.044(0.006)	**0.114**(**0.020**)	0.049(0.002)
CSDI (proposed)	0.298(0.004)	**0.017**(**0.000**)	**0.020**(**0.001**)	0.123(0.003)	**0.047**(**0.003**)

Tashiro, Y., Song, J., Song, Y., & Ermon, S. (2021). Csdi: Conditional score-based diffusion models for probabilistic time series imputation. Advances in Neural Information Processing Systems, 34, 24804-24816

Conclusion

Applications of Diffusion Models

❖ Applications of Diffusion Models

- **Text-to-Image Generation**

  ✓ **GLIDE**: Classifier-free guidance와 CLIP guidance를 통해 text-guided image generation and Editing 수행

- **Anomaly Detection**

  ✓ **AnoDDPM**: A partial diffusion anomaly detection with Gaussian noise and multi-scale simplex noise

- **Natural Language Generation**

  ✓ **DiffuSeq**: Sequence to Sequence Text Generation with Diffusion Models in classifier-free manner

- **Time-series Forecasting & Imputation**

  ✓ **TimeGrad**: Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting

  ✓ **CSDI**: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation