Introduction to Personalization with Diffusion Models

Geonhui Jang

2023.09.15

Introduction

❖ **장건희 (Geonhui Jang)**

- 고려대학교 산업경영공학과 석사 과정 (2023.03 ~ Present)
- Data Mining & Quality Analytics Lab

❖ **Research Interests**

- Deep Generative Models
- Diffusion Models

❖ **Contact**

- csleivear1@korea.ac.kr

목차

Diffusion Models 발전 과정

Image Editing vs. Personalization

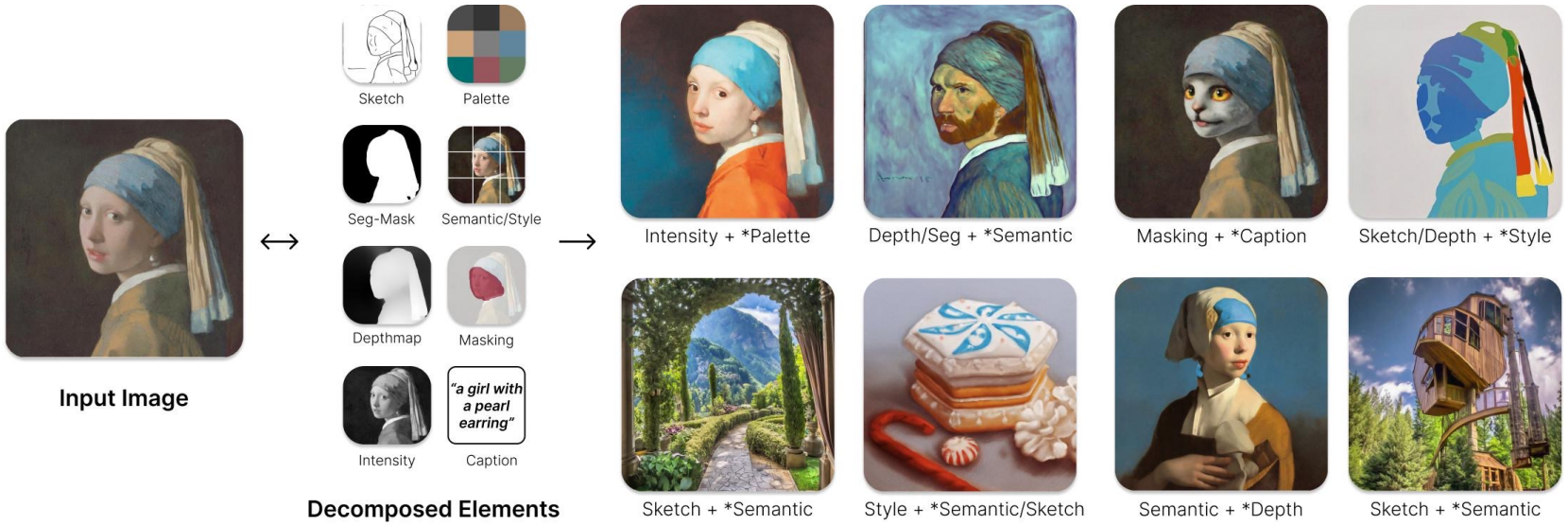


Image Editing

Personalization

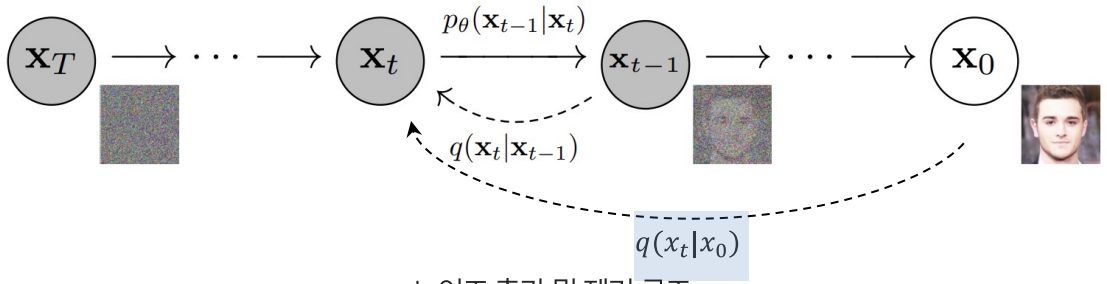
Huang, Lianghua, et al. "Composer: Creative and controllable image synthesis with composable conditions." arXiv preprint arXiv:2302.09778 (2023).
Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. “DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation.” IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Diffusion Models 발전 과정

1. Unconditional Generation → 2. Conditional Generation → 3. Image Editing → 4. Personalization

1. 학습 데이터셋에 따른 무작위 이미지 생성

DDPM (Denoising Diffusion Probabilistic Models 2020.6.19)



<노이즈 추가 및 제거 구조>

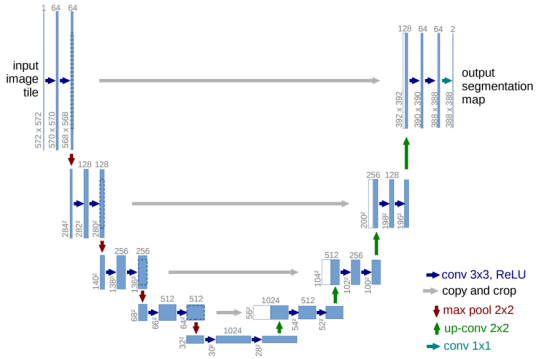
Algorithm 1 Training

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

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** $\mathbf{x}_0$

<학습 및 샘플링 과정>



모델: UNet 사용

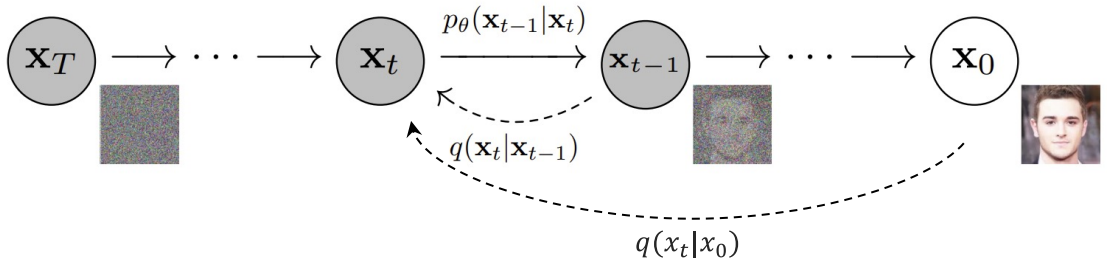
Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Diffusion Models 발전 과정

1. Unconditional Generation → 2. Conditional Generation → 3. Image Editing → 4. Personalization

1. 학습 데이터셋에 따른 무작위 이미지 생성

DDPM (Denoising Diffusion Probabilistic Models 2020.6.19)



<노이즈 추가 및 제거 구조>

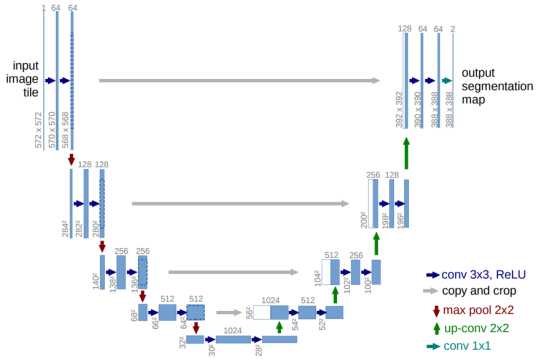
Algorithm 1 Training

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

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** $\mathbf{x}_0$

<학습 및 샘플링 과정>



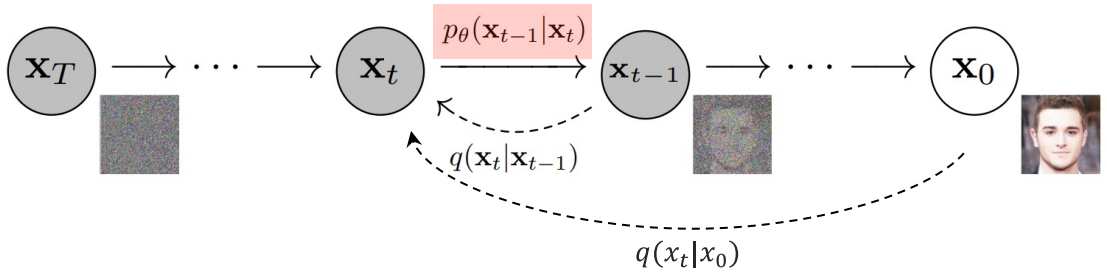
모델: UNet 사용

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Diffusion Models 발전 과정

1. 학습 데이터셋에 따른 무작위 이미지 생성

DDPM (Denoising Diffusion Probabilistic Models 2020.6.19)



<노이즈 추가 및 제거 구조>

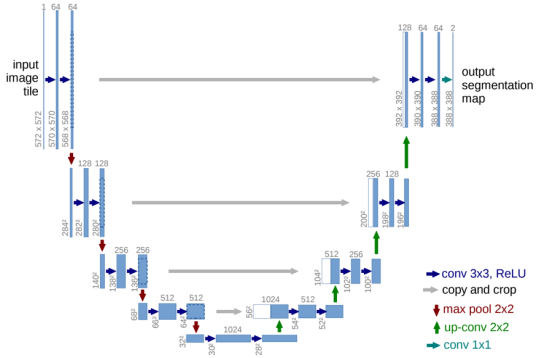
Algorithm 1 Training

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

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** $\mathbf{x}_0$

<학습 및 샘플링 과정>



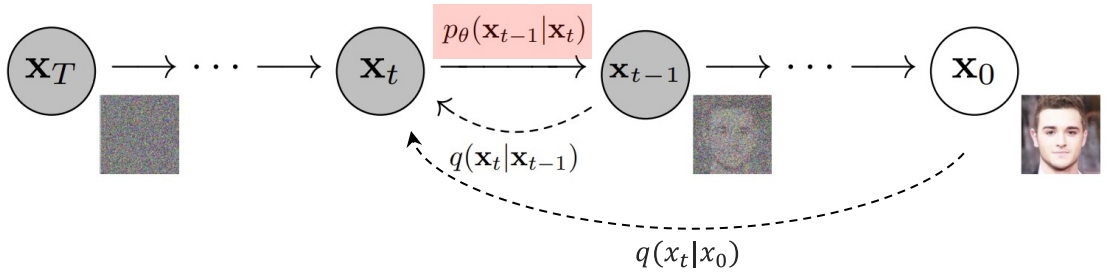
모델: UNet 사용

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Diffusion Models 발전 과정

1. 학습 데이터셋에 따른 무작위 이미지 생성

DDPM (Denoising Diffusion Probabilistic Models 2020.6.19)



