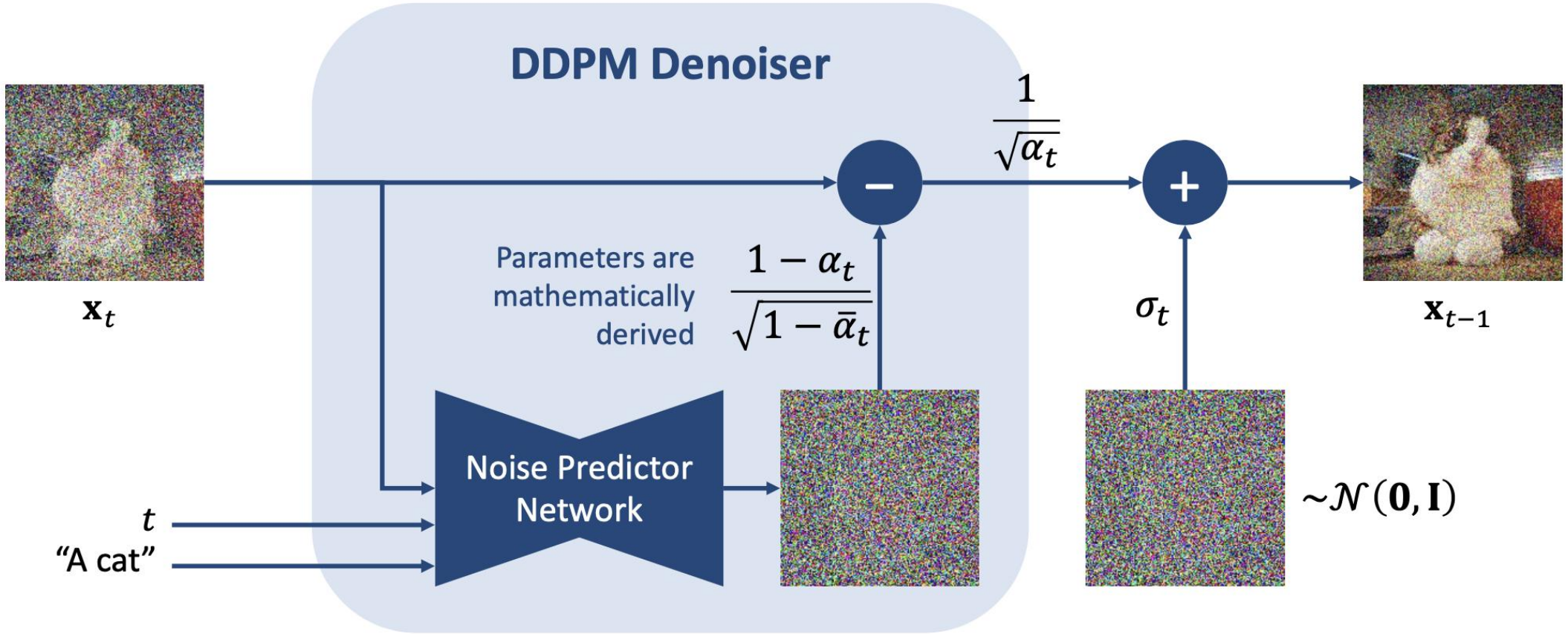
Diffusion Models and Scientific Generative AI

—— Lecture 7: Image Generation

Professor Siyu Zhu
AI3 institute, Fudan University

DDPM (Denoising Diffusion Probabilistic Models)

During generation: denoise step-by-step, in each step, add noise after noise removal

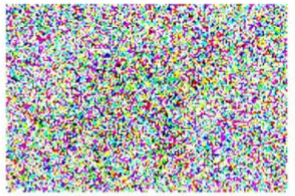


Training

Training

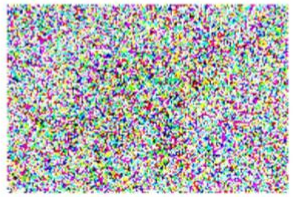
$\bar{\alpha}_1, \bar{\alpha}_2, \ldots \bar{\alpha}_T$

x_0



ε

Sample t

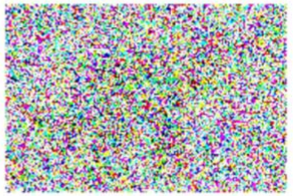
$\sqrt{\bar{\alpha}_t}$  $+ \sqrt{1-\bar{\alpha}_t}$  $=$ 

x_0 ε

t

Noise Predicter

?????

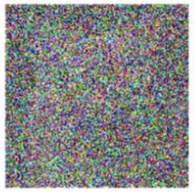


ε

slide from https://speech.ee.ntu.edu.tw/~hylee/ml/2023-spring.php

Inference

Inference



x_T

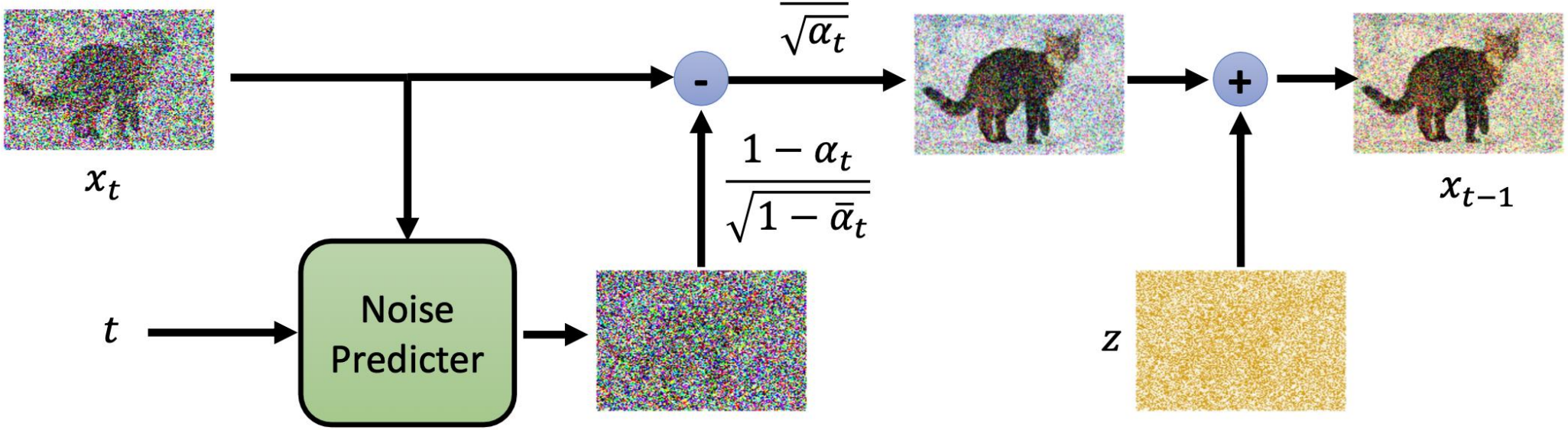
Algorithm 2 Sampling

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: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

sample a noise?!

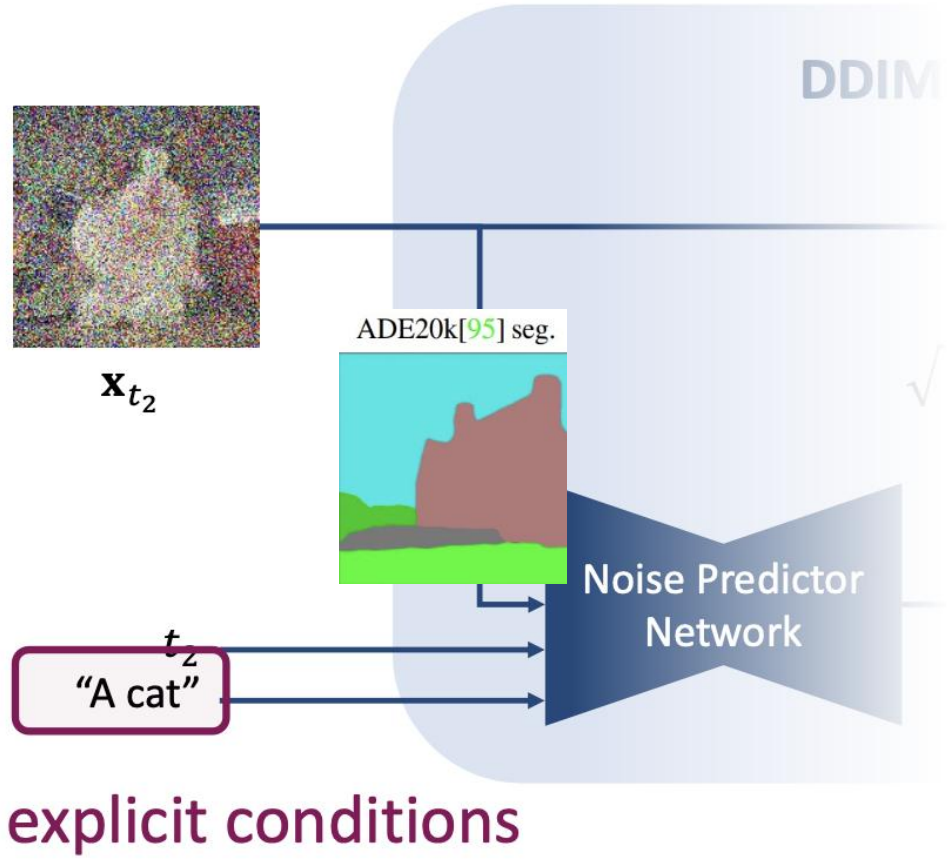
$\bar{\alpha}_1, \bar{\alpha}_2, \ldots \bar{\alpha}_T$

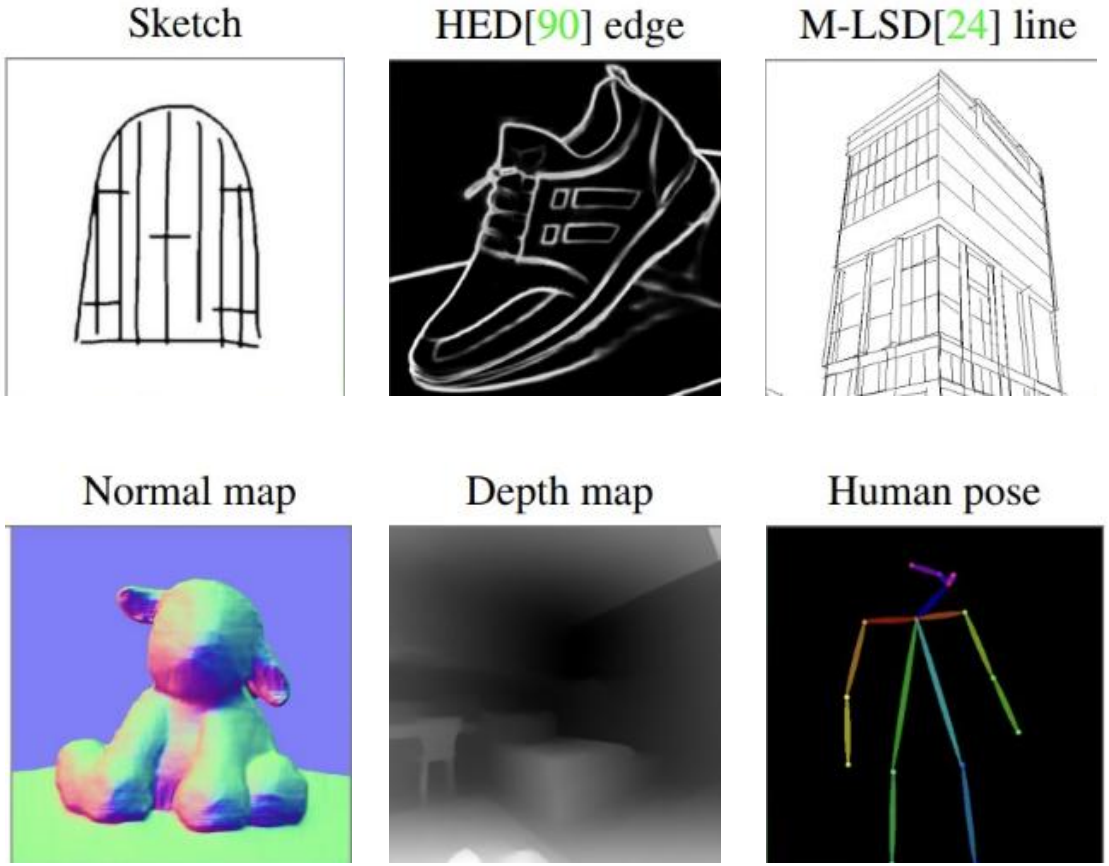
$\alpha_1, \alpha_2, \ldots \alpha_T$



slide from https://speech.ee.ntu.edu.tw/~hylee/ml/2023-spring.php

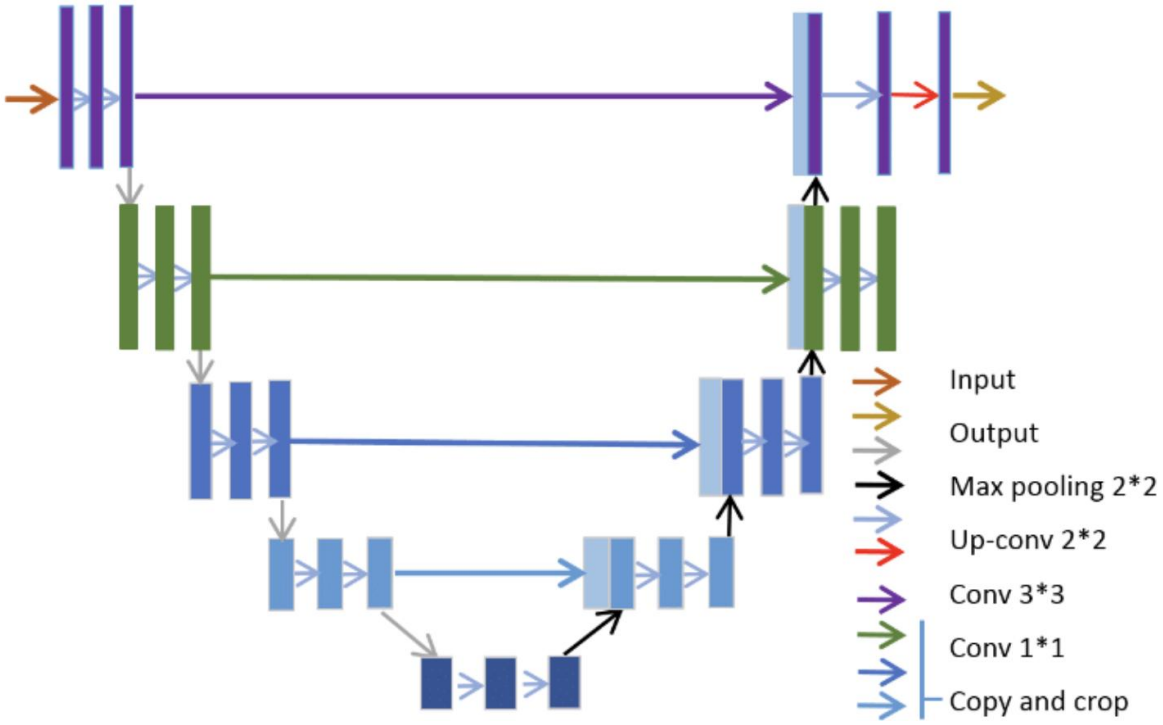
Conditional Generation: Explicit conditions



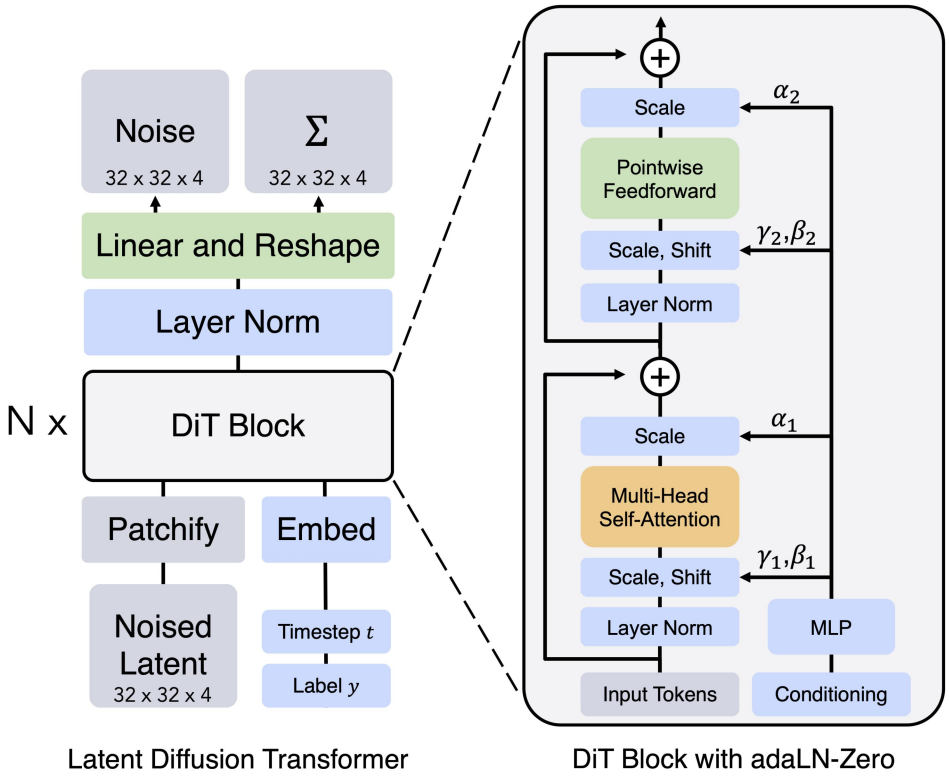


Denoiser

Unet



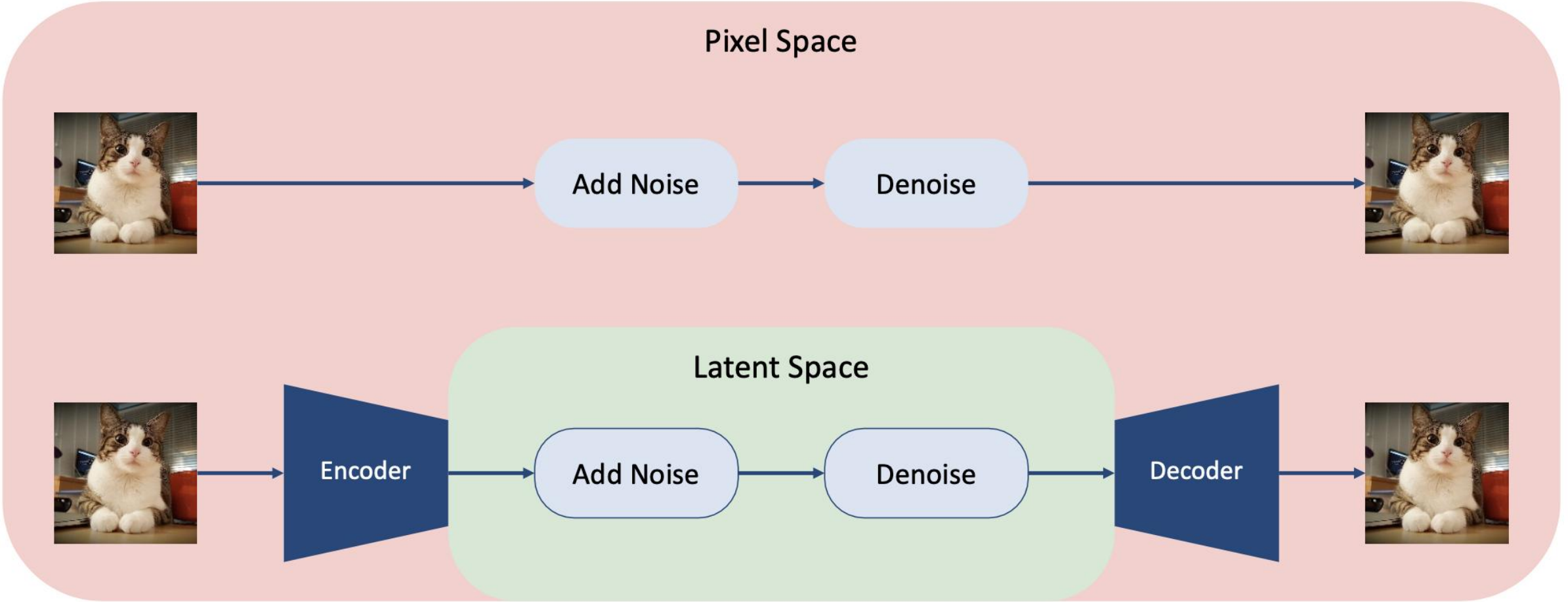
DiT



Latent Diffusion Transformer

DiT Block with adaLN-Zero

Latent Diffusion



Outline

- **Diffusion model architectures**
- **Editing and customization with diffusion models**
- **Other applications of diffusion models**

Problem Definition

Text-Guided Image Generation

Text prompt → image

“Toad practicing karate. ”

Text-Guided Video Generation

Text prompt → video

“Toad practicing karate. ”

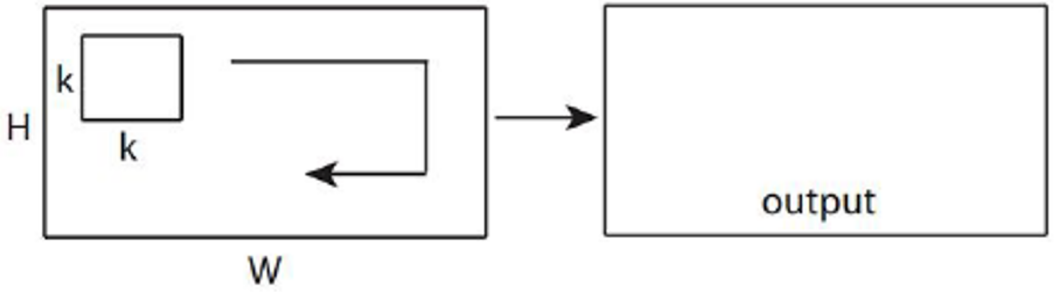
Video from Zhang et al., “Show-1: Marrying Pixel and Latent Diffusion Models for Text-to-Video Generation,” arXiv 2023.

