时光倒转万物生：扩散模型与AIGC

VALSE 2023 扩散模型讲习班

汇报人：李崇轩

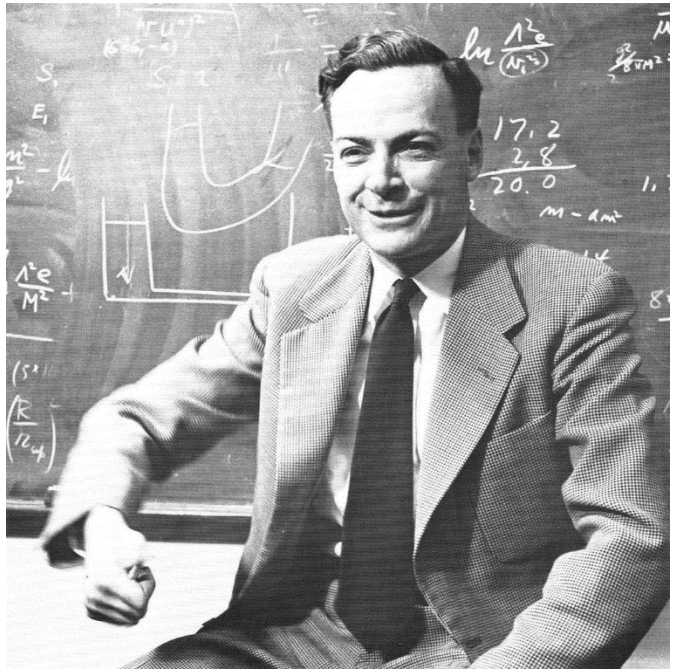
中国人民大学 高瓴人工智能学院

报告提纲

- 概览
- 扩散模型学习算法
- 扩散模型采样算法
- 大规模扩散模型
- 扩散模型与 **AIGC**
- 展望

扩散模型概览

生成模型：一种面向通用智能的技术



Richard Feynman: "*What I cannot create, I do not understand*"

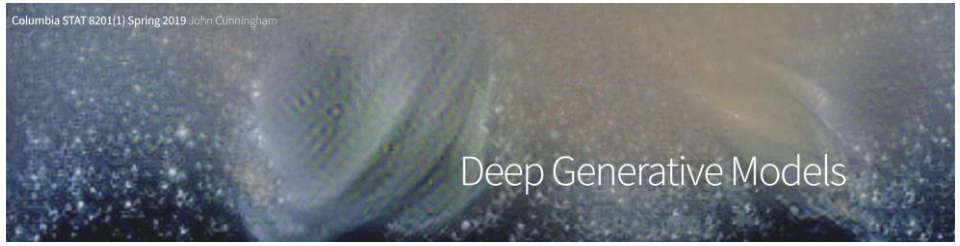
Generative modeling: "*What I understand, I can create*"

基础AI工具

Stanford

Deep Generative Models

CS236 - Fall 2021



Columbia

Cornell

Login

CS6785 › Syllabus

Spring 2021

CS 6785 Deep Probabilistic and Generative Models (2021SP)

UC Berkely

CS294-158-SP20

Deep Unsupervised Learning

活跃的研究领域

2022 Midjourney

深度生成模型的基本原理

- 核心问题：高维、复杂的联合概率分布的表示、学习与推断

经典图像数据：超过十万维的多峰分布

深度生成模型的基本原理

- 核心问题：高维、复杂的联合概率分布的表示、学习与推断

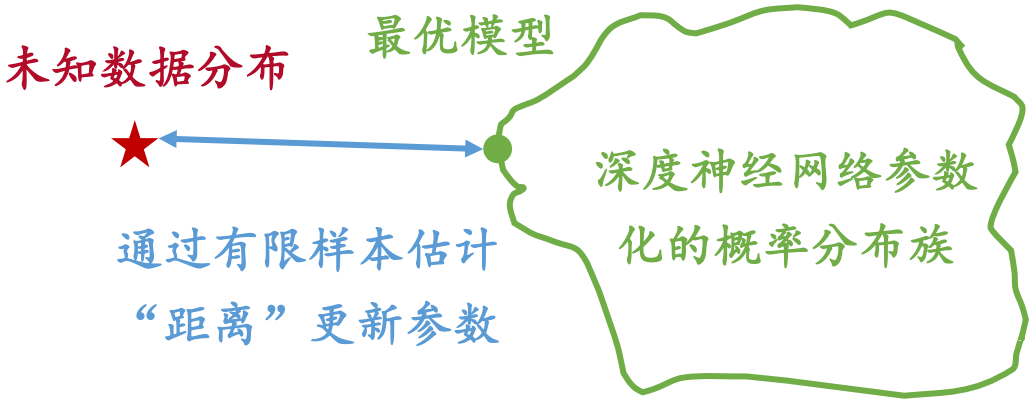
经典图像数据：超过十万维的多峰分布

深度神经网络参数化的概率分布族

深度生成模型的基本原理

- 核心问题：高维、复杂的联合概率分布的表示、学习与推断

经典图像数据：超过十万维的多峰分布



深度生成模型的基本原理

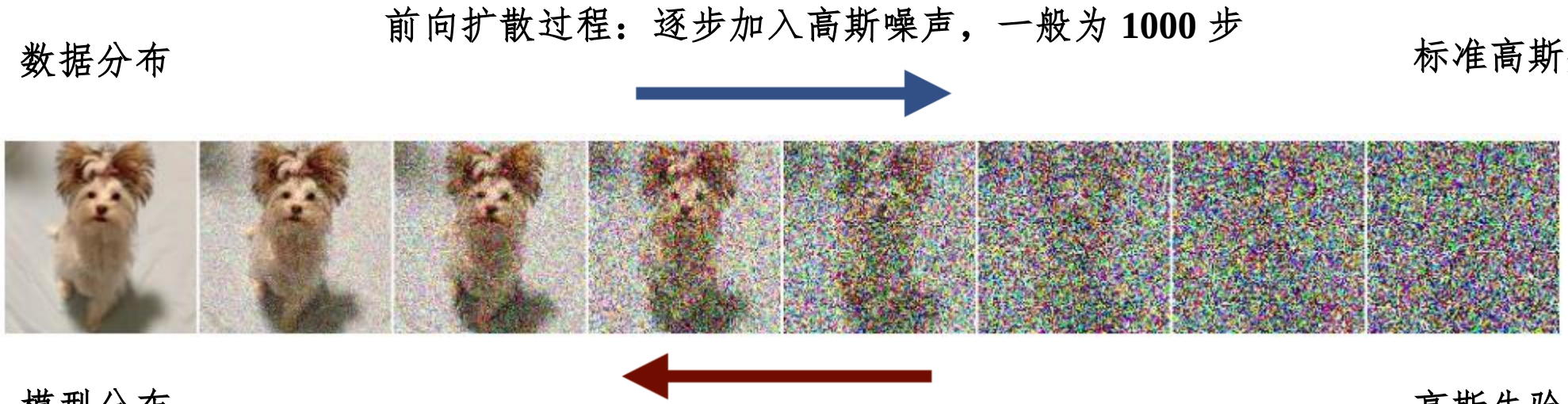
- 核心问题：高维、复杂的联合概率分布的表示、学习与推断

经典图像数据：超过十万维的多峰分布

基于最优模型的采样、似然估计等

扩散模型

- 理论：一定条件下，前向扩散过程对任意输入分布均可逆，且逆过程形式不变
- 直觉：将生成数据这一复杂问题转化为不同噪声级别下的去噪问题（相对简单）



扩散模型赋能 AIGC

Harvest of *vegetables* in a wooden box near the *beds* vegetables grow naturally, summer light background, backlight and *sun rays*, clean sharp focus.

Chinese illustration, *oriental landscape painting*, above super wide angle, magical, romantic, detailed, colorful, *multi-dimensional paper kirigami craft.*

Photography closeup portrait of an adorable *rusty* broken-down *steampunk robot* covered in budding vegetation, surrounded by tall grass, *misty futuristic sci-fi forest environment.*

A cute little matte *low poly* isometric Zelda Breath of the *wild forest island*, *waterfalls*, soft shadows, trending on Artstation, *3d render*, monument valley, fez video game.

The Goddess of high fashion, impressionistic *line art*, *contrasting* earth tones, vibrant, pen and *ink* illustration, ink splatter, *abstract* expressionism superimposed onto majestic space queen.

The *Caped Crusader*, Gotham *skyline,* rooftop, mysterious, powerful, *nighttime*, mixed media, expressionism, *dark tones*, high contrast, in the style of comic book artist *Frank Miller*, modern, gritty and textured, collage technique.

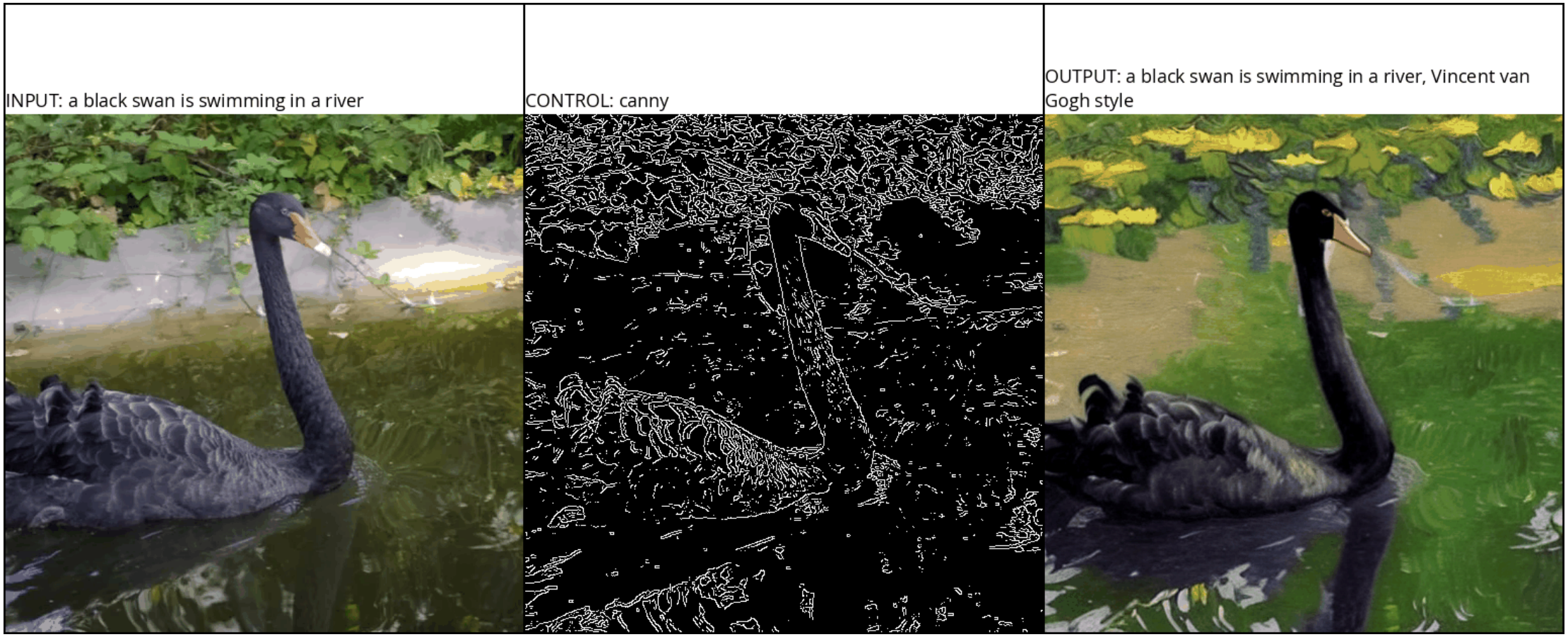
商汤大模型 RAPHAEL

输入文本描述主题“太空歌剧院”合成图像，获美国科罗拉多博览会的年度艺术比赛首奖

扩散模型赋能 AIGC

Stable Diffusion

扩散模型赋能 AIGC



Zhao M, Wang R, Bao F, et al. ControlVideo: Adding Conditional Control for One Shot Text-to-Video Editing[J]. arXiv preprint arXiv:2305.17098, 2023.

扩散模型赋能 AIGC

三维数字人 Rodin (Wang et al. CVPR 2023)

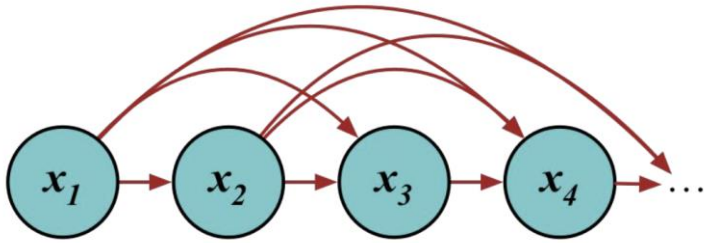
ProlificDreamer (Wang et al. Arxiv 2023)

扩散模型的基本原理

如何表示高维空间中的联合概率分布？

深度神经网络参数化的概率分布族

自回归



链式法则

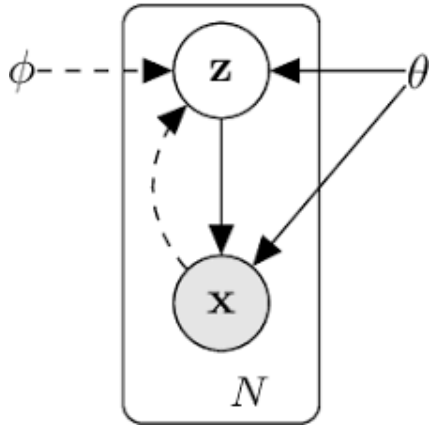
$$p_\theta(\mathbf{x}) = \prod_{i=1}^{d} p_\theta(\mathbf{x}_i \mid \mathbf{x}_{<i})$$

GPT

如何表示高维空间中的联合概率分布？

深度神经网络参数化的概率分布族

变分自编码器



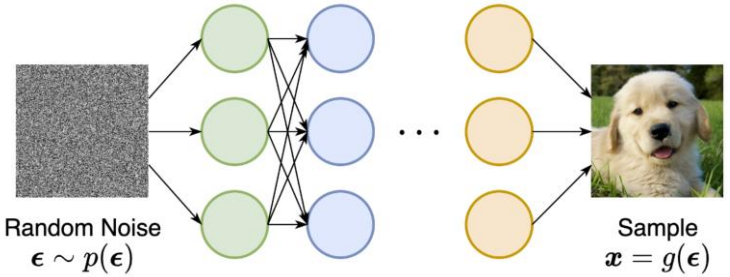
隐变量模型 $p_\theta(x) = \int p_\theta(x|z)p(z)dz$