<노이즈 추가 및 제거 구조>

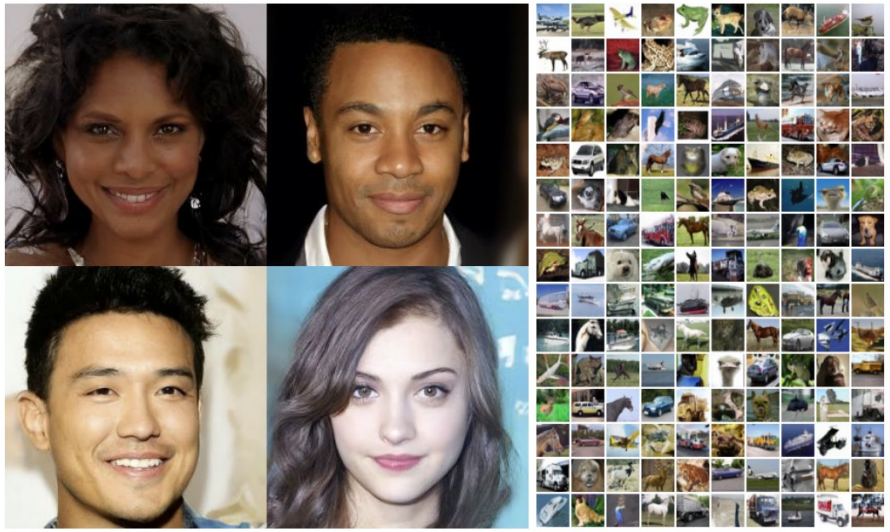
Algorithm 1 Training

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

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** $\mathbf{x}_0$

<학습 및 샘플링 과정>



(1) SelebA-HQ 데이터셋으로 학습한 모델의 output

(2) CIFAR10 데이터셋으로 학습한 모델의 output

<학습한 데이터셋에 따라 무작위로 이미지 생성>

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

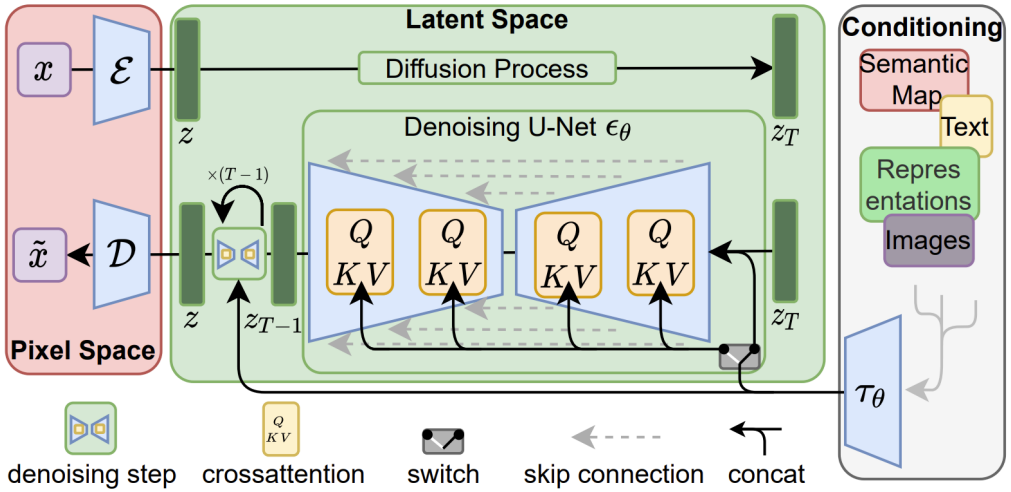
Diffusion Models 발전 과정

2. 입력 조건에 맞는 이미지 생성

CFG (Classifier-free Guidance 2021), **LDM** (Latent Diffusion Models 2021.12.20)

$$\tilde{\epsilon}_\theta = \epsilon_\theta(z_t, t) + \omega \cdot \left(\epsilon_\theta(z_t, t, c) - \epsilon_\theta(z_t, t)\right)$$

<CFG를 이용한 noise prediction>



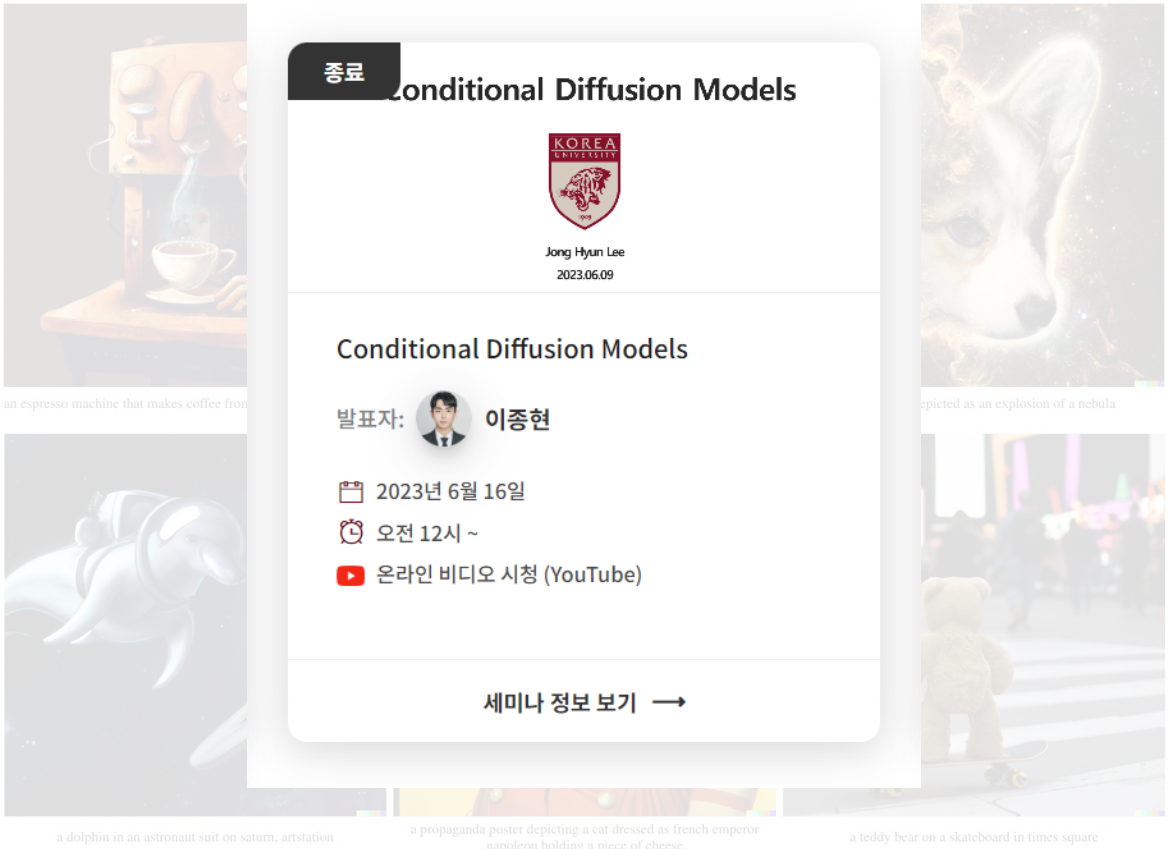
<Latent Diffusion Models 구조>

Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv preprint arXiv:2207.12598* (2022).
Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Diffusion Models 발전 과정

2. 입력 조건에 맞는 이미지 생성

DALL-E 2 (2022.4.13)



<DALL-E 2로 생성한 이미지들>

Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents." *arXiv preprint arXiv:2204.06125* 1.2 (2022): 3.

