2023 하계 세미나

Recent improvements in diffusion models

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
이창헌

Outline

- Background
  - Diffusion models
- Papers
  - Additional model based methods
    - Plug-and-play
    - ControlNet
  - Single model based methods
    - eDiff-I
    - SDXL

Background

- DDPM diffusion models[1]
  - Forward process

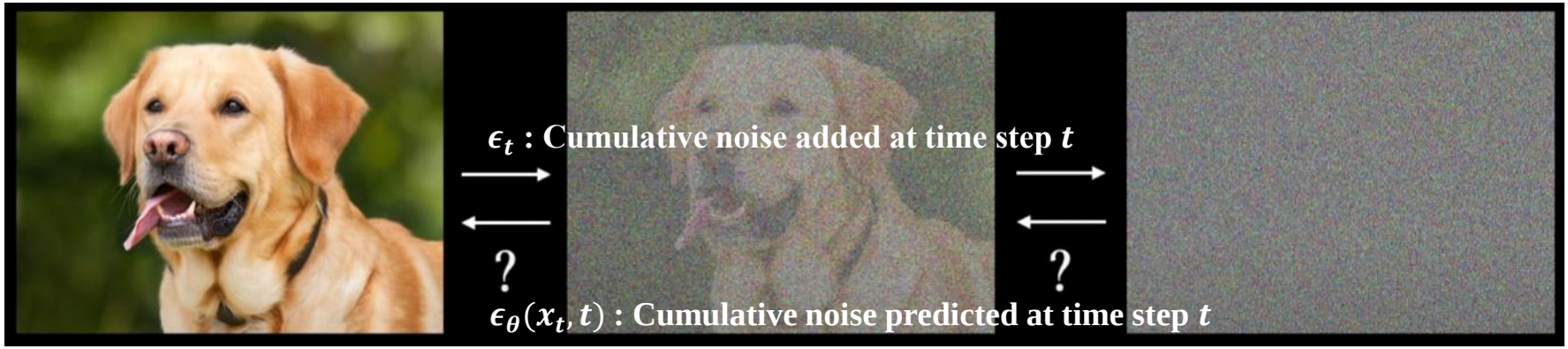
Markov property로 인해서 어떤 time step이던 x_0로부터 cumulative noise를 구할 수 있음

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t} * I, \qquad I \sim N(0,1)$$

  - Reverse process
    - x_t 에 더해진 cumulative noise ϵ_0를 $\epsilon_\theta(x_t, t)$ 으로 예측함
  - Loss
    - Time step t에 대한 실제 cumulative noise와 prediction에 L2 loss function 사용

$$L_t = \|\epsilon_t - \epsilon_\theta(x_t, t)\|^2$$

ϵ_t : 실제 cumulative noise
ϵ_θ: Predicted cumulative noise
x_t : noised 이미지
t : time step

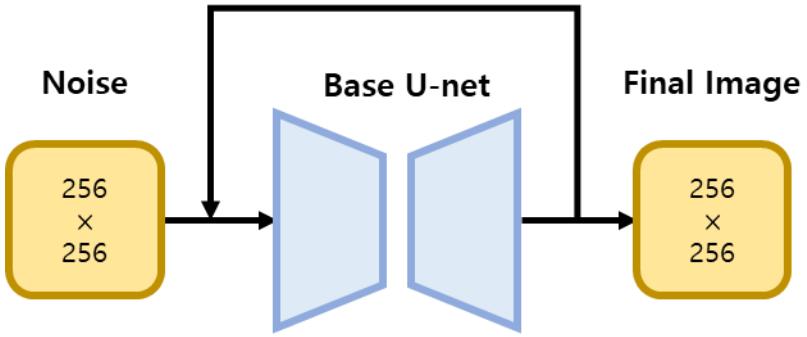


x_0 : Ground truth x_t : Noised image at time step t x_T : White Gaussian noise

1) Schuhmann, Christoph, et al. "Laion-5b: An open large-scale dataset for training next generation image-text models." Advances in Neural Information Processing Systems 35 (2022): 25278-25294.
2) 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.

Background

- Stable diffusion
  - LAION-5B[1]
    - Text-image pair dataset 중 최초의 open dataset
  - Latent diffusion models(LDM)[2]
    - 입출력 image을 작은 크기의 resolution을 가지는 latent image로 축소