DALL E

如何表示高维空间中的联合概率分布？

深度神经网络参数化的概率分布族

对抗网络



采样过程

$$\mathbf{z} \sim p(\mathbf{z})$$
$$\mathbf{x} = g_\theta(\mathbf{z})$$

Style-GAN

如何表示高维空间中的联合概率分布？

两种等价理解

1. 层次化隐变量模型（变分自编码器）
2. 多层次去噪评分匹配（基于评分函数的模型）

自回归 规整流 对抗网络 扩散模型

概率表示

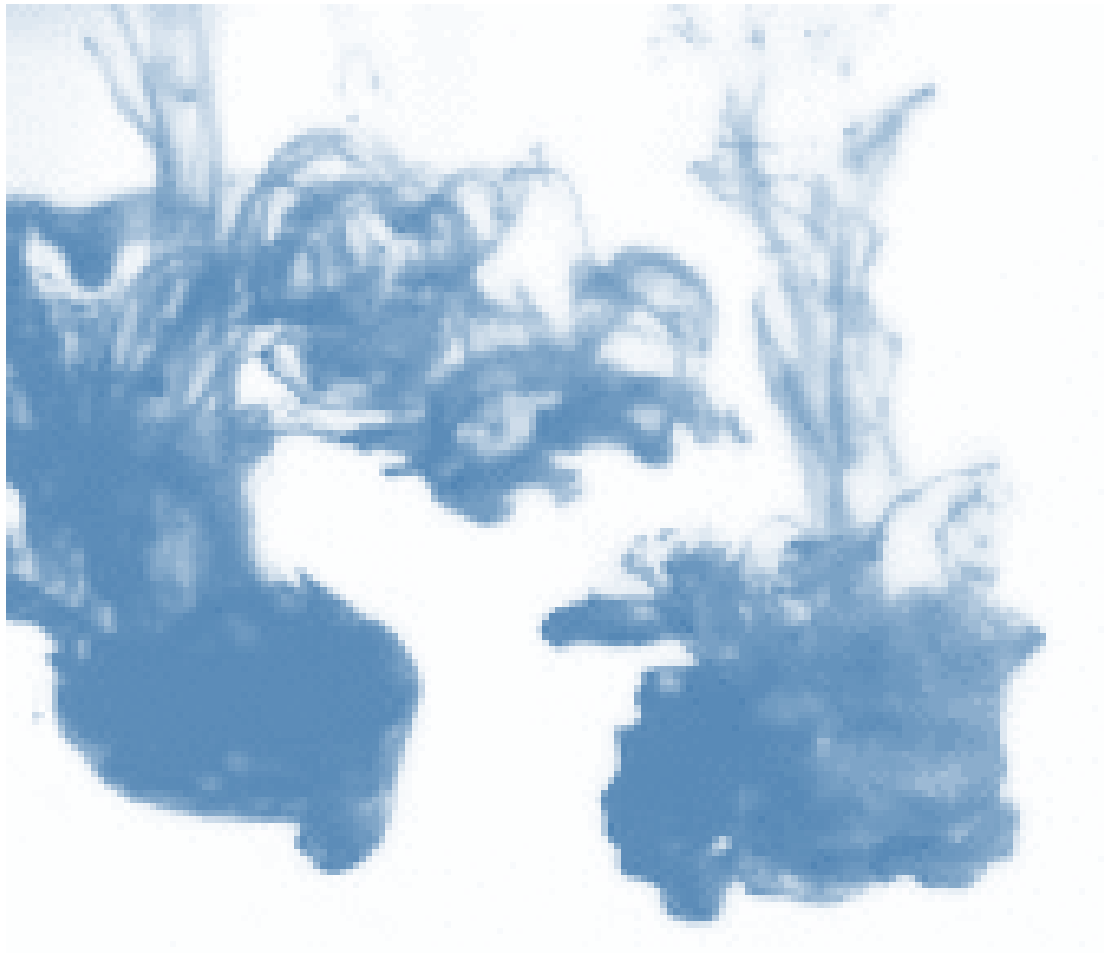
物理世界中的扩散

扩散过程随时间演化破坏结构

物理世界中的扩散

扩散过程随时间演化破坏结构

假如时光倒转？



扩散模型

Sohl-Dickstein et al, ICML 2015

前项链：高斯核马尔科夫链，一般 1000 步

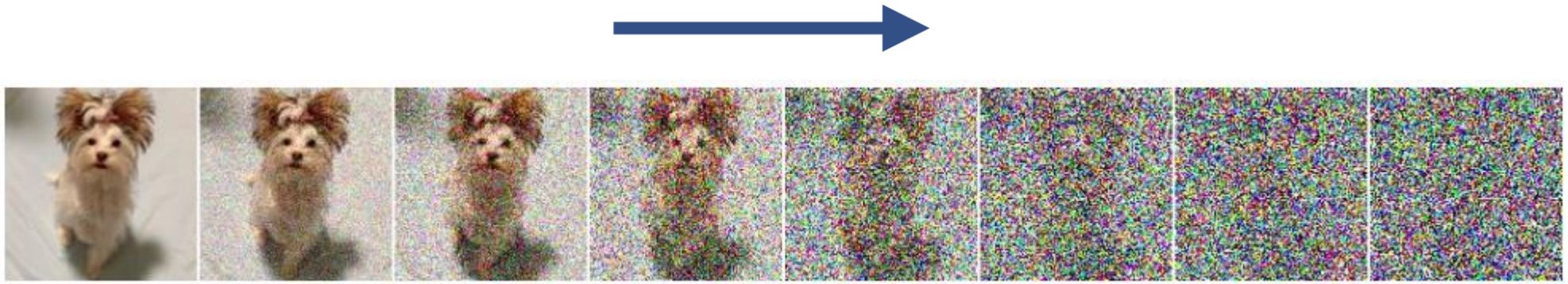
数据分布 $q\left(\mathbf{x}^{(0)}\right)$

加入噪声

高斯噪声 $q\left(\mathbf{x}^{(T)}\right) \approx \mathcal{N}\left(\mathbf{x}^{(T)}; 0, \mathbf{I}\right)$

前向核函数 $q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) = \mathcal{N}\left(\mathbf{x}^{(t)}; \mathbf{x}^{(t-1)}\sqrt{1-\beta_t}, \mathbf{I}\beta_t\right)$

扩散模型



核心思想：学习一个反向过程去噪

- 从任意数据分布出发，均可得到同样的高斯分布（表达能力强）
- 存在唯一的对应反向过程，同样是高斯核马尔科夫链（可学习）

扩散模型

可学习的反向高斯核马尔科夫链

模型分布

$$p\left(\mathbf{x}^{(0)}\right) \approx q\left(\mathbf{x}^{(0)}\right)$$

学习去噪

高斯先验分布

$$p\left(\mathbf{x}^{(T)}\right) = \mathcal{N}\left(\mathbf{x}^{(T)}; 0, \mathbf{I}\right)$$

参数化高斯的均值和方差

$$p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right) = \mathcal{N}\left(\mathbf{x}^{(t-1)}; f_\mu\left(\mathbf{x}^{(t)}, t\right), f_\Sigma\left(\mathbf{x}^{(t)}, t\right)\right)$$

理解一：层次化隐变量模型，定义了概率密度 $p(x_0) = \int p(x_0|x_1)p(x_1|x_2)\dots p(x_{T-1}|x_T)p(x_T)dx_{1\dots T}$

模型训练

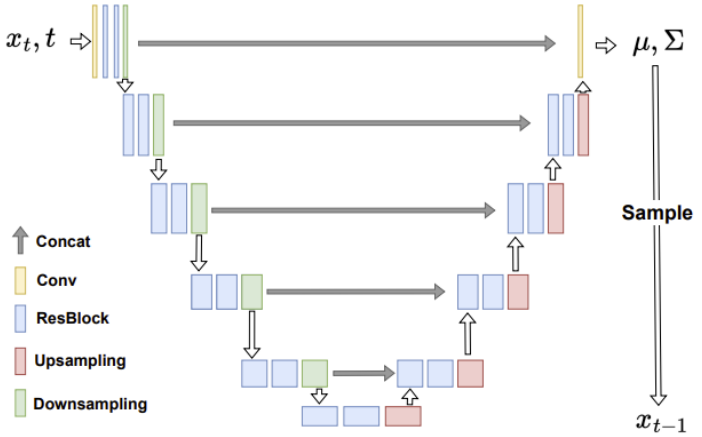
Jonathon et al., NeurIPS 2021

理解二：手工设置方差，做最大似然估计，等价于同时处理 **1000** 个不同层级的去噪任务

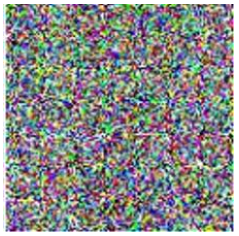
$$\mathrm{E}_{p_D(x_0),\epsilon}\,\mathrm{E}_{t\sim U[1,2,3\ldots,T]}\|\epsilon_\theta(x_t,t)-\epsilon\|^2$$

随机噪声大小　噪声预测网络　高斯噪声

带噪图片，噪声大小与采样的 t 有关

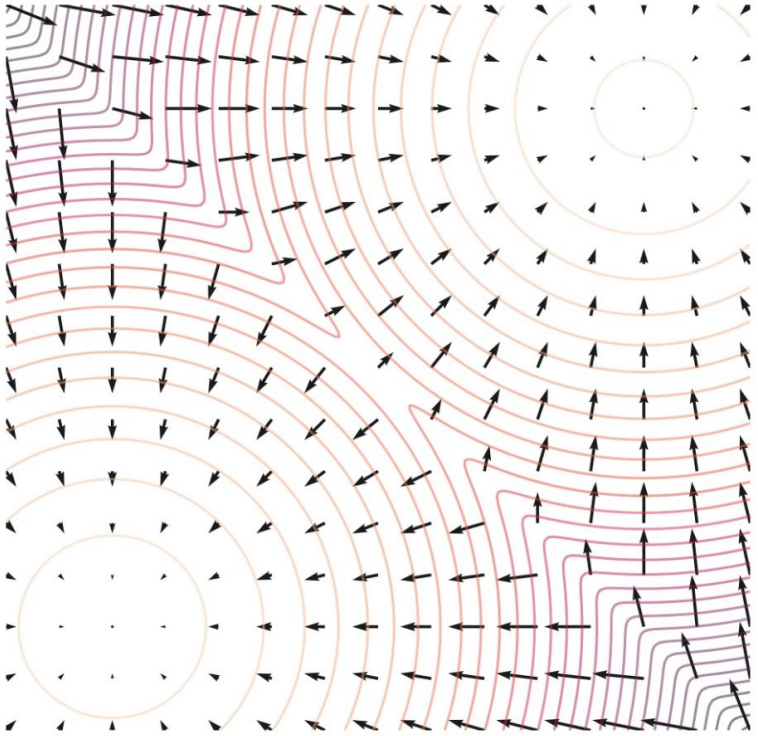


接收带噪图片和时间，预测噪声



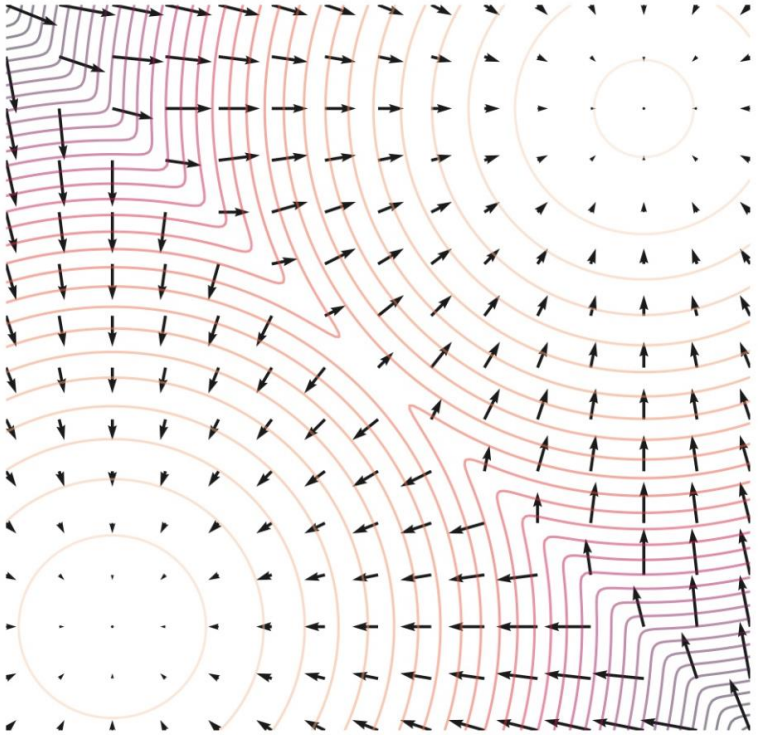
高斯噪声

评分函数



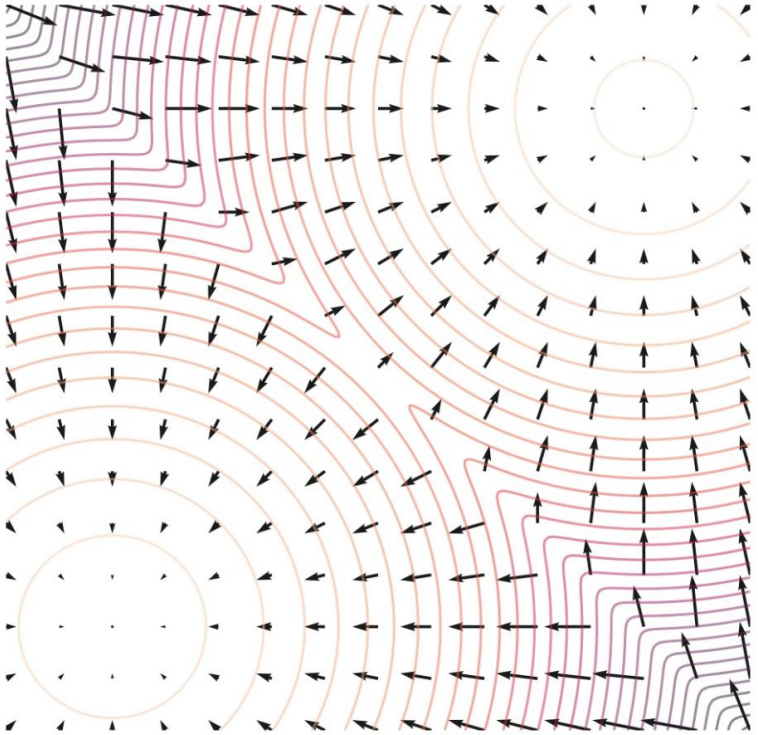
概率密度（圆环）：样本出现的可能性

评分函数

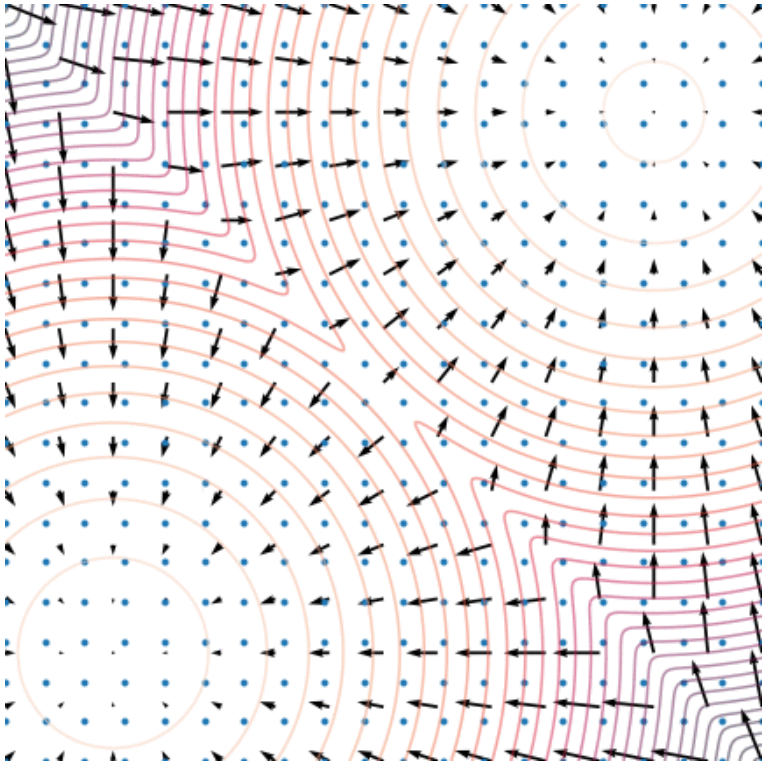


评分函数 $\nabla_x \log p(x)$（箭头），指向局部增大概率密度的方向

评分函数

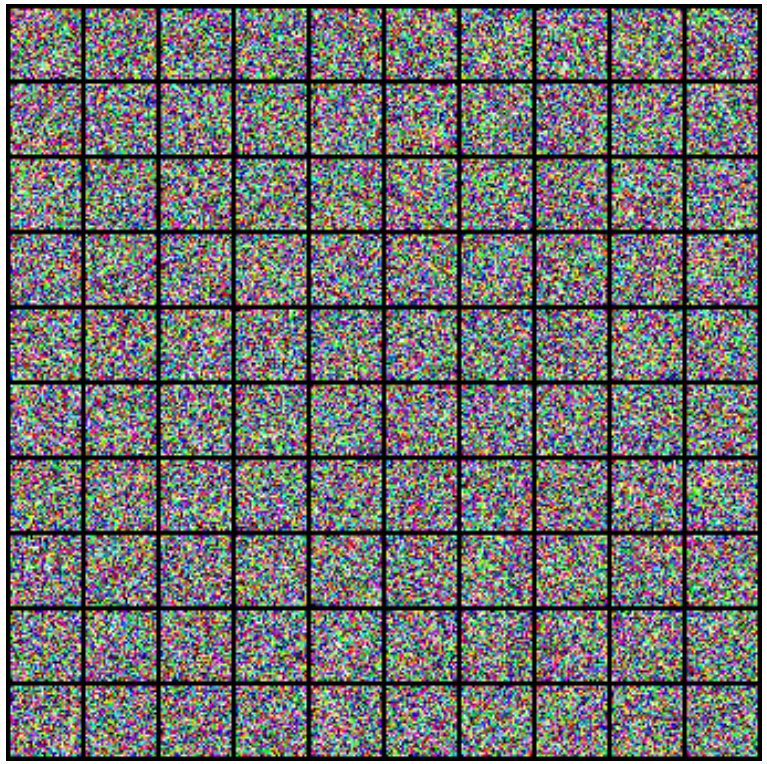


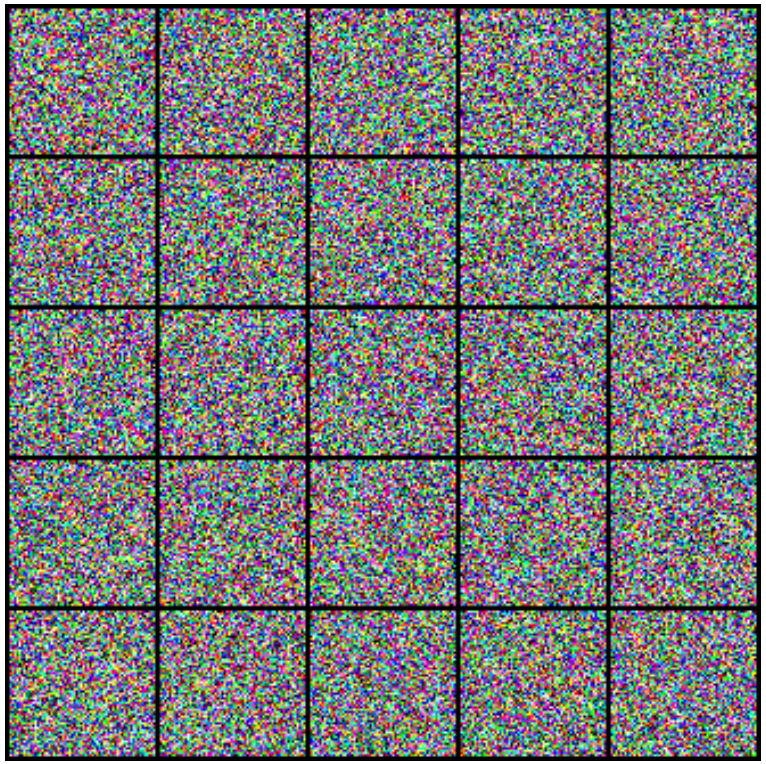
评分函数 $\nabla_x \log p(x)$（箭头），指向局部增大概率密度的方向