Problem Definition

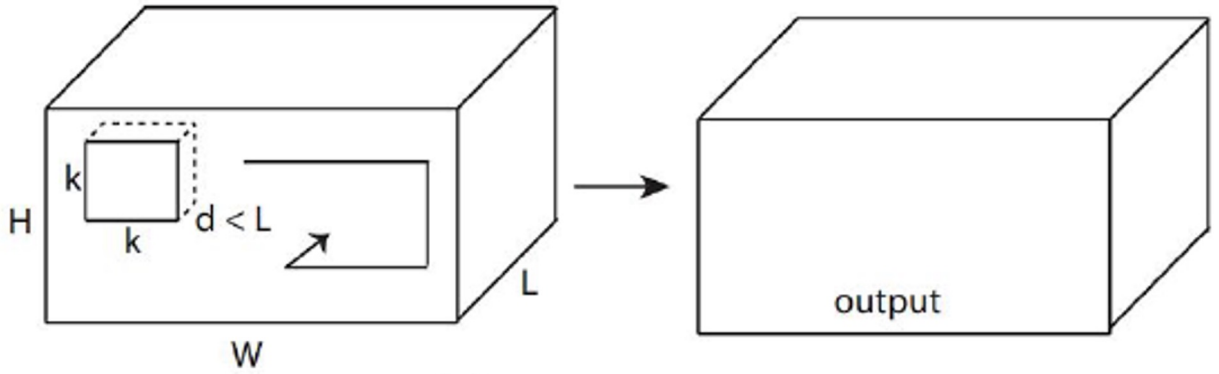
2D output → 3D output

Video Diffusion Models

Recap 3D Conv



(a) 2D convolution

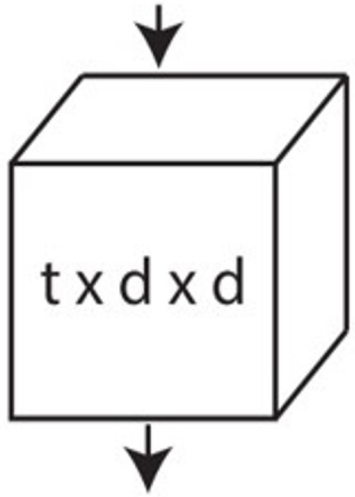


(c) 3D convolution

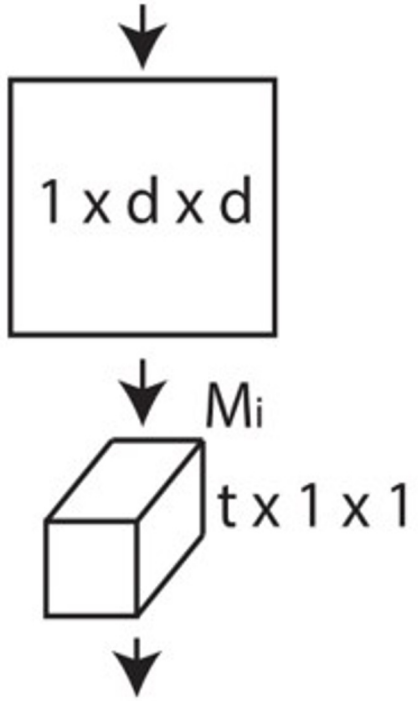
Du et al., "Learning Spatiotemporal Features with 3D Convolutional Networks," ICCV 2015.

Video Diffusion Models

Recap (2+1)D Conv



3D Conv

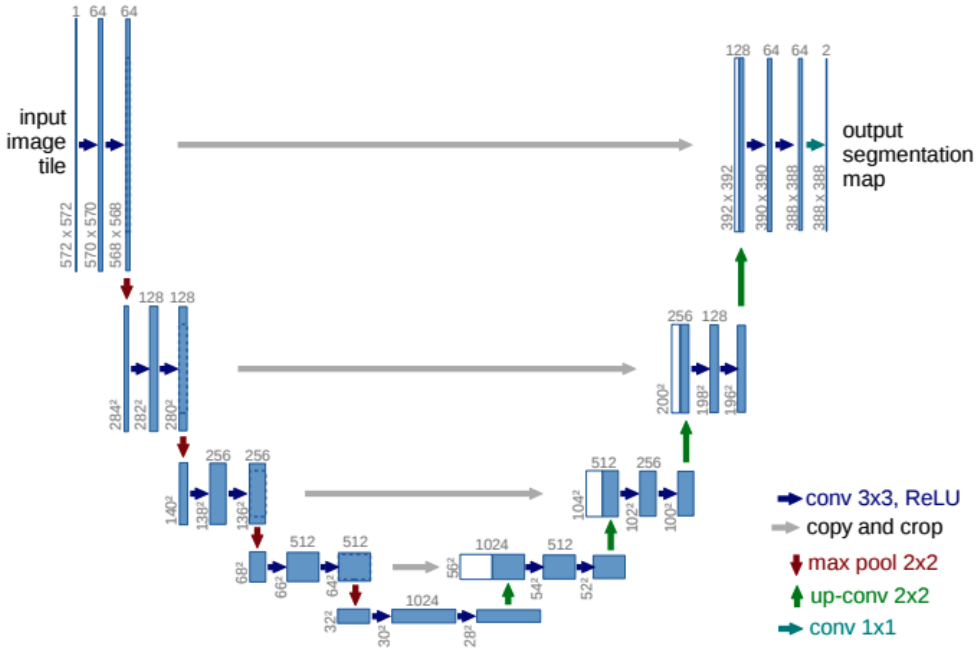


(2+1)D Conv

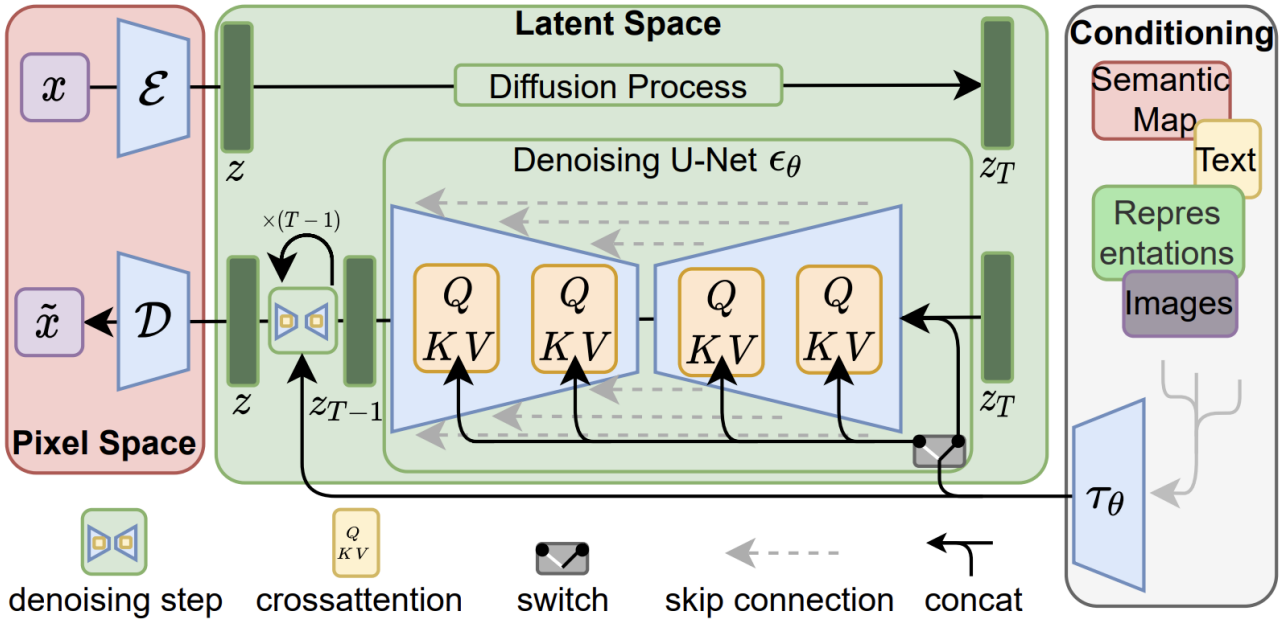
Du et al., "A Closer Look at Spatiotemporal Convolutions for Action Recognition," CVPR 2018.

Architecture

U-Net Architecture



U-Net architecture
Image source: Ronneberger et al.



U-Net based diffusion architecture
Image source: Rombach et al.

Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI 2015
Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models", CVPR 2022

U-Net Architecture

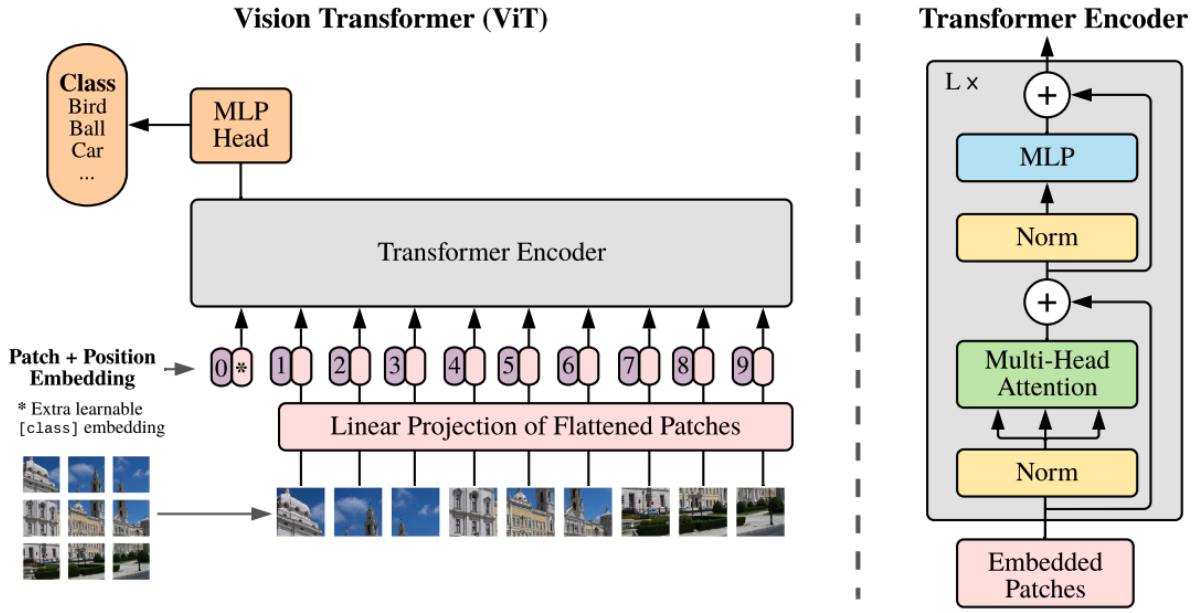
Imagen
Saharia et al.

Stable Diffusion
Rombach et al.

eDiff-I
Balaji et al.

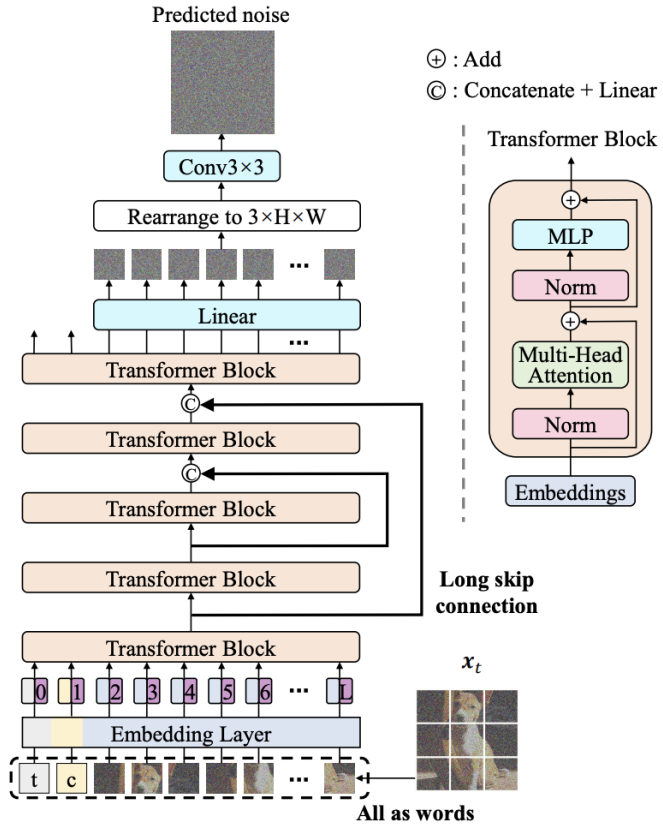
Saharia et al. "Photorealistic text-to-image diffusion models with deep language understanding", NeurIPS 2022
Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models", CVPR 2022
Balaji et al.," ediffi: Text-to-image diffusion models with an ensemble of expert denoisers", arXiv 2022

Transformer Architecture



Vision transformer

Image source: Dosovitskiy et al.



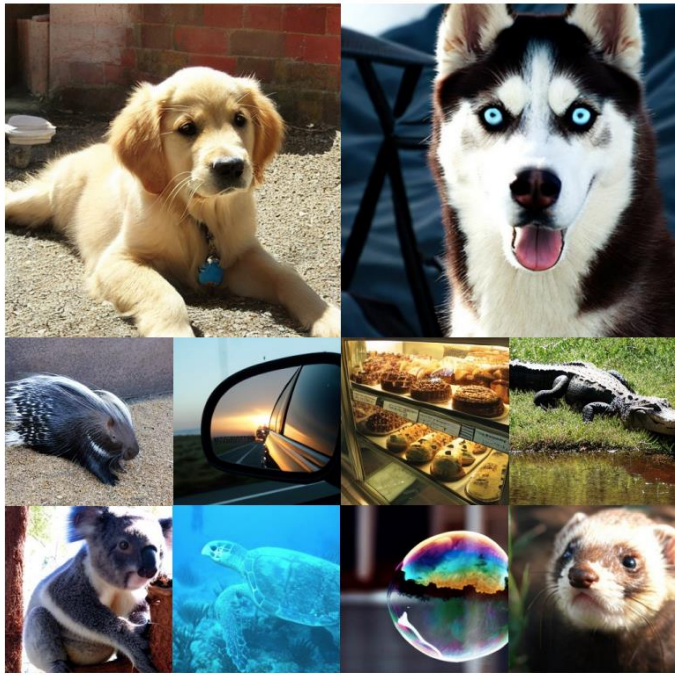
Transformer based diffusion model

Image source: Bao et al.

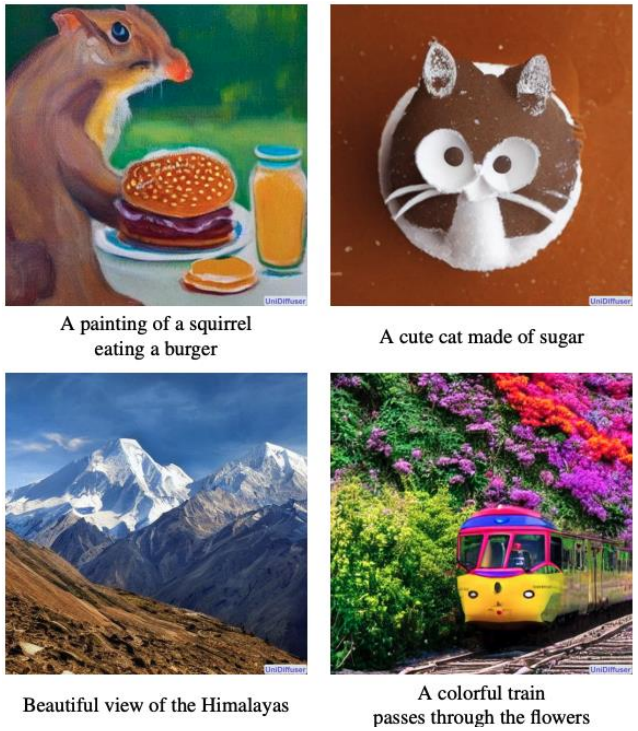
Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", ICLR 2021

Bao et al., "All are Worth Words: a ViT Backbone for Score-based Diffusion Models", arXiv 2022

Transformer Architecture



Scalable Diffusion Models with Transformers



One Transformer Fits All Distributions in Multi-Modal Diffusion at Scale

simple diffusion: End-to-end diffusion for high resolution images

Peebles and Xie, "Scalable Diffusion Models with Transformers", arXiv 2022
Bao et al., "One Transformer Fits All Distributions in Multi-Modal Diffusion at Scale", arXiv 2023
Hoogeboom et al., "simple diffusion: End-to-end diffusion for high resolution images", arXiv 2023

Autoregression Architecture

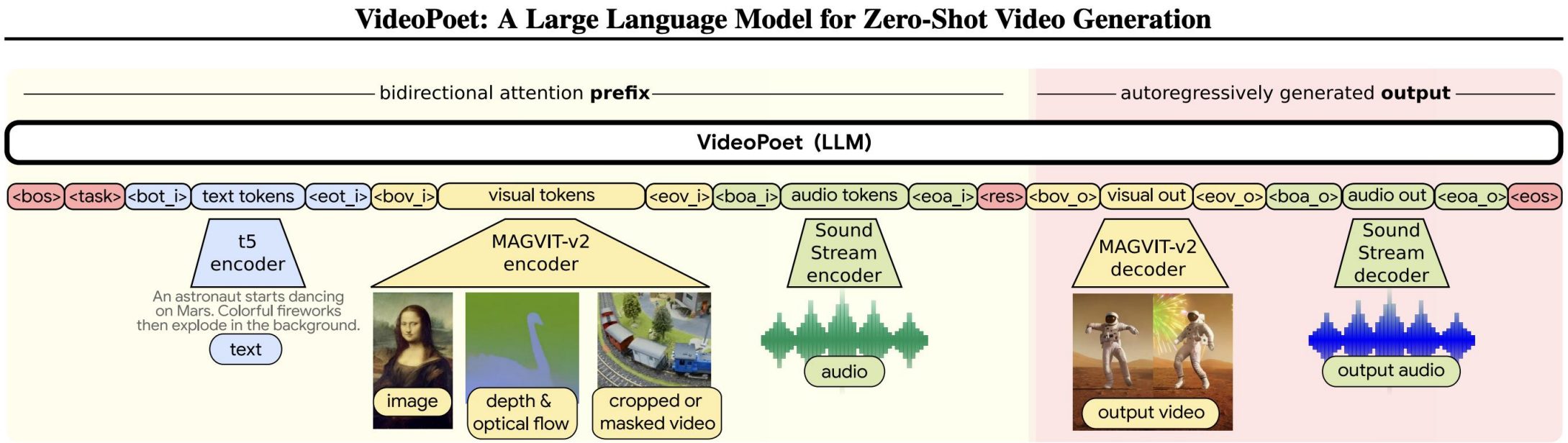


Figure 2: **Sequence layout for VideoPoet**. We encode all modalities into the discrete token space, so that we can directly use large language model architectures for video generation. We denote special tokens in <> (see Table 4 for definitions). The modality agnostic tokens are in darker red; the text related components are in blue; the vision related components are in yellow; the audio related components are in green. The left portion of the layout on light yellow represents the bidirectional prefix inputs. The right portion on darker red represents the autoregressively generated outputs with causal attention.

Autoregression Architecture

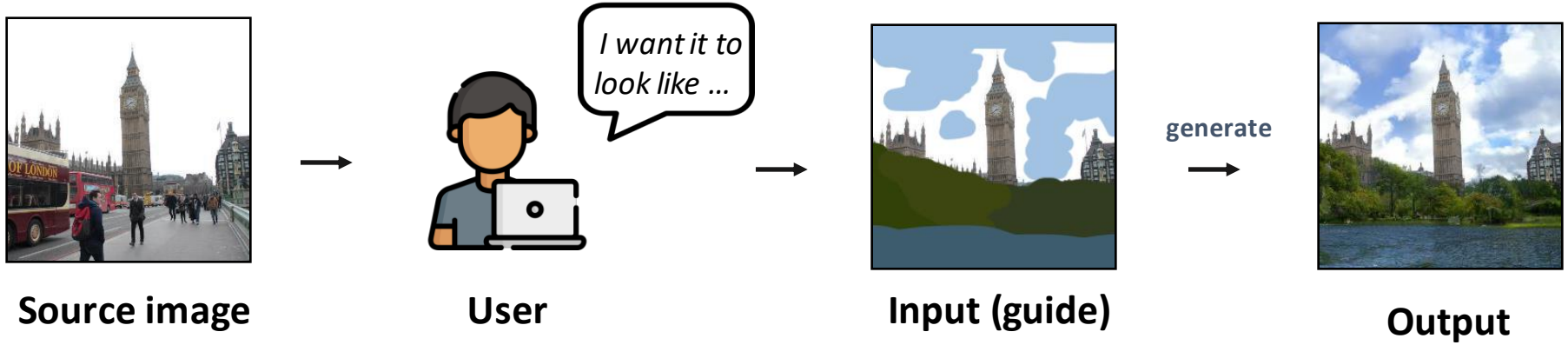
Figure 1: **Emu3** is trained to predict the next token with a single Transformer on a mix of video, image, and text tokens. **Emu3** achieves state-of-the-art performance compared to well-established task-specific models in generation and perception tasks.