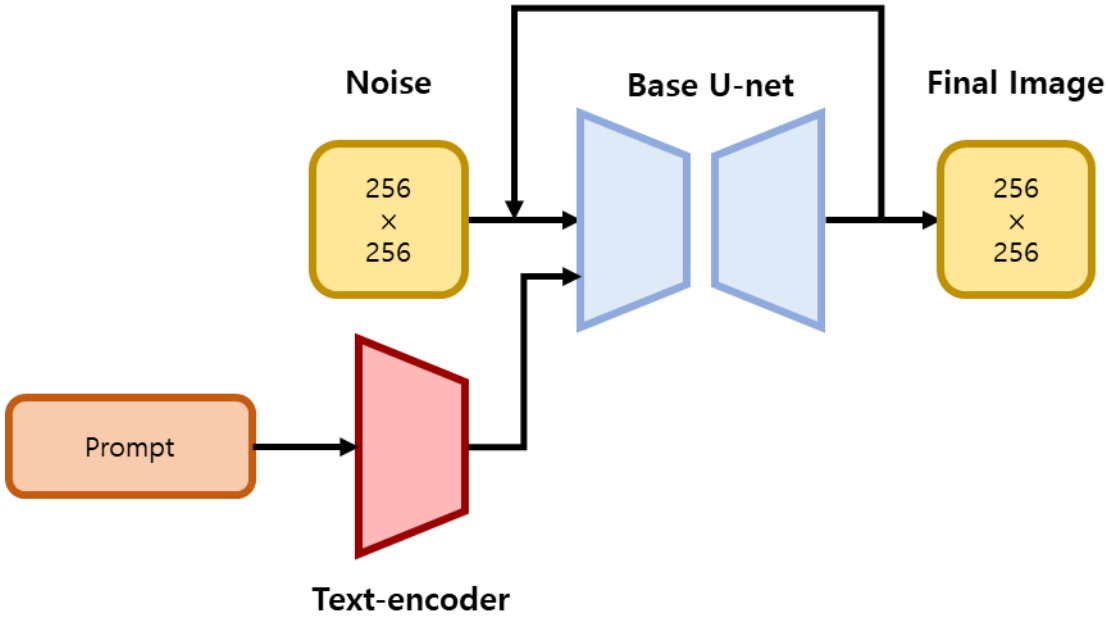


<Diffusion model 구조도>

1) Schuhmann, Christoph, et al. "Laion-5b: An open large-scale dataset for training next generation image-text models." Advances in Neural Information Processing Systems 35 (2022): 25278-25294.
2) 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.

Background

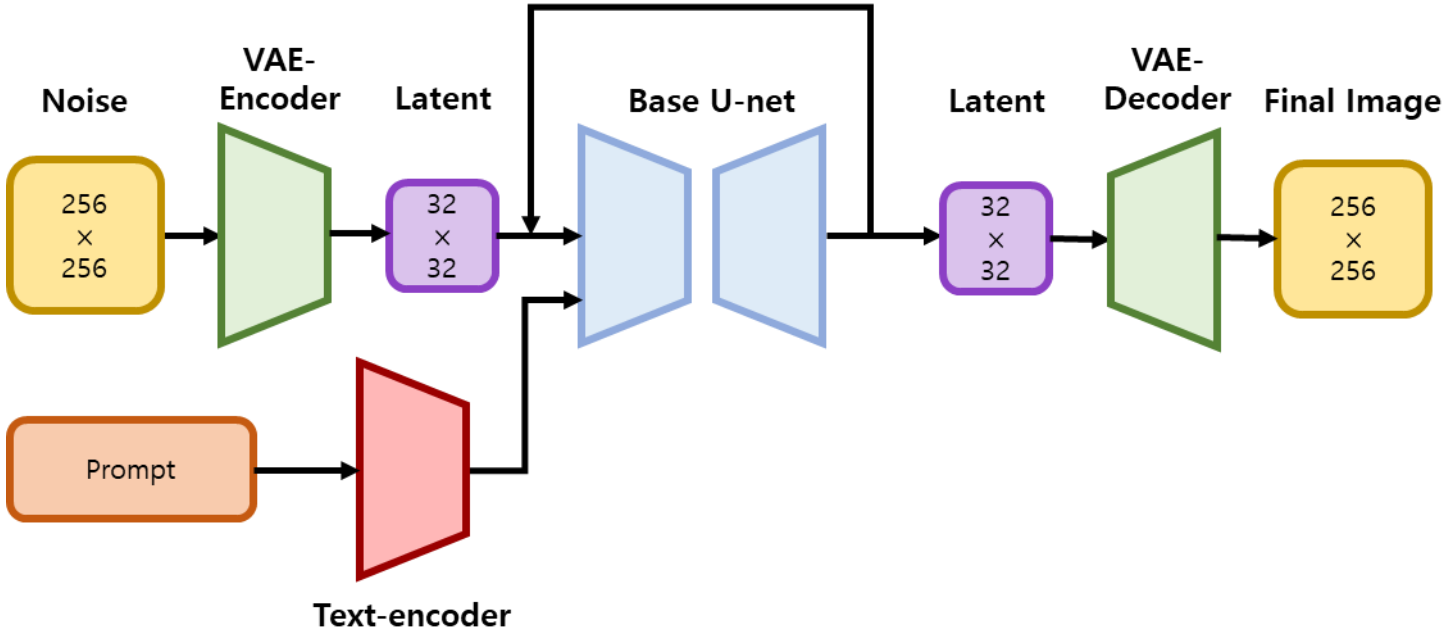
- Stable diffusion
  - LAION-5B[1]
    - Text-image pair dataset 중 최초의 open dataset
  - Latent diffusion models(LDM)[2]
    - 입출력 image을 작은 크기의 resolution을 가지는 latent image로 축소