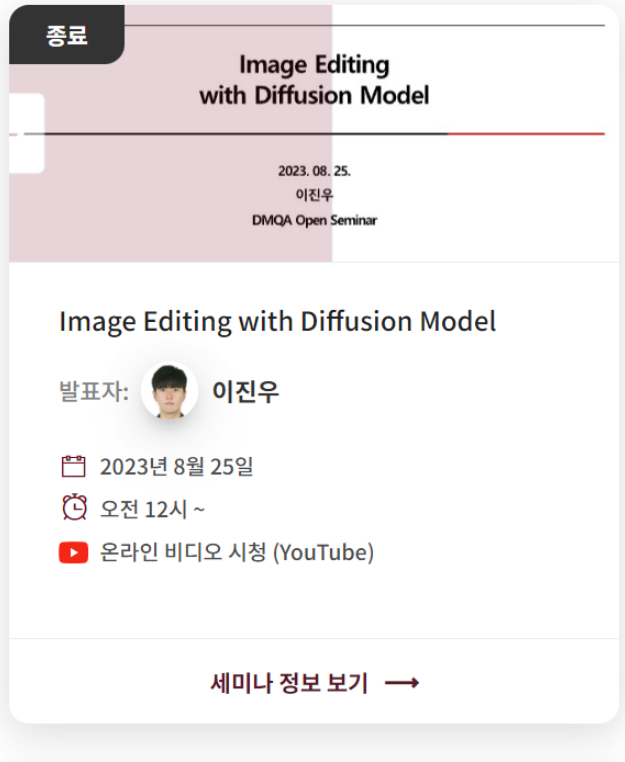
Diffusion Models 발전 과정

3. 이미지 편집

P2P (Prompt-to-prompt 2022.8.2), Null-text Inversion (2022.11.17), Plug-and-Play (2022.11.22)

<P2P으 synthe

<Null-te real ima

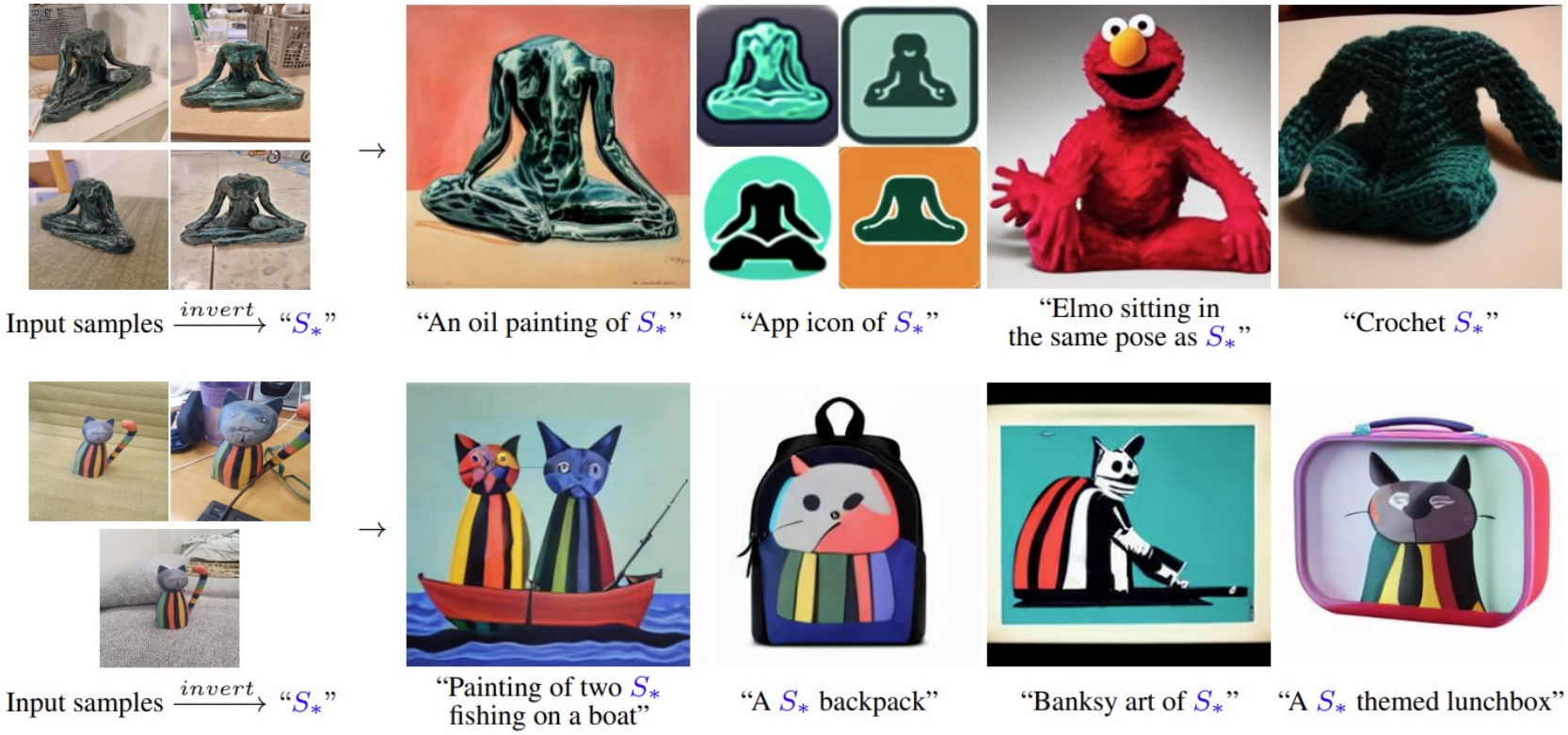


<Plug-and-Play의 real image 편집>

Ho, Jonathan, and Tim Salimans. "Classifier-free diffusion guidance." *arXiv preprint arXiv:2207.12598* (2022).
Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Textual Inversion – Personalization의 시작

An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion (2022.8.2)



<이미지 내 대상을 스페셜 토큰 S_*에 담아 이를 텍스트 프롬프트에 사용>

Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. "An image is worth one word: Personalizing text-to-image generation using textual inversion." International Conference on Learning Representations, 2023.

Textual Inversion – Personalization의 시작

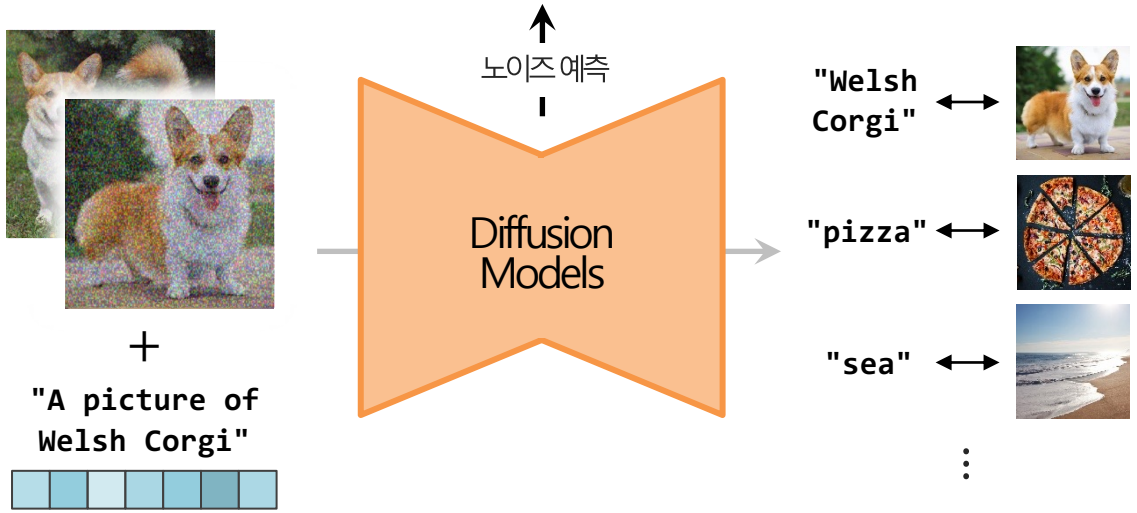
An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion (2022.8.2)

Diffusion Models 학습 방식

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

$$L = \mathbb{E}_{t,\mathbf{x}_0,y,\boldsymbol{\epsilon}} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\mathbf{x}_t, t, y \right) \right\|^2 \right], \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$$



모델 ϵ_θ가 이미지에 추가된 노이즈를 예측하도록 학습

Textual Inversion – Personalization의 시작

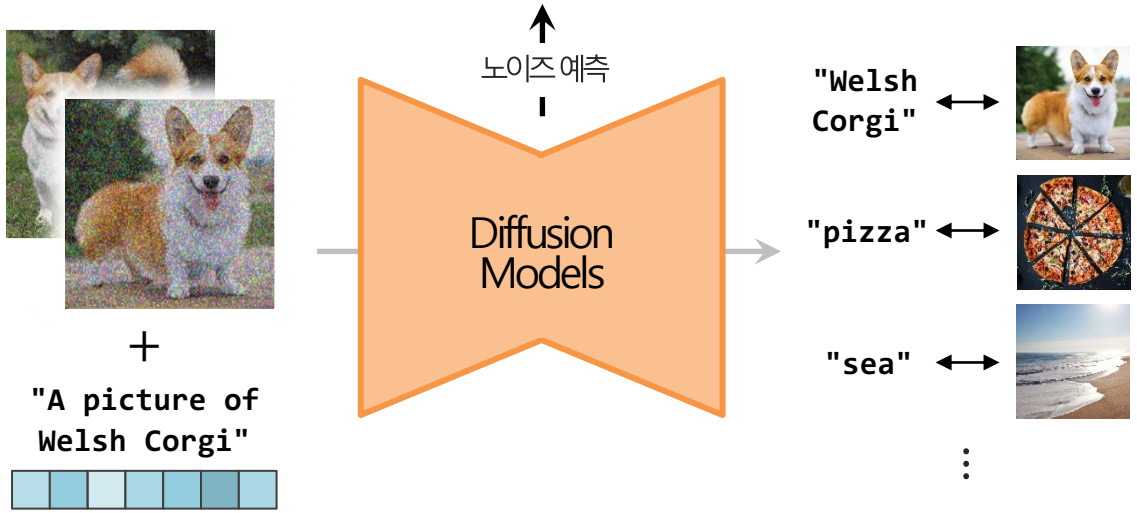
An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion (2022.8.2)

Diffusion Models 학습 방식

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

$$L = \mathbb{E}_{t,\mathbf{x}_0,y,\boldsymbol{\epsilon}} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\mathbf{x}_t, t, y \right) \right\|^2 \right], \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$$