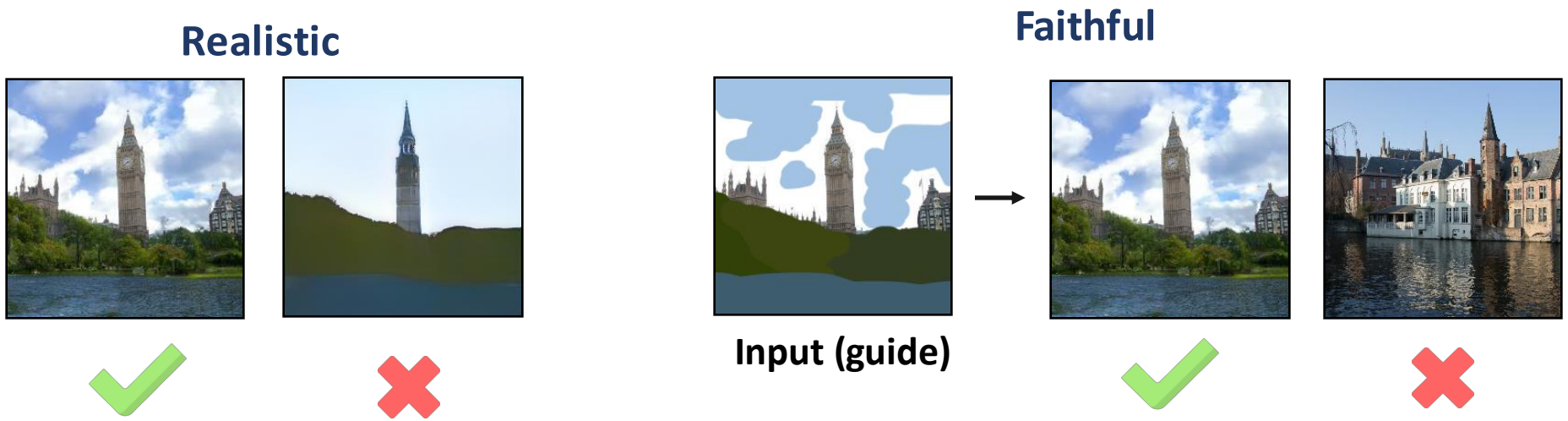
Image editing and customization with diffusion models

Image editing and customization with diffusion models

- RGB pixel guidance
- Text guidance
- Reference image guidance

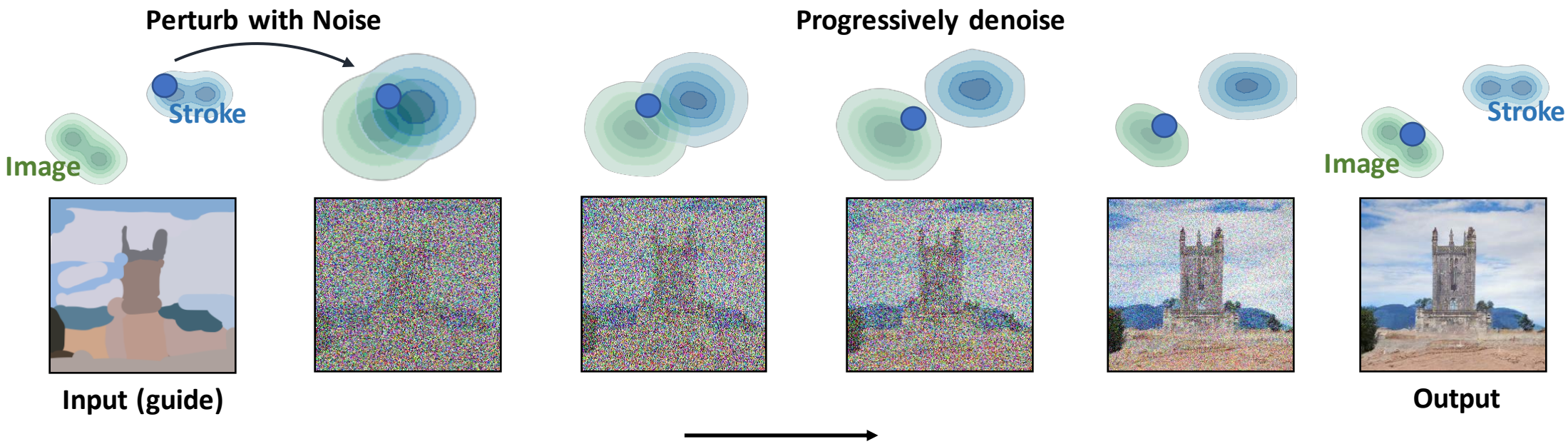
How to perform guided synthesis/editing?





SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations

First perturb the input with **Gaussian noise** and then progressively remove the noise using a pretrained diffusion model.

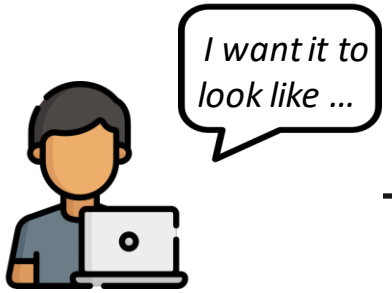


Gradually projects the input to the manifold of natural images.

Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", ICLR 2022

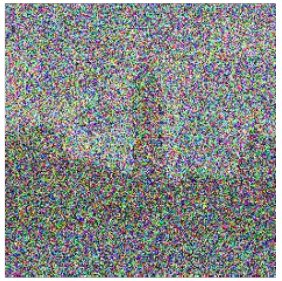
Fine-grained control using strokes

Original



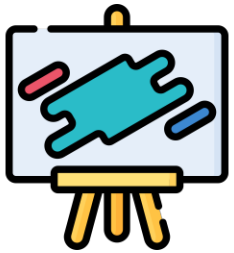
User

Input (guide)

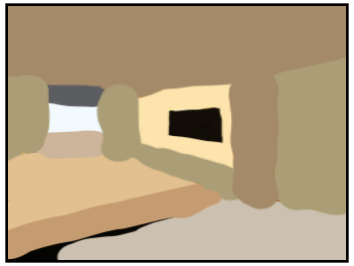


SDEdit

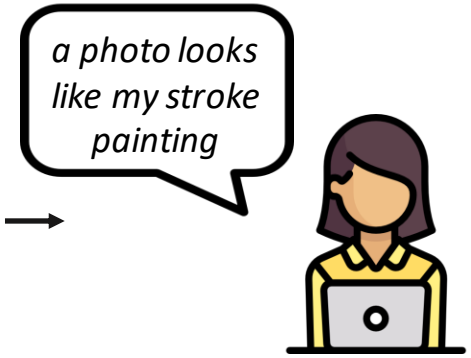
Output



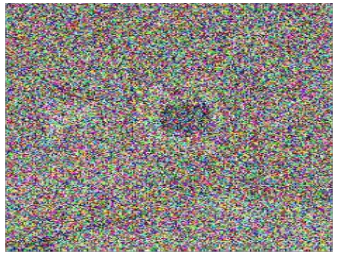
Empty Canvas



Painting



User

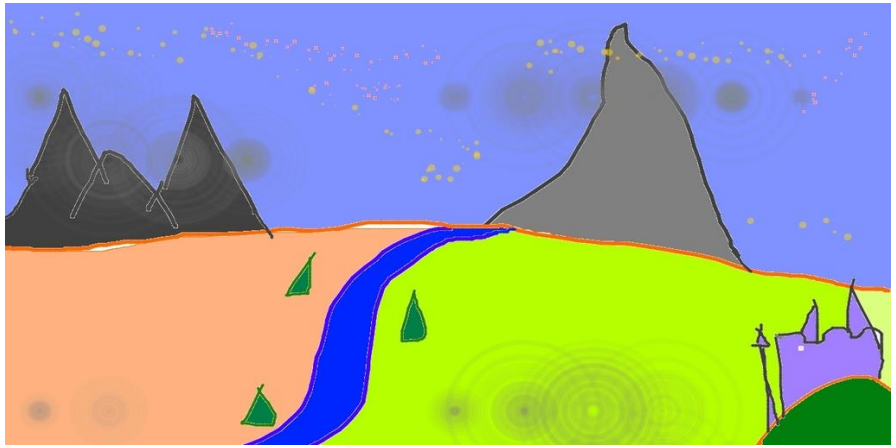


SDEdit

Output

Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", ICLR 2022

Application to Stable Diffusion (img2img)



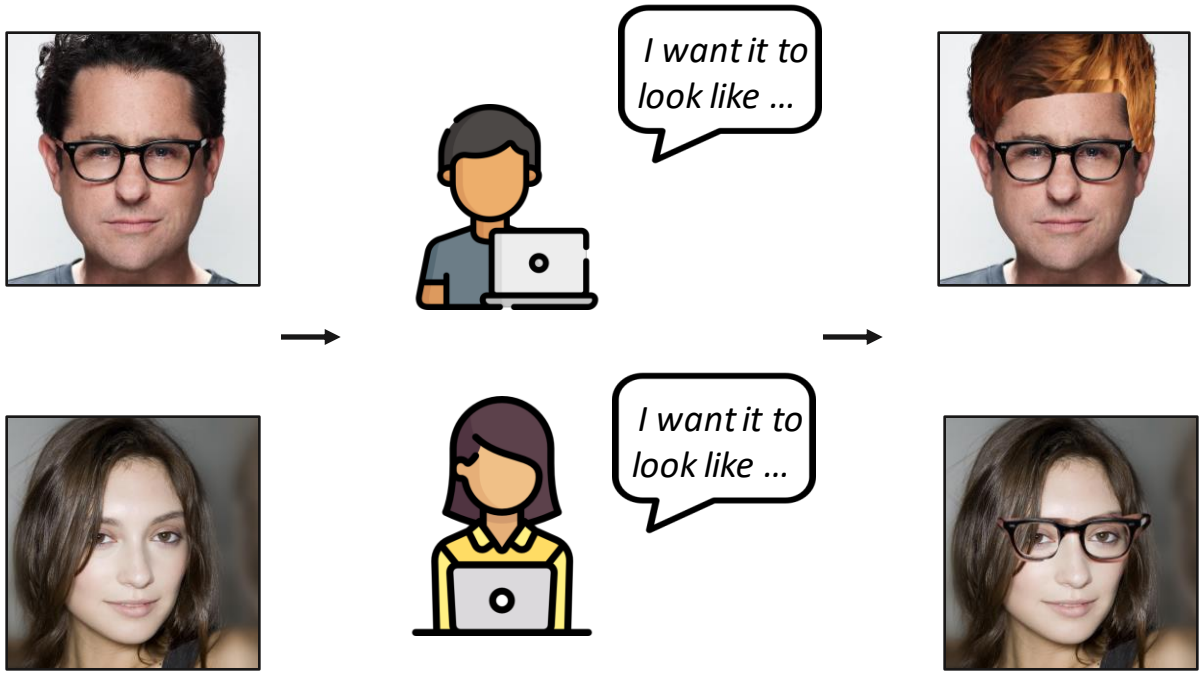
Input

SDEdit/img2img →

Outputs

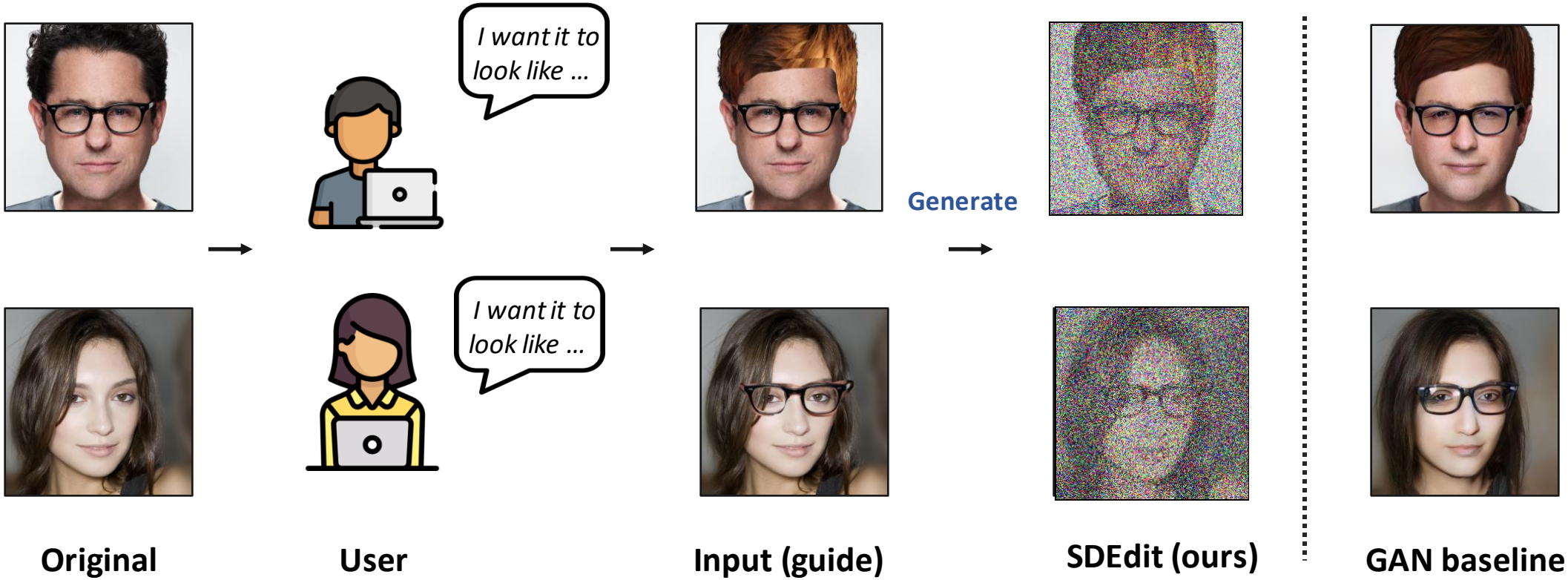
Source: https://github.com/CompVis/stable-diffusion

Image compositing



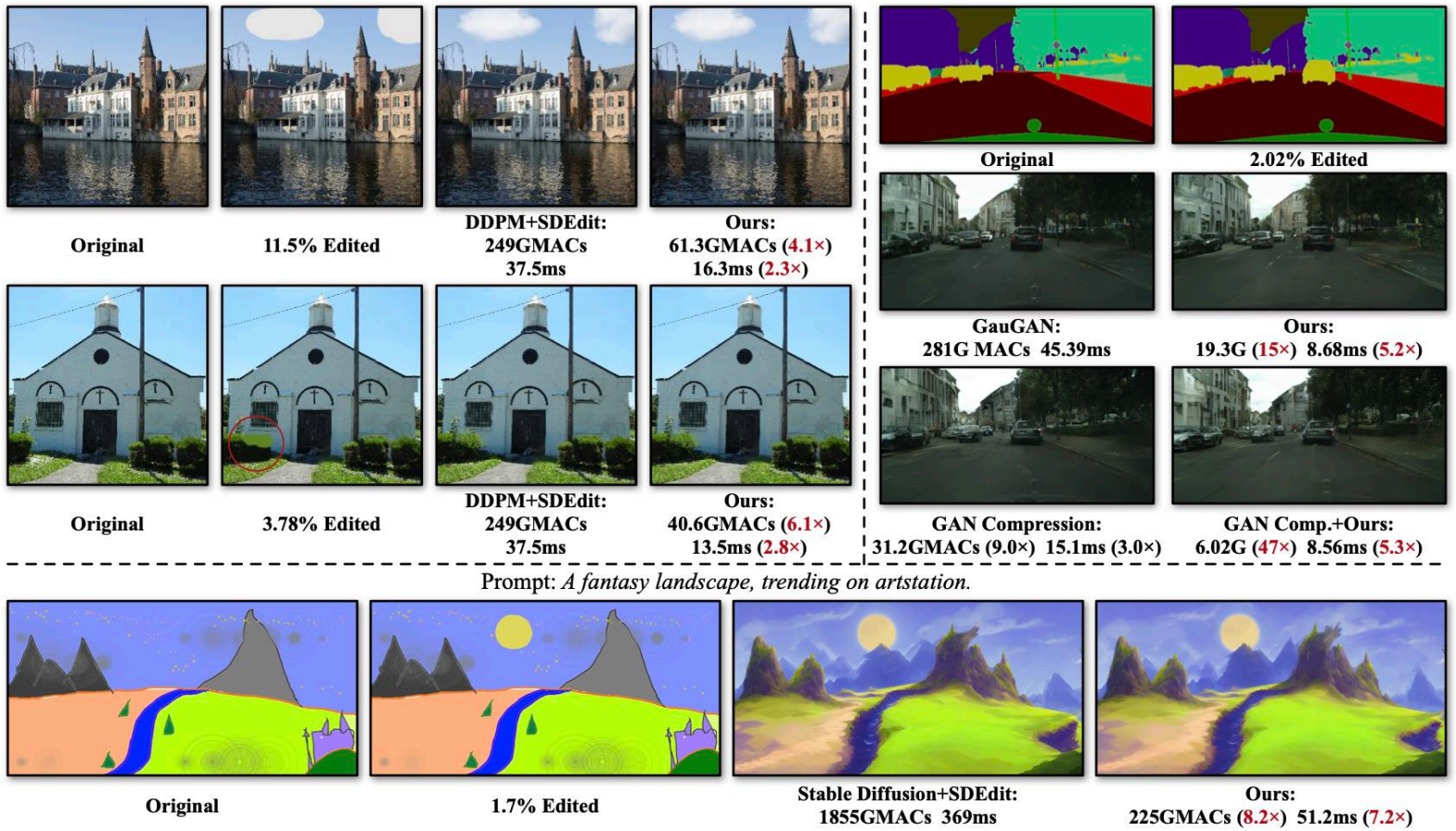
Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", ICLR 2022

Image compositing



Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", ICLR 2022

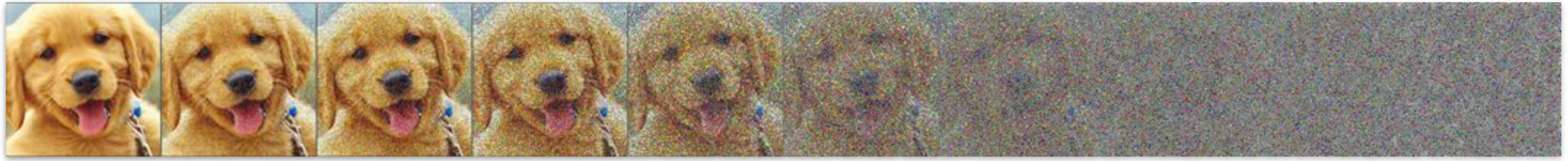
Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models



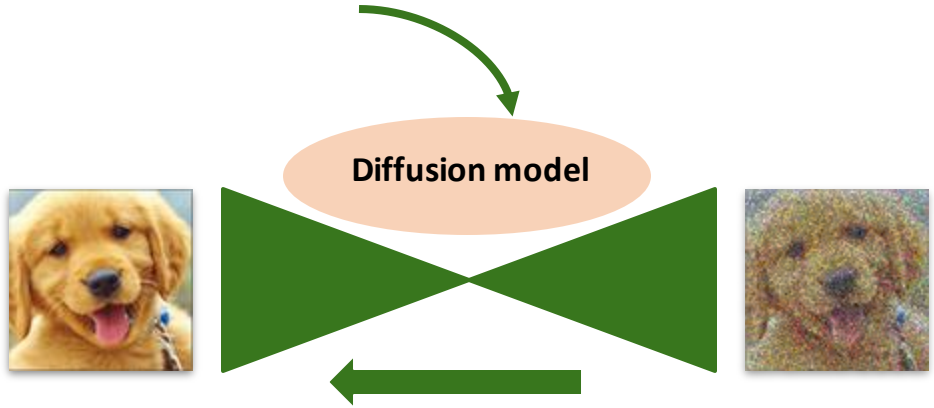
Li et al., "Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models", NeurIPS 2022

DDIM Inversion

Simple **forward process** slowly maps data to noise

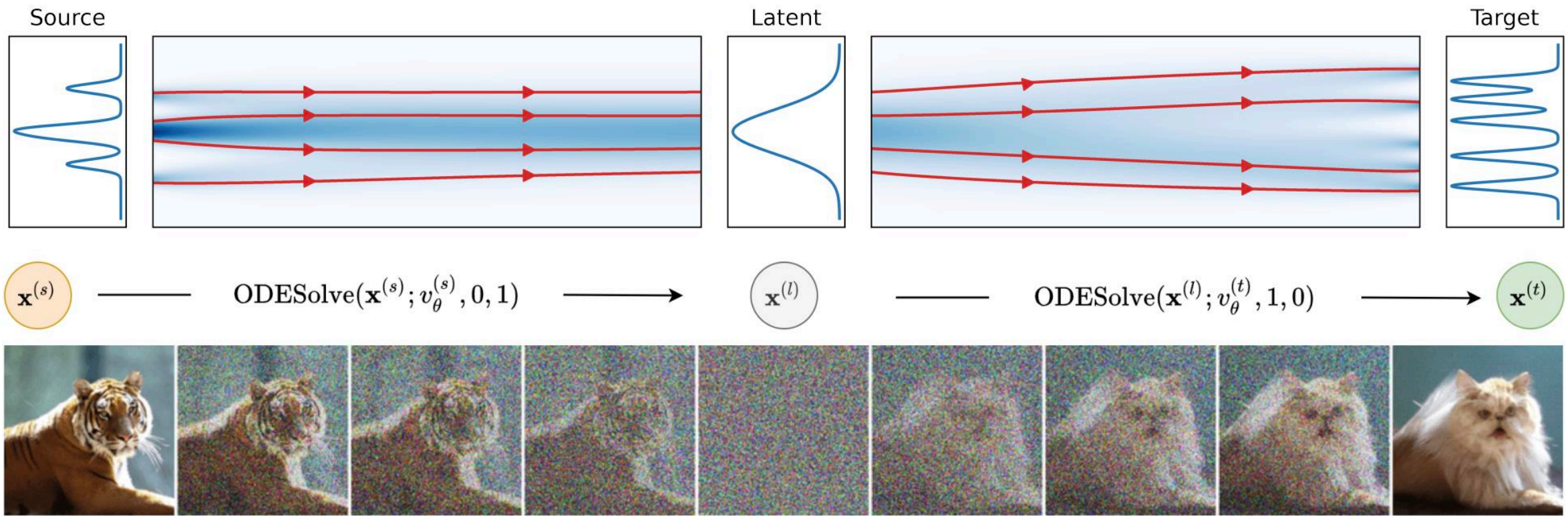


Reverse process maps noise back to data where **diffusion model** is trained



Song et al., "Denoising Diffusion Implicit Models", ICLR 2021

Style transfer with DDIM inversion



Su et al., "Dual diffusion implicit bridges for image-to-image translation", ICLR 2023

Style transfer with DDIM inversion

Multi-domain translation

Reference Image

Source Image 1 Target Image 1

Example-Guided Color Transfer

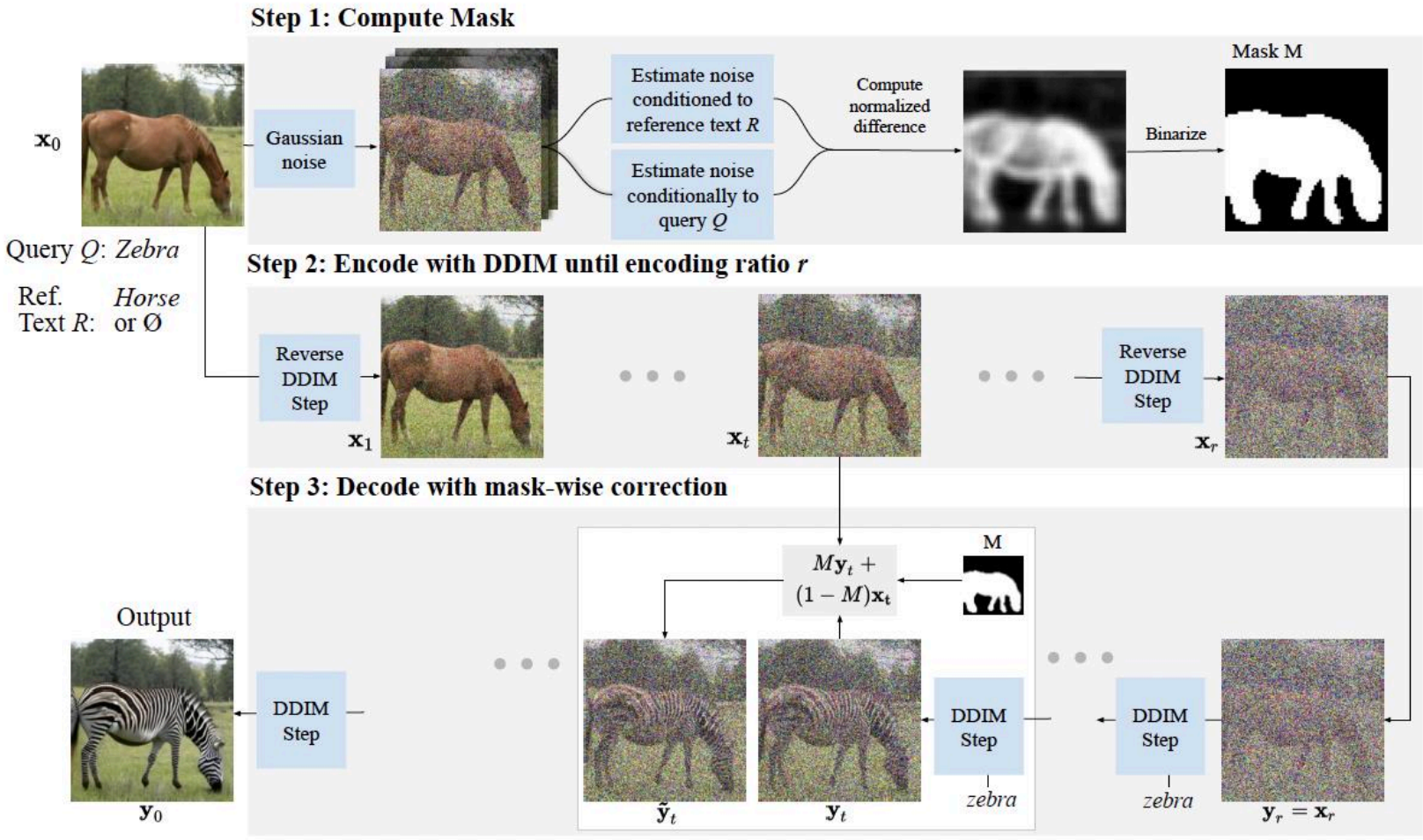
Su et al., "Dual diffusion implicit bridges for image-to-image translation", ICLR 2023

DiffEdit: Diffusion-based semantic image editing with mask guidance

Instead of asking users to provide the mask, the model will generate the mask itself based on the caption and query.