<Text-conditional diffusion model 구조도>

Background

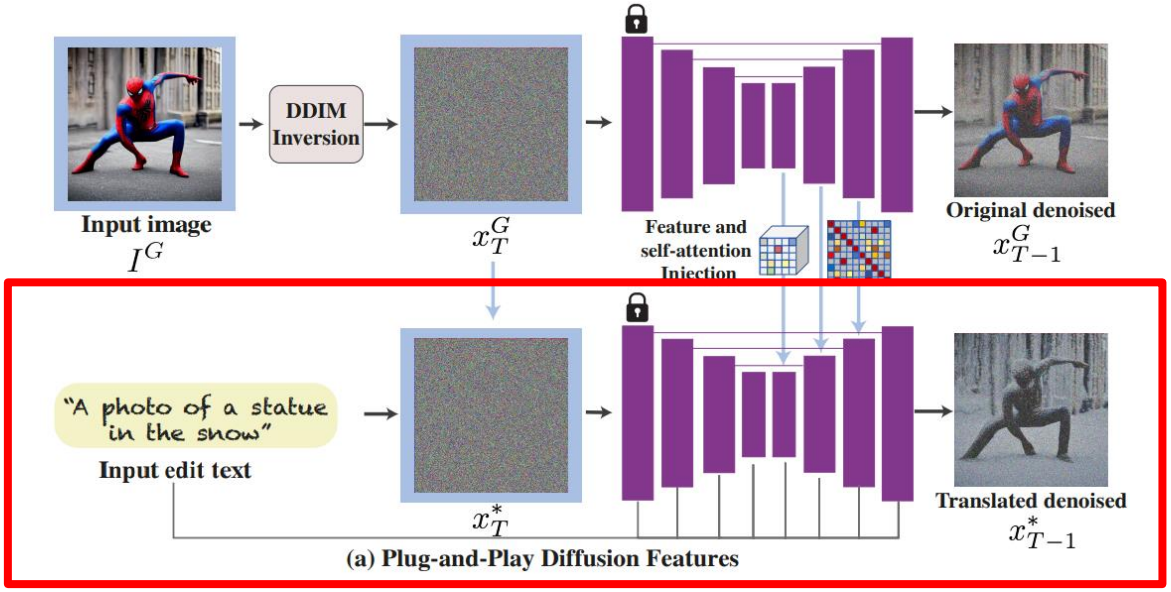
- Stable diffusion
  - LAION-5B[1]
    - Text-image pair dataset 중 최초의 open dataset
  - Latent diffusion models(LDM)[2]
    - 입출력 image을 작은 크기의 resolution을 가지는 latent image로 축소



<LDM의 text-conditional diffusion model 구조도>

Trends in 2023

- Plug-and-play[1]
  - 기존 prompt를 사용하는 main diffusion model에 추가로 image condition을 사용하는 diffusion model을 학습해서 활용함
    - Coarse한 부분은 Image condition I^G가 제공하는 structure 정보로부터 생성됨
      - ✓세부적인 detail은 prompt에 따라서 생성됨

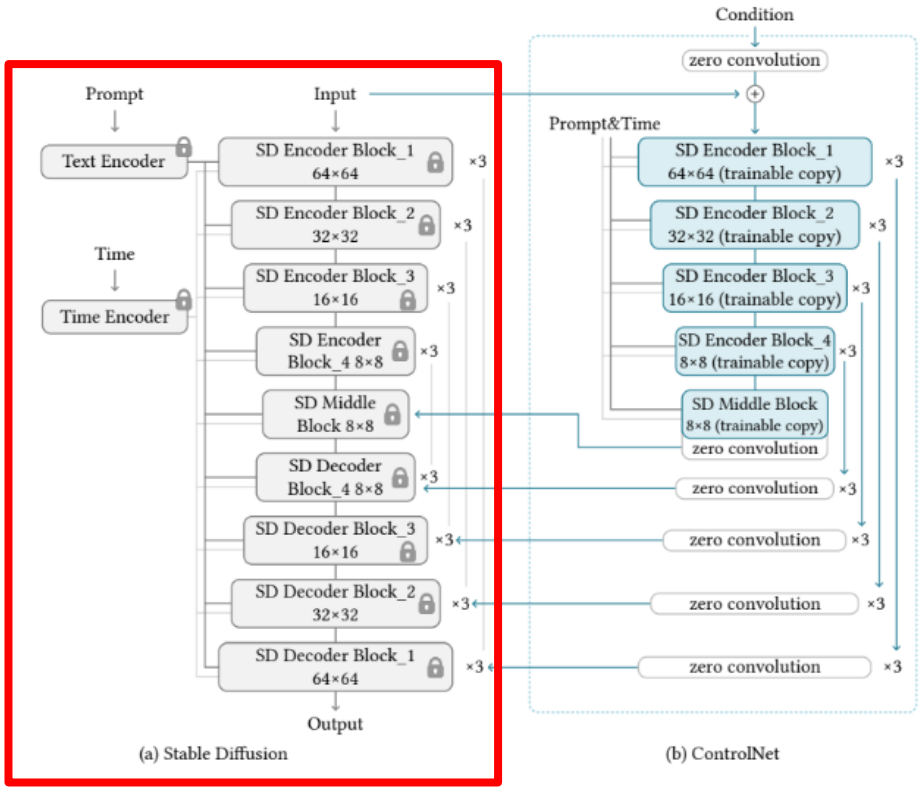


Main diffusion model (Stable diffusion)