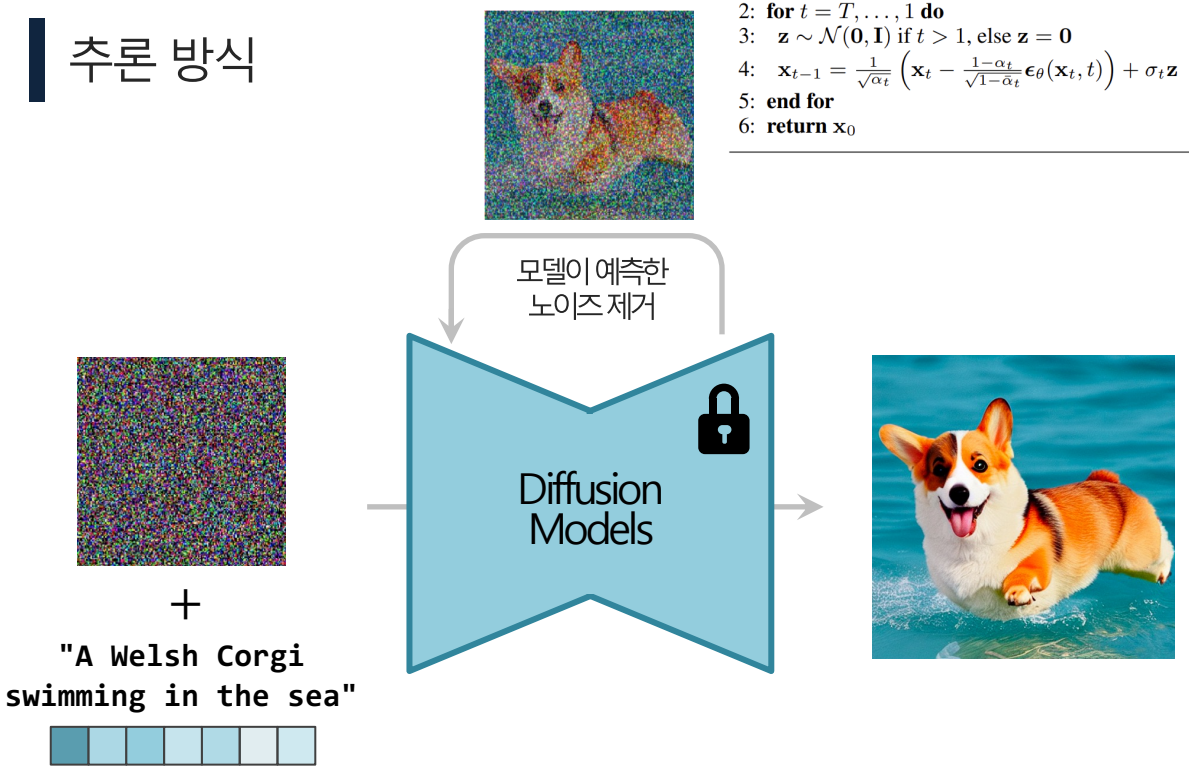


모델 ϵ_θ가 이미지에 추가된 노이즈를 예측하도록 학습

추론 방식

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** $\mathbf{x}_0$



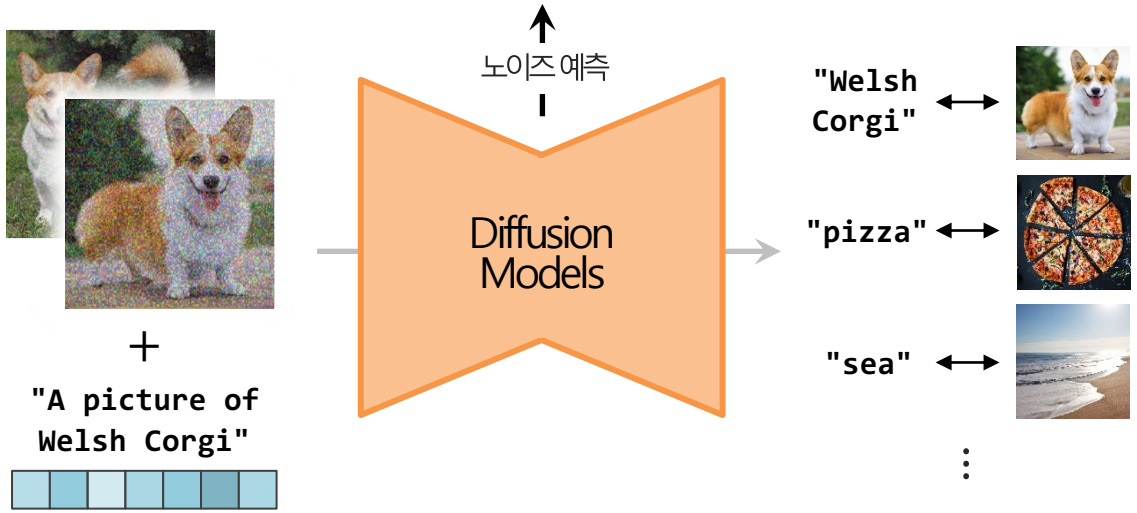
가우시안 노이즈로부터 프롬프트의 방향으로 노이즈를 반복적으로 제거

Textual Inversion – Personalization의 시작

An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion (2022.8.2)

Diffusion Models 학습 방식

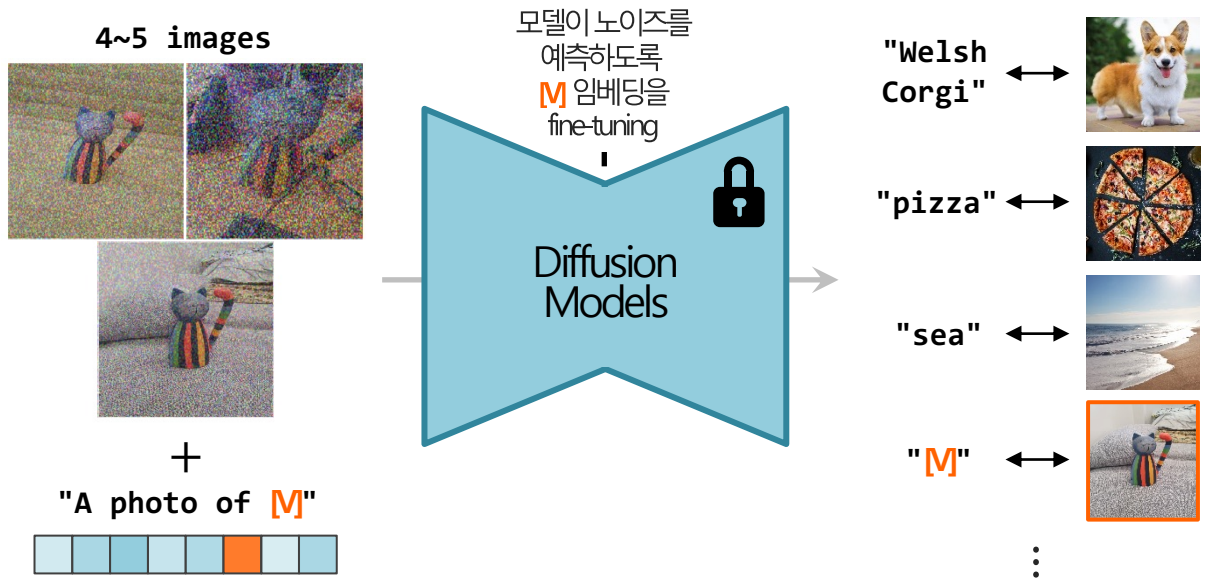
$$L = \mathbb{E}_{t,\mathbf{x}_0,y,\epsilon}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t, y)\|^2\right], \quad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$



모델 ϵ_θ가 이미지에 추가된 노이즈를 예측하도록 학습

Textual Inversion 학습 방식

$$v_* = \arg\min_v \mathbb{E}_{t,\mathbf{x}_0,y,\epsilon}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t, y)\|^2\right], \quad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$



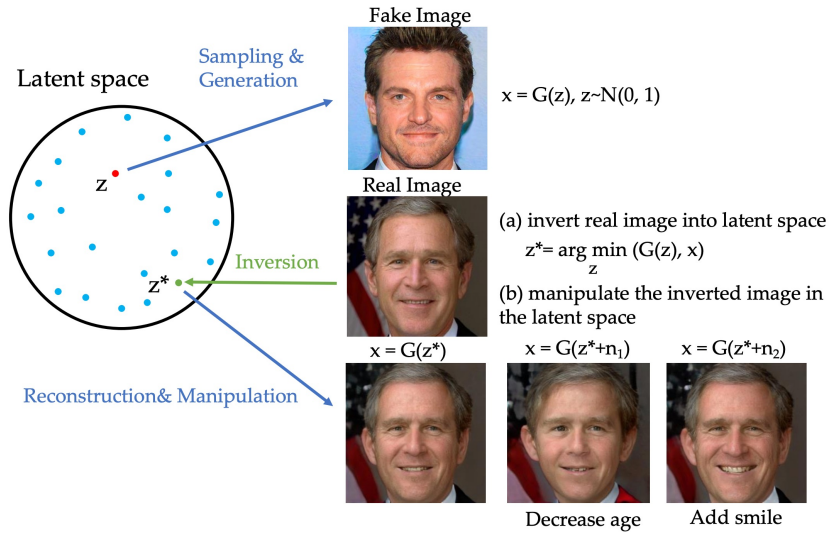
스페셜 토큰 [M]가 텍스트 공간 상에서
대상을 잘 나타낼 수 있도록 [M]를 fine-tuning

Textual Inversion – Personalization의 시작

An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion (2022.8.2)

GAN Inversion과의 차이점

- GAN Inversion은 주어진 **이미지**에 대응되는 latent representation을 찾아 이를 이미지 편집에 사용
- Textual Inversion은 입력 이미지 내 대상의 **concept**을 invert
  → 이 concept을 모델의 vocabulary에 추가하여 text representation을 활용해 더 직관적으로 이미지 편집 가능



<GAN Inversion>

<Textual Inversion>

Xia, Weihao, et al. "Gan inversion: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.3 (2022): 3121-3138.

Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. "An image is worth one word: Personalizing text-to-image generation using textual inversion." International Conference on Learning Representations, 2023.