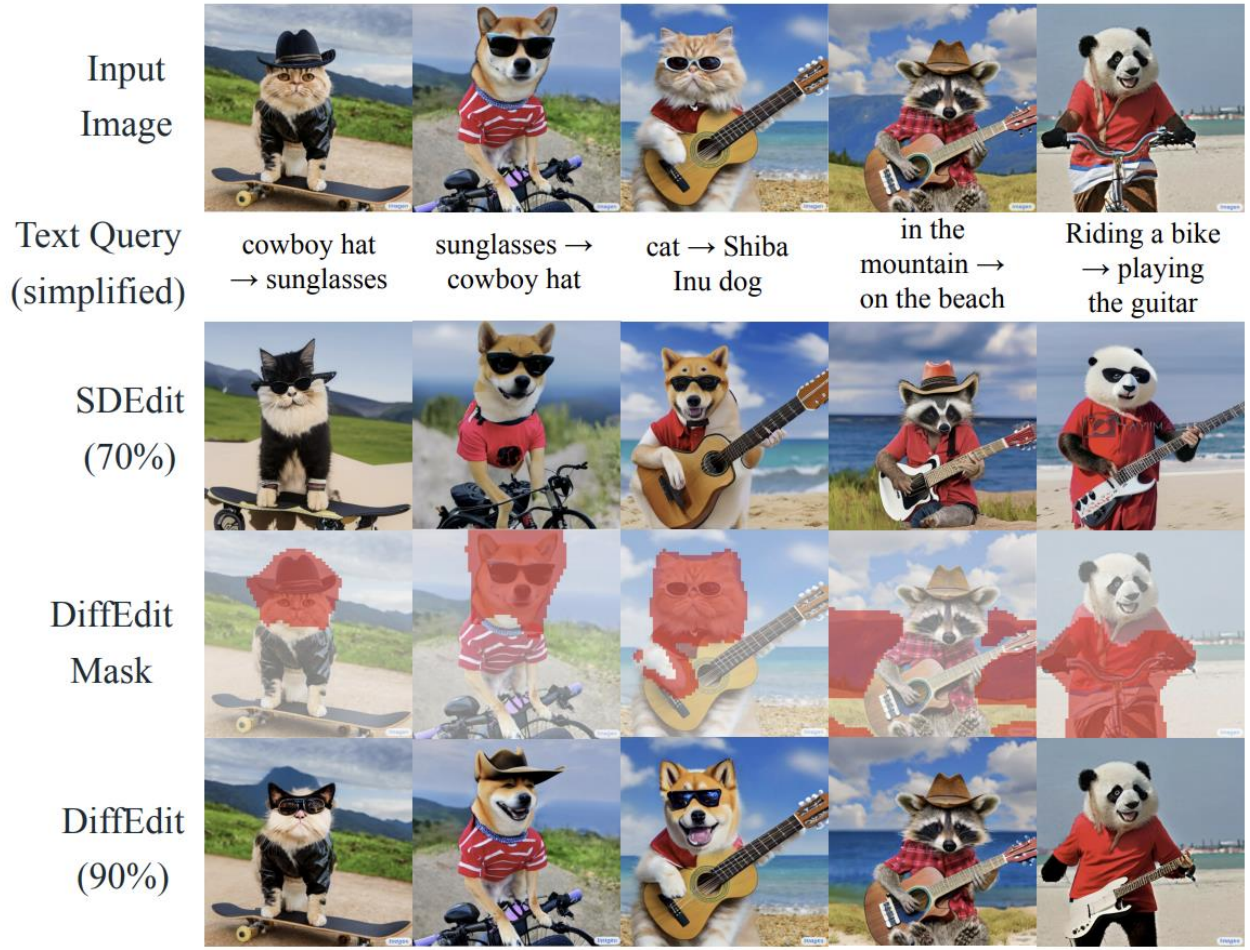
Couairon et al., "DiffEdit: Diffusion-based semantic image editing with mask guidance", ICLR 2023

DiffEdit: Diffusion-based semantic image editing with mask guidance



Couairon et al., "DiffEdit: Diffusion-based semantic image editing with mask guidance", ICLR 2023

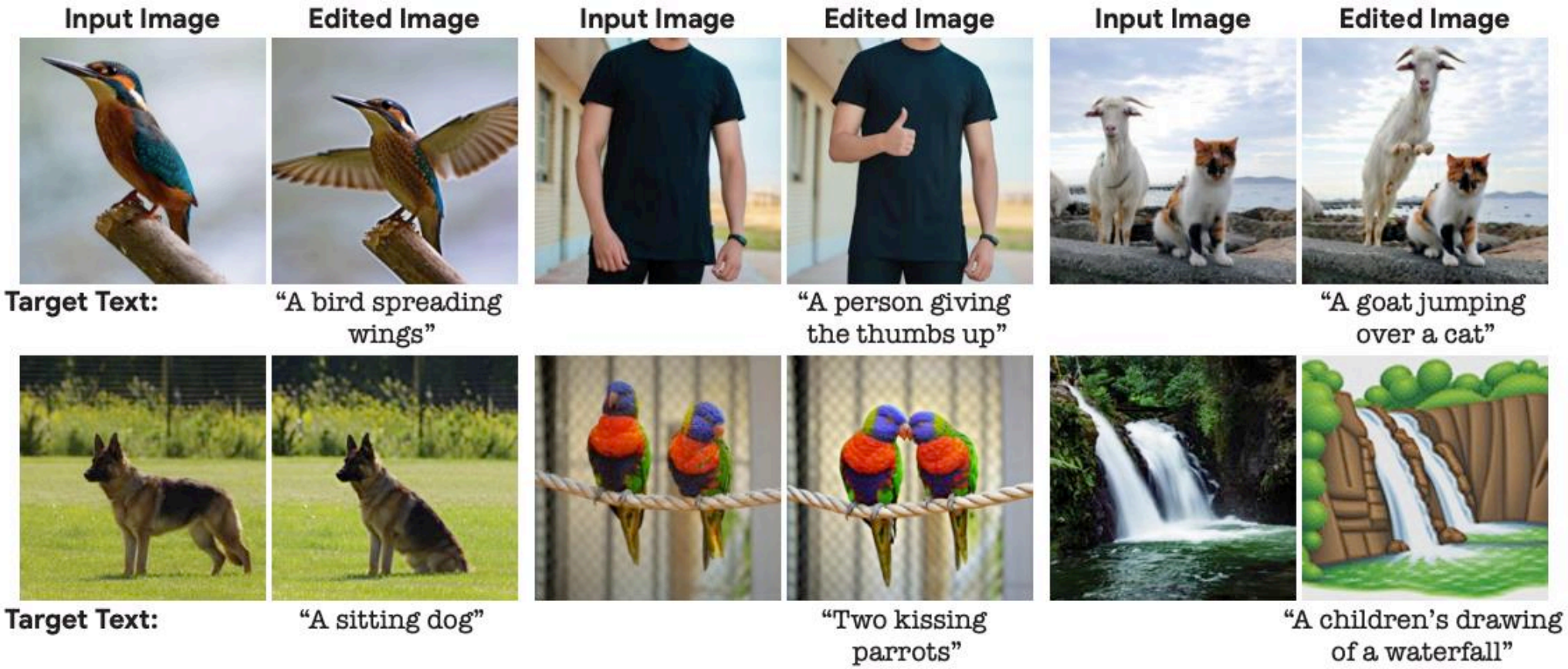
DiffEdit: Diffusion-based semantic image editing with mask guidance



Edits on Imagen dataset

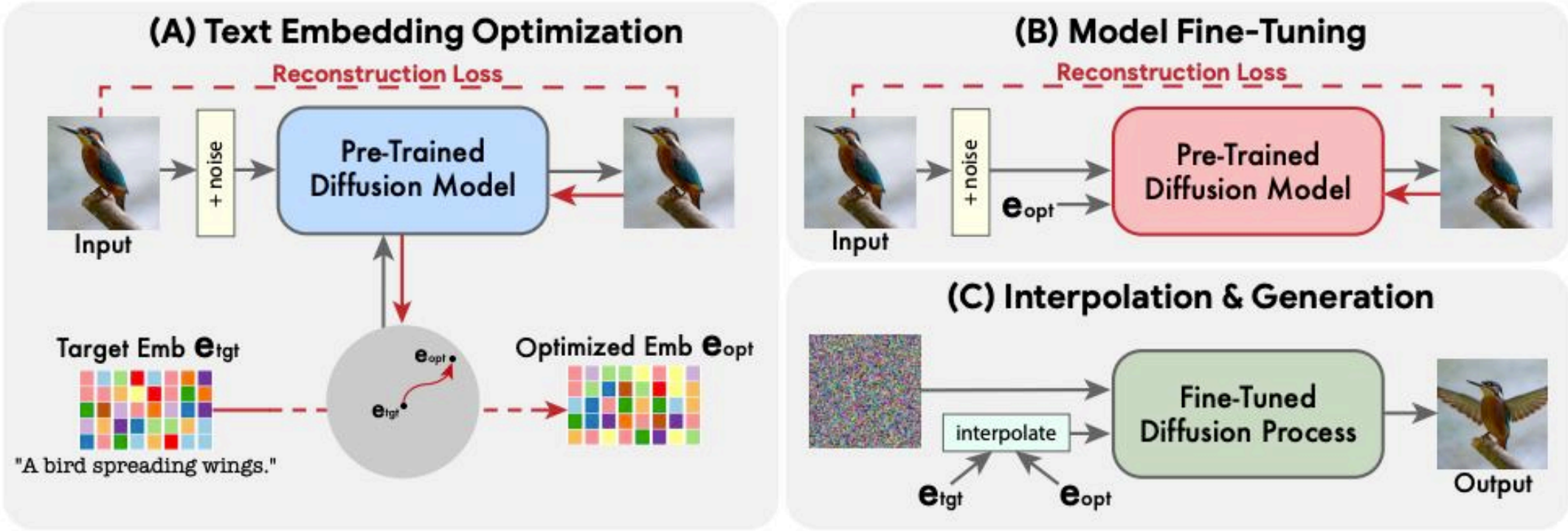
Couairon et al., "DiffEdit: Diffusion-based semantic image editing with mask guidance", ICLR 2023

Imagic: Text-Based Real Image Editing with Diffusion Models



Kawar et al., "Imagic: Text-Based Real Image Editing with Diffusion Models", CVPR 2023

Imagic: Text-Based Real Image Editing with Diffusion Models

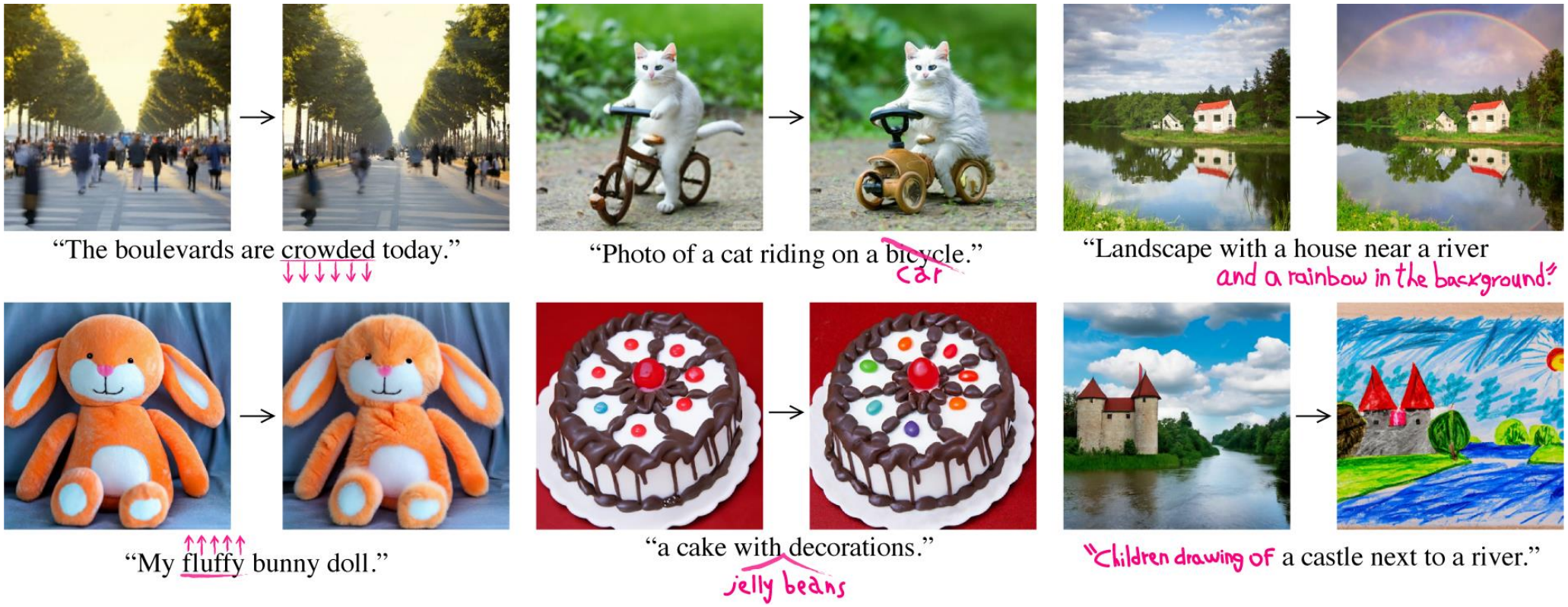


Kawar et al., ["Imagic: Text-Based Real Image Editing with Diffusion Models"](), CVPR 2023

Imagic: Text-Based Real Image Editing with Diffusion Models

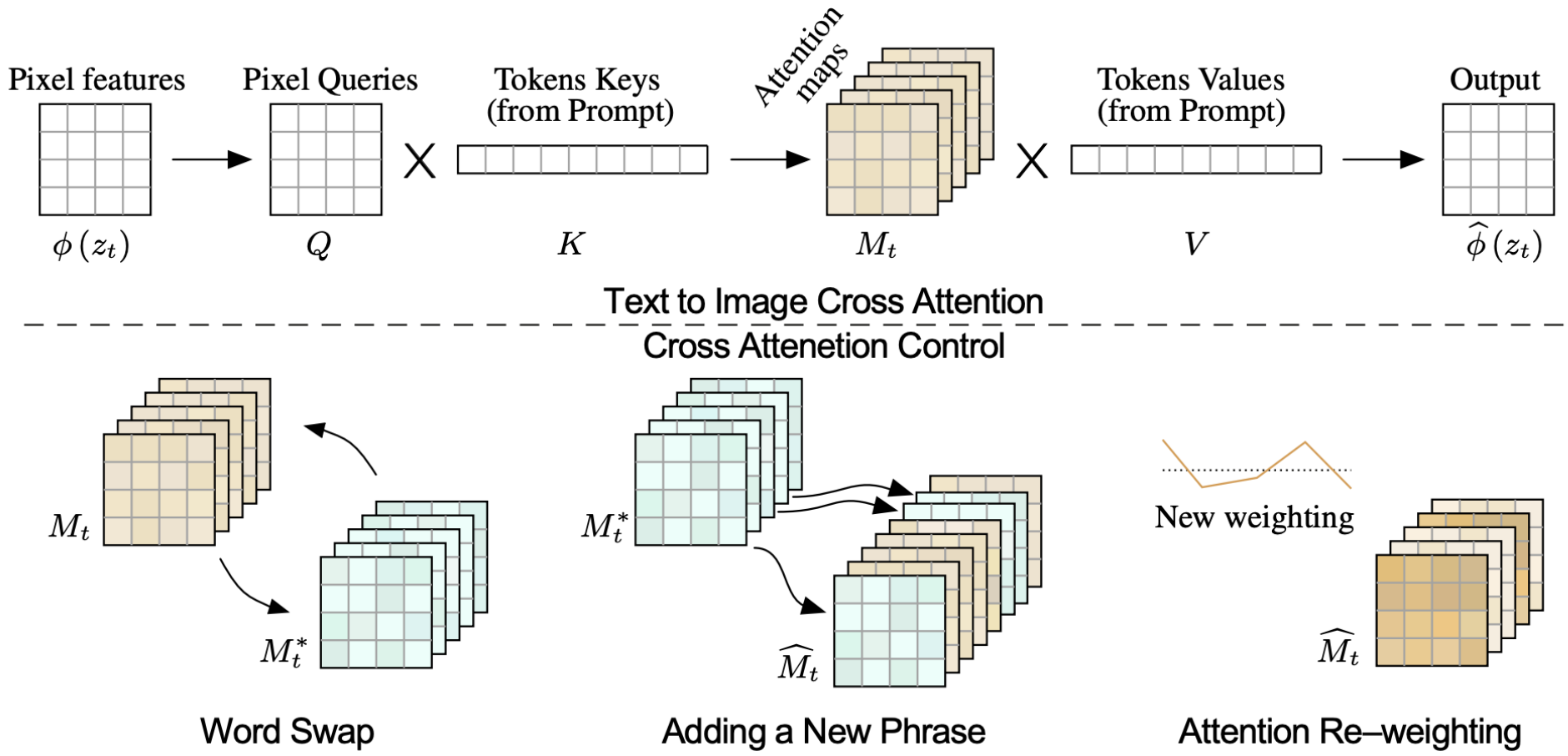
Kawar et al., "Imagic: Text-Based Real Image Editing with Diffusion Models", CVPR 2023

Prompt-to-Prompt Image Editing with Cross-Attention Control



Hertz et al., "Prompt-to-Prompt Image Editing with Cross-Attention Control", ICLR 2023

Prompt-to-Prompt Image Editing with Cross-Attention Control



Hertz et al., "Prompt-to-Prompt Image Editing with Cross-Attention Control", ICLR 2023

Prompt-to-Prompt Image Editing with Cross-Attention Control

"A photo of a birthday(↓) cake next to an apple."

"The picnic is ready under a blossom(↓) tree."

"A photo of a house on a snowy(↑) mountain."