Trends in 2023

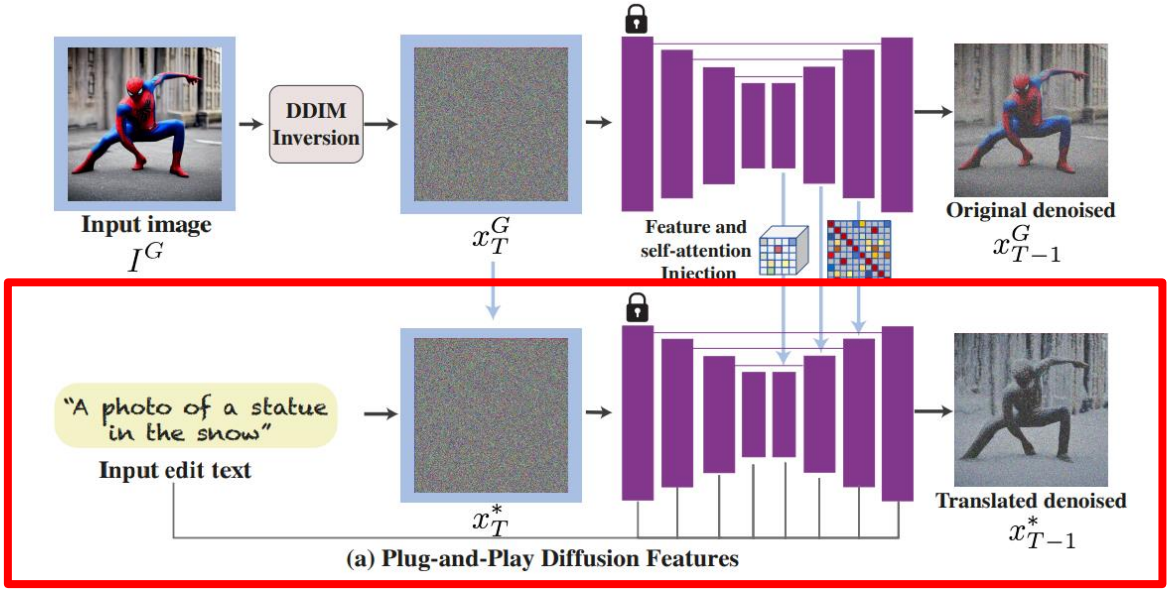
- ControlNet[1]
  - 기존 prompt를 사용하는 main diffusion model에 추가로 extra condition을 사용하는 diffusion model을 학습해서 활용함
    - Main diffusion model의 출력에 신규 conditioning 모델의 출력을 더하는 구조

Main diffusion model (Stable diffusion) →



Trends in 2023

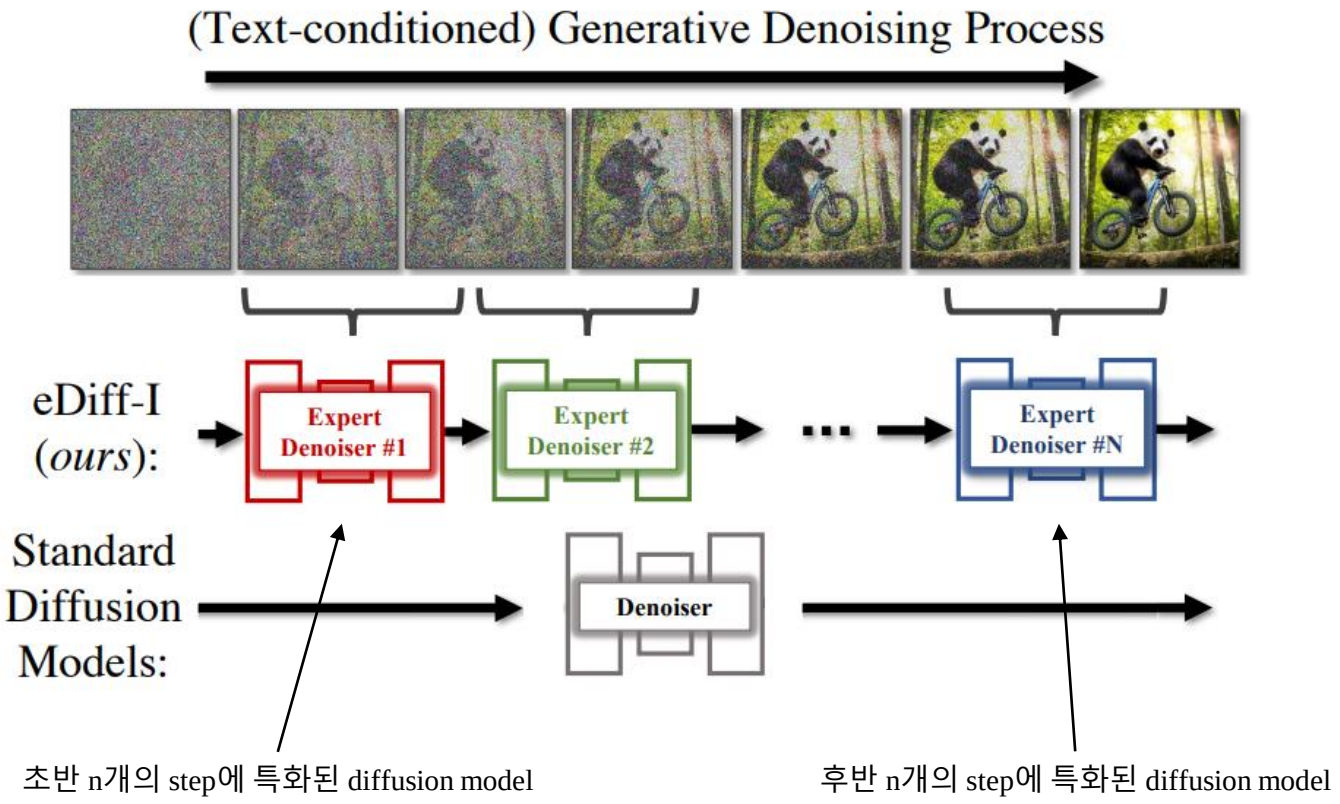
- Plug-and-play[1]
  - 기존 prompt를 사용하는 main diffusion model에 추가로 image condition을 사용하는 diffusion model을 학습해서 활용함
    - Coarse한 부분은 Image condition I^G가 제공하는 structure 정보로부터 생성됨
      - ✓세부적인 detail은 prompt에 따라서 생성됨



Improving main diffusion model?