LD MCMC：按照评分函数方向更新并注入合适的噪声，可以采样

从扩散模型中采样



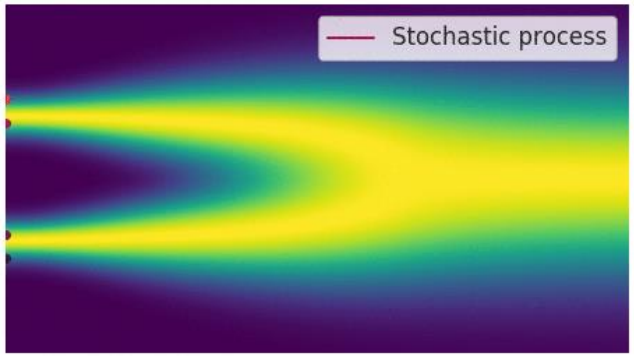


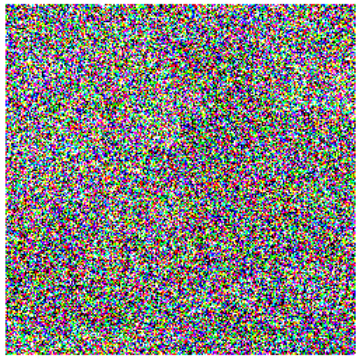
隐变量模型的祖先采样 或 退火郎之万动力学

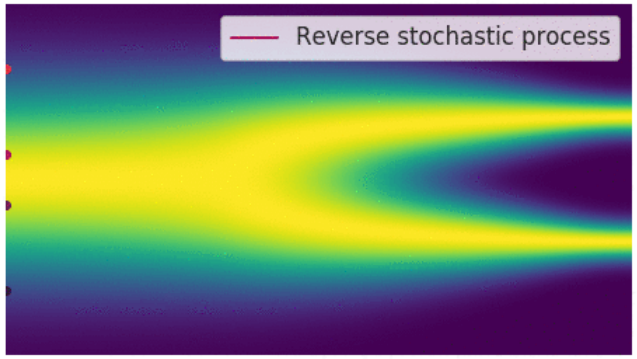
随机微分方程视角

Song et al, ICLR 2021

$$\mathbf{x}_i = \sqrt{1-\beta_i}\mathbf{x}_{i-1} + \sqrt{\beta_i}\mathbf{z}_{i-1} \quad \Rightarrow \quad \mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$$



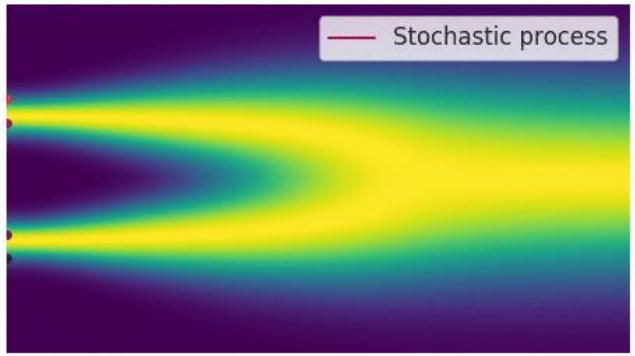


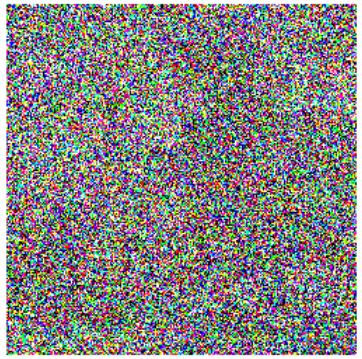


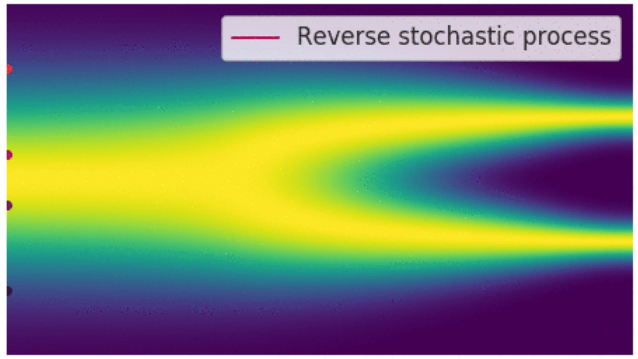
随机微分方程视角：连续时间训练

Song et al, ICLR 2021

$\mathrm{E}_{p_D(x_0),\epsilon}\,\mathrm{E}_{t\sim U[1,2,3\ldots,T]}\|\epsilon_\epsilon$ ➡ $\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$





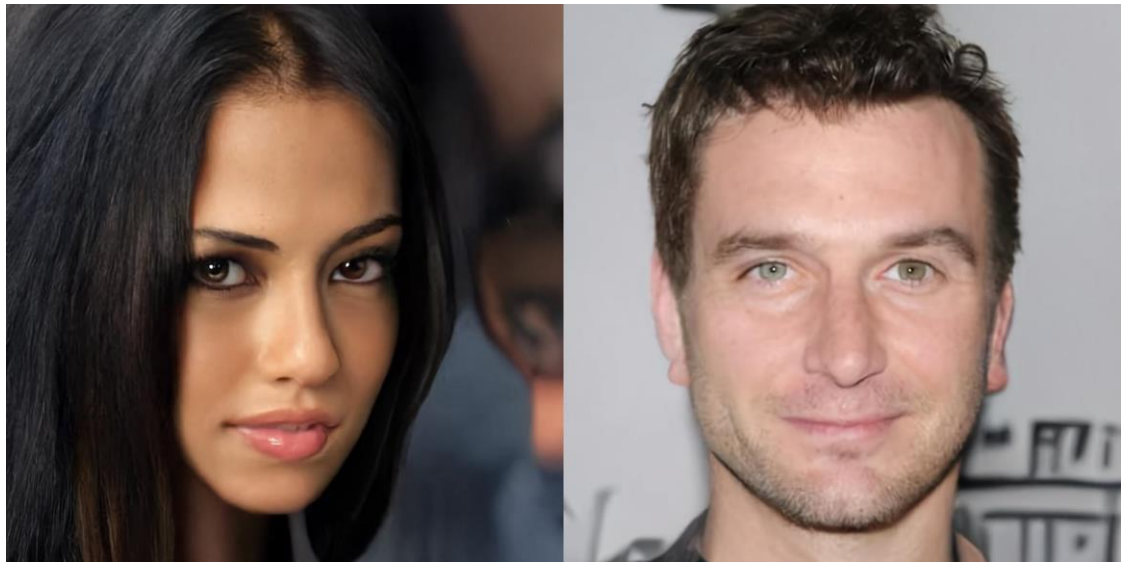


离散时间 $\mathrm{E}_{p_D(x_0),\epsilon}\,\mathrm{E}_{t\sim U[1,2,3\ldots,T]}\|\epsilon_\theta(x_t,t)-\epsilon\|^2$

连续时间 $\boxed{\frac{1}{2}\int_0^T}\omega(t)\mathbb{E}_{q_0(\boldsymbol{x}_0)}\mathbb{E}_{q(\boldsymbol{\epsilon})}\left[\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t)-\boldsymbol{\epsilon}\|_2^2\right]\mathrm{d}t$

在多个指标下领先的结果

Song et al, ICLR 2021



1024 高清图片

Table 2: NLLs and FIDs (ODE) on CIFAR-10.

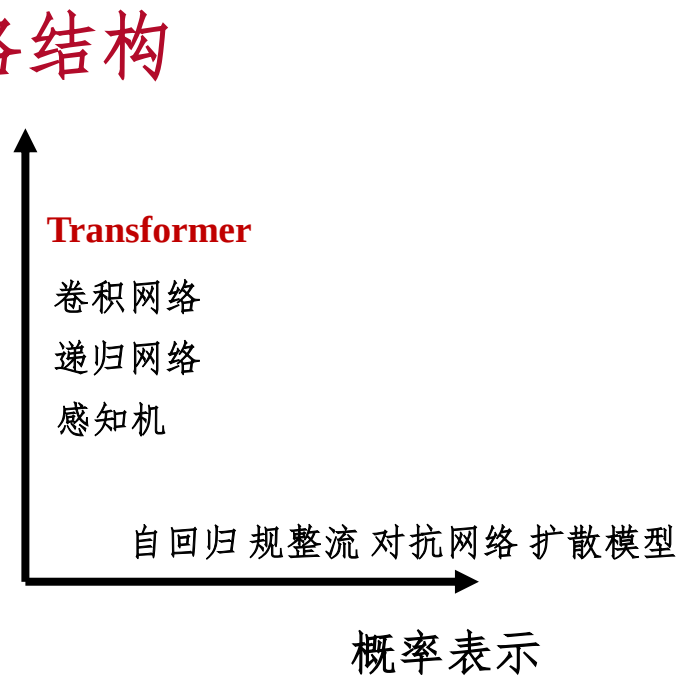
Model	NLL Test ↓	FID ↓
RealNVP (Dinh et al., 2016)	3.49	-
iResNet (Behrmann et al., 2019)	3.45	-
Glow (Kingma & Dhariwal, 2018)	3.35	-
MintNet (Song et al., 2019b)	3.32	-
Residual Flow (Chen et al., 2019)	3.28	46.37
FFJORD (Grathwohl et al., 2018)	3.40	-
Flow++ (Ho et al., 2019)	3.29	-
DDPM (L) (Ho et al., 2020)	$\leqslant 3.70^{*}$	13.51
DDPM (L_{simple}) (Ho et al., 2020)	$\leqslant 3.75^{*}$	3.17
DDPM	3.28	3.37
DDPM cont. (VP)	3.21	3.69
DDPM cont. (sub-VP)	3.05	3.56
DDPM++ cont. (VP)	3.16	3.93
DDPM++ cont. (sub-VP)	3.02	3.16
DDPM++ cont. (deep, VP)	3.13	3.08
DDPM++ cont. (deep, sub-VP)	**2.99**	**2.92**