Hertz et al., "Prompt-to-Prompt Image Editing with Cross-Attention Control", ICLR 2023

InstructPix2Pix: Learning to Follow Image Editing Instructions

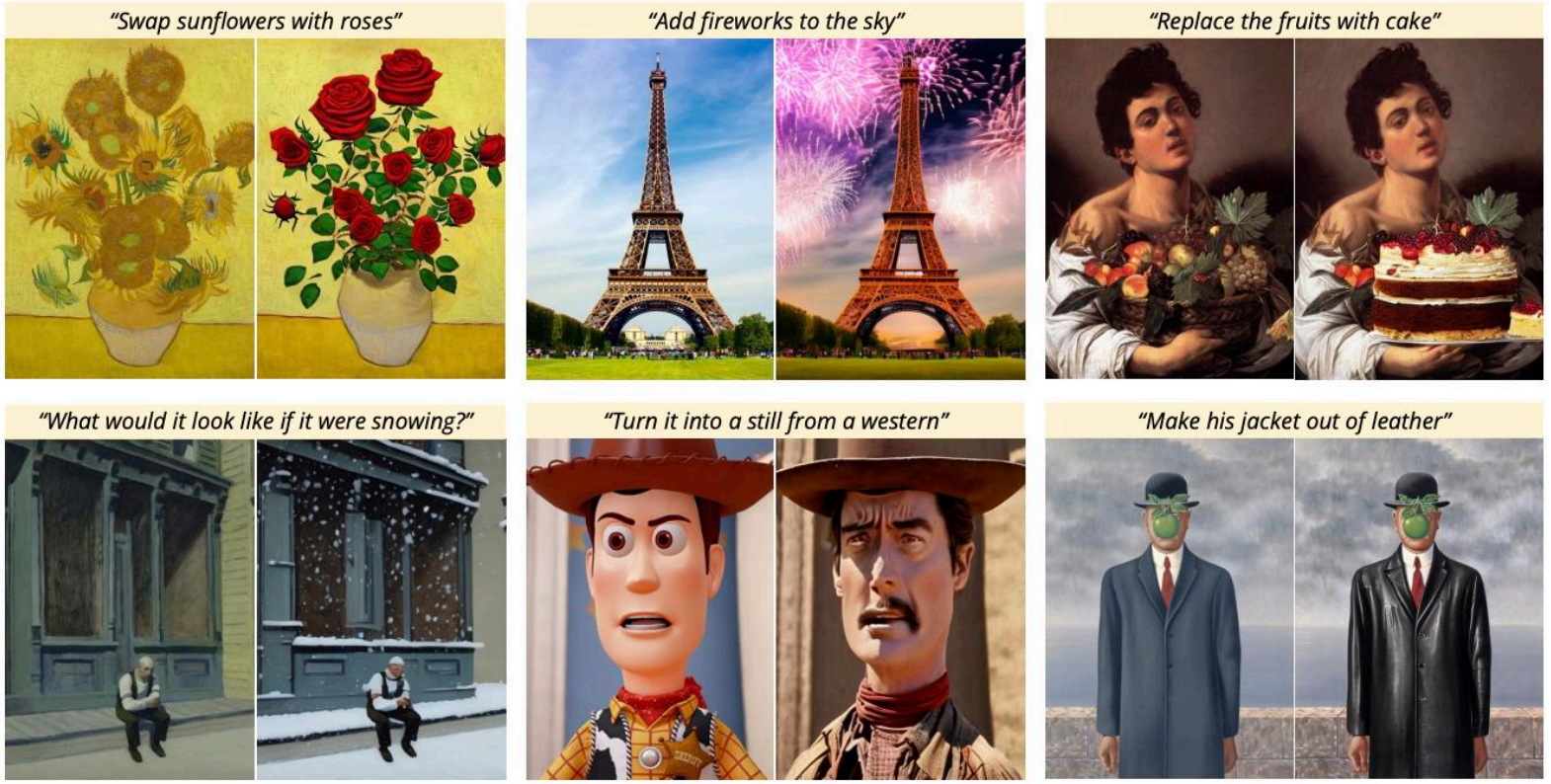
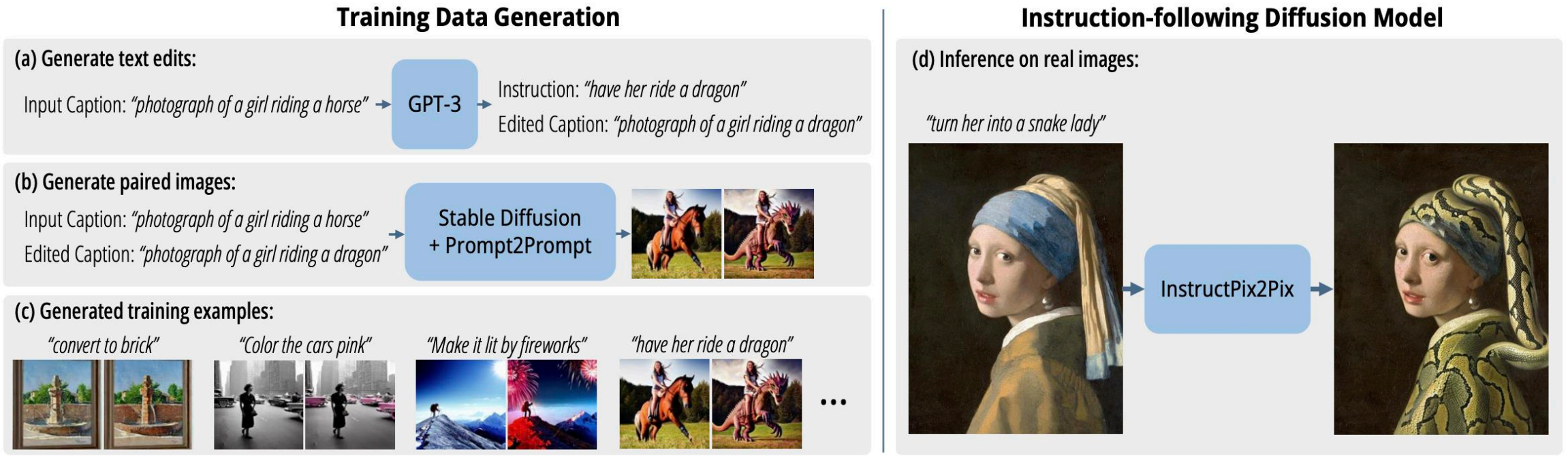


Figure 1. Given **an image** and **an instruction** for how to edit that image, our model performs the appropriate edit. Our model does not require full descriptions for the input or output image, and edits images in the forward pass without per-example inversion or fine-tuning.

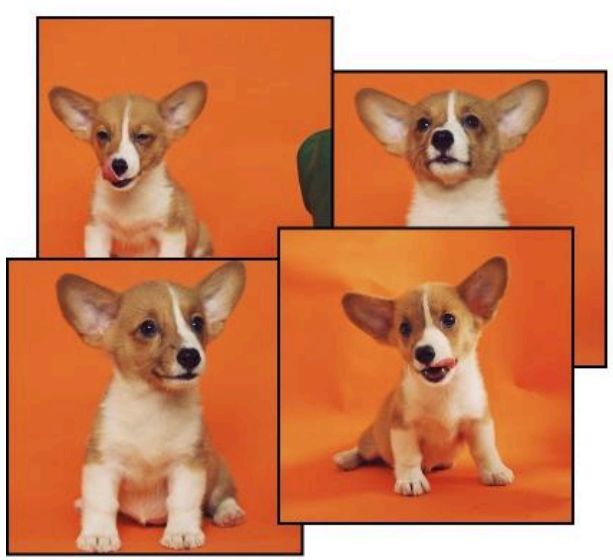
Brooks et al., "Instructpix2pix: Learning to follow image editing instructions", CVPR 2023

InstructPix2Pix: Learning to Follow Image Editing Instructions



Brooks et al., "Instructpix2pix: Learning to follow image editing instructions", CVPR 2023

Personalization with diffusion models



Input images

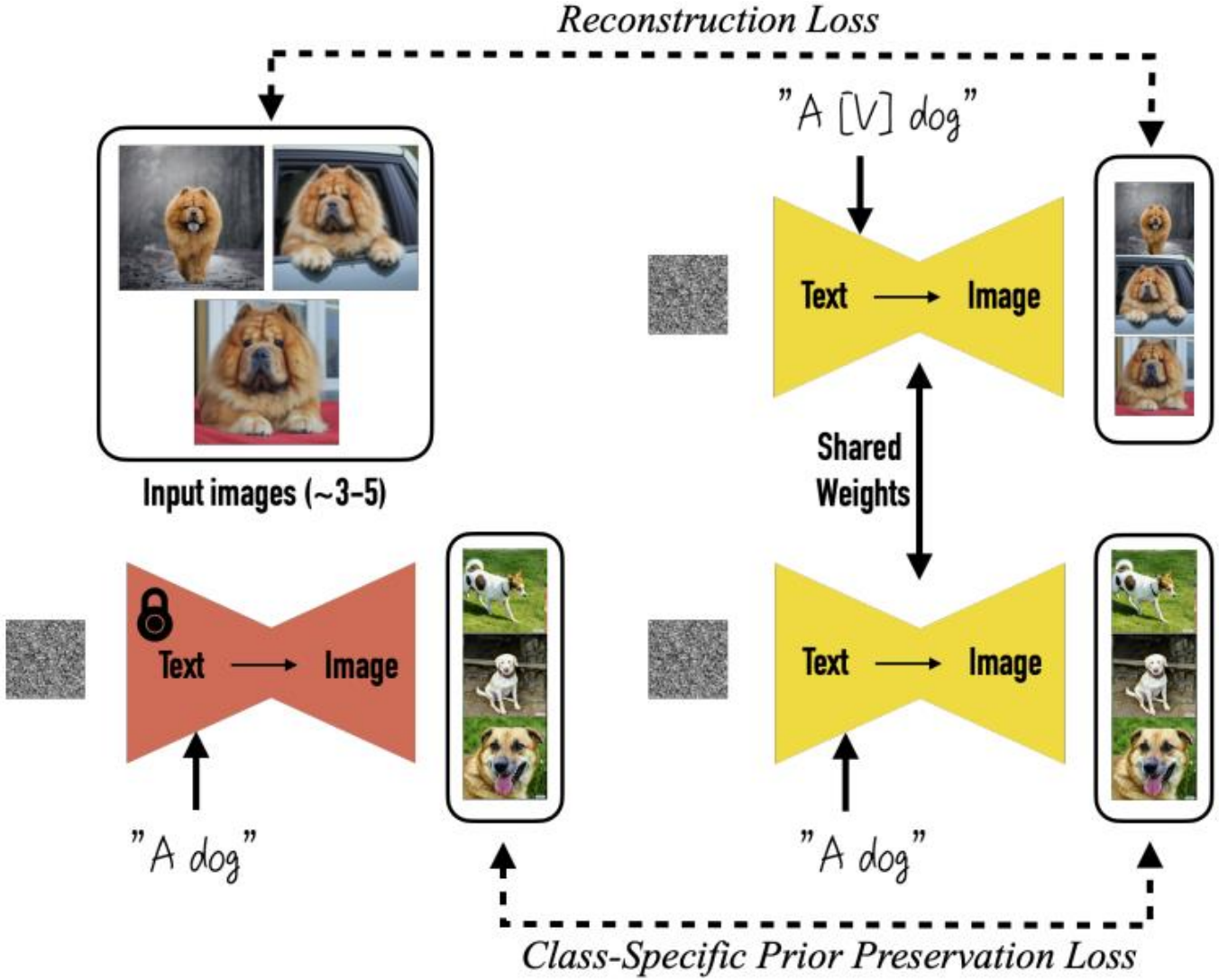
Generated images

Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023

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

Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023

The DreamBooth Method

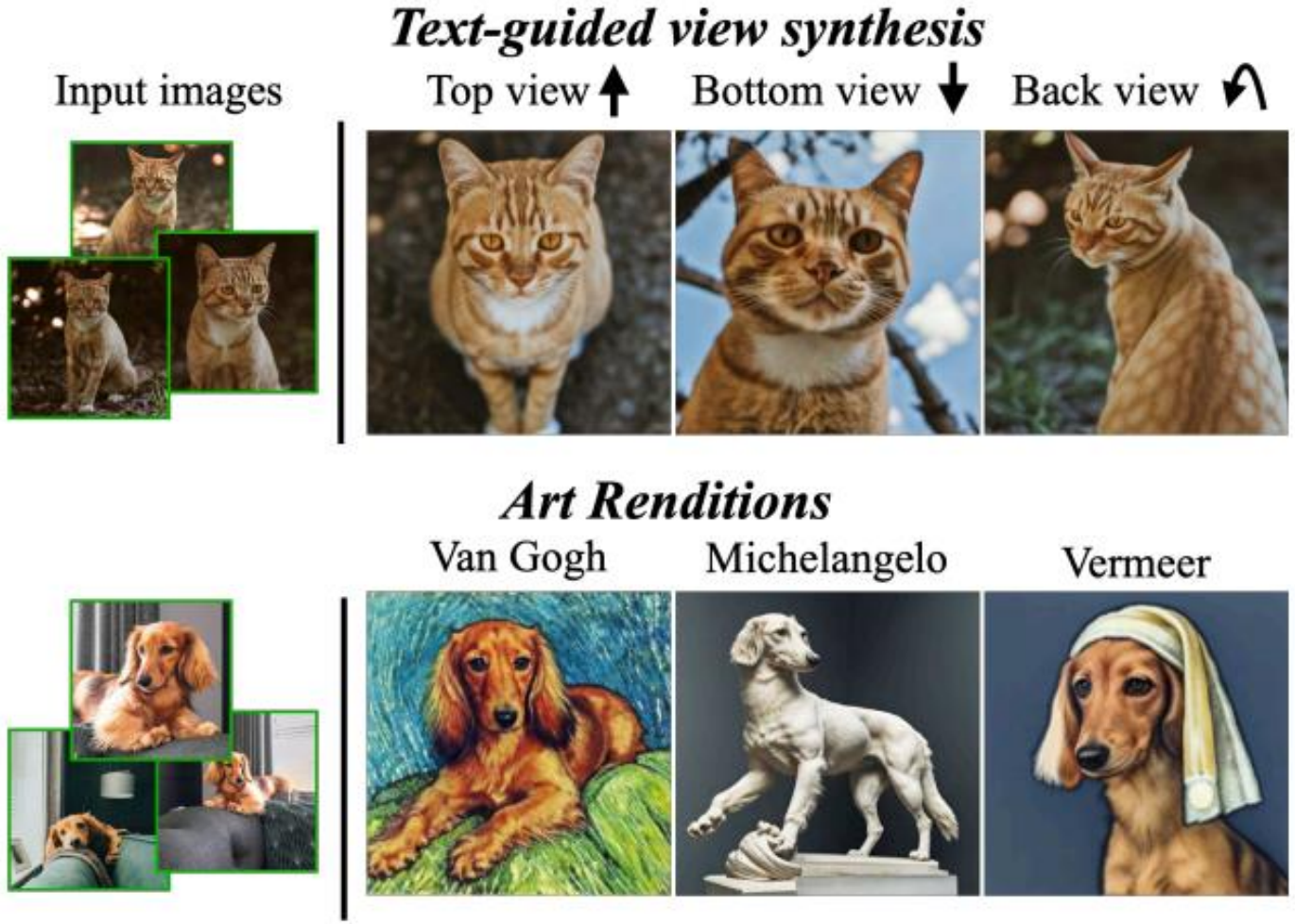


Ruiz et al., ["DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation"](), CVPR 2023

DreamBooth Results

Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023

DreamBooth Applications

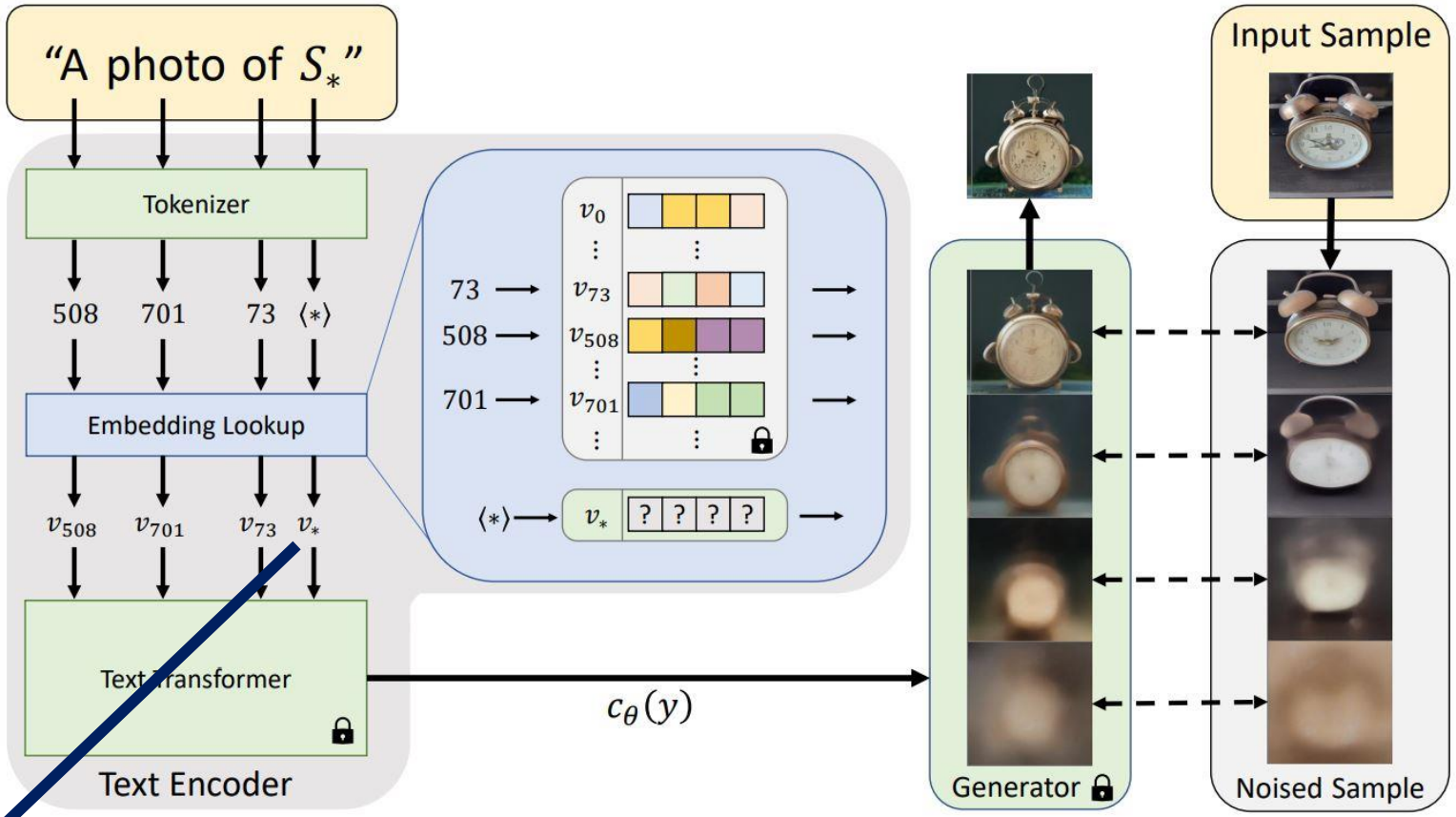


Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023

Textual Inversion: Optimizing Text Embedding

Gal et al., "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion", ICLR 2023

Textual Inversion: Optimizing Text Embedding



$$v_* = \arg\min_v \mathbb{E}_{z\sim\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t, c_\theta(y))\|_2^2\right]$$

Gal et al., "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion", ICLR 2023

Textual Inversion Results

Input samples

"S_* sports car"

"S_* made of lego"

"S_* onesie"

"da Vinci sketch of S_*"

Input samples

"Manga drawing of a steaming S_*"

"A S_* watering can"

"S_* Death Star"

"A poster for the movie 'The Teapot' starring S_*"

Gal et al., "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion", ICLR 2023

Works well for artistic styles

Input samples

"The streets of Paris in the style of S_*"

"Adorable corgi in the style of S_*"

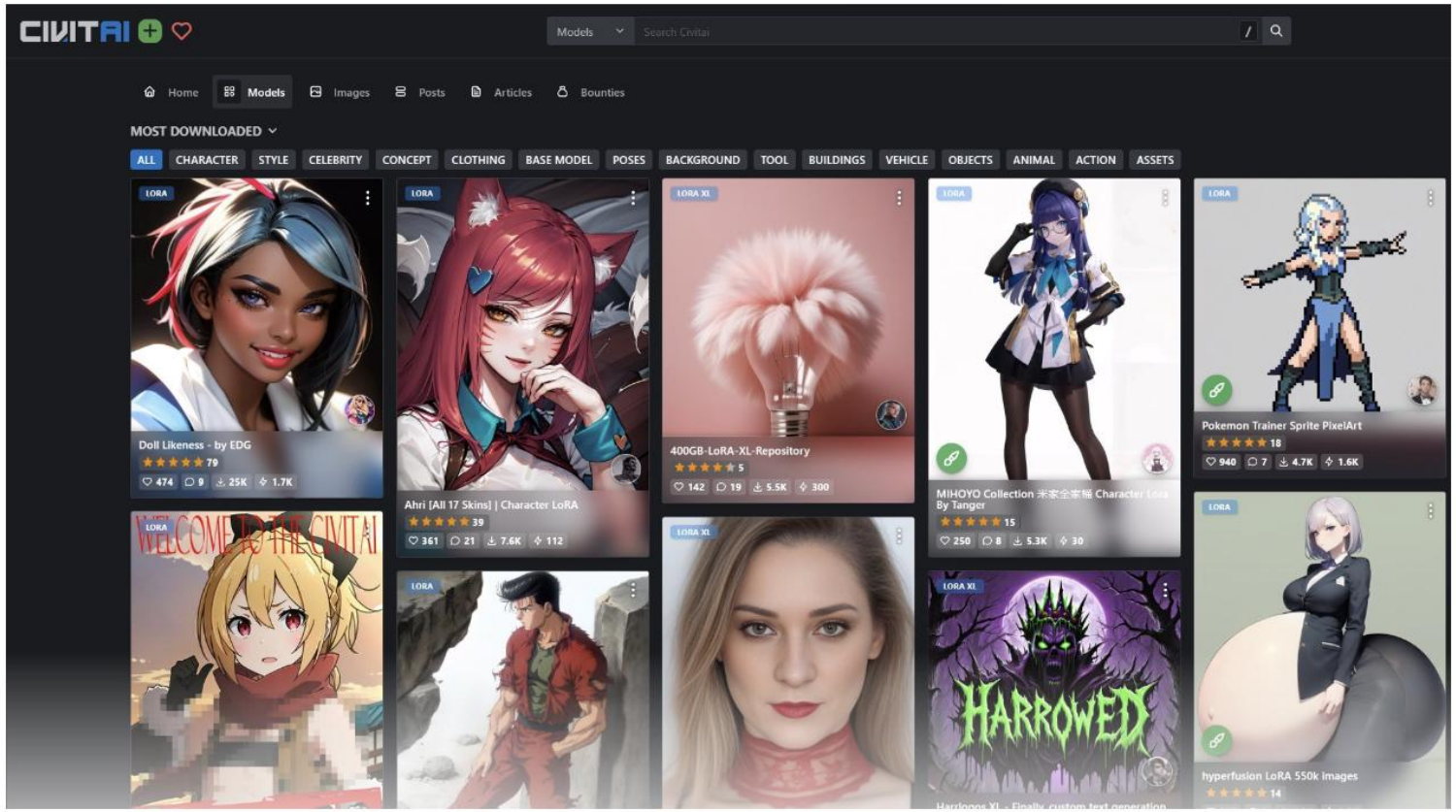
"Painting of a black hole in the style of S_*"

"Times square in the style of S_*"

Gal et al., "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion", ICLR 2023

LoRA: Low-Rank Adaptation

Few-shot finetuning of large models for personalized generation.