Papers

- eDiff-I: Text-to-Image Diffusion Models with Ensemble of Expert Denoisers[1]
  - Ensemble of Diffusion models
    - 생성 step을 일정한 주기로 나눠서 특정 구간의 생성에 특화된 모델을 여러 개 학습함



Papers

- eDiff-I: Text-to-Image Diffusion Models with Ensemble of Expert Denoisers[1]
  - Ensemble of Diffusion models
    - 생성 step을 일정한 주기로 나눠서 특정 구간의 생성에 특화된 모델을 여러 개 학습함

두번 째 Prompt에 전혀 영향 받지 않음

후반 60%의 denoising에서 받은 text input의 영향이 초반 40%의 것을 뛰어넘음

Papers

- eDiff-I: Text-to-Image Diffusion Models with Ensemble of Expert Denoisers[1]
  - Ensemble of Text encoders
    - CLIP text encoder, T5 text encoder의 출력을 concatenate한 text embedding을 사용함

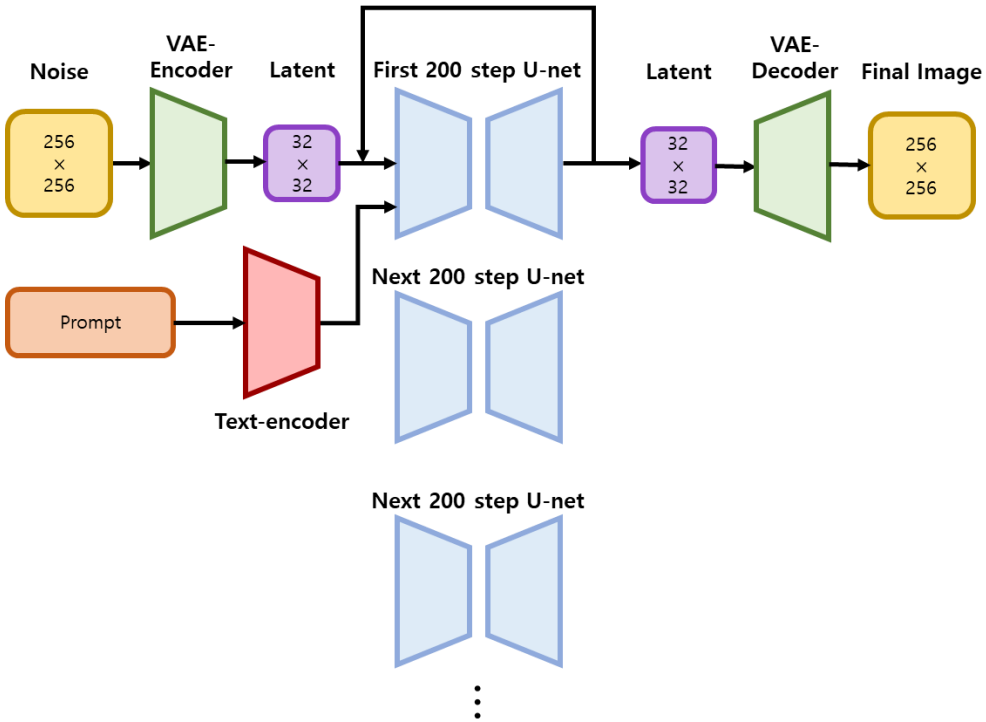
Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Ensemble of Text encoders
    - CLIP, OpenCLIP text encoder의 출력을 concatenate한 text embedding을 사용함
      - CLIP
        - ✓YFCC100M dataset의 subset인 1400만장의 text-image pair dataset으로 학습함
      - OpenCLIP
        - ✓LAION-5B text-image pair dataset으로 학습함

Model	*SDXL*	SD 1.4/1.5
# of UNet params	2.6B	860M
Text encoder	CLIP ViT-L & OpenCLIP ViT-bigG	CLIP ViT-L
Context dim.	2048	768