图像数据上终结了不同生成模型处理不同任务的时代

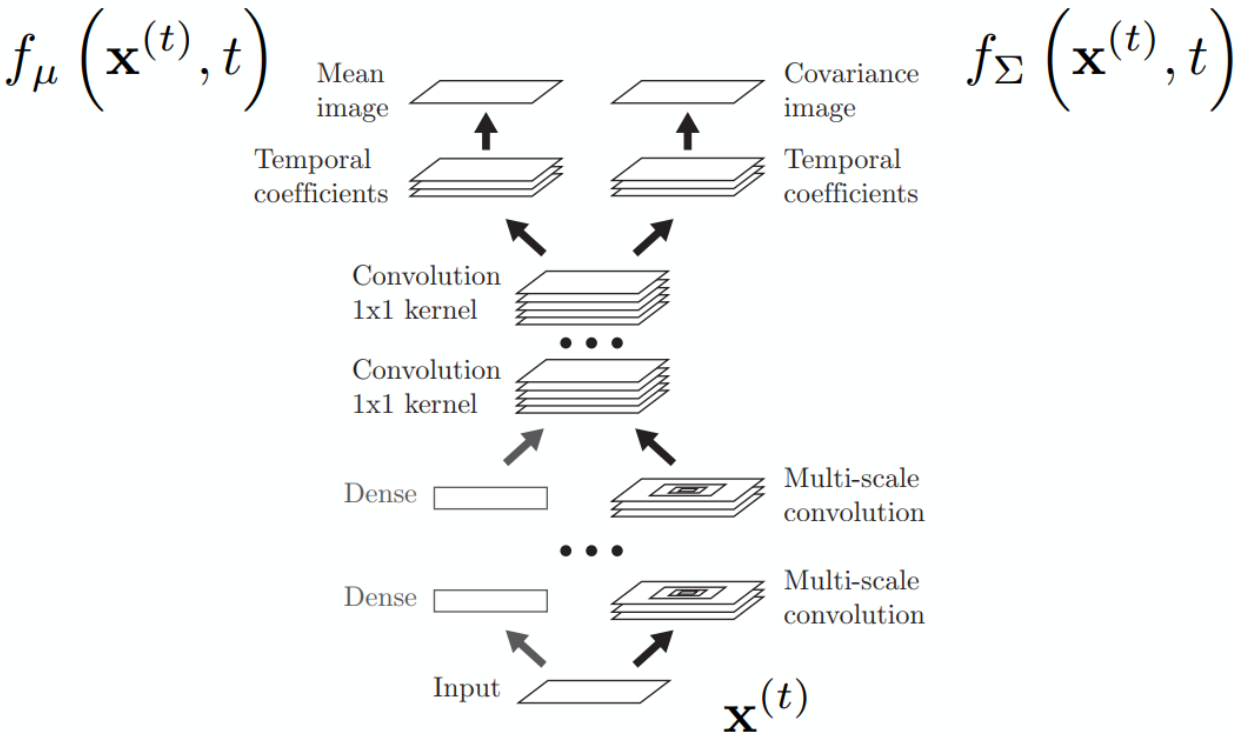
网络结构

网络结构设计是概率分布表示的一个重要维度

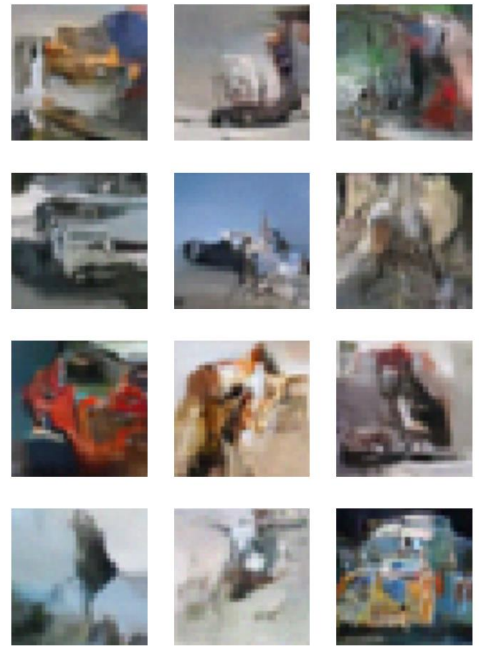
深度神经网络参数化的概率分布族



结果深受网络结构的影响

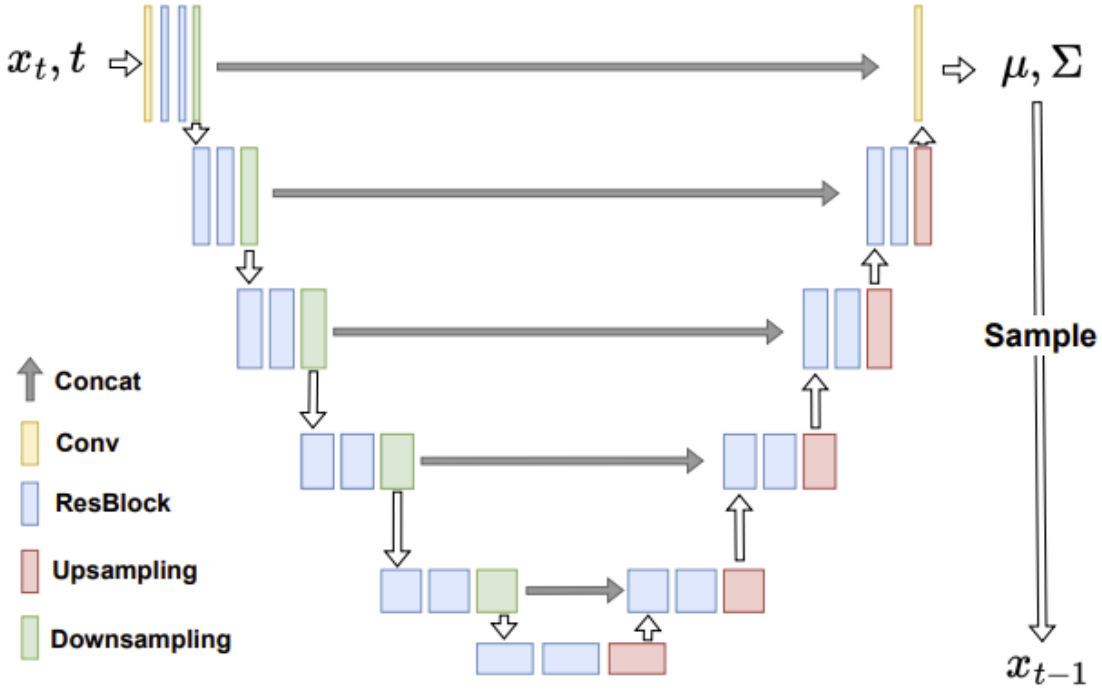


2015 多尺度多通路的卷积网络

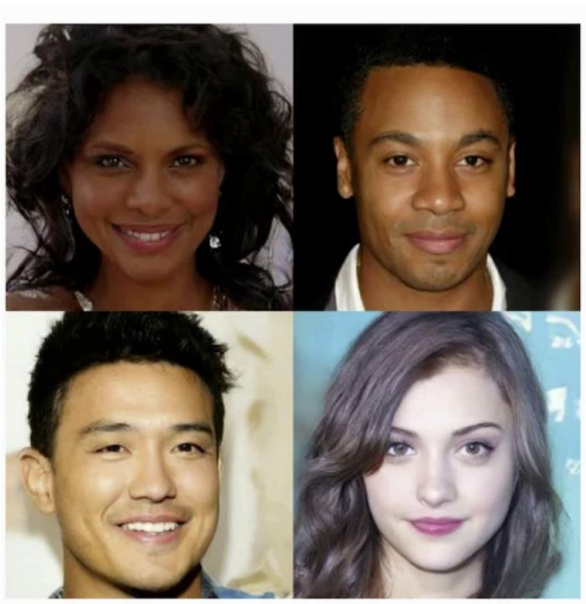


生成结果

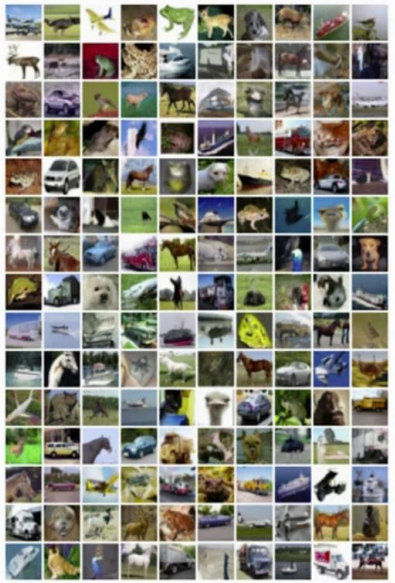
结果深受网络结构的影响



U-Net 网络 2021 年之后变成主流



CelebA-HQ 256x256

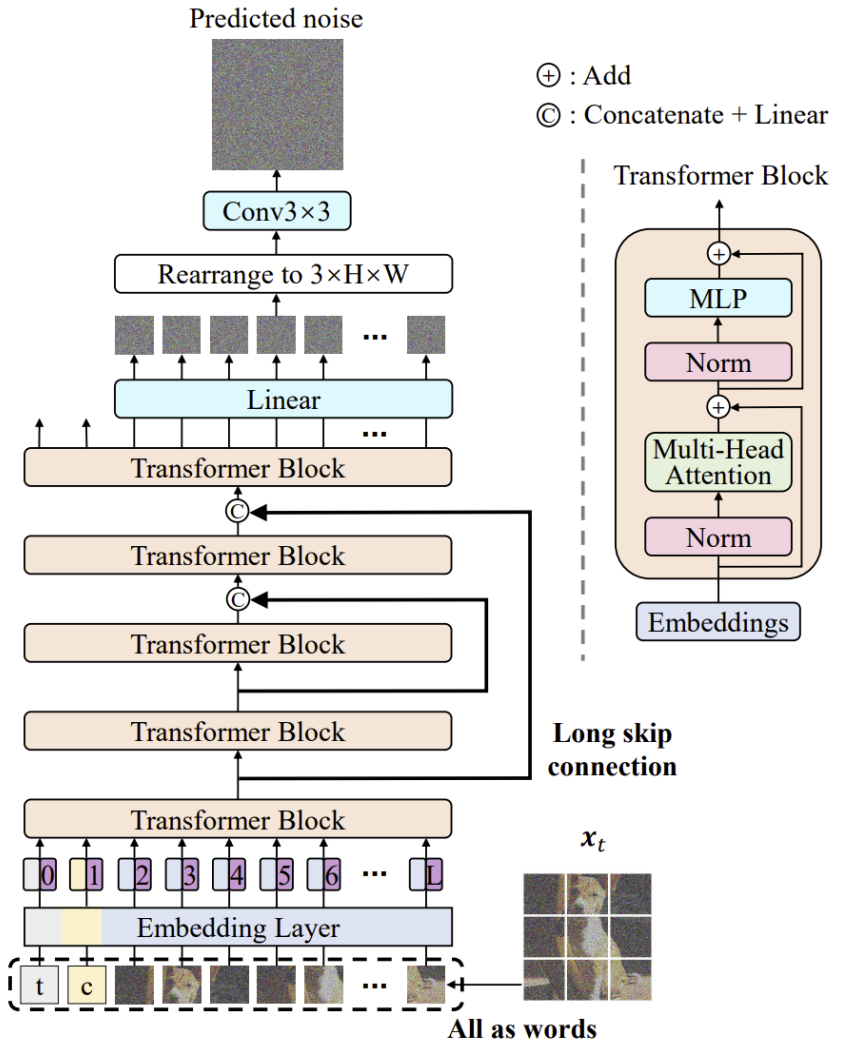


CIFAR-10 FID = 3.17 (SOTA)

生成结果

U-ViT

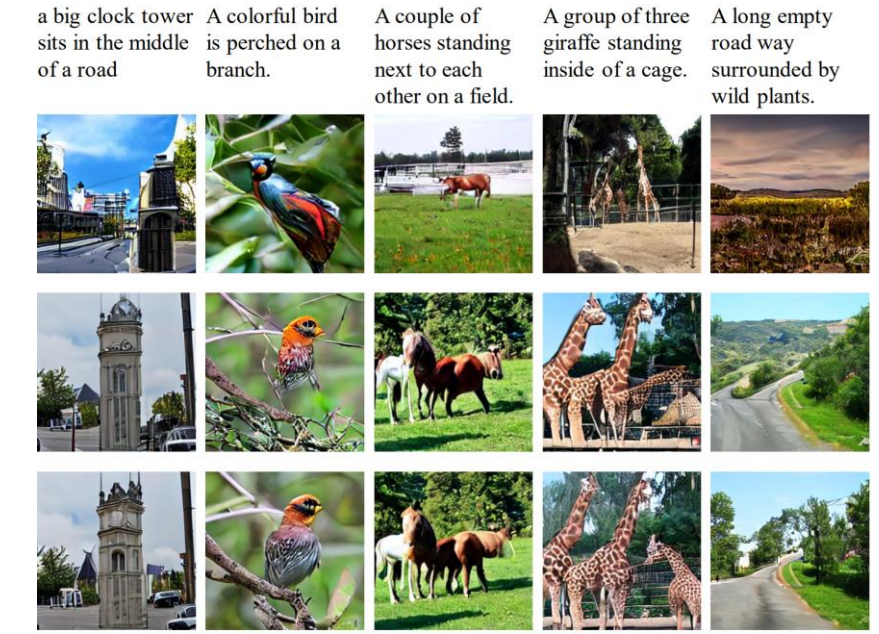
- 基于 **Transformer** 的新结构 **U-ViT**
- 探究 **U-Net** 哪些模块是必要/不必要的
- 构造一个简单易扩展的结构
  - 方便大规模训练
  - 方便跨模态训练



Bao F et al. All are Worth Words: A ViT Backbone for Diffusion Models. CVPR 2023.

通用多模态扩散模型：网络结构

在 **512** 高清图像生成任务上超越 **U-Net**

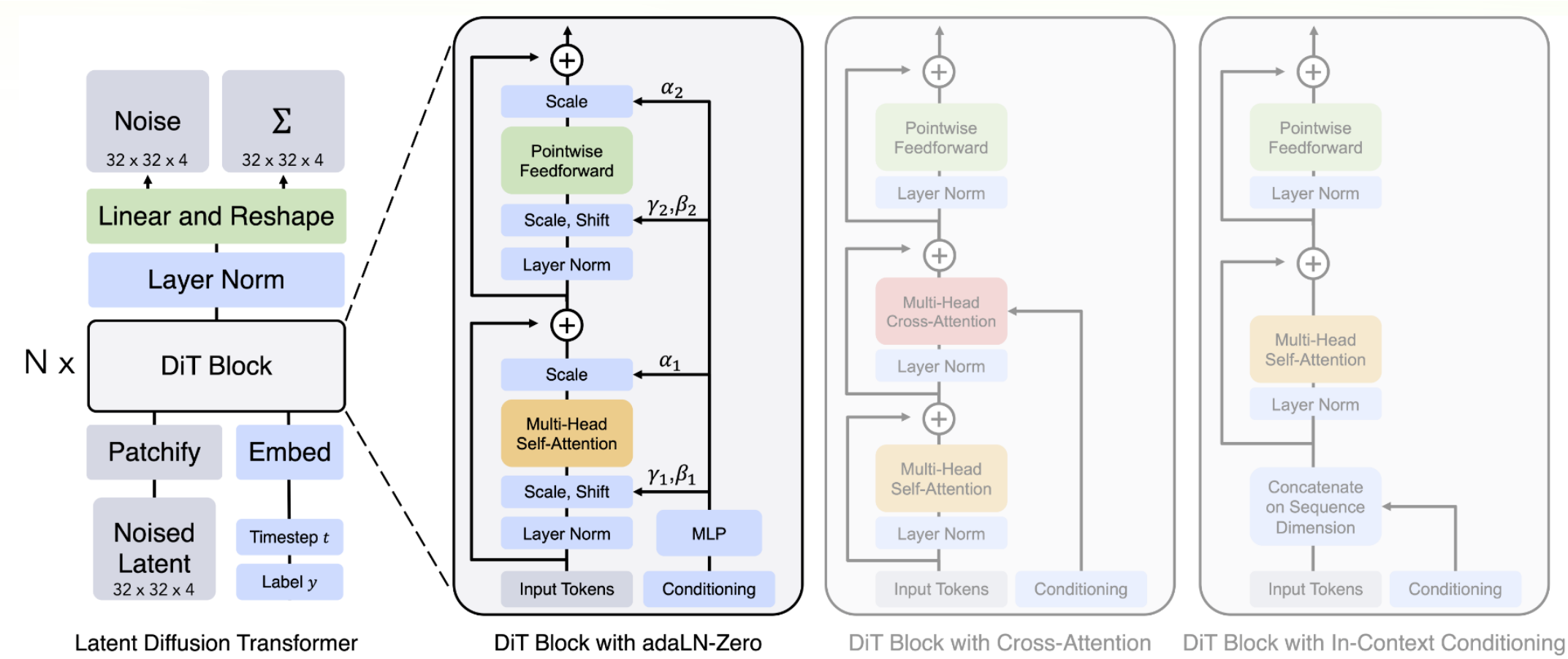


MS-COCO 文到图领先的生成结果

Bao F et al. All are Worth Words: A ViT Backbone for Diffusion Models. CVPR 2023.

DiT

Peebles and Xie, CVPR 2023



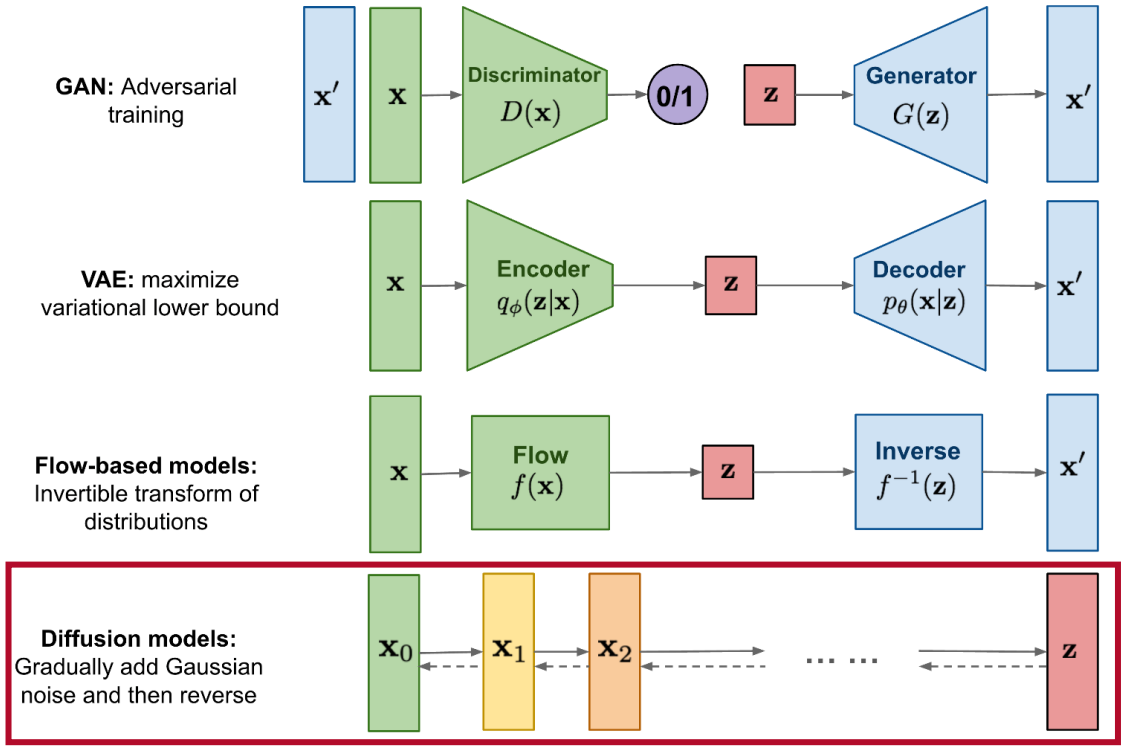
精心设计了条件融合的方式，并探究了模型大小对性能的影响

结果

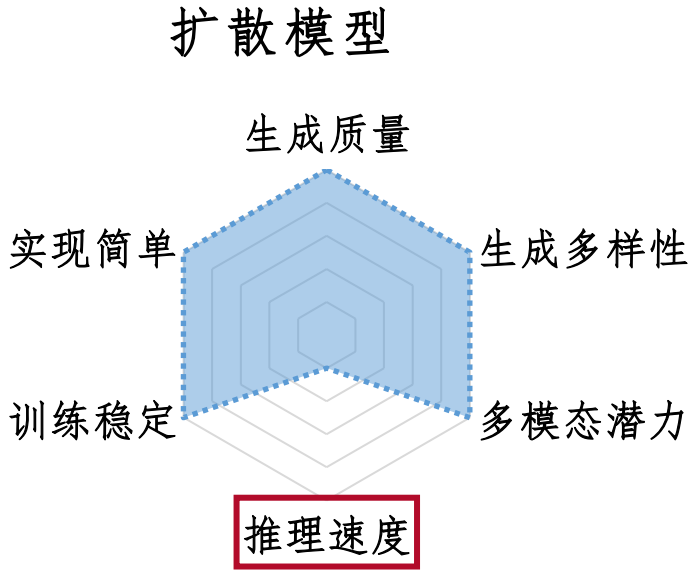
同参数量下，和 **U-ViT** 几乎完全一样的定量、定性结果；专注于图像生成本身