DreamBooth

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (2022.8.25)

<이미지 내 대상을 스페셜 토큰 [V]에 담아 이를 텍스트 프롬프트에 사용>

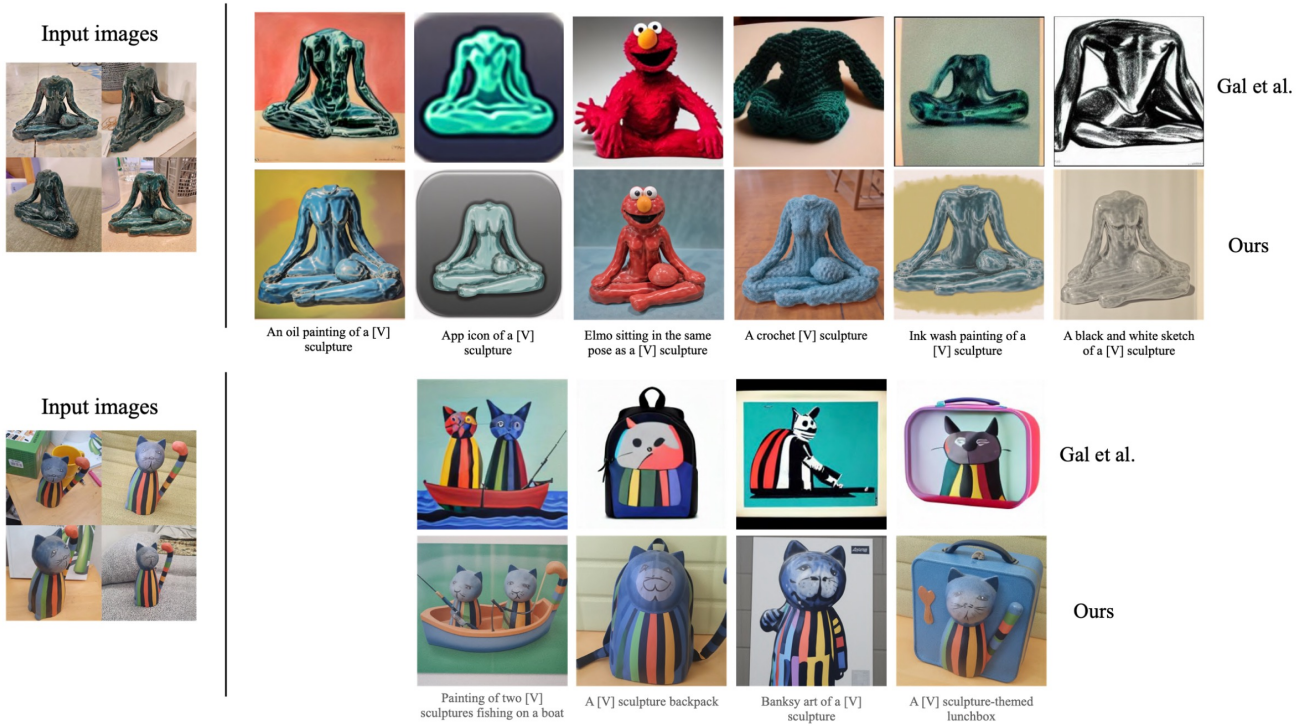
Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. “DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation.” IEEE Conference on Computer Vision and Pattern Recognition, 2023.

DreamBooth

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (2022.8.25)

학습 방식

- **Textual Inversion**은 스페셜 토큰 [V]의 임베딩만 fine-tuning
  ↔ **DreamBooth**는 Diffusion Models 레이어 전체 fine-tuning
- 768차원의 fine-tuned 임베딩 하나를 text-image 공간에 끼워넣는 것보다 UNet, 텍스트 인코더 등 모델의 레이어까지 fine-tuning 하는 것이 유리
- 모델의 모든 레이어를 fine-tuning했을 때 maximum subject fidelity 기록



<Textual Inversion과 DreamBooth의 비교>

Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. “DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation.” IEEE Conference on Computer Vision and Pattern Recognition, 2023.

DreamBooth

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (2022.8.25)

Class-specific Prior Preservation Loss

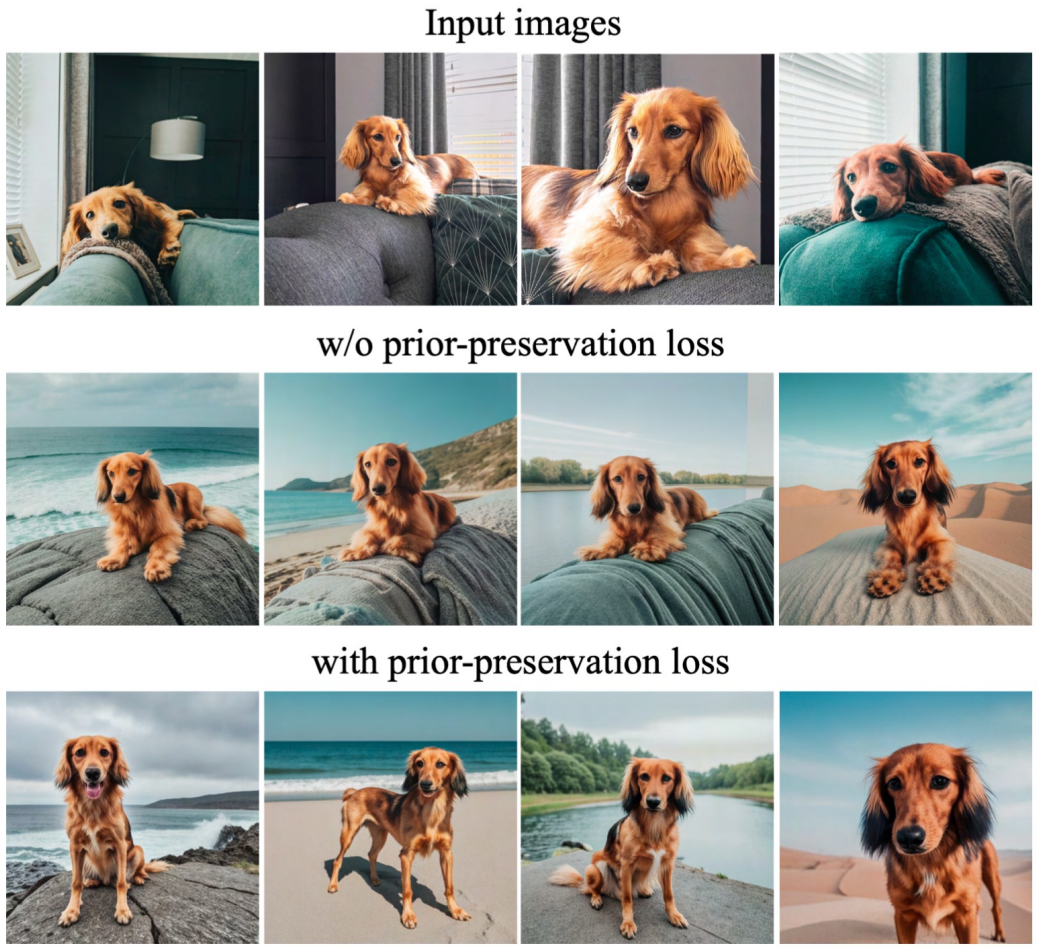
- **모델의 모든 레이어를 fine-tuning했을 때 발생하는 문제점**

  (1) 텍스트 임베딩에 영향을 받는 레이어 역시 입력 이미지에 대해 fine-tuned 되는데, 이때 **Language Drift** 문제 발생

  * Language Drift : large text corpus에 대해 학습한 언어 모델이 이후 특정 task를 위해 fine-tuned 될 때 syntactic & semantic knowledge를 잊는 현상

  (2) 샘플링 시 모델이 생성하는 이미지의 **다양성 감소**

- 논문에서는 **Class-specific Prior Preservation Loss**를 도입하여 이러한 문제 해결



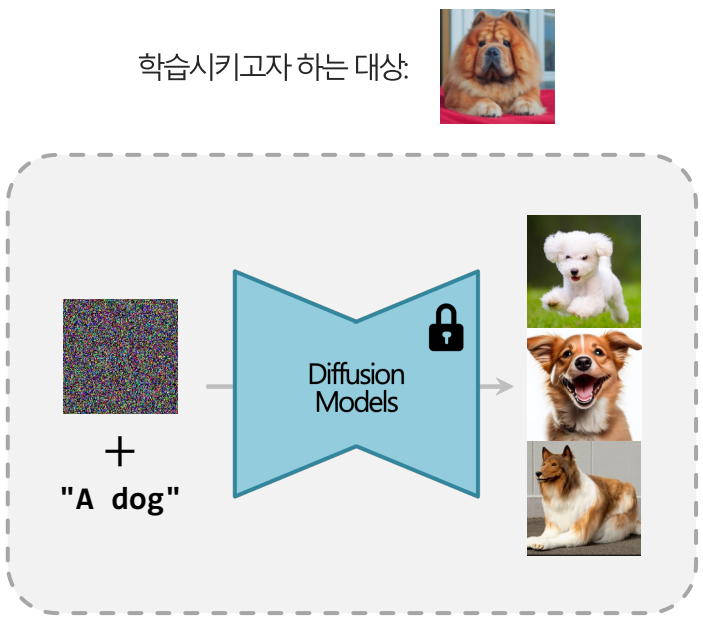
(2) **생성 이미지 다양성 감소** (두 번째 열)

Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. “DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation.” IEEE Conference on Computer Vision and Pattern Recognition, 2023.