A Parameter-Efficient Fine-Tuning Method

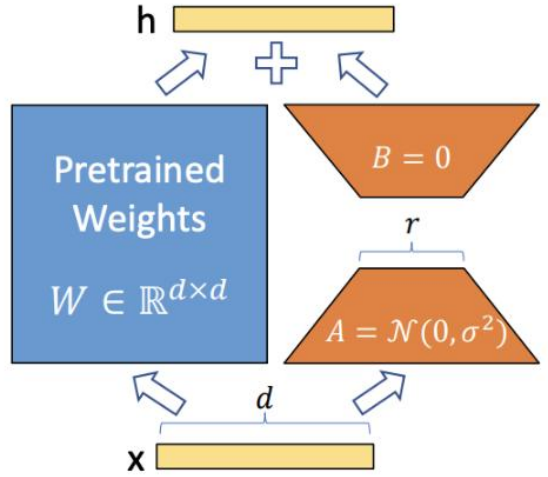
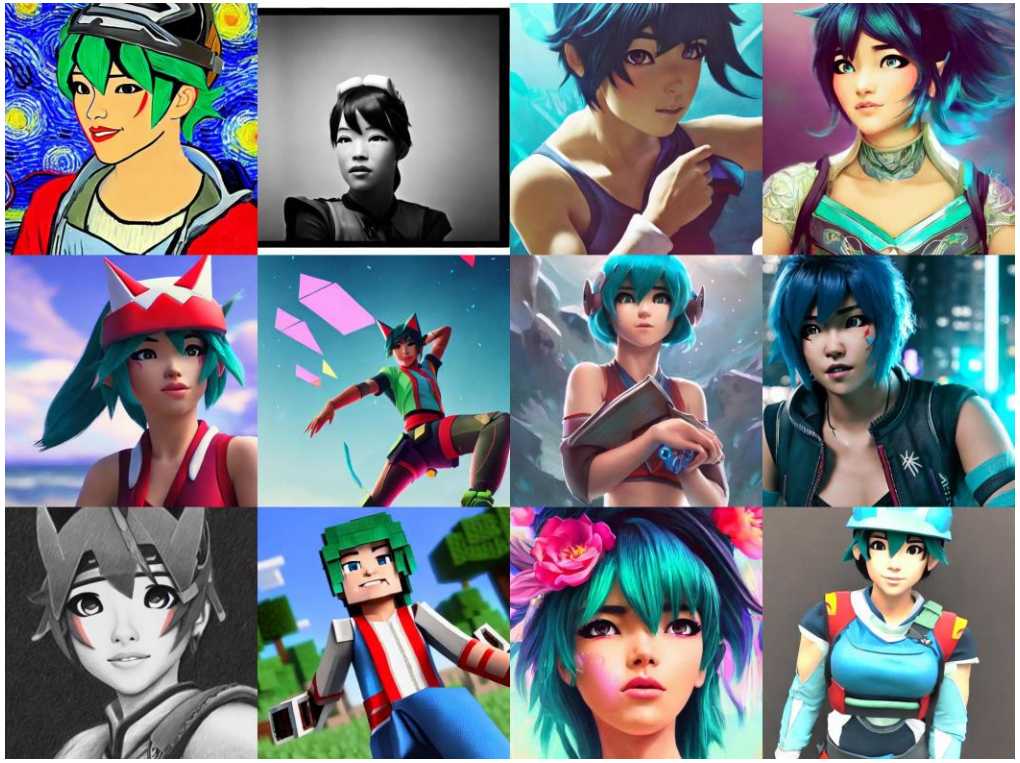


Figure 1: Our reparametrization. We only train A and B.

Low-rank Adaptation (LoRA)

- Lora: Low-rank adaptation of large language models



Original weights

$$W = W_0 + BA$$

Low-rank difference

Lora [Edward J. Hu*, Yelong Shen*, et al., ICLR 2022]

Lora + Dreambooth (by Simo Ryu): https://github.com/cloneofsimo/lora

Low-rank Adaptation (LoRA)

Finetuned with only 9 images
Visualized every 500 steps

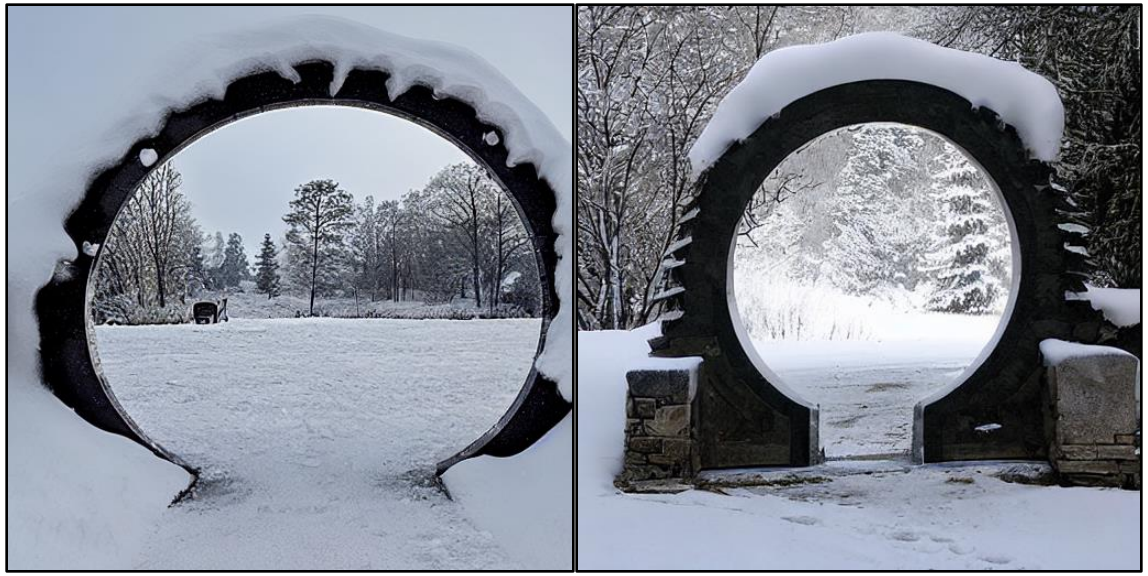
Lora [Edward J. Hu*, Yelong Shen*, et al., ICLR 2022]

Lora + Dreambooth (by Simo Ryu): https://github.com/cloneofsimo/lora

Fine-tuning all model weights

Photo of a moongate

Moongate in snowy ice

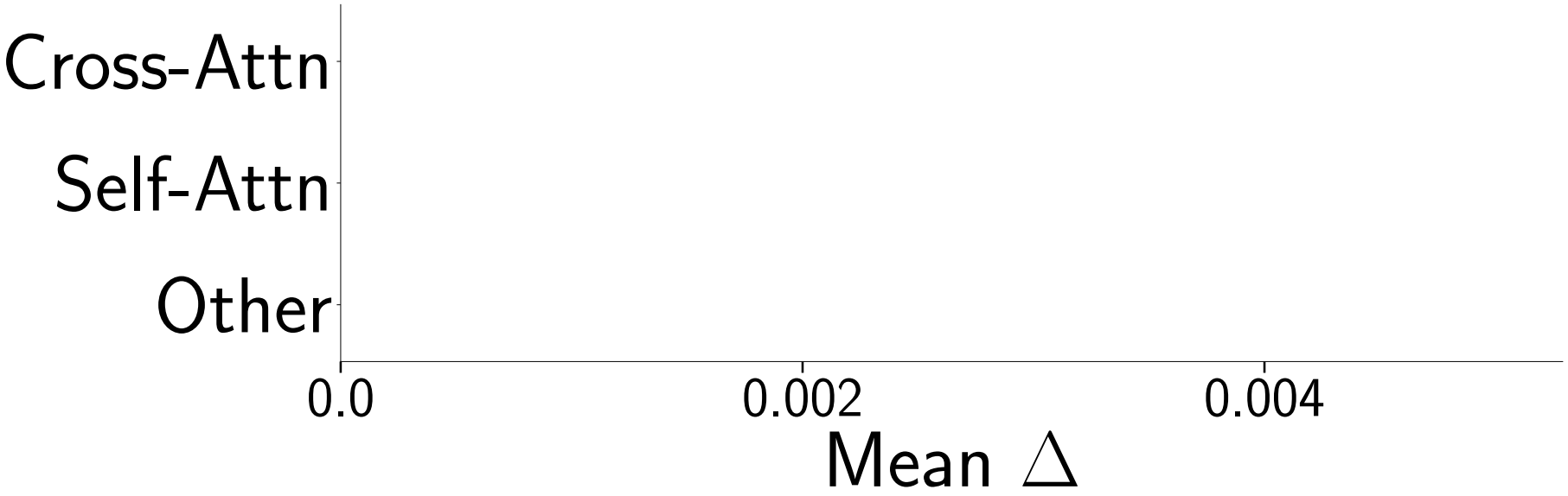


Storage requirement. 4GB storage for each fine-tuned model.
Compute requirement. It requires more VRAM/training time.
Compositionality. Hard to combine multiple models.

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Analyze change in weights

$$\Delta_l = \frac{||\theta_l' - \theta_l||}{||\theta_l||}$$ where θ_l' : updated weights, θ_l : pretrained weights



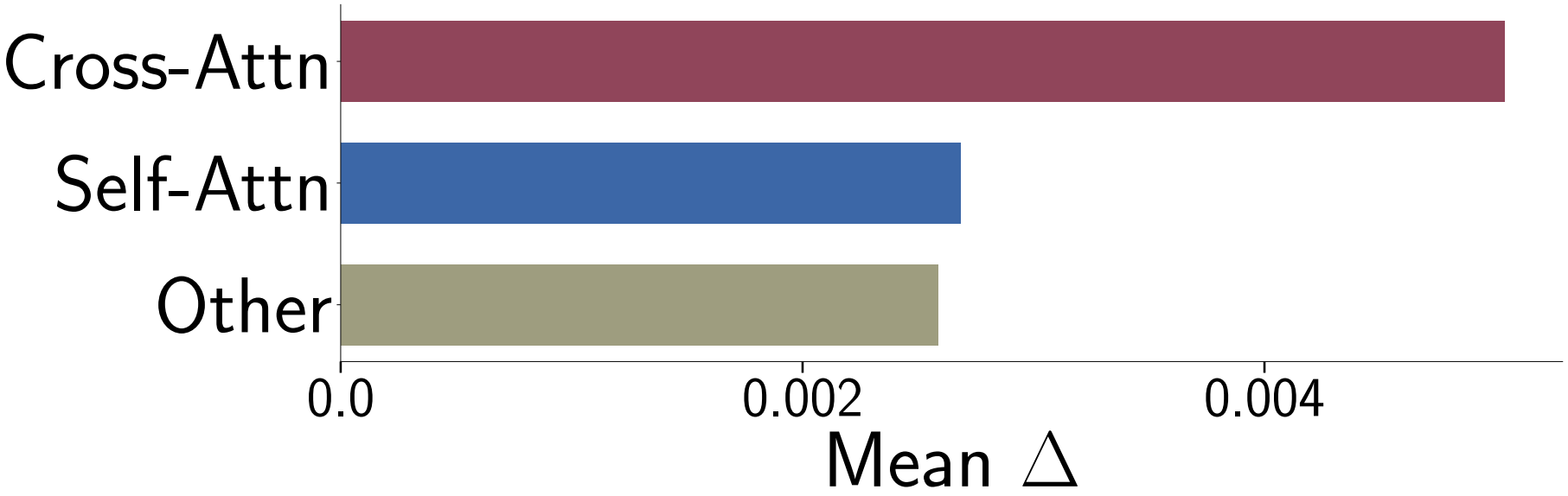
Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Analyze change in weights

$$\Delta_l = \frac{||\theta'_l - \theta_l||}{||\theta_l||}$$

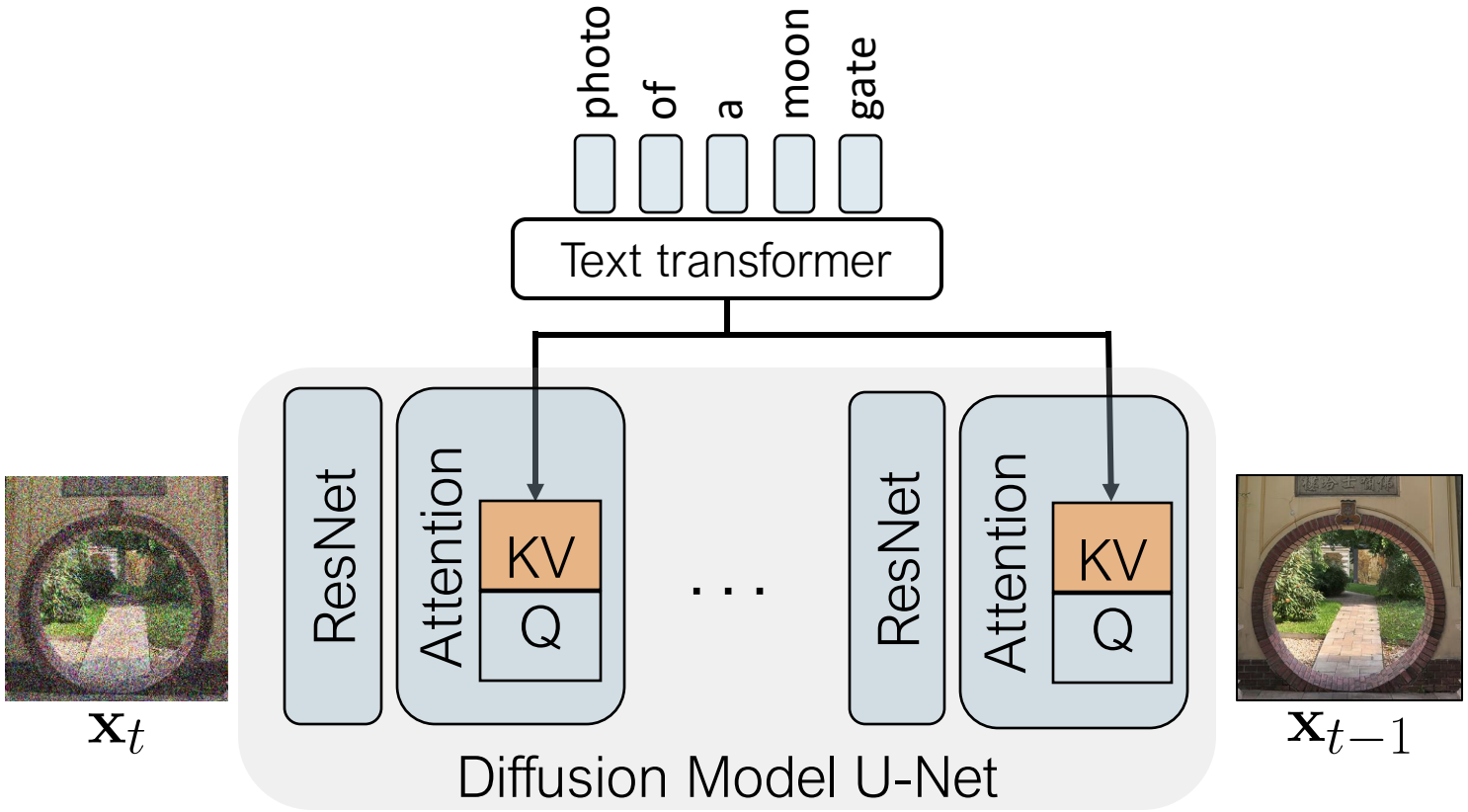
where θ'_l : updated weights

θ_l : pretrained weights



Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Only fine-tune cross-attention layers



Trainable Frozen

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

How to prevent overfitting?

Photo of a {moongate}

Photo of a {moongate}

. . .

Target images

+

sky full of stars and the moon

Blood moon

. . .

Add regularization images

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Personalized concepts

Jun-Yan's dog, Stark

How to describe personalized concepts?

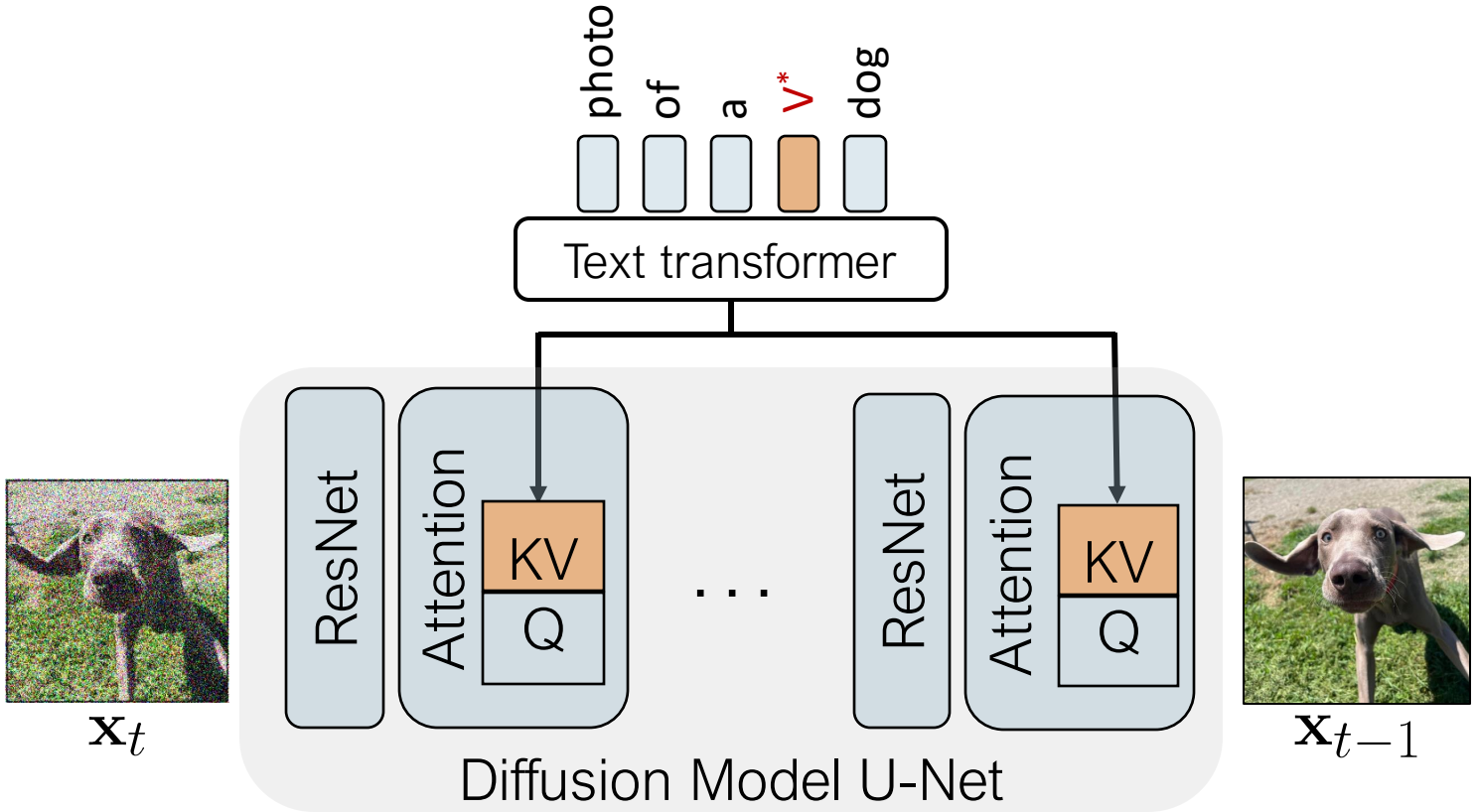
V^* dog

Where V^* is a modifier token in the text embedding space

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Personalized concepts

Also fine-tune the modifier token V* that describes the personalized concept



Trainable Frozen

[Nupur Kumari et al., CVPR 2023]

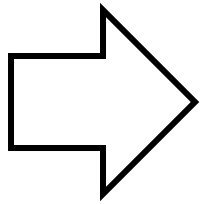
Single concept results

V* dog wearing headphones

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Multiple new concepts?

+



?