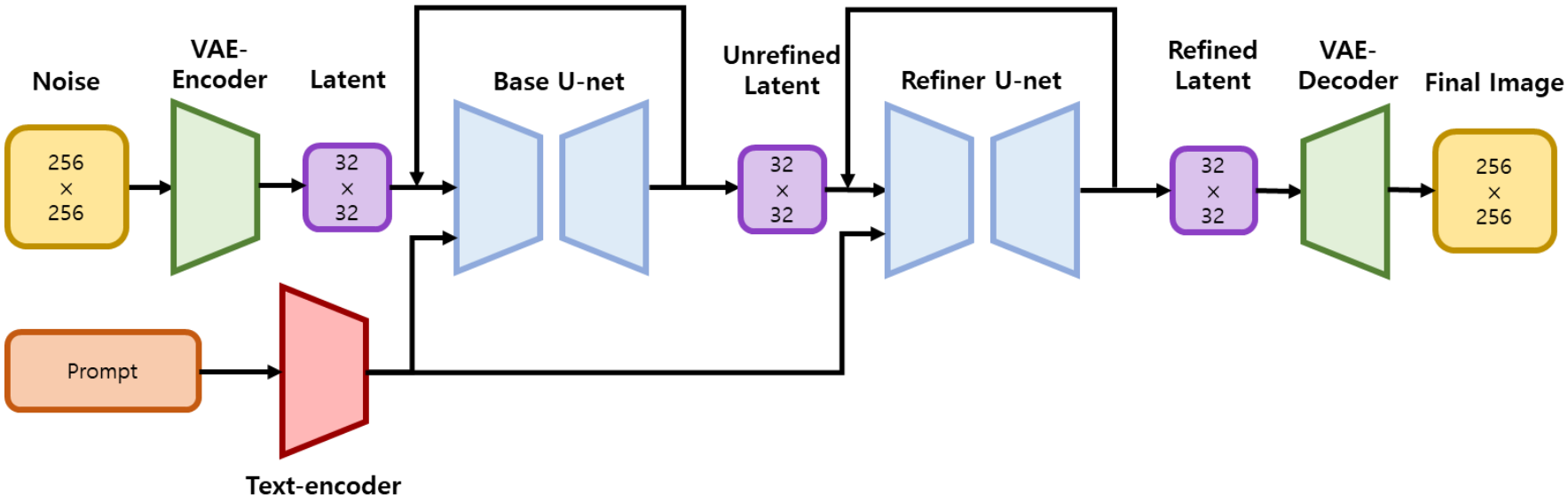
Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Ensemble of Diffusion models
    - Base diffusion model이 출력한 latent를 개선하는 refiner diffusion model을 추가함
      - Refiner diffusion model은 첫 200 step 구간의 생성에 특화된 모델임



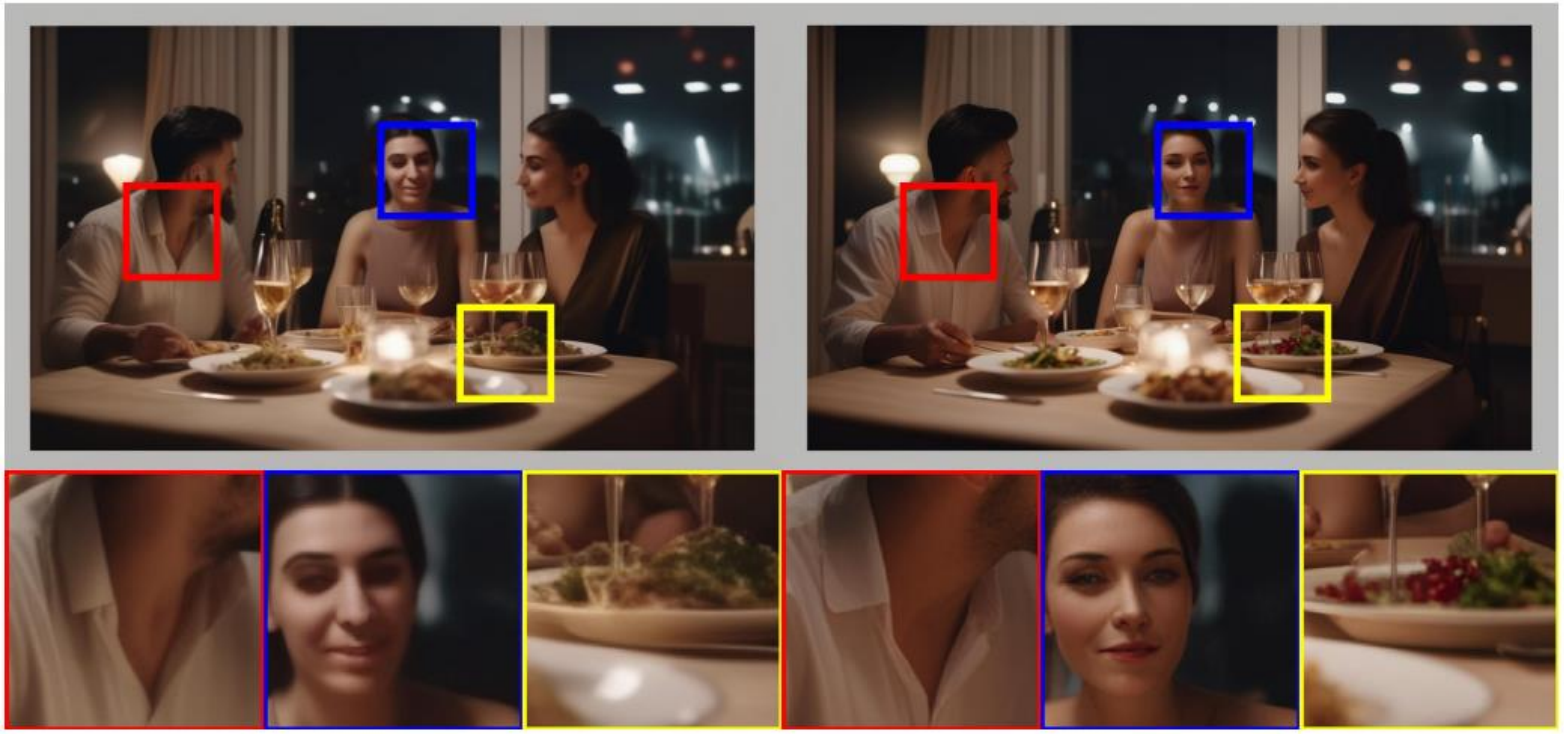
Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Ensemble of Diffusion models
    - Base diffusion model이 출력한 latent를 개선하는 refiner diffusion model을 추가함
      - Refiner diffusion model은 첫 200 step 구간의 생성에 특화된 모델임



Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Ensemble of Diffusion models
    - Base diffusion model이 출력한 latent를 개선하는 refiner diffusion model을 추가함
      - Refiner diffusion model은 첫 200 step 구간의 생성에 특화된 모델임

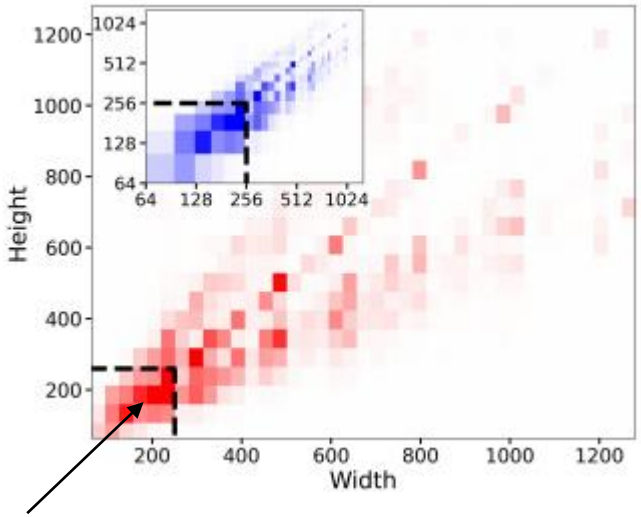


<Stable diffusion 생성 결과>

<SDXL 생성 결과>

Papers

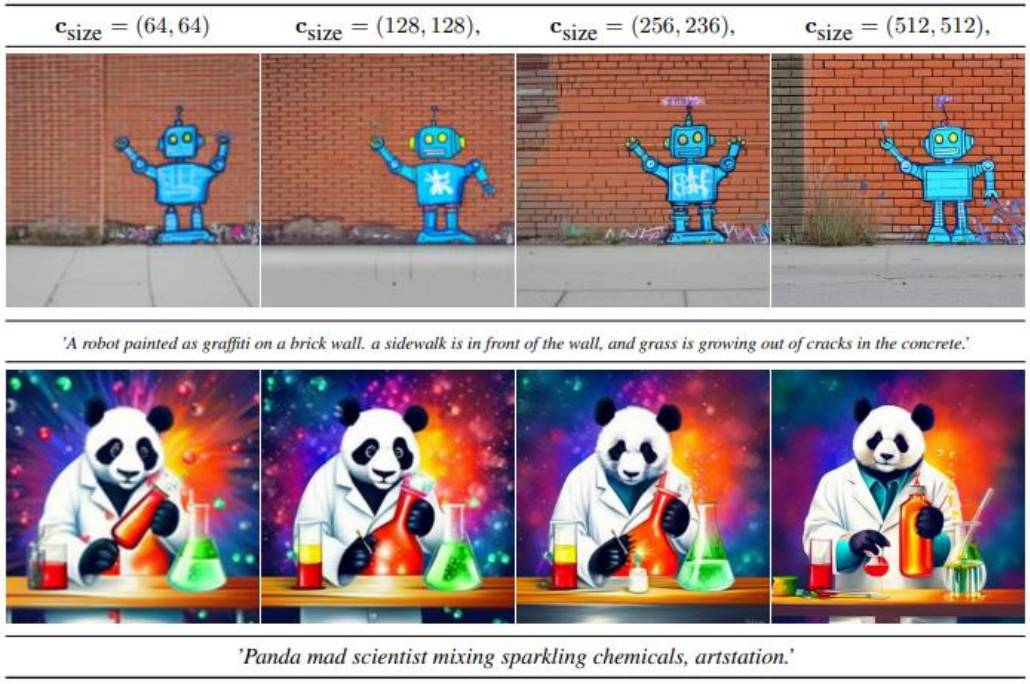
- SDXL – A drastically improved version of Stable diffusion[1]
  - Micro-conditioning
    - 추가 supervision 없이 2개의 condition을 제공할 수 있는 방법을 제안함
      - Conditioning the Model on Image Size
        - ✓원본 이미지의 해상도 정보를 diffusion model의 condition으로 제공함



생성 크기 256×256보다 작은 학습 데이터로 인해 모델 성능이 떨어짐

Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Micro-conditioning
    - 추가 supervision 없이 2개의 condition을 제공할 수 있는 방법을 제안함
      - Conditioning the Model on Image Size
        - ✓원본 이미지의 해상도 정보를 diffusion model의 condition으로 제공함



model	FID-5k ↓	IS-5k ↑
CIN-512-only	43.84	110.64
CIN-nocond	39.76	211.50
CIN-size-cond	**36.53**	**215.34**