DreamBooth

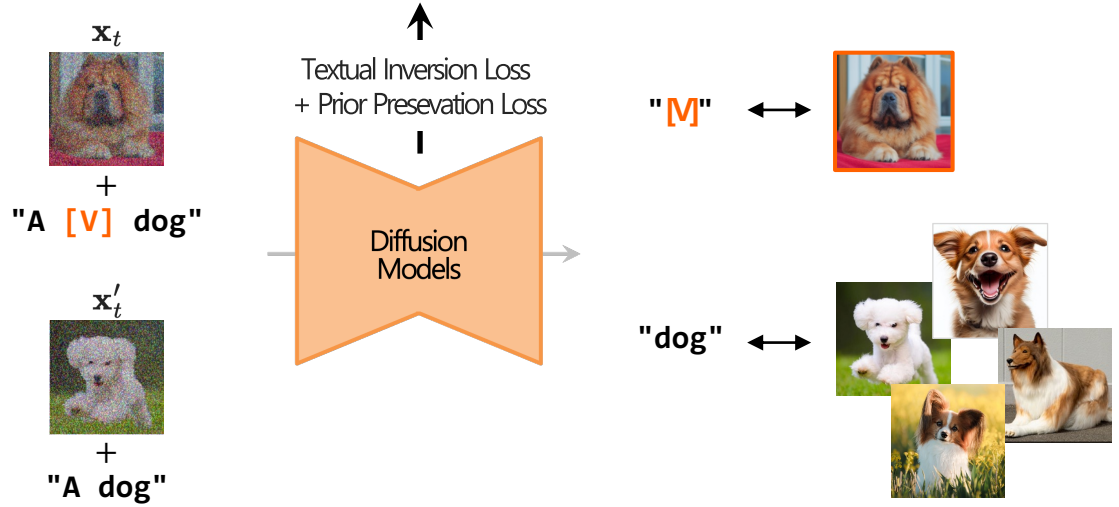
DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (2022.8.25)

Class-specific Prior Preservation Loss를 반영한 모델 학습



<Prior Presevation Loss를 위한 데이터셋 준비>

$$\mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}'_0,y,y',\epsilon,\epsilon'}\left[\|\epsilon-\epsilon_\theta(\mathbf{x}_t,t,y)\|^2+\lambda\|\epsilon'-\epsilon_\theta(\mathbf{x}'_t,t,y')\|^2\right]$$

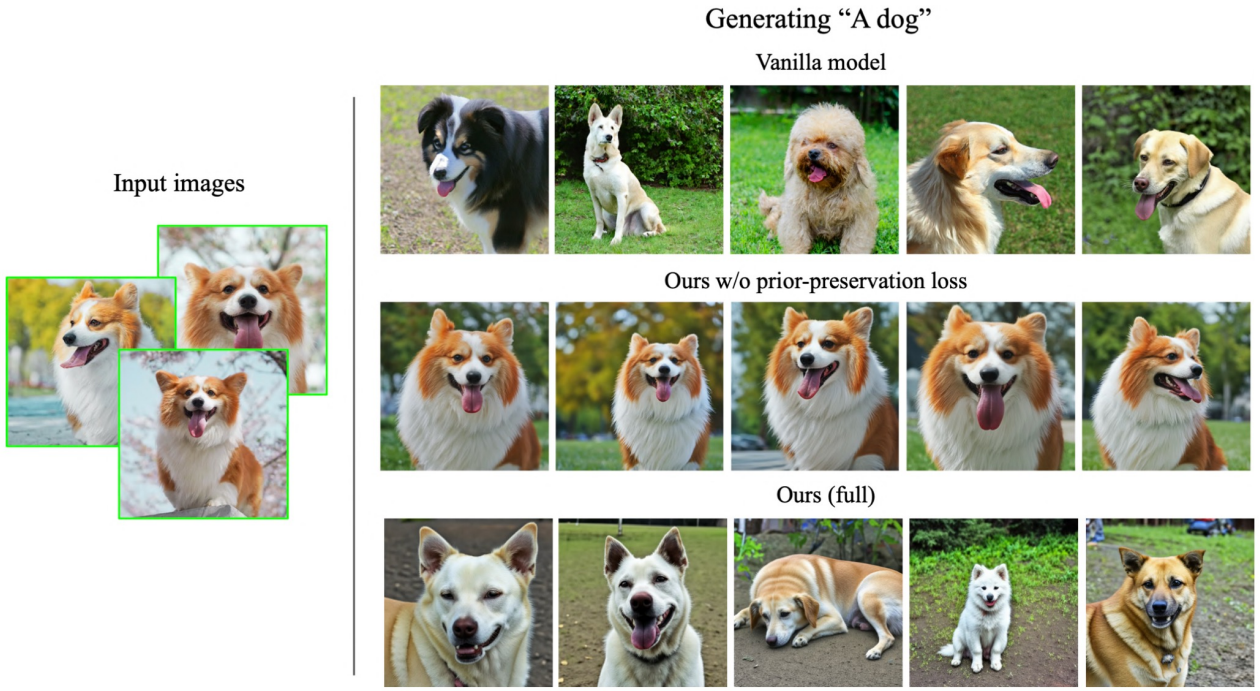


<DreamBooth 학습 과정>

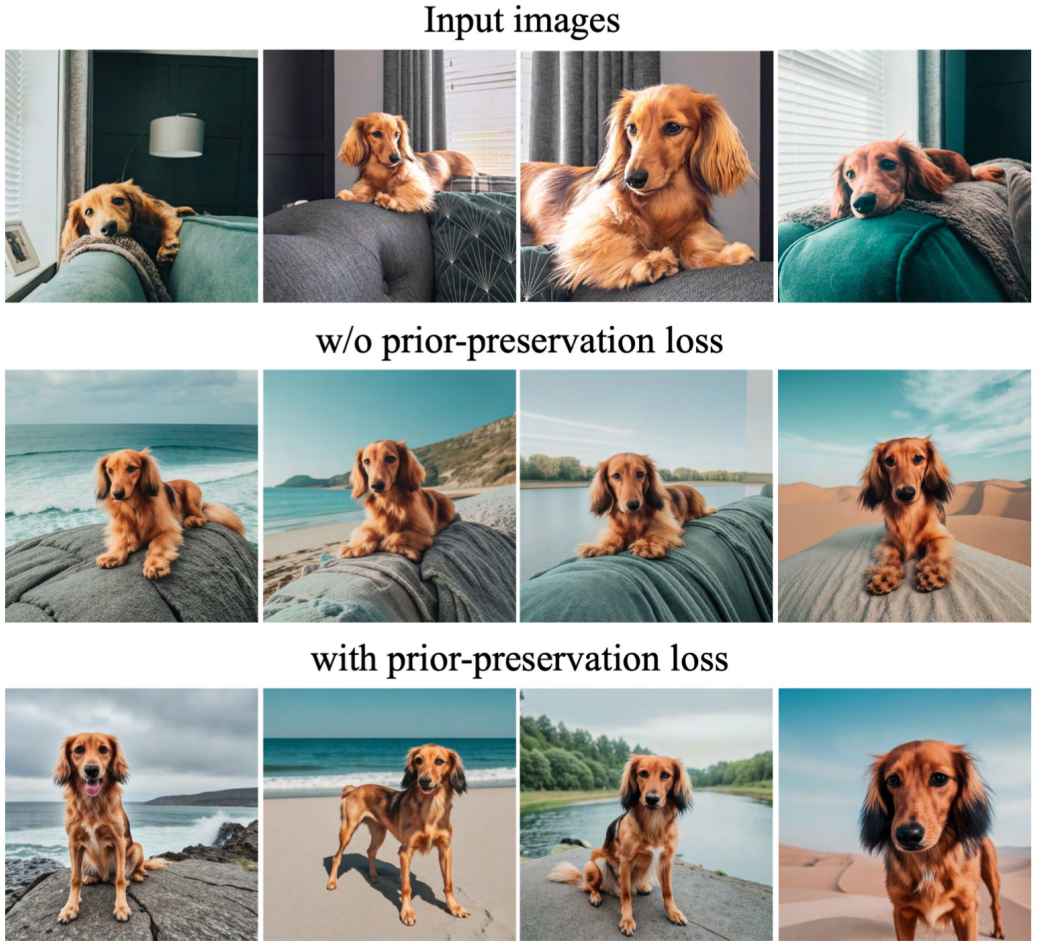
DreamBooth

DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (2022.8.25)

Class-specific Prior Preservation Loss를 반영한 모델 학습



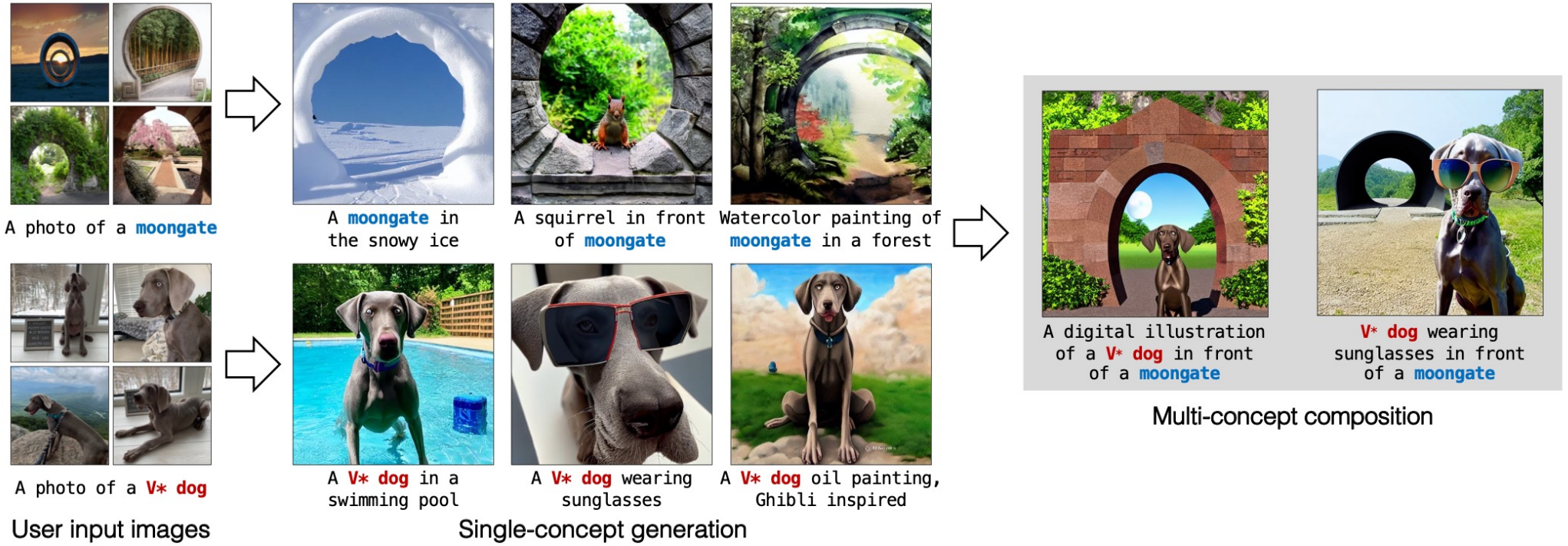
(1) **Language Drift 완화** (세 번째 열)