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Joint training

1. Combine the training dataset of multiple concepts

Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Two concept results

The V_1^* cat is sitting inside a V_2^* wooden pot and looking up

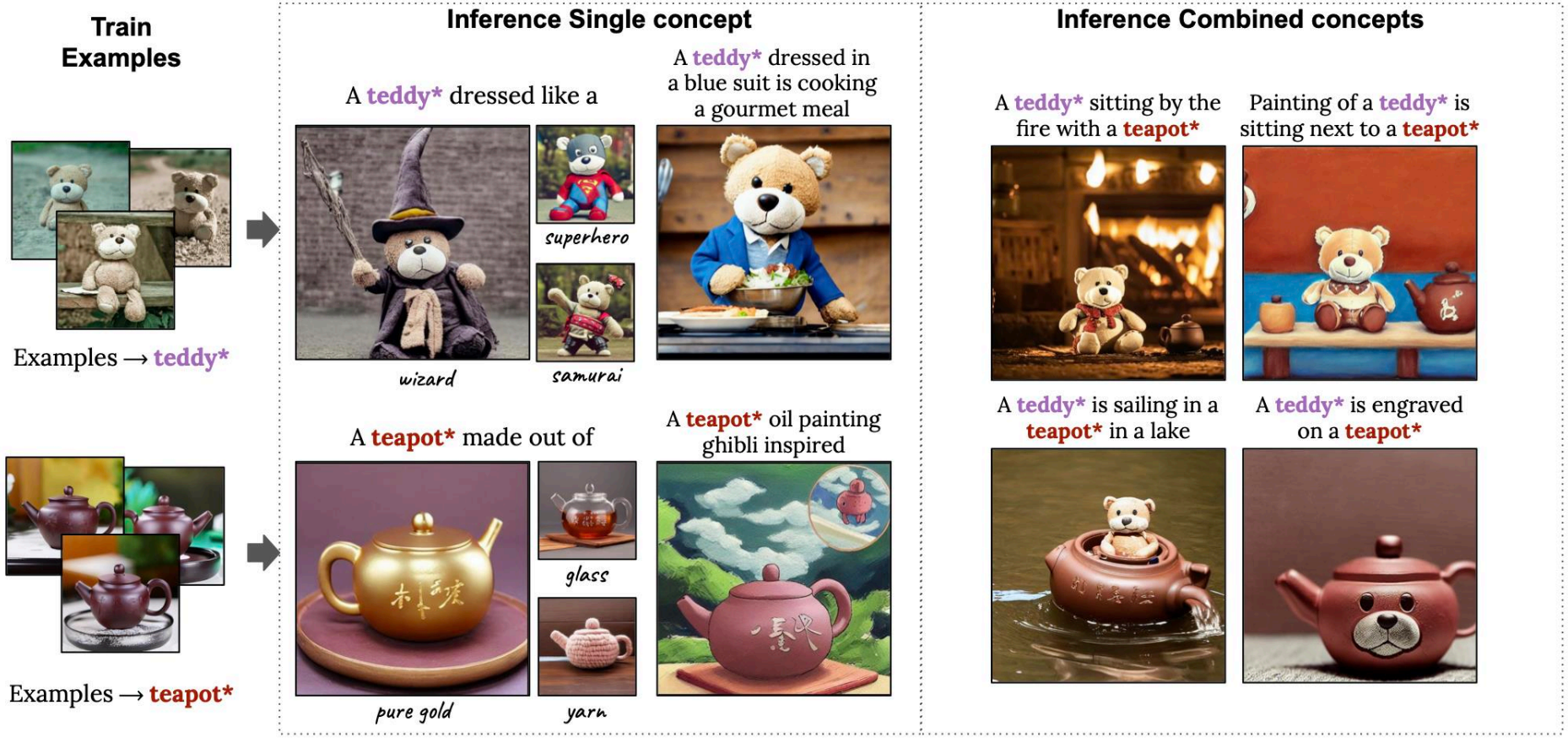
Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Two concept results

V_1^* chair with the V_2^* cat sitting on it near a beach

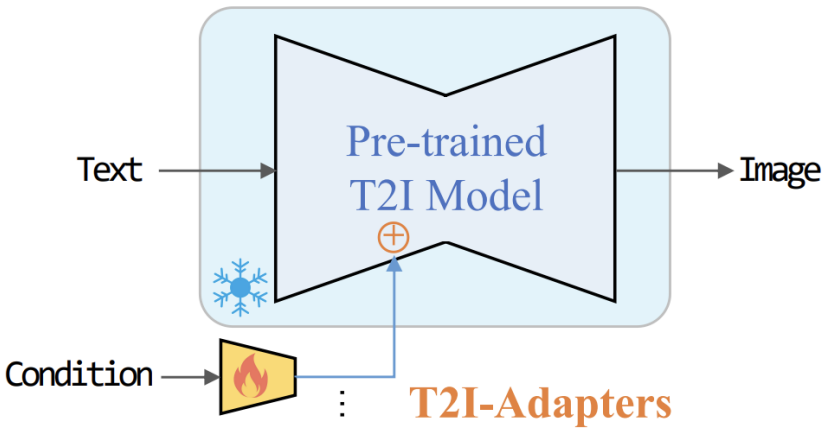
Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023

Key-Locked Rank One Editing for Text-to-Image Personalization

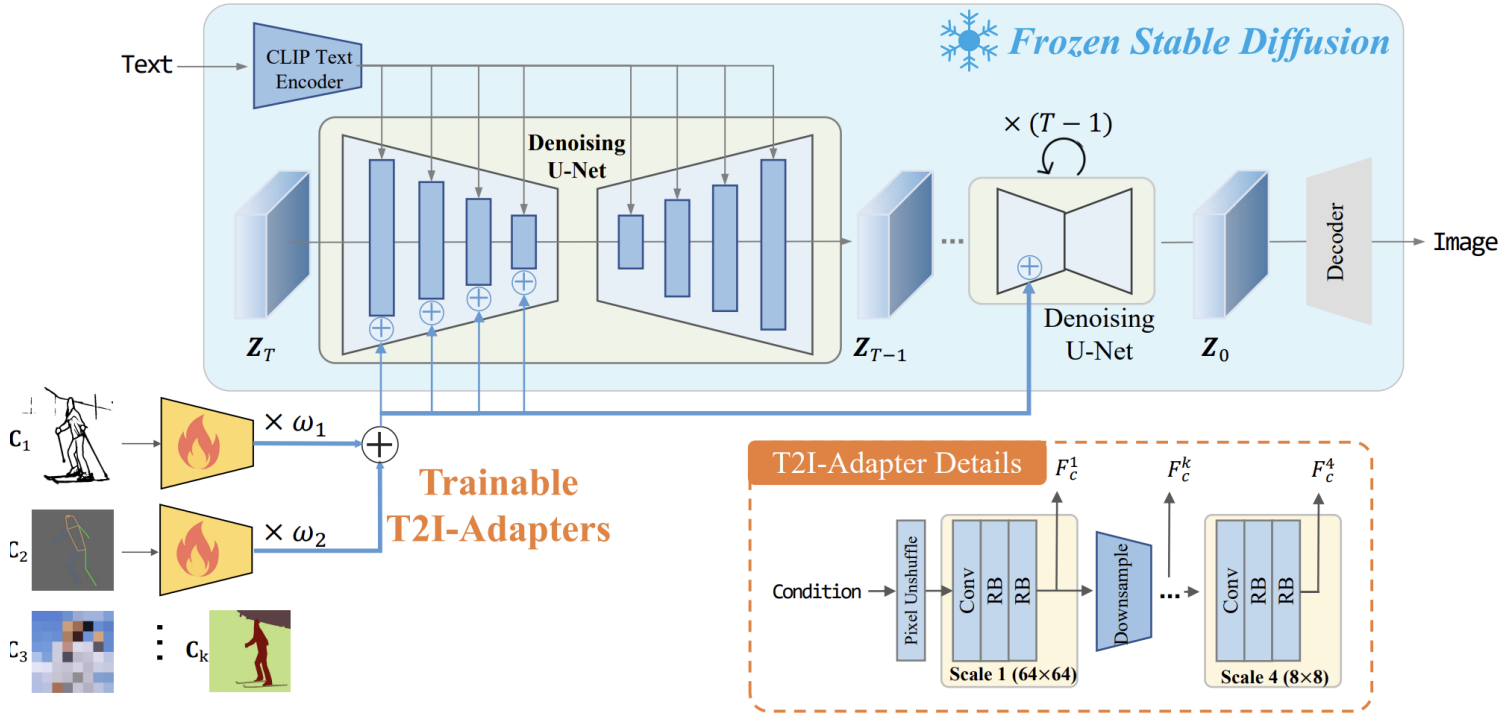


Tewel et al., "Key-Locked Rank One Editing for Text-to-Image Personalization", SIGGRAPH 2023

T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models

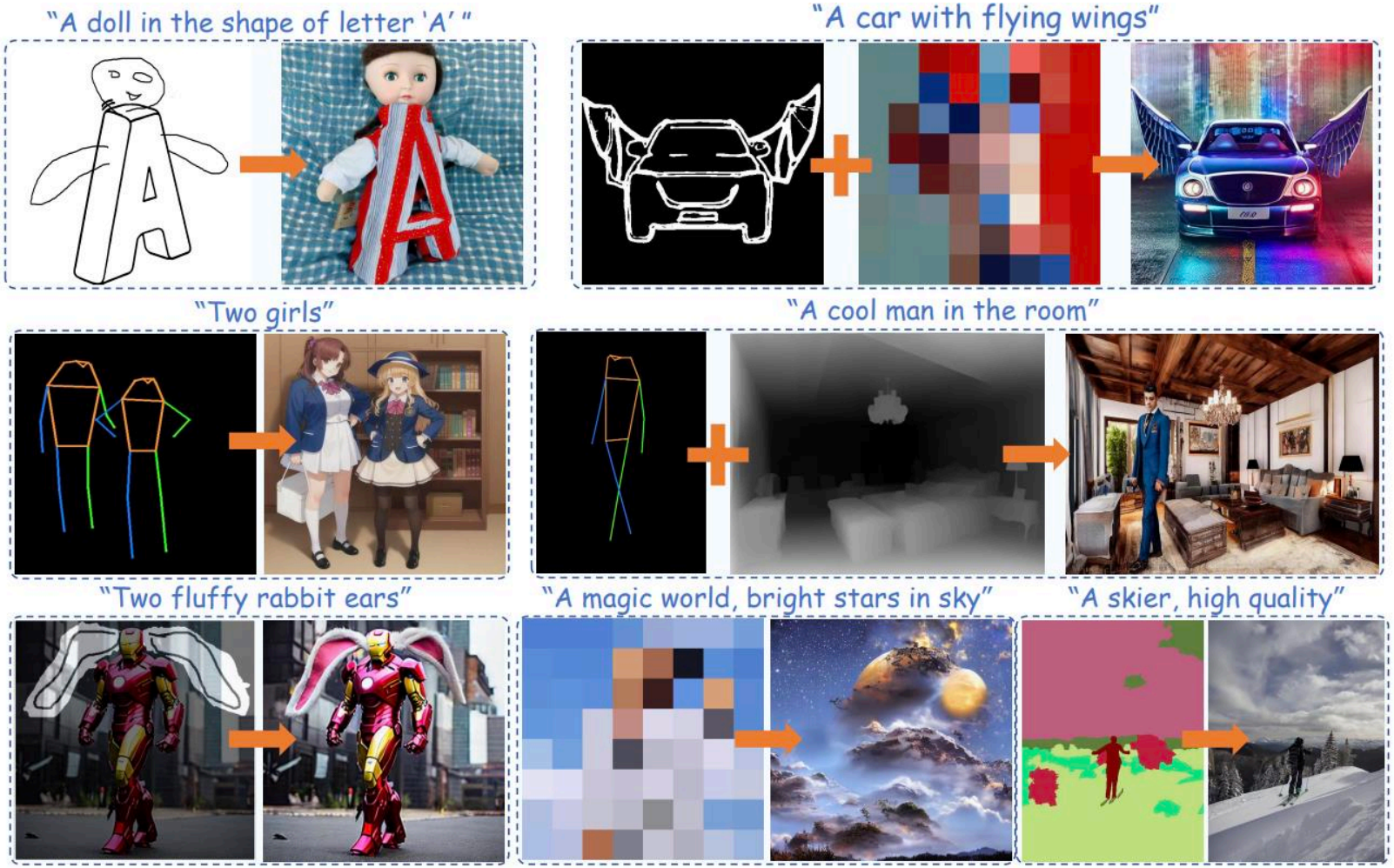


- ✓ ***Plug-and-play.*** *Not affect original network topology and generation ability*
- ✓ ***Simple and small.*** *~77M parameters and ~300M storage*
- ✓ ***Flexible.*** *Various adapters for different control conditions*
- ✓ ***Composable.*** *Several adapters to achieve multi-condition control*
- ✓ ***Generalizable.*** *Can be directly used on customed models*



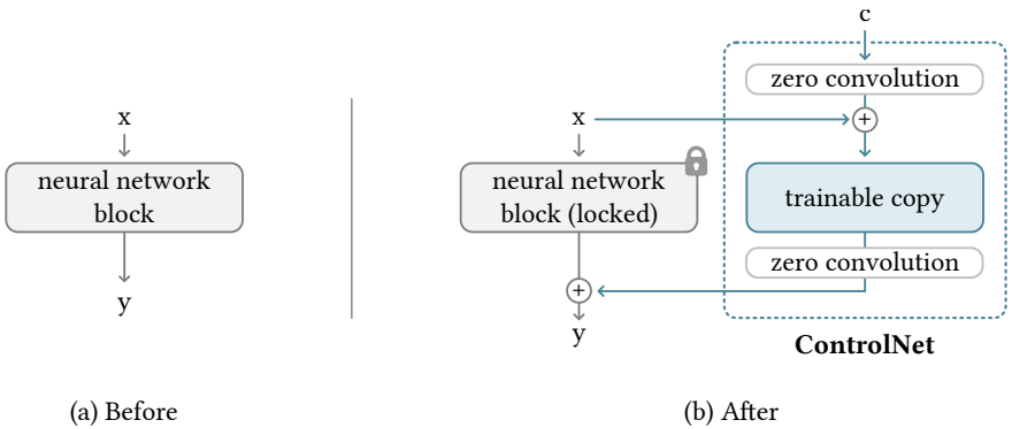
Mou et al., "T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models", arXiv 2023

T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models



Mou et al., "T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models", arXiv 2023

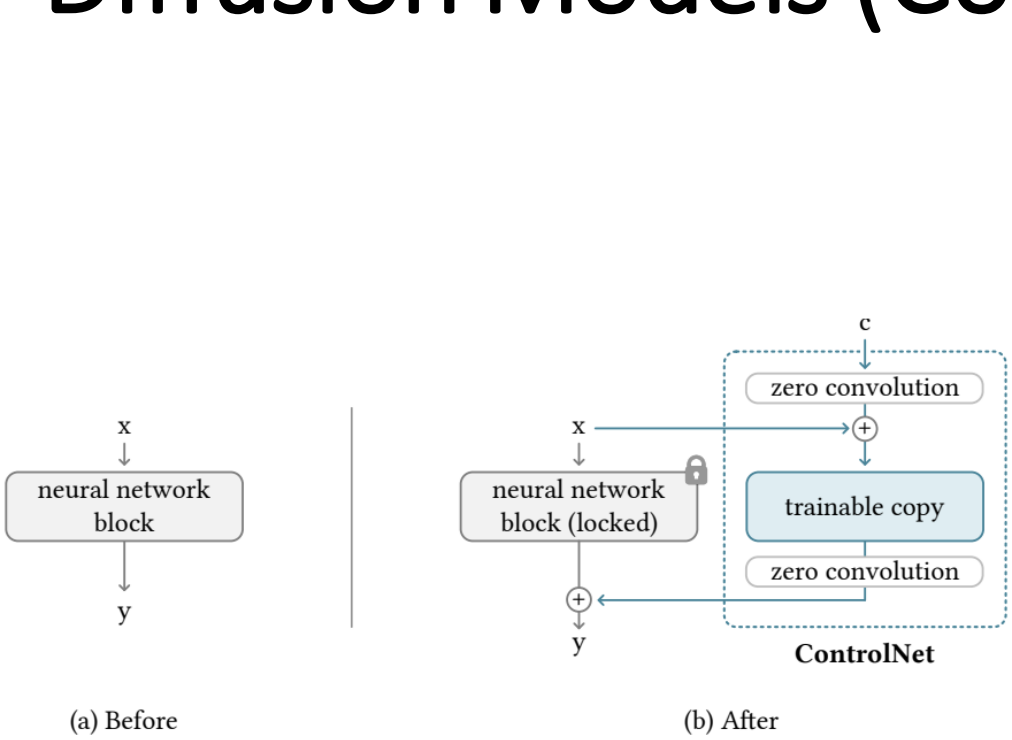
Adding Conditional Control to Text-to-Image Diffusion Models (ControlNet)



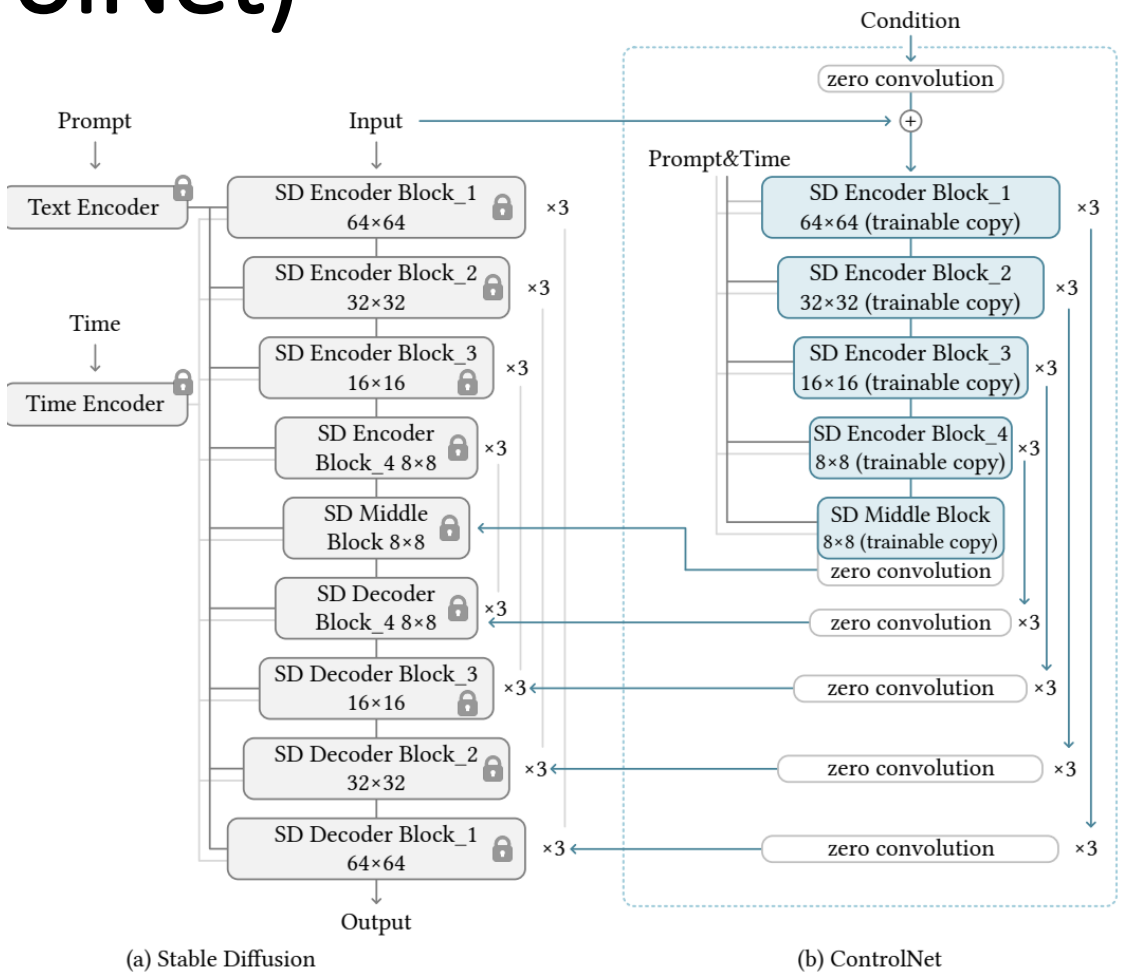
ControlNet

Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

Adding Conditional Control to Text-to-Image Diffusion Models (ControlNet)



ControlNet



Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

Adding Conditional Control to Text-to-Image Diffusion Models

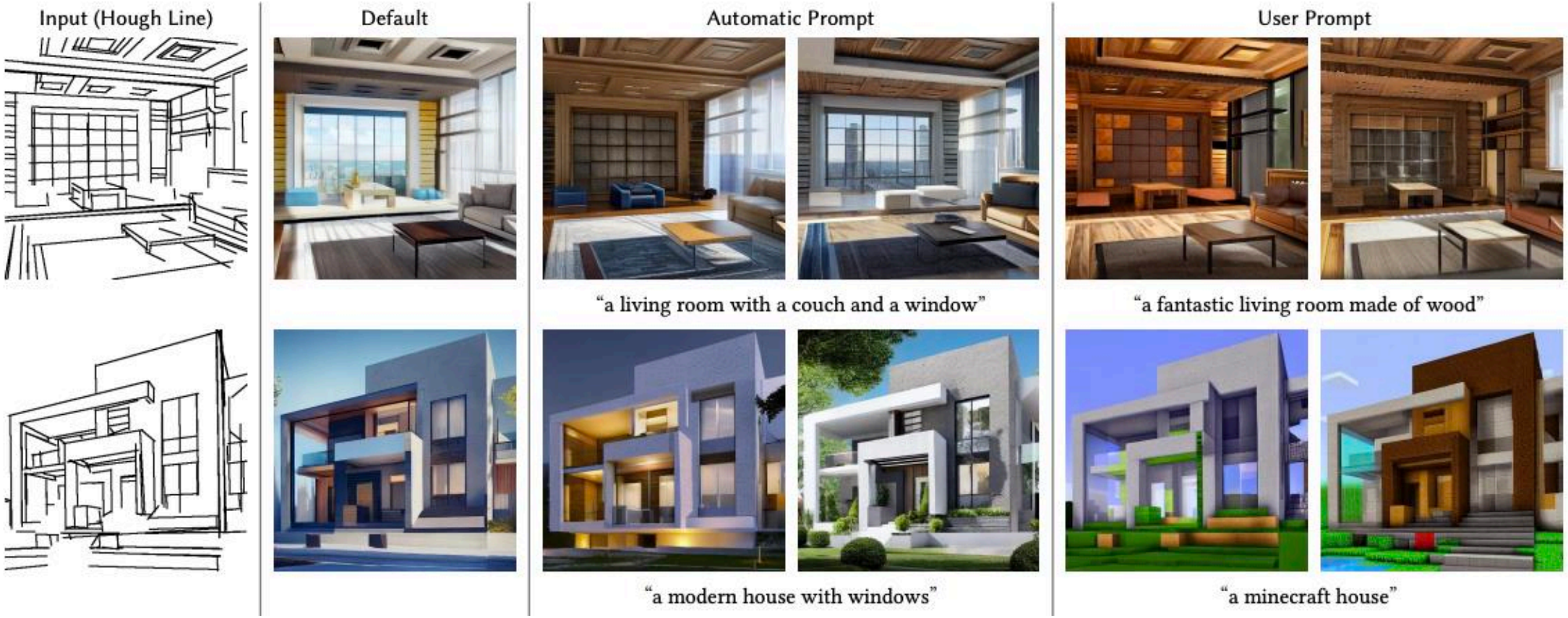
Train objective

$$\mathcal{L} = \mathbb{E}_{\boldsymbol{z}_0, t, \boldsymbol{c}_t, \boldsymbol{c}_\mathrm{f}, \epsilon \sim \mathcal{N}(0,1)} \left[\| \epsilon - \epsilon_\theta(z_t, t, \boldsymbol{c}_t, \boldsymbol{c}_\mathrm{f}) \|_2^2 \right]$$

where t is the time step, c_t is the text prompts, c_f is the task-specific conditions

Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

ControlNet

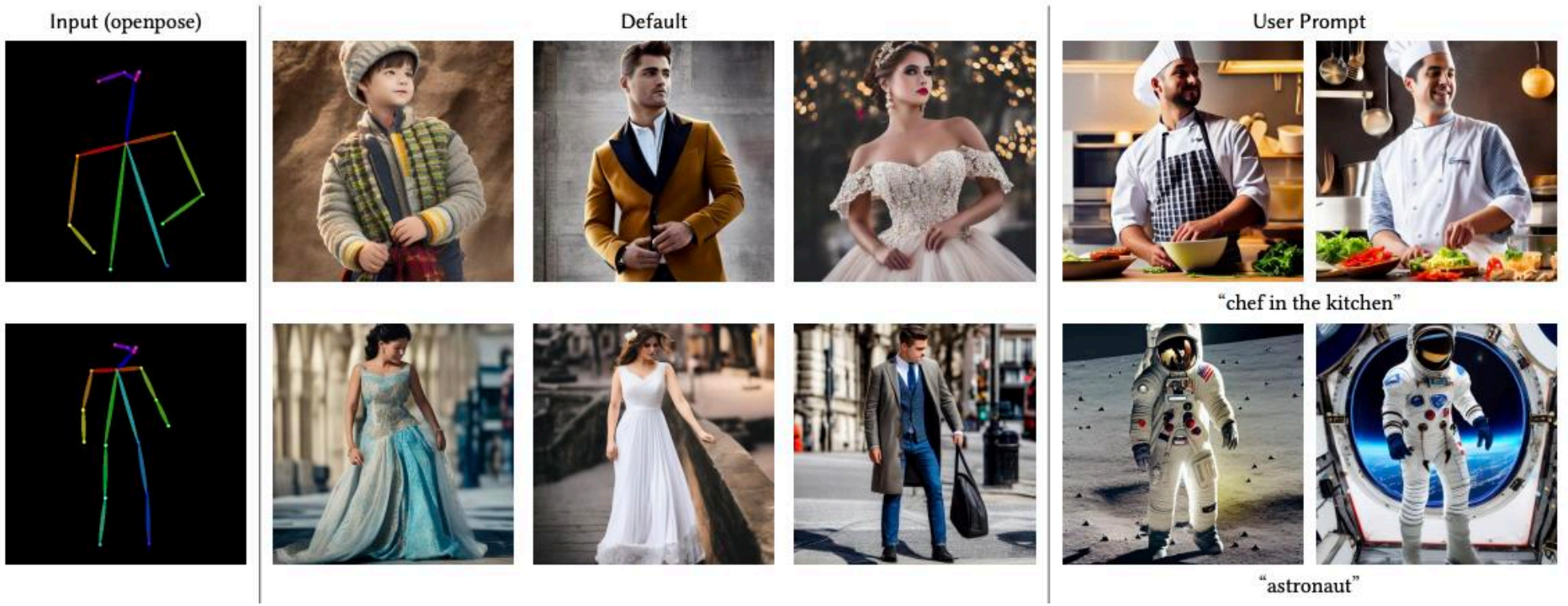


Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

ControlNet

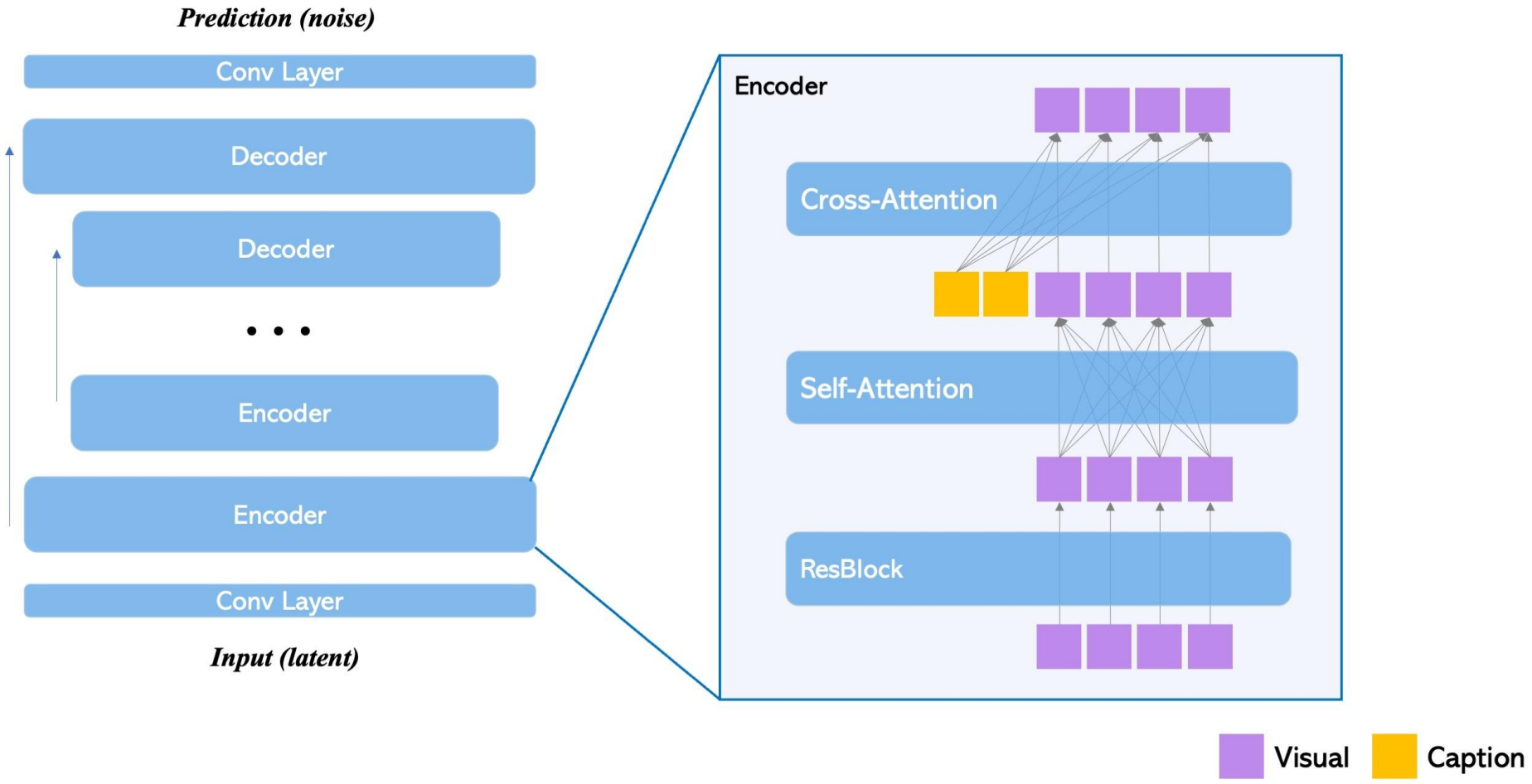
Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

ControlNet



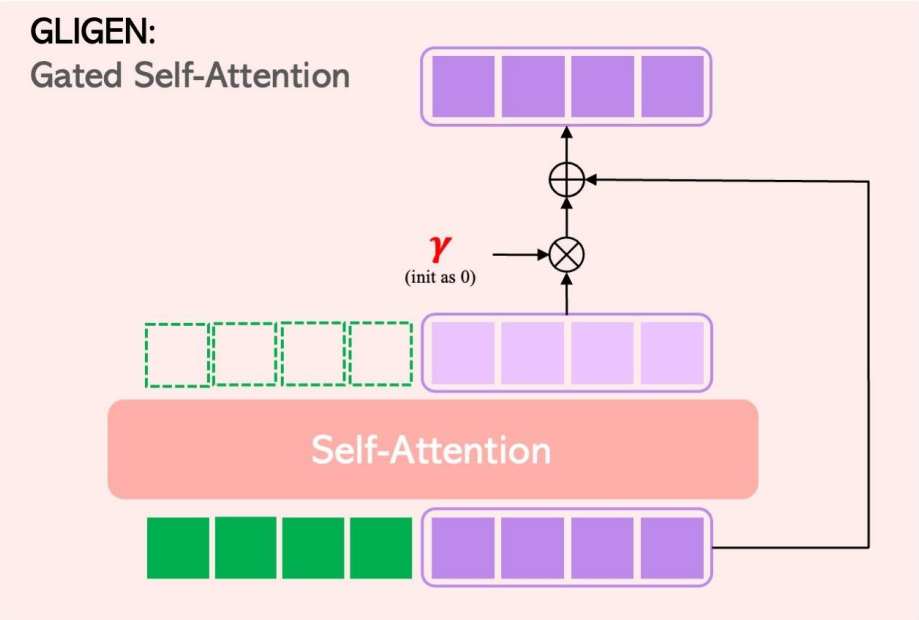
Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

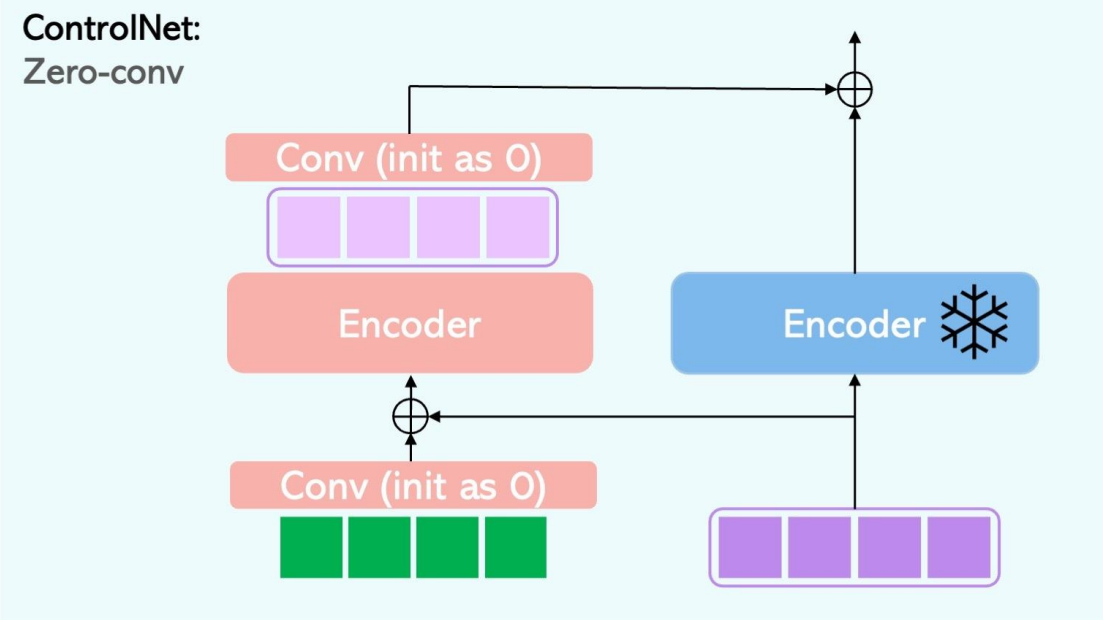
GLIGEN: Open-Set Grounded Text-to-Image Generation



Li et al., "GLIGEN: Open-Set Grounded Text-to-Image Generation", CVPR 2023

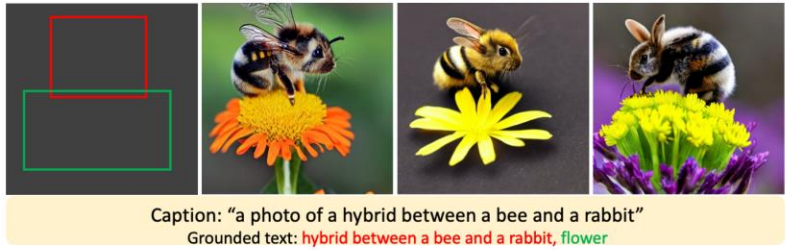
GLIGEN: Open-Set Grounded Text-to-Image Generation





Li et al., "GLIGEN: Open-Set Grounded Text-to-Image Generation", CVPR 2023

GLIGEN: Open-Set Grounded Text-to-Image Generation

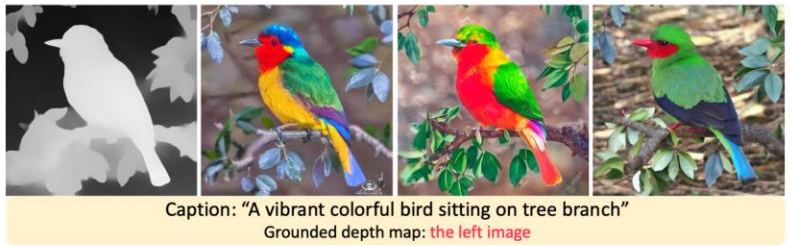


Caption: "a photo of a hybrid between a bee and a rabbit"
Grounded text: hybrid between a bee and a rabbit, flower

Caption: "Pikachu is under a tree, digital art"
Grounded text: Pikachu, tree; Grounded style image: blue inset

Caption: "A dog / bird / helmet / backpack is on the grass"
Grounded image: red inset

Caption: "superman / monkey / Hormer Simpson / is scratching its head"
Grounded keypoints: plotted dots on the left image



Caption: "A vibrant colorful bird sitting on tree branch"
Grounded depth map: the left image

Caption: "The beautiful scenery of a clam village near the sea"
Grounded HED map: the left image

Caption: "Cars park on the snowy street"
Grounded normal map: the left image

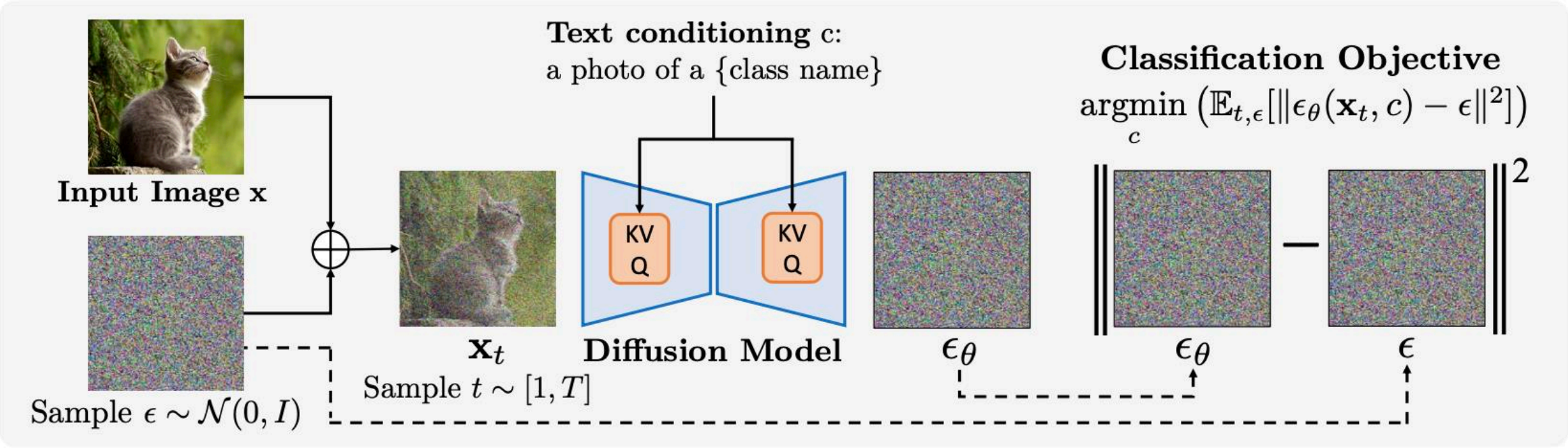
Caption: "A living room filled with lots of furniture and plants"
Grounded semantic map: the left image

Diverse form of input

- Box+Text
- Box+Text+Image
- Keypoint
- Box+Text
- Box+Text+Image
- Hed map
- Canny map
- Depth map
- Semantic map
- Normal map

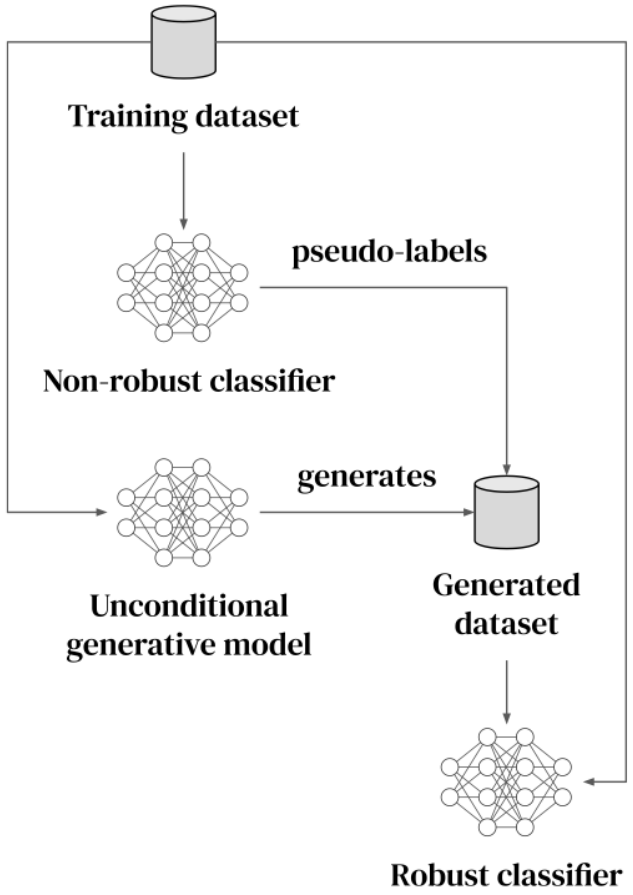
Other applications

Your Diffusion Model is Secretly a Zero-Shot Classifier



Li et al., "Your Diffusion Model is Secretly a Zero-Shot Classifier", arXiv 2023

Improving Robustness using Generated Data

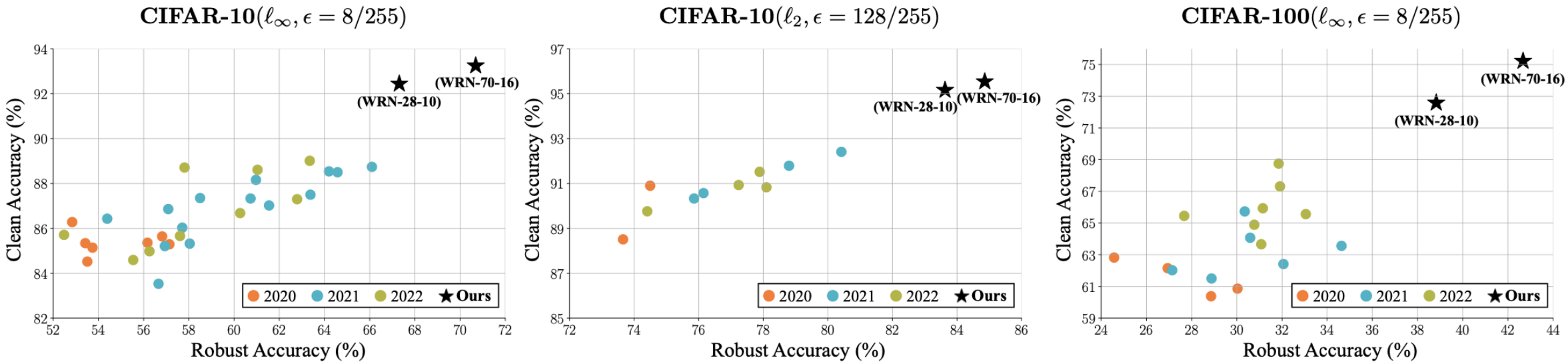


Overview of the approach:

1. train a generative model and a non-robust classifier, which are used to provide pseudo-labels to the generated data.
2. The generated and original training data are combined to train a robust classifier.

Gowal et al., "Improving Robustness using Generated Data", NeurIPS 2021

Better Diffusion Models Further Improve Adversarial Training



Wang et al., "Better Diffusion Models Further Improve Adversarial Training", ICML 2023

Acknowledgement

- Thanks everyone for suggesting the papers. Topics of the tutorial are based on *https://github.com/cvpr2023-tutorial-diffusion-models/papers*
- Part of the slides are adapted from Jun-Yan Zhu's talk on image customization.

Reference

- Bao et al., "All are Worth Words: a ViT Backbone for Score-based Diffusion Models", arXiv 2022
- Peebles and Xie, "Scalable Diffusion Models with Transformers", arXiv 2022
- Bao et al., "One Transformer Fits All Distributions in Multi-Modal Diffusion at Scale", arXiv 2023
- Jabri et al., "Scalable Adaptive Computation for Iterative Generation", arXiv 2022
- Hoogeboom et al., "simple diffusion: End-to-end diffusion for high resolution images", arXiv 2023
- Meng et al., "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations", ICLR 2022
- Li et al., "Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models", NeurIPS 2022
- Avrahami et al., "Blended Diffusion for Text-driven Editing of Natural Images", CVPR 2022
- Hertz et al., "Prompt-to-Prompt Image Editing with Cross-Attention Control", ICLR 2023
- Kawar et al., "Imagic: Text-Based Real Image Editing with Diffusion Models", CVPR 2023
- Couairon et al., "DiffEdit: Diffusion-based semantic image editing with mask guidance", ICLR 2023

- Sarukkai et al., "Collage Diffusion", arXiv 2023
- Bar-Tal et al., "MultiDiffusion: Fusing Diffusion Paths for Controlled Image Generation", ICML 2023
- Gal et al., "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion", ICLR 2023
- Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023
- Kumari et al., "Multi-Concept Customization of Text-to-Image Diffusion", CVPR 2023
- Tewel et al., "Key-Locked Rank One Editing for Text-to-Image Personalization", SIGGRAPH 2023
- Zhao et al., "A Recipe for Watermarking Diffusion Models", arXiv 2023
- Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models", ICLR 2022
- Li et al., "GLIGEN: Open-Set Grounded Text-to-Image Generation", CVPR 2023
- Avrahami et al., "SpaText: Spatio-Textual Representation for Controllable Image Generation", CVPR 2023
- Zhang and Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models", arXiv 2023

- Mou et al., "T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models", arXiv 2023
- Orgad et al., "Editing Implicit Assumptions in Text-to-Image Diffusion Models", arXiv 2023
- Han et al., "SVDiff: Compact Parameter Space for Diffusion Fine-Tuning", arXiv 2023
- Xie et al., "DiffFit: Unlocking Transferability of Large Diffusion Models via Simple Parameter-Efficient Fine-Tuning", arXiv 2023
- Saharia et al., "Palette: Image-to-Image Diffusion Models", SIGGRAPH 2022
- Whang et al., "Deblurring via Stochastic Refinement", CVPR 2022
- Xu et al., "Open-Vocabulary Panoptic Segmentation with Text-to-Image Diffusion Models", arXiv 2023
- Saxena et al., "Monocular Depth Estimation using Diffusion Models", arXiv 2023
- Li et al., "Your Diffusion Model is Secretly a Zero-Shot Classifier", arXiv 2023
- Gowal et al., "Improving Robustness using Generated Data", NeurIPS 2021
- Wang et al., "Better Diffusion Models Further Improve Adversarial Training", ICML 2023

Thanks!