扩散模型的高效采样

扩散模型的瓶颈问题：采样效率低



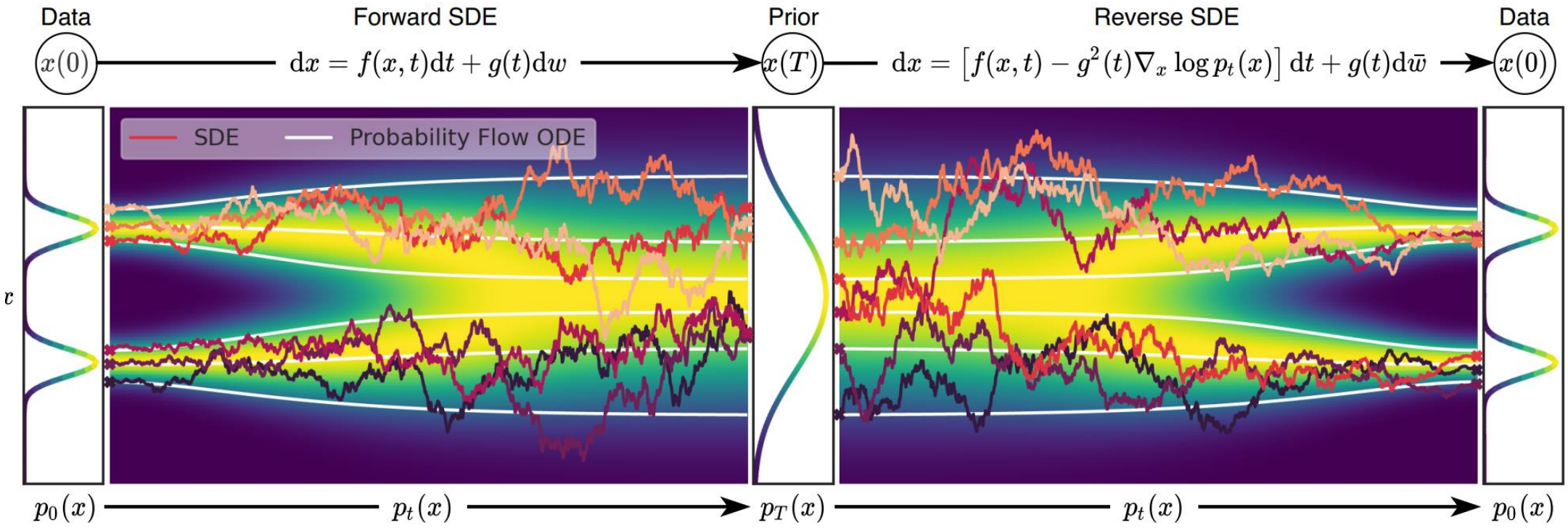
原则上需要迭代1000次，每次需调用神经网络



“五边形战士”：采样效率是瓶颈问题

扩散模型采样等价于微分方程的离散化

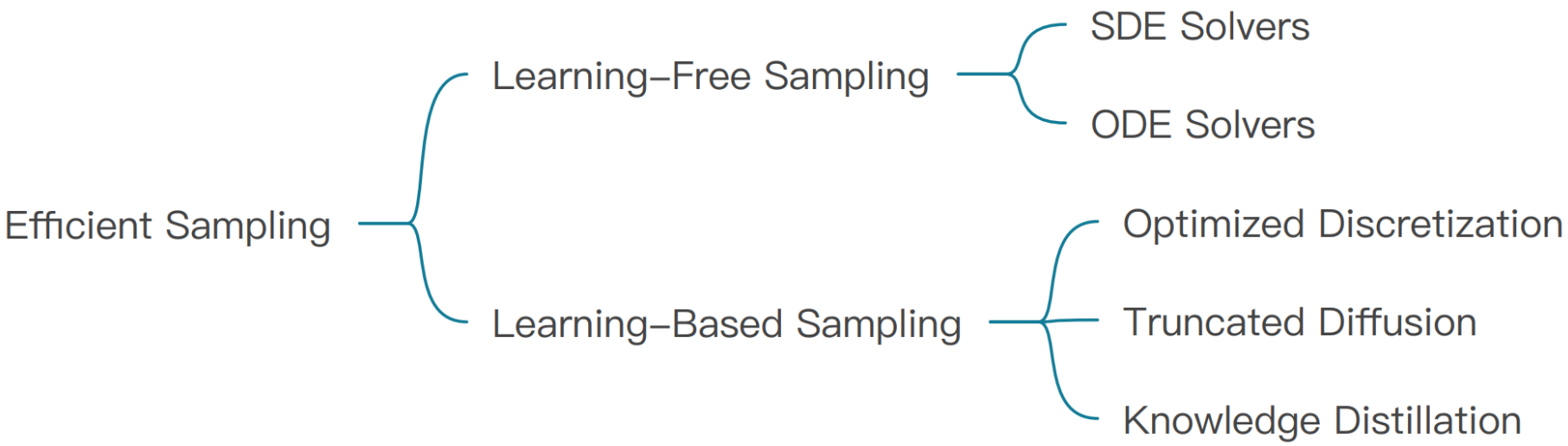
Song et al, ICLR 2021



扩散模型的采样是随机微分方程/等价常微分方程的离散化

高效采样

Review of diffusion models: Yang et al, arxiv 2022

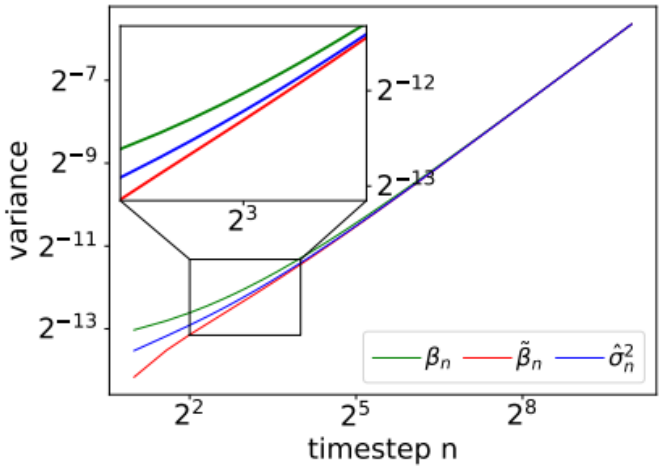


扩散模型的最优采样方差理论（隐变量模型视角出发）

- 证明最大似然意义下最优采样方差闭式解，改变了手工设计方差的范式
- 发表于机器学习领域旗舰国际会议 **ICLR 2022**，获杰出论文奖（接收率 **0.15%**）

定理：扩散概率模型在最大似然意义下关于评分函数/去噪函数的最优采样方差闭式解如下：

$$\sigma_t^{*2} = \frac{\beta_t}{1-\beta_t}\left(1 - \beta_t \mathrm{E}_{q_t(x_t)} \frac{\|\nabla \log q_t(x_t)\|^2}{d}\right).$$

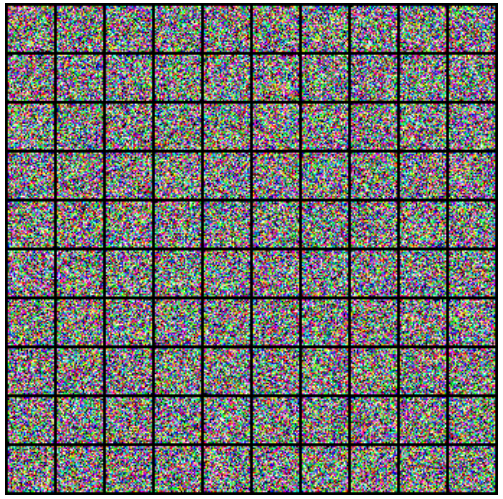


最优方差（蓝色）与手工方差在零时刻（即数据分布）附近有显著区别

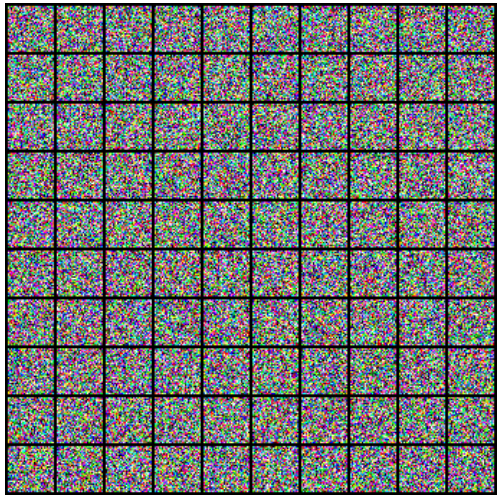
Bao F, Li C, Zhu J, et al. Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. ICLR 2022.

Analytic-DPM

- 无需额外训练，保证合成样本质量不变，加速 **20-80** 倍
- 作为核心技术部署于 **OpenAI** 公司发布的领先文到图生成大模型 **DALLE·2**



经典方法 **1000** 步



所提方法 **50** 步

To obtain a full generative model of images, we combine the CLIP image embedding *decoder* with a *prior* model, which generates possible CLIP image embeddings from a given text caption. We compare our text-to-image system with other systems such as DALL-E [40] and GLIDE [35], finding that our samples are comparable in quality to GLIDE, but with greater diversity in our generations. We also develop methods for training diffusion priors in latent space, and show that they achieve comparable performance to autoregressive priors, while being more compute-efficient. We refer to our full text-conditional image generation stack as *unCLIP*, since it generates images by inverting the CLIP image encoder.

For the AR prior, we use a Transformer text encoder with width 2048 and 24 blocks and a decoder with a causal attention mask, width 1664, and 24 blocks. For the diffusion prior, we use a Transformer with width 2048 and 24 blocks, and sample with Analytic DPM [2] with 64 strided sampling steps. To reuse hyperparameters tuned for diffusion noise schedules on images from Dhariwal and Nichol [11], we scale the CLIP embedding inputs by 17.2 to match the empirical variance of RGB pixel values of ImageNet images scaled to $[-1, 1]$.

	AR prior	Diffusion prior	64	$64 \rightarrow 256$	$256 \rightarrow 1024$
Diffusion steps	-	1000	1000	1000	1000
Noise schedule	-	cosine	cosine	cosine	linear
Sampling steps	-	64	250	27	15
Sampling variance method	-	analytic [2]	learned [34]	DDIM [47]	DDIM [47]
Crop fraction	-	-	-	0.25	0.25
Model size	1B	1B	3.5B	700M	300M
Channels	-	-	512	320	192
Depth	-	-	3	3	2
Channels multiple	-	-	1,2,3,4	1,2,3,4	1,1,2,2,4,4
Heads channels	-	-	64	-	-
Attention resolution	-	-	32,16,8	-	-
Text encoder context	256	256	256	-	-
Text encoder width	2048	2048	2048	-	-
Text encoder depth	24	24	24	-	-
Text encoder heads	32	32	32	-	-

显著加速 **DALLE·2**

“a painting of a fox sitting in a field at sunrise in the style of Claude Monet”

扩散概率模型的常微分方程离散化

- 针对扩散概率模型半线性等特点，设计等价常微分方程的离散化解析形式
- 发表于机器学习领域旗舰国际会议 **NeurIPS 2022**，口头报告（接收率 **1.7%**）

经典龙格库塔法 $$\boldsymbol{x}_t = \boldsymbol{x}_s + \int_s^t \left(f(\tau)\boldsymbol{x}_\tau + \frac{g^2(\tau)}{2\sigma_\tau}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_\tau, \tau) \right) \mathrm{d}\tau$$ 整体黑盒泰勒展开并做差分近似

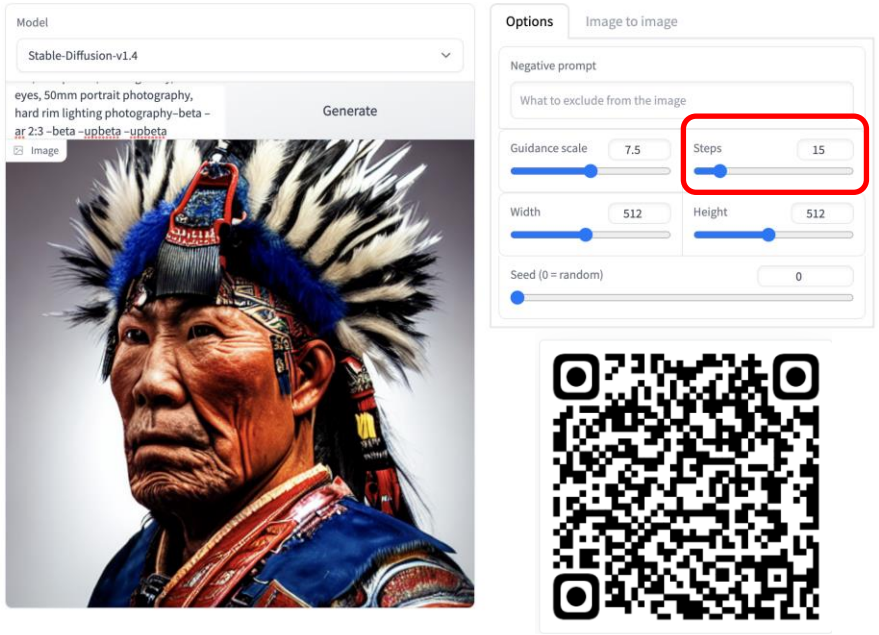
所提 **DPM-Solver** $$\boldsymbol{x}_{t_{i-1}\to t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}} - \alpha_{t_i}\sum_{n=0}^{k-1}\hat{\boldsymbol{\epsilon}}_\theta^{(n)}(\hat{\boldsymbol{x}}_{\lambda_{t_{i-1}}}, \lambda_{t_{i-1}})\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\frac{(\lambda-\lambda_{t_{i-1}})^n}{n!}\mathrm{d}\lambda + \mathcal{O}(h_i^{k+1})$$

解析形式　　神经网络部分差分近似　　解析形式　　高阶小量

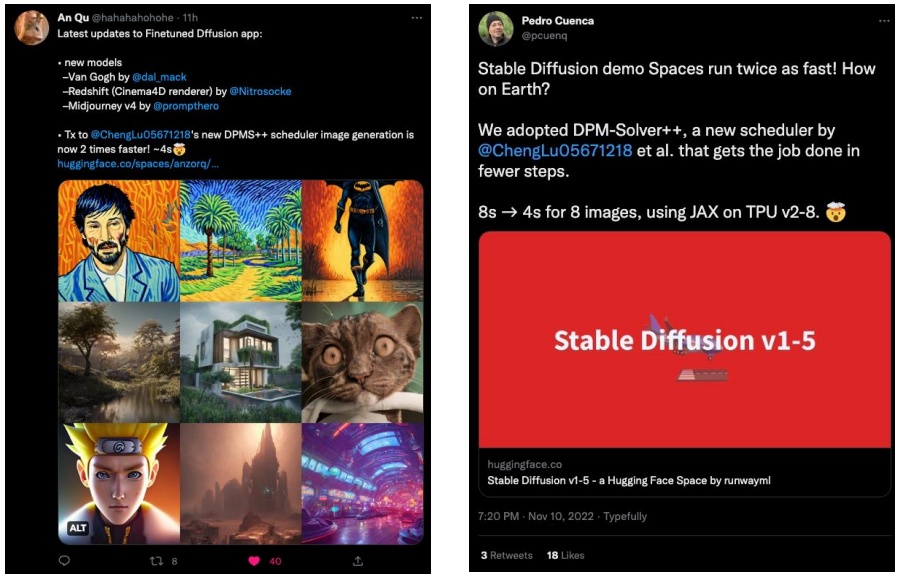
Lu C, et al. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. NeurIPS 2022.

DPM-Solver

- 是当前最快的无需额外学习的扩散概率模型采样算法，15 步生成高清图像
- 被多个主流开源社区（**Github** 累计星标 **6** 万余次）支持/设为默认算法

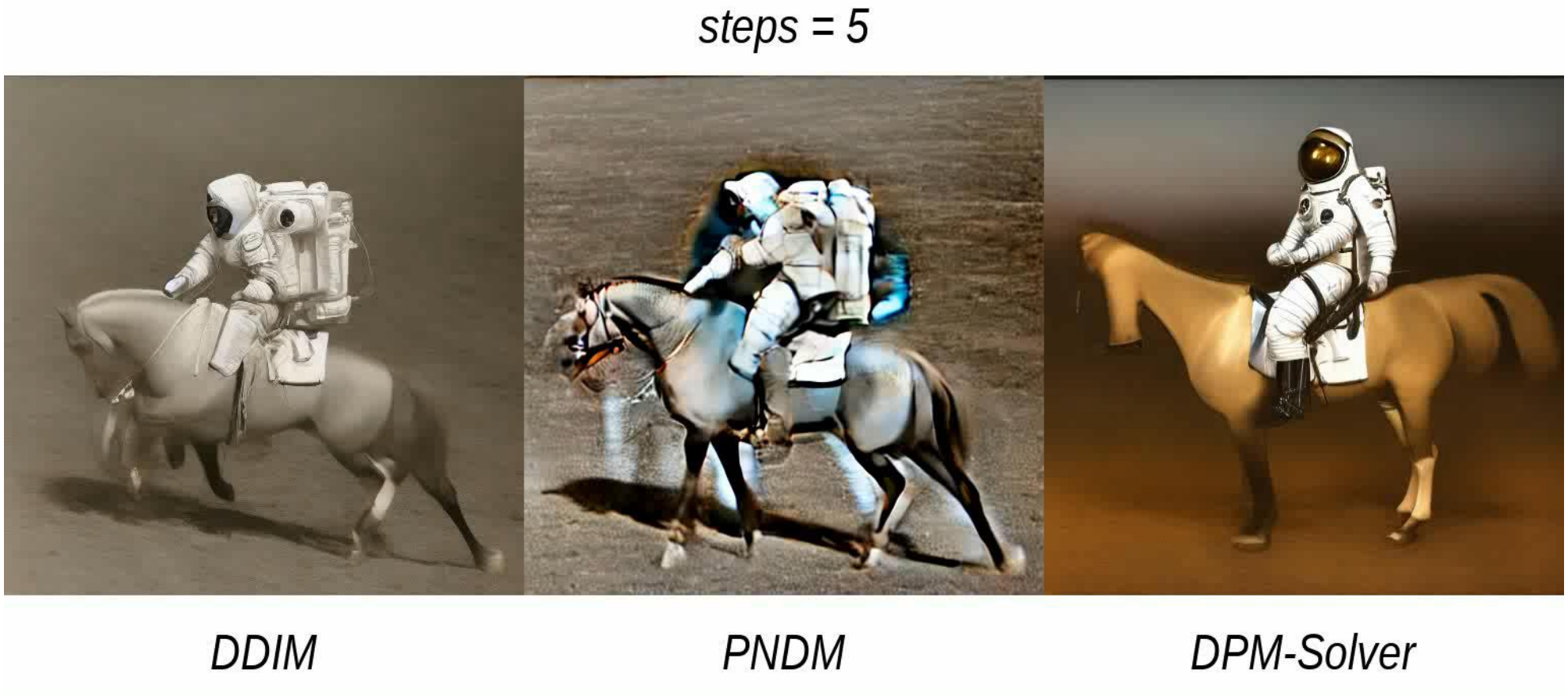


根据文本输入 **15** 步生成 **512x512** 高清图像



著名开源模型 **Stable Diffusion** 等官方宣传

扩散模型的加速采样方法 (Stable Diffusion 官方代码)



一行代码加速采样

Stable-Diffusion

The second-order multistep DPM-Solver++ is the default solver for Stable-Diffusion online demos (e.g., see example) and can also be used in LoRA (e.g., see example). Here is an example:

```
import torch
from diffusers import StableDiffusionPipeline, DPMSolverMultistepScheduler

model_id = "stabilityai/stable-diffusion-2-1"

# Use the DPMSolverMultistepScheduler (DPM-Solver++) scheduler here
pipe = StableDiffusionPipeline.from_pretrained(model_id, torch_dtype=torch.float16)
pipe.scheduler = DPMSolverMultistepScheduler.from_config(pipe.scheduler.config)
pipe = pipe.to("cuda")

prompt = "a photo of an astronaut riding a horse on mars"
image = pipe(prompt).images[0]

image.save("astronaut_rides_horse.png")
```