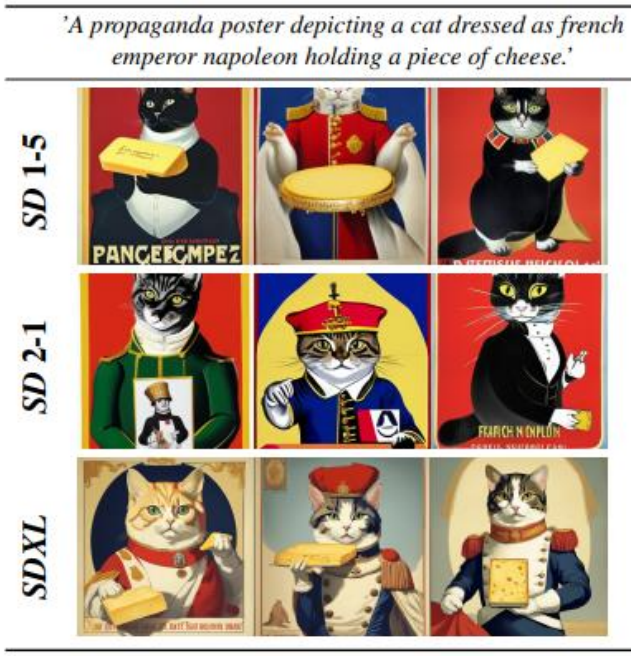
CIN-512-only : 512 크기 이하의 이미지는 학습에서 제외

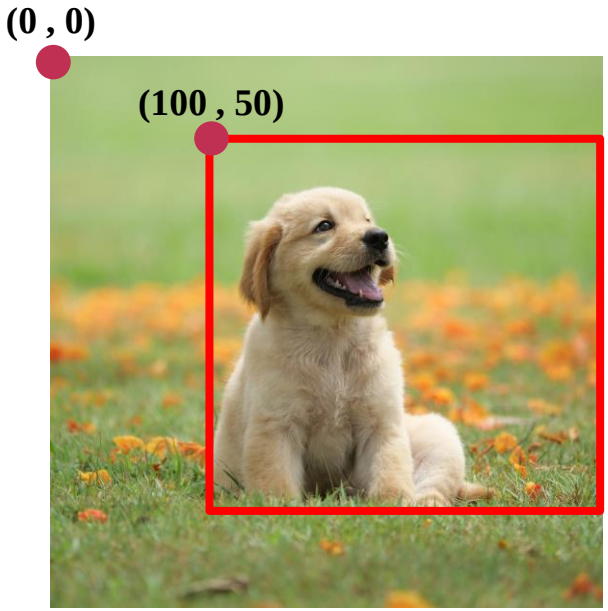
CIN-nocond : 512 크기 이하의 이미지는 upscale 후 학습

CIN-size-cond : 이미지의 원본 크기를 condition으로 제공

Papers

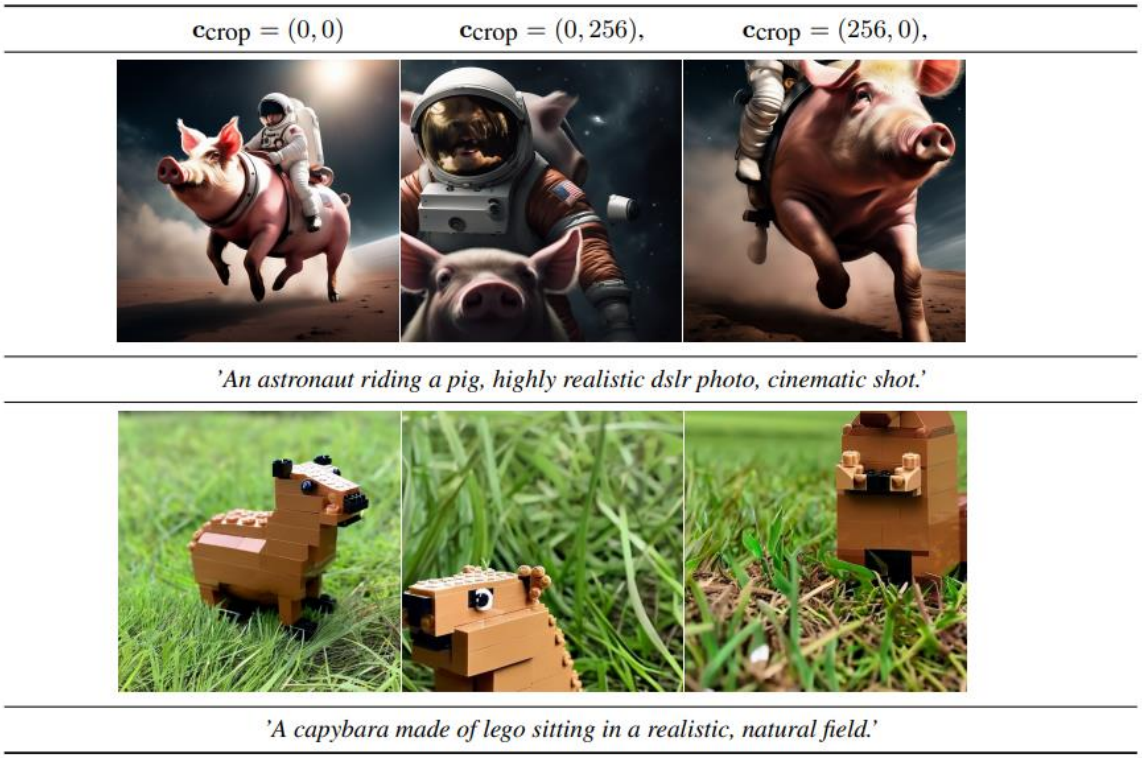
- SDXL – A drastically improved version of Stable diffusion[1]
  - Micro-conditioning
    - 추가 supervision 없이 2개의 condition을 제공할 수 있는 방법을 제안함
      - Conditioning the Model on Cropping parameters
        - ✓학습 시 적용된 random crop의 parameter를 condition으로 제공함





Papers

- SDXL – A drastically improved version of Stable diffusion[1]
  - Micro-conditioning
    - 추가 supervision 없이 2개의 condition을 제공할 수 있는 방법을 제안함
      - Conditioning the Model on Cropping parameters
        - ✓학습 시 적용된 random crop의 parameter를 condition으로 제공함



Thank you! ☺