(2) **생성 이미지 다양성 유지** (세 번째 열)

Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)



<이미지 내 대상을 기존 단어 또는 스페셜 토큰 [V]에 담아 이를 텍스트 프롬프트에 사용>

Kumari, Nupur, et al. "Multi-concept customization of text-to-image diffusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)

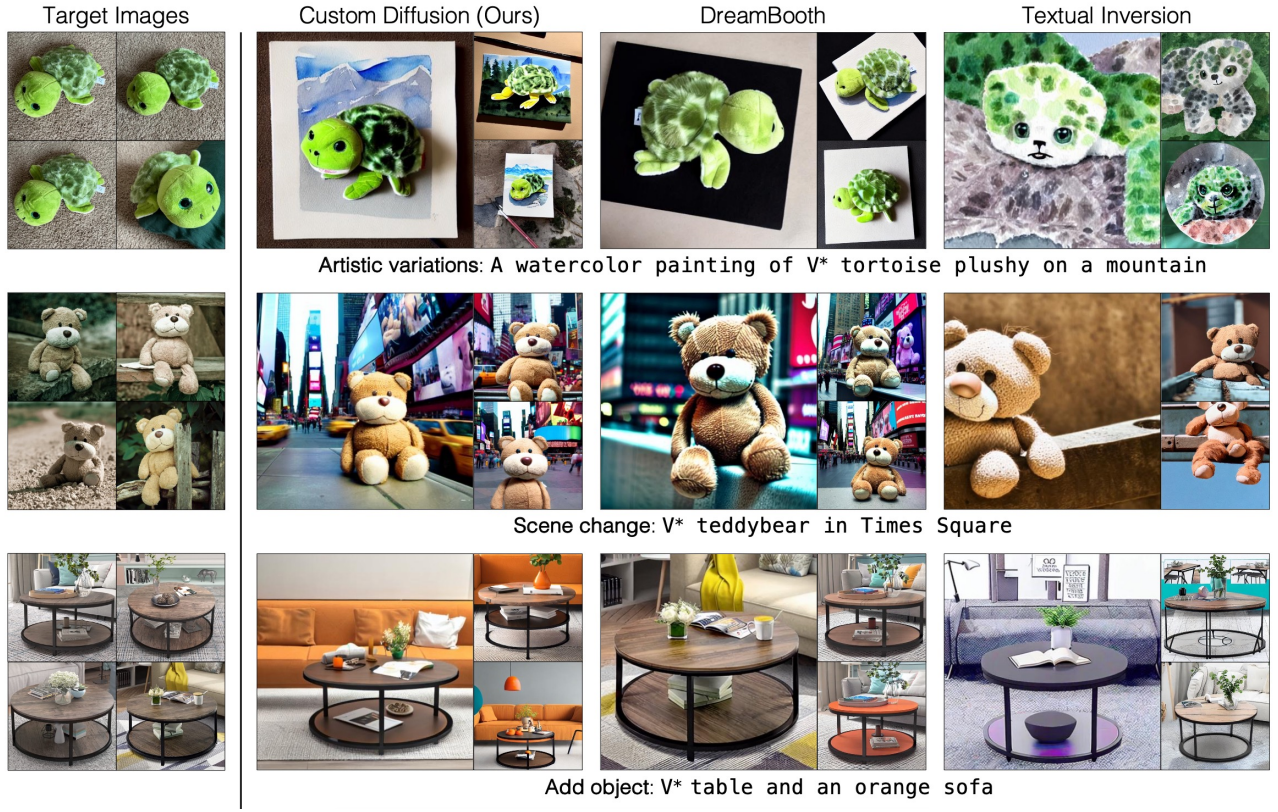
학습 방식

- **Textual Inversion**은 스페셜 토큰 [V]의 임베딩만 fine-tuning
  DreamBooth는 Diffusion Models 레이어 전체 fine-tuning
  ↔ **Custom Diffusion**은 Diffusion Models의 Cross attention blocks 중 W^K, W^V matrices만 fine-tuning

- fine-tuning의 목적: 주어진 텍스트가 이미지 분포에 mapping 되도록 파라미터를 업데이트 하는 것

  + text features는 Cross attention blocks의 W^K, W^V matrices를 거침

  → 따라서 **fine-tuning 과정에서 W^K와 W^V의 파라미터만 업데이트**



<Custom Diffusion, DreamBooth,Textual Inversion 비교>

Kumari, Nupur, et al. "Multi-concept customization of text-to-image diffusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)

학습 방식

- Cross attention layers의 파라미터는 UNet 전체 파라미터의 5%에 불과 → 그러나 fine-tuning 시 변화량은 가장 큼
- Fine-tuning 시 cross attention layers만 업데이트해도 효과적으로 모델을 학습시킬 수 있음

UNet의 convolution layer, normalization layer 등

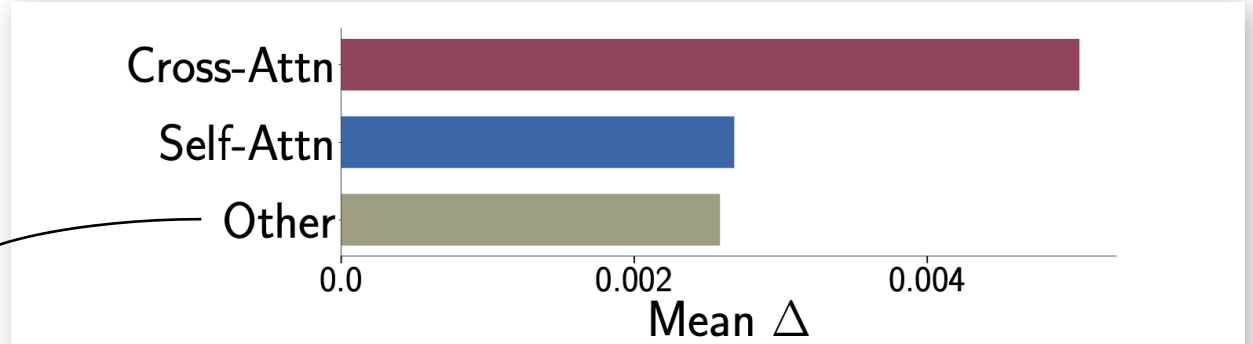


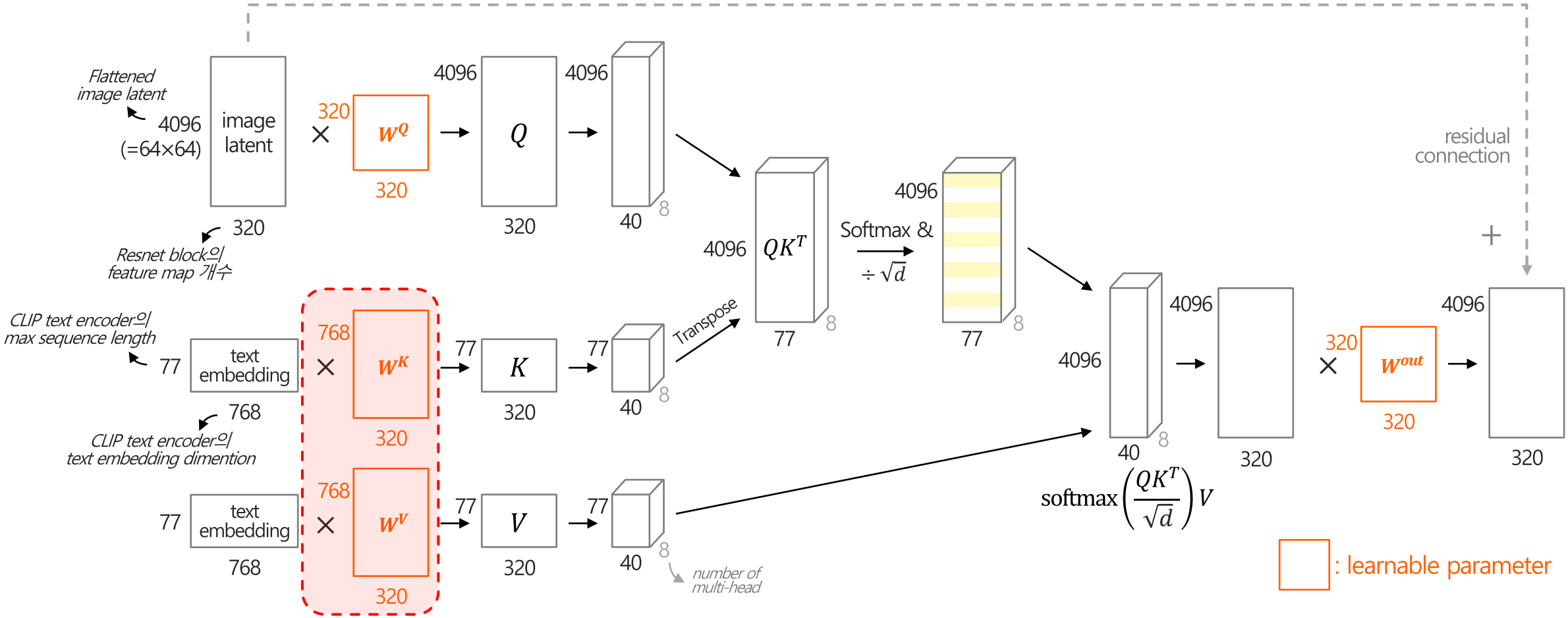
Figure 3. **Analysis of change in weights** on updating all network weights during fine-tuning. The mean change in the cross-attention layers is significantly higher than other layers even though they only make up 5% of the total parameter count.