总结：扩散模型的表示、学习与推断

- 表示
  - 两种等价概率建模方式：隐变量模型 **vs.** 评分函数估计
  - 网络结构：卷积 **vs. Transformer**
- 学习
  - 噪声预测：离散化训练 **vs.** 连续化训练 （也有其他等价预测目标）
- 推断
  - 迭代去噪：随机微分方程离散化 **vs.** 常微分方程离散化
  - 加速推断：针对扩散模型对应微分方程的结构得到解析解

条件模型与指引函数

条件生成

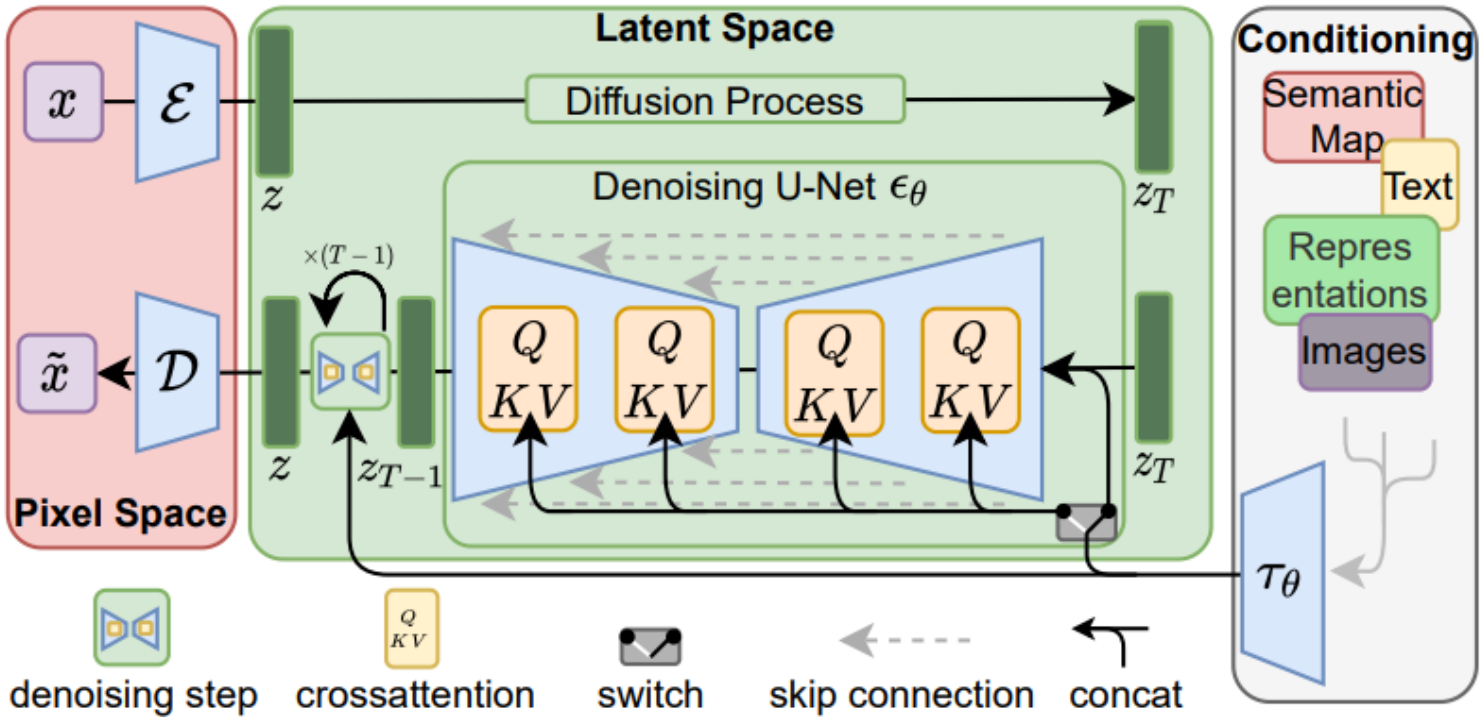
随机生成逼真的图像

给定控制条件（类别、文本、图像等）生成对应图像

A blue jay standing on a large basket of rainbow macarons.

文本描述

条件扩散概率模型



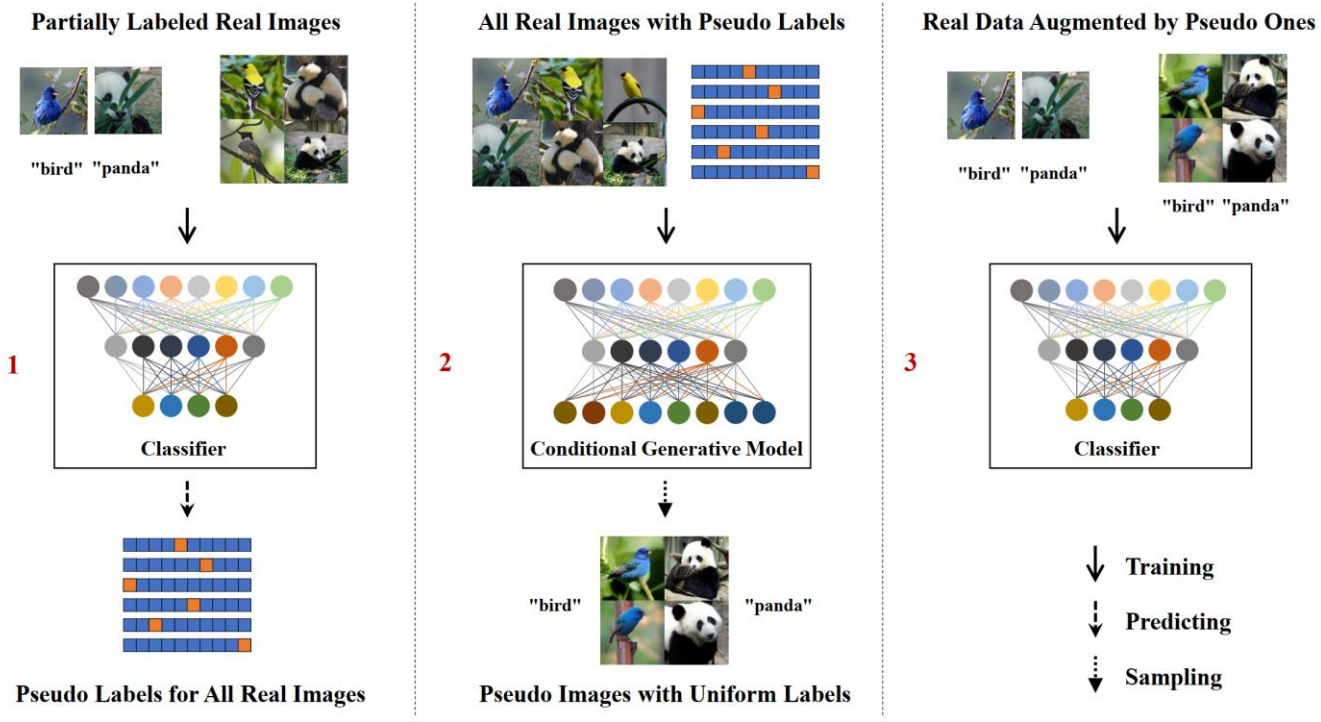
需要成对数据

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon-\epsilon_\theta(z_t,t,\tau_\theta(y))\|_2^2\right]$$

DPT：半监督扩散概率模型

- 在少量标注下，如何训练条件扩散概率模型并控制生成样本的语义？

半监督分类器与条件扩散模型的协同训练方法



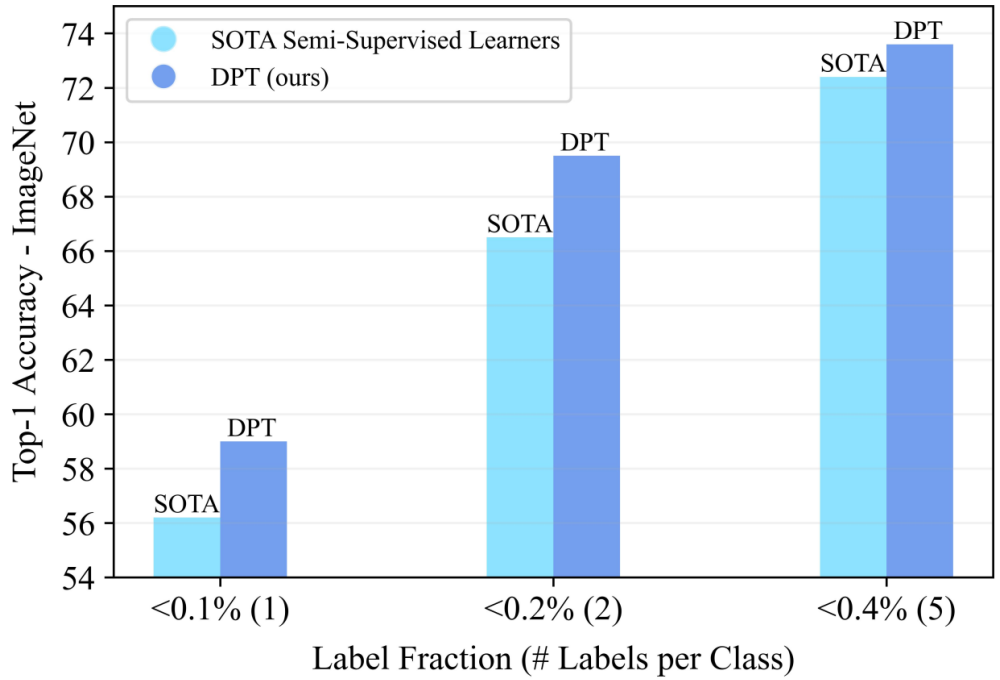
You Z, et al. Diffusion Models and Semi-Supervised Learners Benefit Mutually with Few Labels. Arxiv preprint 2023.

少量标注下控制生成样本的语义

(a) Random samples with *one* label per class. *Left:* "Gondola". *Right:* "Yellow lady's slipper".

(b) Random samples with *two* labels per class. *Left:* "Triceratops". *Right:* "Echidna".

每类使用 **1 （< 0.1%）个标注**，生成高清、可控的图像



标杆数据 **ImageNet** 上领先的半监督分类结果

Classifier guidance

A Nichol et al., ICML 2021

原始出发点：如何把无条件扩散模型转为条件模型

似然函数　　先验（扩散模型）

贝叶斯公式　$$p(x \mid c) = \frac{p(c \mid x)\, p(x)}{p(c)}$$

后验（目标分布）　　证据（常数）

Classifier guidance

A Nichol et al., ICML 2021

贝叶斯公式 $p(x \mid c) = \dfrac{p(c \mid x)\, p(x)}{p(c)} \quad \Rightarrow \quad \nabla_x \log p(x \mid c) = \nabla_x \log p(x) + \nabla_x \log p(c \mid x)$

Classifier guidance

A Nichol et al., ICML 2021

贝叶斯公式 $$p(x \mid c) = \frac{p(c \mid x)\, p(x)}{p(c)} \quad \Rightarrow \quad \nabla_x \log p(x \mid c) = \nabla_x \log p(x) + \nabla_x \log p(c \mid x)$$

温度

分类器指引 $$\tilde{\epsilon}_{\theta,\phi}(x_t, t, c) \coloneqq \epsilon_\theta(x_t) - s\ \sigma_t \nabla_{x_t} \log p_\phi(c|x_t, t)$$

采样方向　　预训练扩散模型　　预训练分类器

Classifier guidance

A Nichol et al., ICML 2021

贝叶斯公式 $$p(x \mid c) = \frac{p(c \mid x)\, p(x)}{p(c)} \quad \Rightarrow \quad \nabla_x \log p(x \mid c) = \nabla_x \log p(x) + \nabla_x \log p(c \mid x)$$

温度

分类器指引 $$\tilde{\epsilon}_{\theta,\phi}(x_t, t, c) := \epsilon_\theta(x_t) - s\ \sigma_t \nabla_{x_t} \log p_\phi(c|x_t, t)$$

采样方向　　预训练扩散模型　　预训练分类器

迭代版本：引入不同噪声层次下的指引，并作泰勒展开近似

Classifier guidance

A Nichol et al., ICML 2021

贝叶斯公式 $p(x \mid c) = \dfrac{p(c \mid x)\, p(x)}{p(c)} \quad \Rightarrow \quad \nabla_x \log p(x \mid c) = \nabla_x \log p(x) + \nabla_x \log p(c \mid x)$

温度

分类器指引 $\tilde{\epsilon}_{\theta,\phi}(x_t, t, c) \coloneqq \epsilon_\theta(x_t, c) - s\ \sigma_t \nabla_{x_t} \log p_\phi(c|x_t, t)$

采样方向　　预训练扩散模型　　预训练分类器

即使训练了条件模型，使用较大的温度也可以更好地权衡多样性和与条件的匹配关系

Classifier free guidance

Ho and Salimans, Arxiv preprint 2022

分类器指引

$$\nabla_x \log p_s(x \mid c) = \nabla_x \log p(x) + s\, \nabla_x \log p(c \mid x)$$

如何在不需要训练分类器的情况下得到同样的权衡效果？

贝叶斯公式

$$p(c \mid x) = \frac{p(x \mid c)\, p(c)}{p(x)} \quad \Rightarrow \quad \nabla_x \log p(c \mid x) = \nabla_x \log p(x \mid c) - \nabla_x \log p(x)$$

Classifier free guidance

Ho and Salimans, Arxiv preprint 2022

分类器指引

$$\nabla_x \log p_s(x \mid c) = \nabla_x \log p(x) + s\, \nabla_x \log p(c \mid x)$$

贝叶斯公式

$$p(c \mid x) = \frac{p(x \mid c)\, p(c)}{p(x)} \quad \Rightarrow \quad \nabla_x \log p(c \mid x) = \nabla_x \log p(x \mid c) - \nabla_x \log p(x)$$

带入分类器指引采样公式

无分类器指引

$$\nabla_x \log p_s(x \mid c) = (1 - s)\nabla_x \log p(x) + s\, \nabla_x \log p(x \mid c)$$

采样方向　　无条件评分函数　　条件评分函数

Classifier free guidance

Ho and Salimans, Arxiv preprint 2022

分类器指引

$$\nabla_x \log p_s(x \mid c) = \nabla_x \log p(x) + s\, \nabla_x \log p(c \mid x)$$

无分类器指引

$$\nabla_x \log p_s(x \mid c) = (1 - s)\nabla_x \log p(x) + s\, \nabla_x \log p(x \mid c)$$

采样方向　　无条件评分函数　　条件评分函数

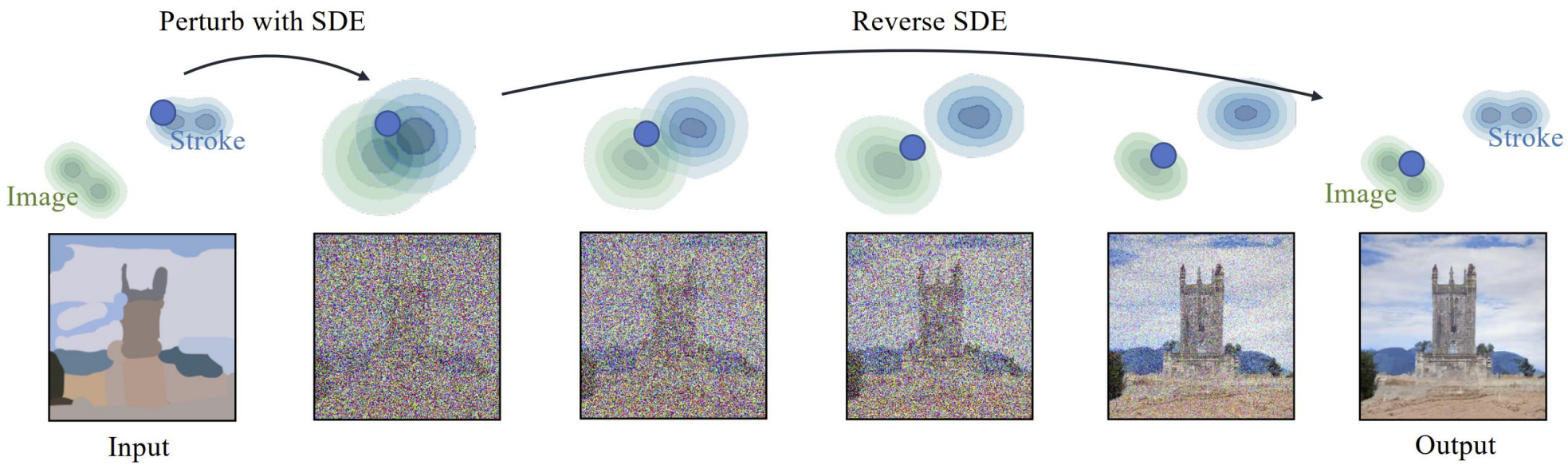
- 训练两个评分函数模型，共享参数
- 从一个“极端条件模型”采样，即 $s > 1$
- 需要成对数据训练但是参数高效，训练稳定
- 目前对于各类条件生成都非常有效，在文到图生成等任务中应用广泛

条件不止是类别

SDEdit

Meng et al. ICLR 2022

基于预训练模型的图像编辑：零样本采样方法生成目标域的样本



前向过程若干步

从带噪图片而非高斯噪声出发反向若干步

SDEdit

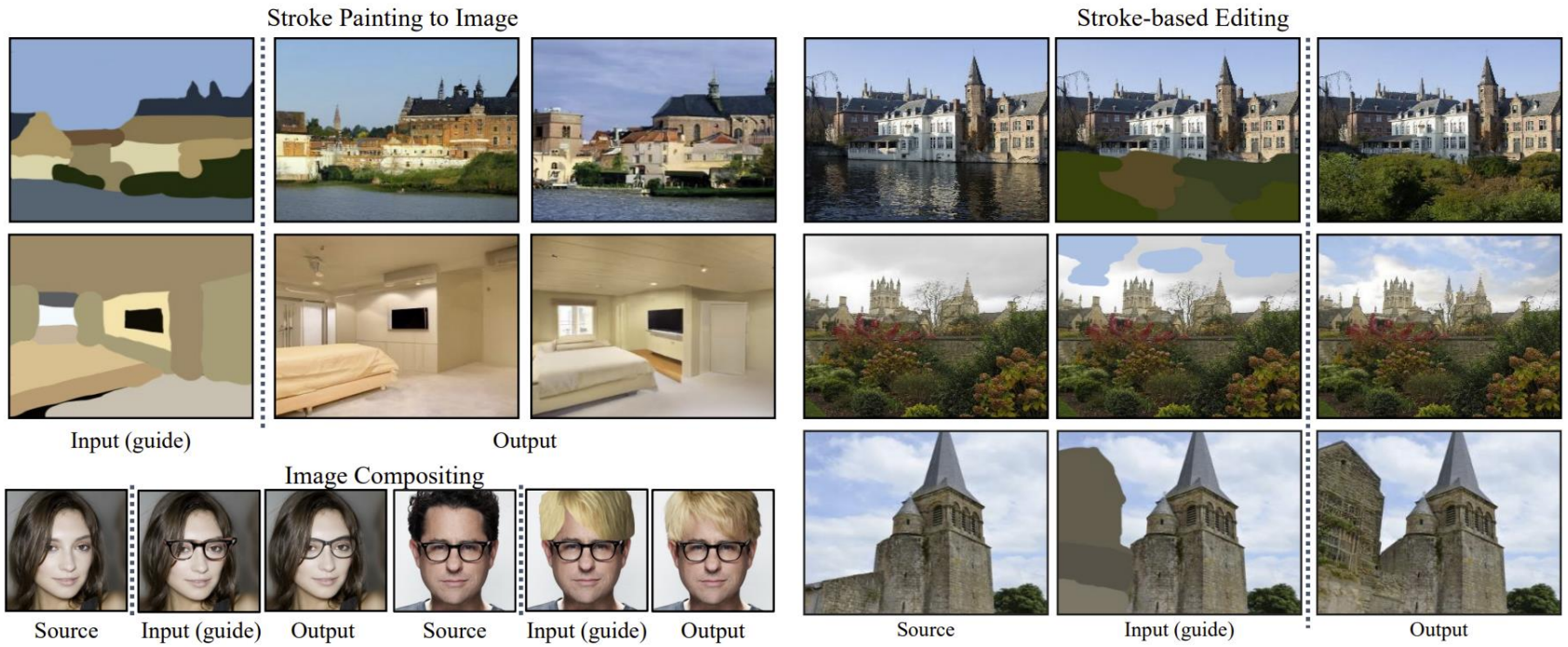


Figure 1: Stochastic Differential Editing (SDEdit) is a **unified** image synthesis and editing framework based on stochastic differential equations. SDEdit allows stroke painting to image, image compositing, and stroke-based editing **without** task-specific model training and loss functions.

能量函数指引框架

能量指引

Zhao et al, NeurIPS 2022

- 一种加入知识的一般性框架

能量指引

$$\tilde{\epsilon}_{\theta,\phi}(x_t, t, c) \coloneqq \epsilon_\theta(x_t) - s\,\nabla_{x_t}\varepsilon_\phi(x_t, t, c)$$

采样方向　　预训练扩散模型　　预训练能量函数

- 只需能量函数可微
- 分类器指引和无分类器指引是能量指引的特例
- 可以灵活组合各种能量函数

能量指引采样分布

Zhao et al, NeurIPS 2022

- 能量函数定义了如下的概率密度

$$q_\phi(x|c) = \frac{1}{Z(\phi)} e^{-\varepsilon_\phi(x_t,t,c)}, \qquad Z(\phi) = \int e^{-\varepsilon_\phi(x_t,t,c)} dx.$$

- 可以证明，能量指引近似地从如下乘积专家模型中采样

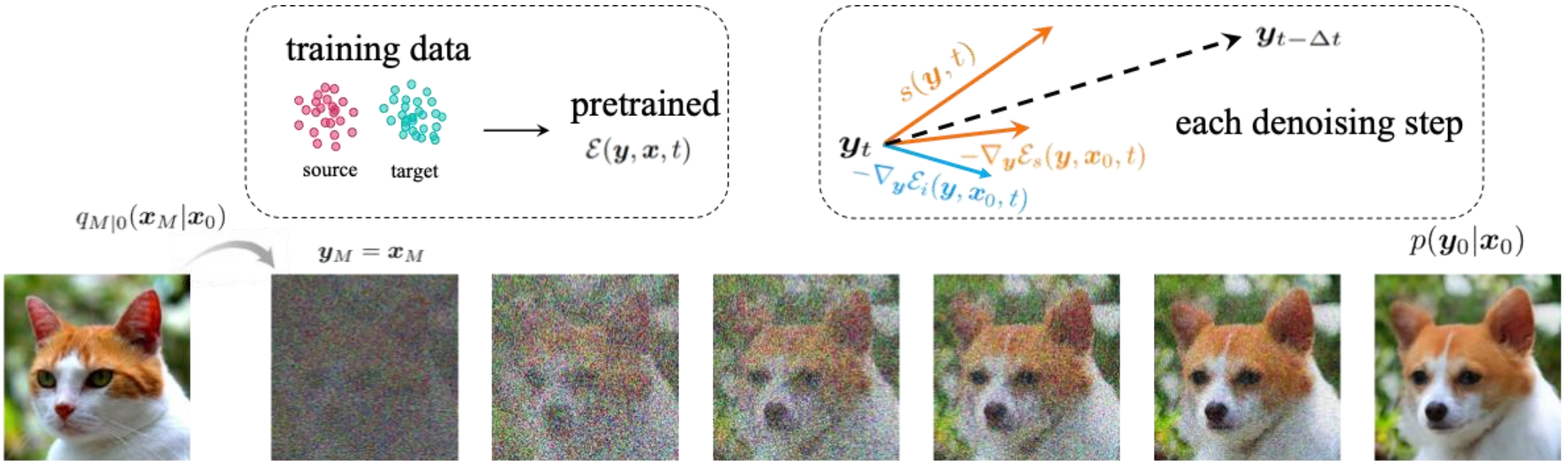
$$p_{\theta,\phi}(x \mid c) \propto p_\theta(x)\, q_\phi(x \mid c)$$

- 贝叶斯公式、**RLHF** 均为乘积专家模型的特例

应用一：不成对图到图的翻译

Zhao et al.，NeurIPS 2022

SDE 采样中加入预训练的自定义能量函数控制样本：零样本乘积专家模型采样



EGSDE：权衡不同专家/能量函数

Input

$\lambda_i = 0$

realistic expert $\mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t)$

realistic

$\lambda_s = 0$

$\lambda_s = 1500$

faithful expert $\mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t)$

faithful

$\lambda_i = 0$

$\lambda_i = 150$

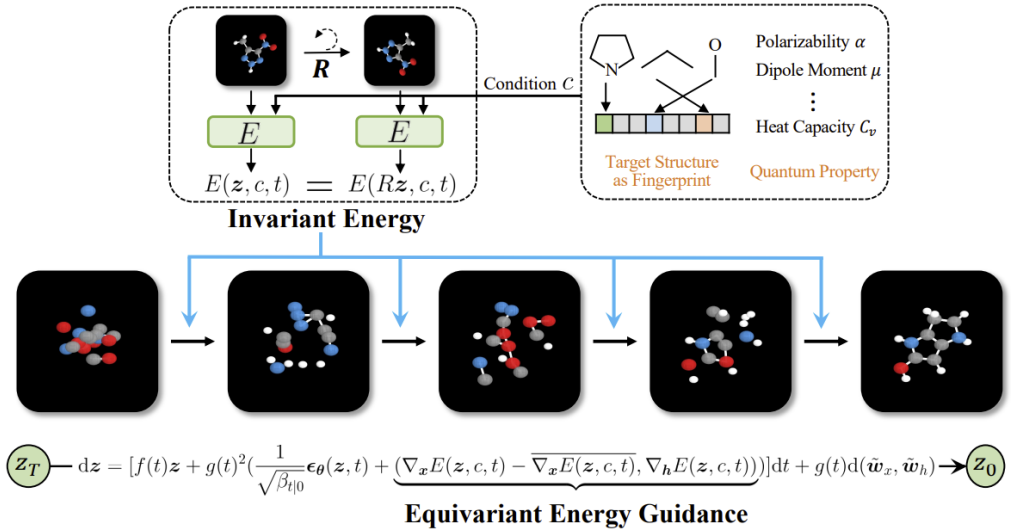
Output

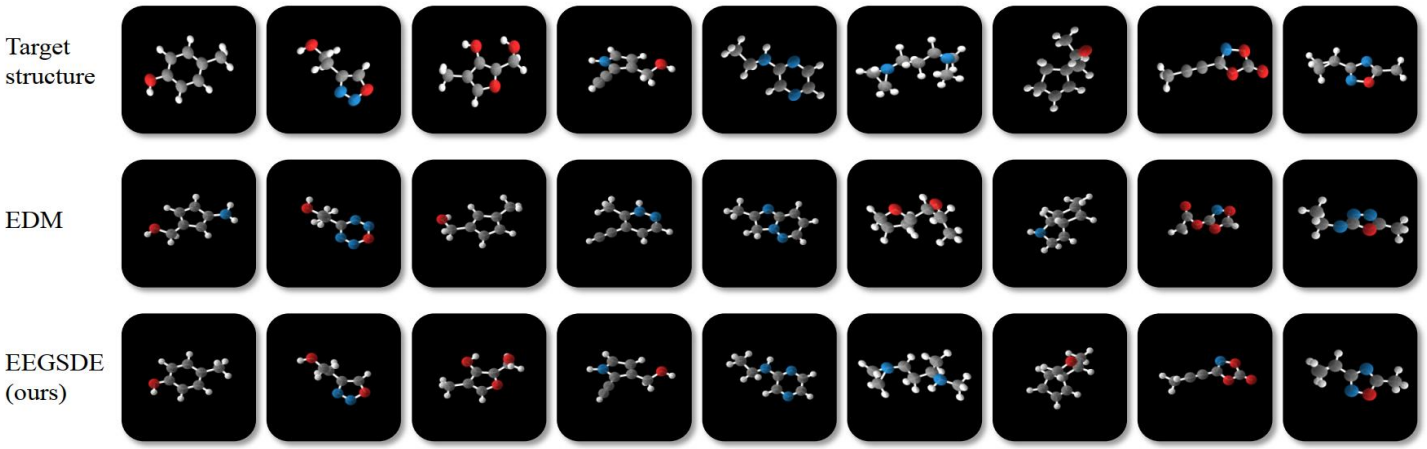
$$\mathcal{E}(\boldsymbol{y}, \boldsymbol{x}_0, t)$$
$$\|$$
$$\lambda_s \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) + \lambda_i \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t)$$

应用二：可控分子生成

Bao & Zhao et al., ICLR 2023

- 几何等变的能量函数指引的扩散模型
- 有效控制官能团、量子性质等合成分子性质，可以同时控制多种属性





总结：条件扩散模型与 Guidance

- 条件模型
  - 类别半监督：**DPT**
  - 同模态跨域无监督：**SDEdit**、**EGSDE**
- 采样中的指引
  - 指引方式：分类器指引、无分类器指引、能量函数指引
  - 采样分布：条件模型（贝叶斯公式）、乘积专家模型

大规模扩散模型

大规模深度生成模型：人工智能新范式逐步改变研究、生产与生活方式

CH 为什么学习概率图模型

学习概率图模型（Probabilistic Graphical Models）的好处有很多，以下是一些主要原因：

1. 建模能力：概率图模型提供了一种建模框架，可以用于描述大量实际问题，包括计算机视觉、自然语言处理、医疗诊断、金融风险管理等各种领域。
2. 推理能力：概率图模型可以进行推理和预测，通过对模型进行推理，可以推断出某些变量的概率分布，这对于决策和优化问题非常有用。
3. 知识表示和推理：概率图模型可以有效地表示和推理不确定性知识，从而使得模型更加准确和稳健。
4. 机器学习应用：概率图模型是机器学习的一种重要方法，它可以应用于监督学习、无监督学习和半监督学习等不同场景。
5. 交叉学科应用：概率图模型是交叉学科研究的一种重要工具，可以应用于计算机科学、统计学、人工智能、心理学、物理学等不同领域。

总之，学习概率图模型可以帮助我们更好地理解和应用概率论，提高我们的建模和推理能力，以及解决一些实际问题。

ChatGPT: 基于语言大模型的对话机器人

零样本 (Zero-Shot) 图像精确理解

GPT-4 visual input example, Extreme Ironing:

User What is unusual about this image?

Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

GPT-4 The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi.