<Fine-tuning 시 UNet 내부 3종류 레이어의 평균 파라미터 변화량>

Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)

Cross Attention in UNet



Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)

Multiple Concepts Fine-tuning

1. **Joint training on multiple concepts**
   - 단순히 각 concept i에 대해 스페셜 토큰 [V]i를 각각 정의하고 fine-tuning 시 이들을 모두 학습에 이용

2. **Constrained optimization to merge concepts**
   - 각 concept i에 대한 text features C를 concat하여 라그랑주 승수법을 이용해 closed-form 해를 구함
   - 각 concept i에 학습된 matrix W_i가 필요

- W_0: pre-trained model parameters
- W_i: fine-tuned parameters for concept i

$$\hat{W} = \arg\min_{W} ||WC_{\text{reg}}^{\top} - W_0 C_{\text{reg}}^{\top}||_F$$
$$\text{s.t. } WC^{\top} = V, \text{ where } C = [\mathbf{c}_1 \cdots \mathbf{c}_N]^{\top}$$
$$\text{and } V = [W_1\mathbf{c}_1^{\top} \cdots W_N\mathbf{c}_N^{\top}]^{\top}$$

$$\rightarrow \quad \hat{W} = W_0 + \mathbf{v}^{\top}\mathbf{d}, \text{ where } \mathbf{d} = C(C_{\text{reg}}^{\top}C_{\text{reg}})^{-1}$$
$$\text{and } \mathbf{v}^{\top} = (V - W_0C^{\top})(\mathbf{d}C^{\top})^{-1}$$

Custom Diffusion

Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)

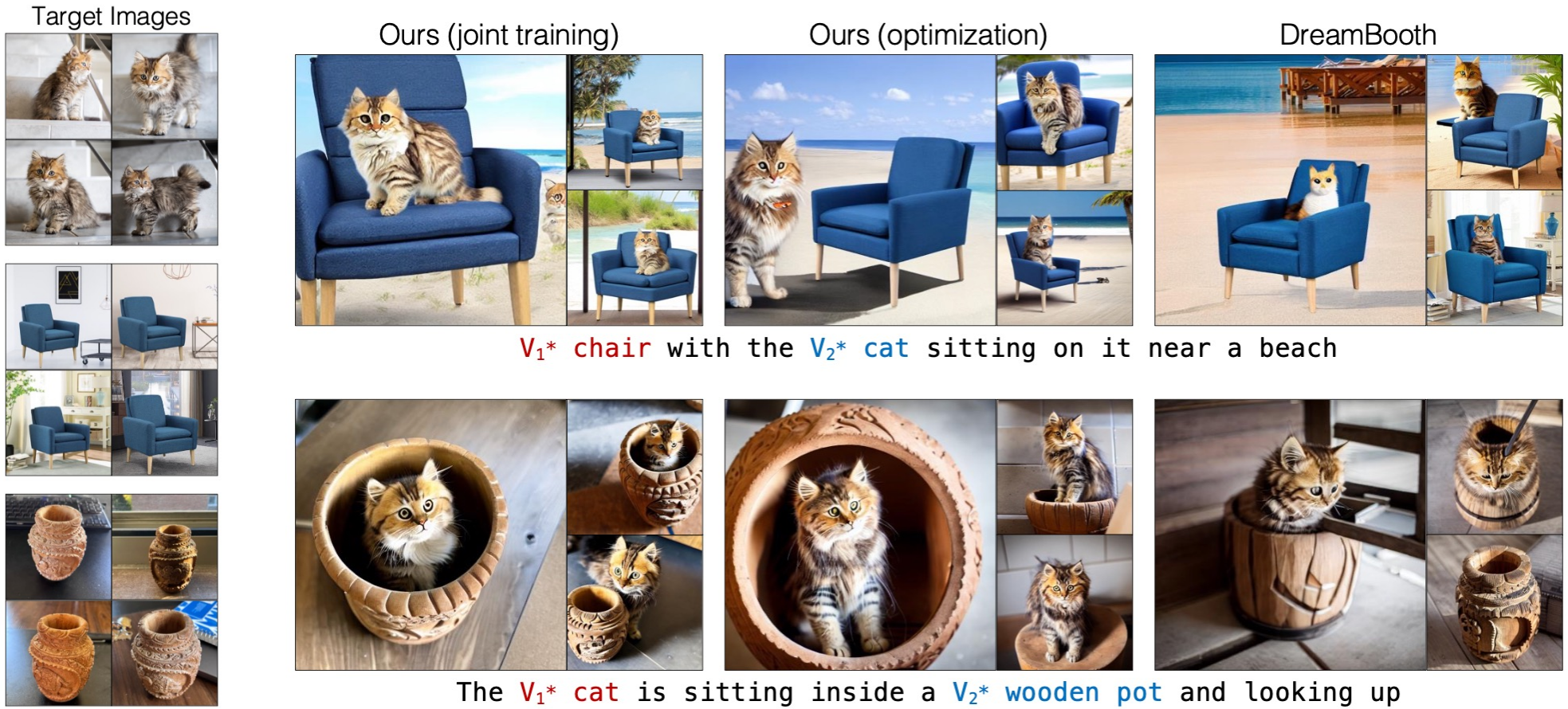
Regularization

- DreamBooth와 유사하게 language drift 문제 제기
- fine-tuning 시 대상 단어를 그대로 이용하거나 스페셜 토큰 [V] 도입
  → 이때 대상 단어를 그대로 사용할 경우 **language drift** 발생
- 논문에서는 **regularization dataset**을 도입하여 이러한 문제 해결
- target text prompt와 CLIP text score가 0.85 이상인 캡션을 갖는 이미지를 LAION-400M 데이터셋에서 200개 선택
  → 이를 DreamBooth와 같은 방식으로 fine-tuning 시 regularization에 사용

<"moongate"라는 단어를 학습에 그대로 사용할 경우 "moon"에 대해 language drift 발생>

Custom Diffusion

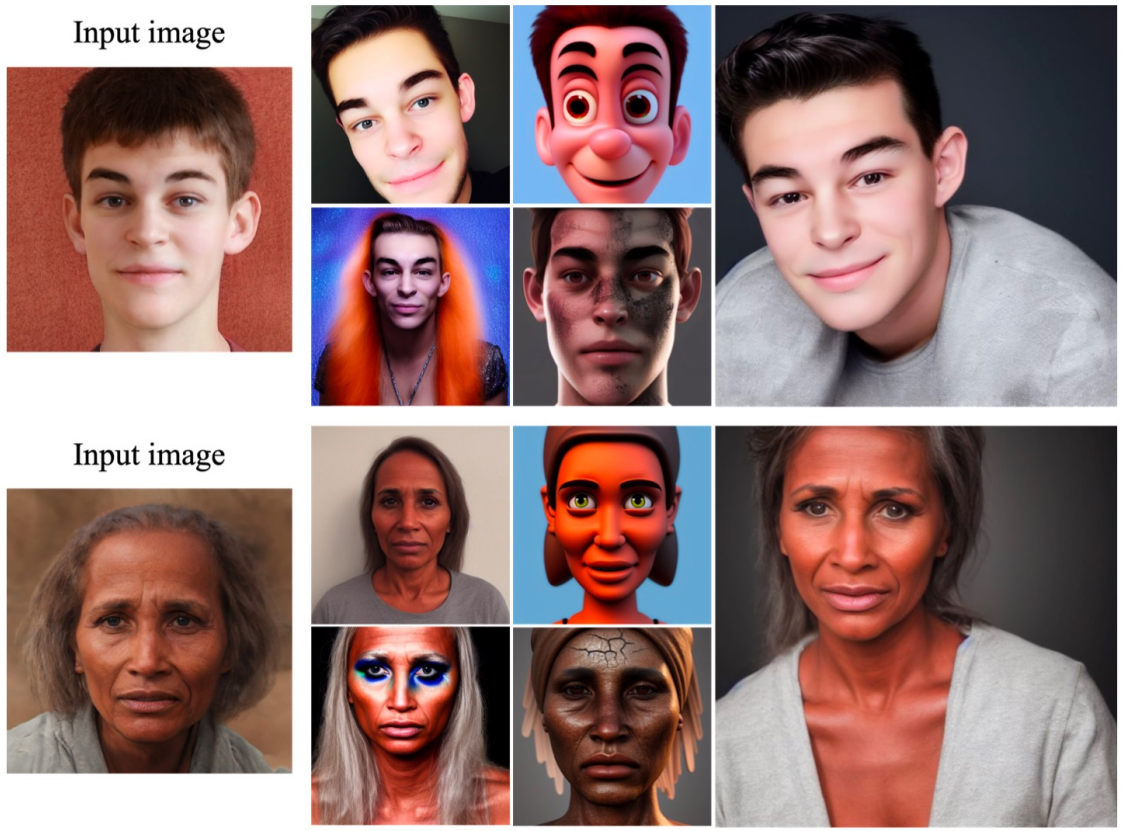
Multi-Concept Customization of Text-to-Image Diffusion (2022.12.8)



<Custom Diffusion의 joint training, optimization, 그리고 DreamBooth 결과 비교>

Kumari, Nupur, et al. "Multi-concept customization of text-to-image diffusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

Personalization



<face에 특화된 personalization>

<style에 특화된 personalization>

Ruiz, Nataniel, et al. "Hyperdreambooth: Hypernetworks for fast personalization of text-to-image models." *arXiv preprint arXiv:2307.06949* (2023).
Sohn, Kihyuk, et al. "StyleDrop: Text-to-Image Generation in Any Style." *arXiv preprint arXiv:2306.00983* (2023).

고맙습니다

<cat과 tiger의 텍스트 임베딩을 절반씩 섞어 생성한 이미지>