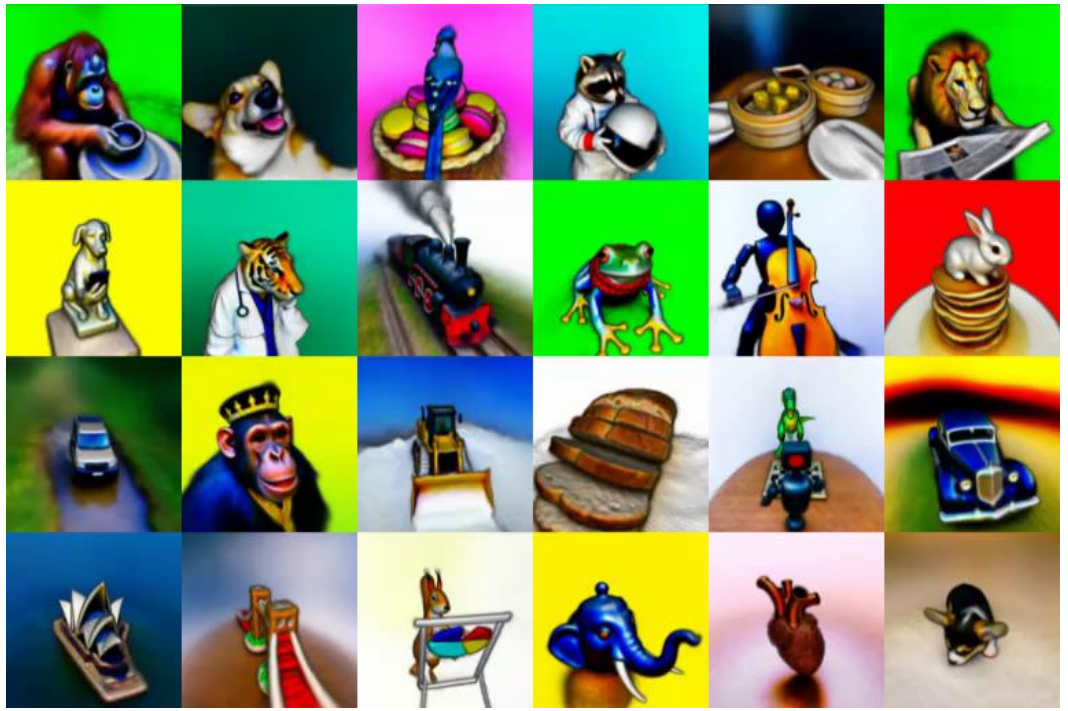
GPT-4: 以语言模型为核心的多模态内容理解

大规模深度生成模型：人工智能新范式逐步改变研究、生产与生活方式

输入文本描述主题“太空歌剧院”合成图像，获美国科罗拉多博览会的年度艺术比赛首奖

Stable Diffusion：文到图生成扩散概率大模型

输入文本，直接渲染 3D 场景

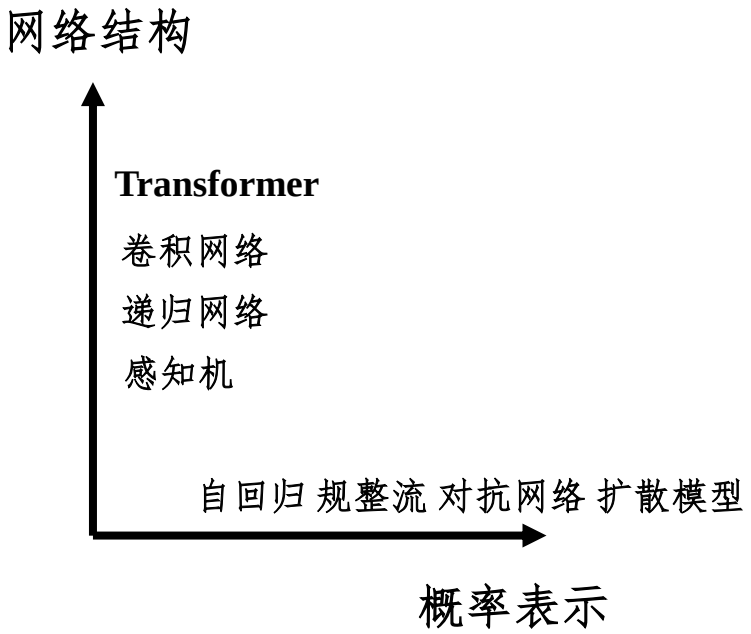


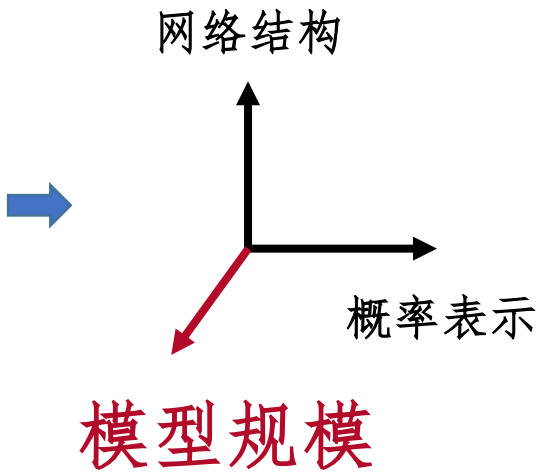
Dreamfusion：基于2D大模型的零样本3D场景建模

模型规模

- **模型规模**是大规模深度生成模型中联合概率分布表示的第三维度

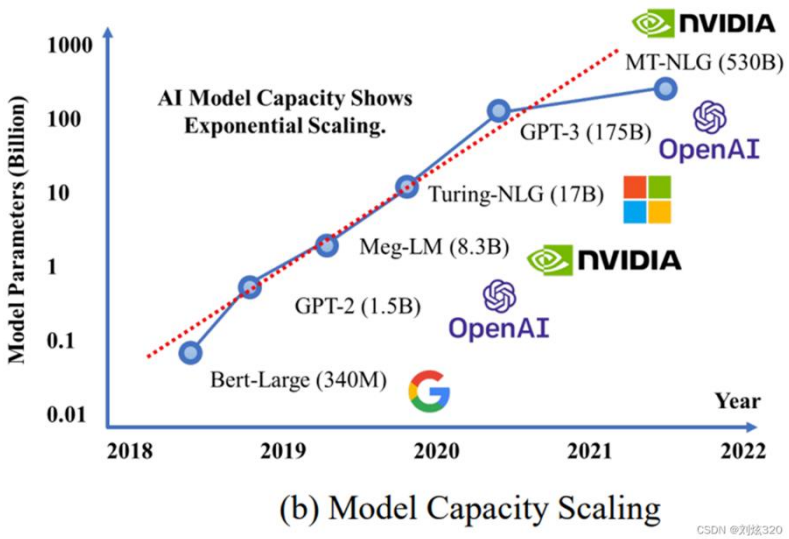
深度神经网络参数化的概率分布族

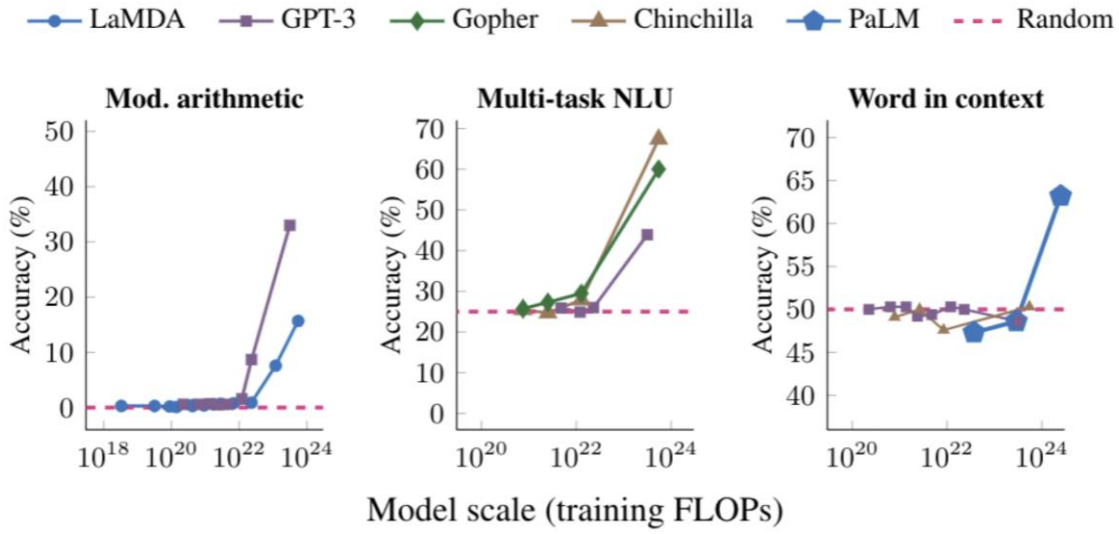




模型规模

- 模型规模是大规模深度生成模型中联合概率分布表示的第三维度

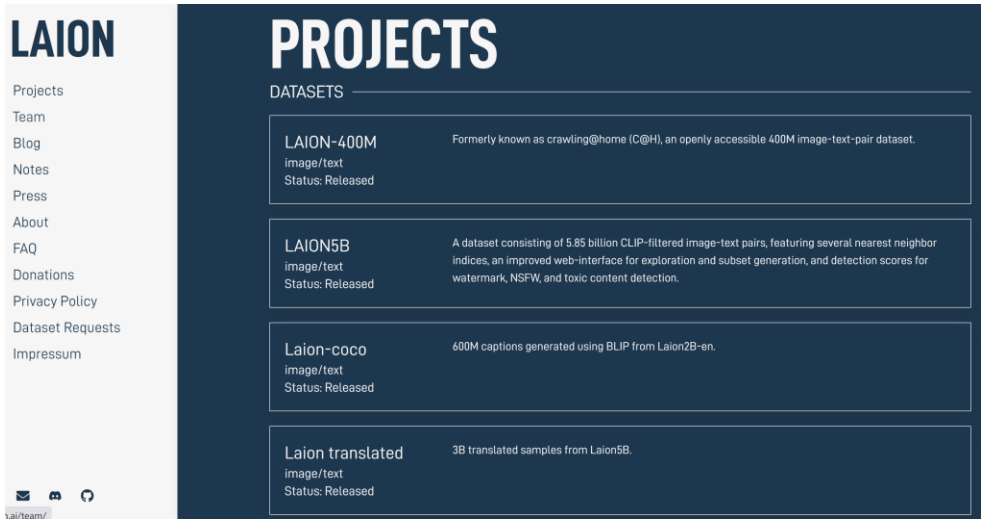




随着模型规模增大，表现显著提升

大规模概率建模：从算法为核心转化为数据和模型为核心

- 充分利用“大模型 + 大数据” = 开放域强泛化、任务通用



- 网络结构、算法设计服务于“大模型+大数据”

LAION：开放域，高噪声，大规模

数亿甚至十亿样本


Animals Square Cute Kitten Cat Diamond Painting Ki...


Kitten by Michael Creese


Удивительные котики. Автор - Ольга Бессогонова


Imagenes A Lapiz De Gatos | como dibujar un gato r...


Pet Portrait: Digital Custom Cat Portrait Painting


Cat Watercolor Painting original 5 x 7. Small cat ...


Милый котенок, фото 1


Artista russa cria gatinhos de feltro que parecem ...


Art: Flower Picking ACEO by Artist Carmen Medlin


Watercolor MaineCoon Cat Hand Drawn Pet Portrait I...


Kitten Watercolor Painting Art Print - Cat Paintin...


5d Cat Diamond Painting Kit Premium-54


rainbow cat


Kitten


Kitteh - coloured pencil


8 трикови како да се зближите со вашата нова мачка


Koci ta koty kot. Cat


ingefära katt päls bildbanksfoto


Схема вязания такого


Фото Картины на

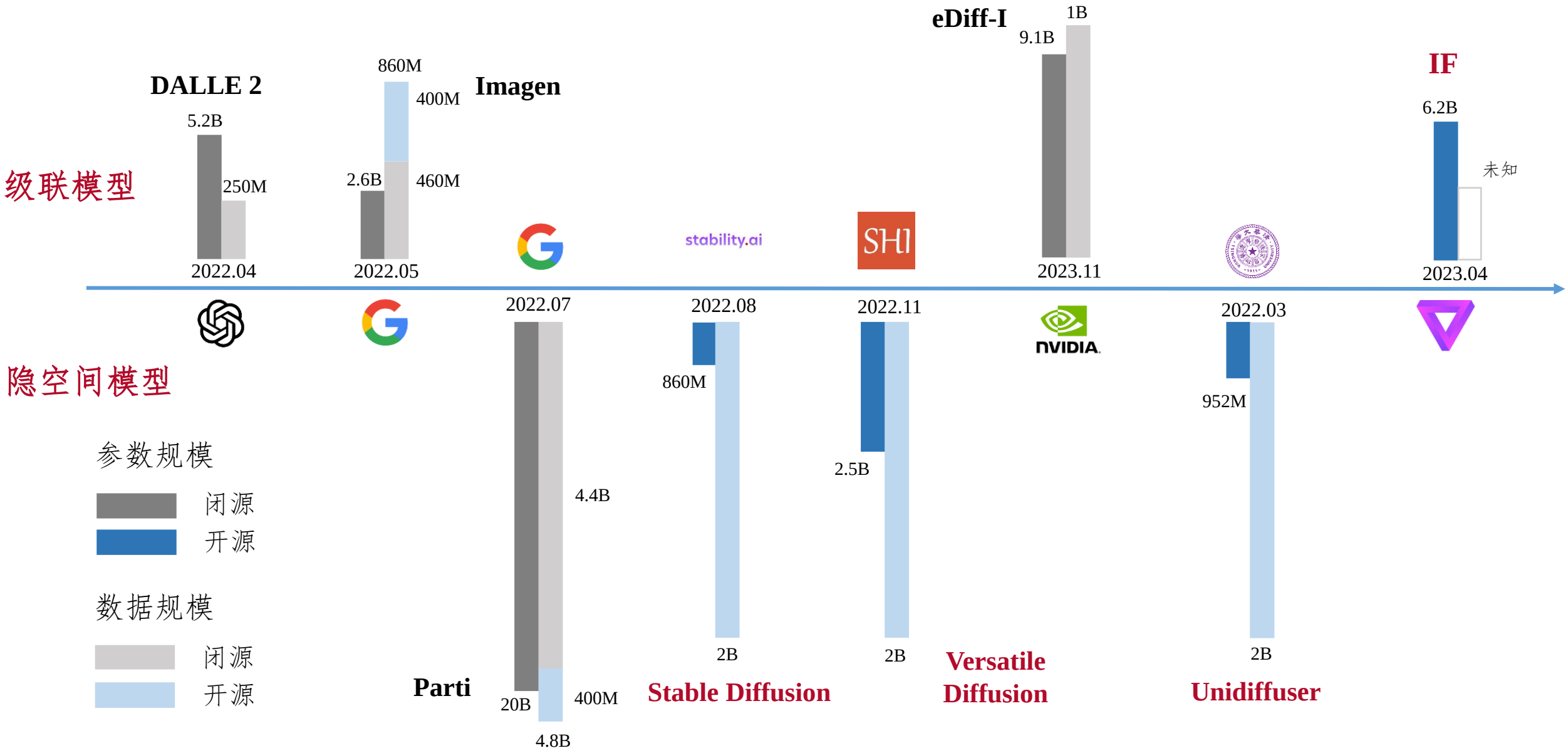

gatto

ImageNet：特定域，低噪声，小规模

精准类别标签

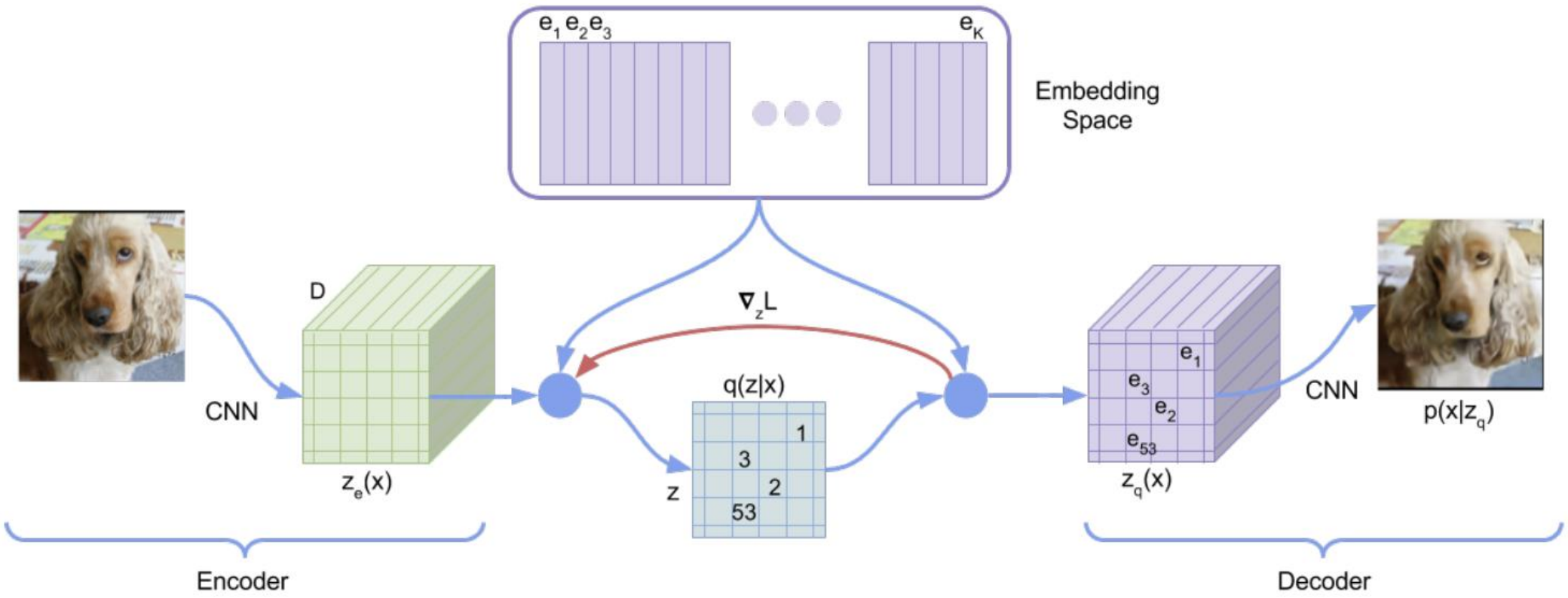
约一百万样本

大规模文到图扩散模型


级联模型
隐空间模型
DALLE 2
5.2B
250M
2022.04
860M
400M
Imagen
2.6B
460M
2022.05
2022.07
Parti
20B
4.4B
400M
4.8B
stability.ai
2022.08
860M
2B
Stable Diffusion
SHI
2022.11
2.5B
2B
Versatile Diffusion
eDiff-I
1B
9.1B
2023.11
NVIDIA
2022.03
952M
2B
Unidiffuser
IF
6.2B
未知
2023.04
参数规模
闭源
开源
数据规模
闭源
开源

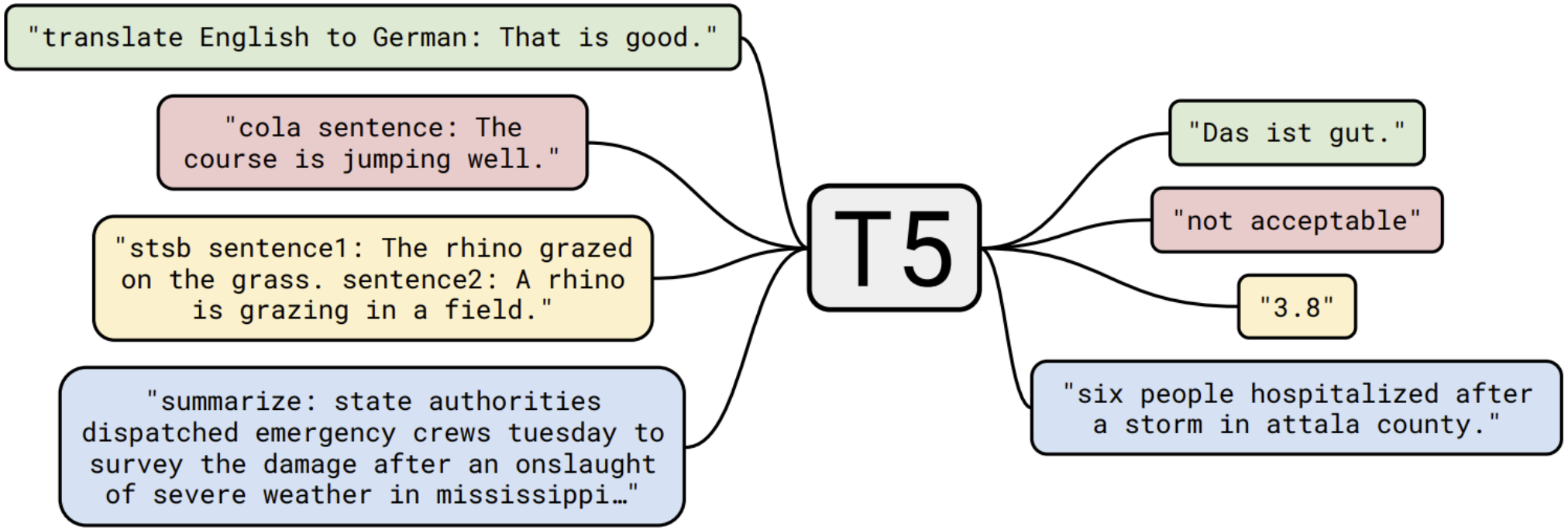
文到图扩散概率大模型：隐空间模型

VQVAE



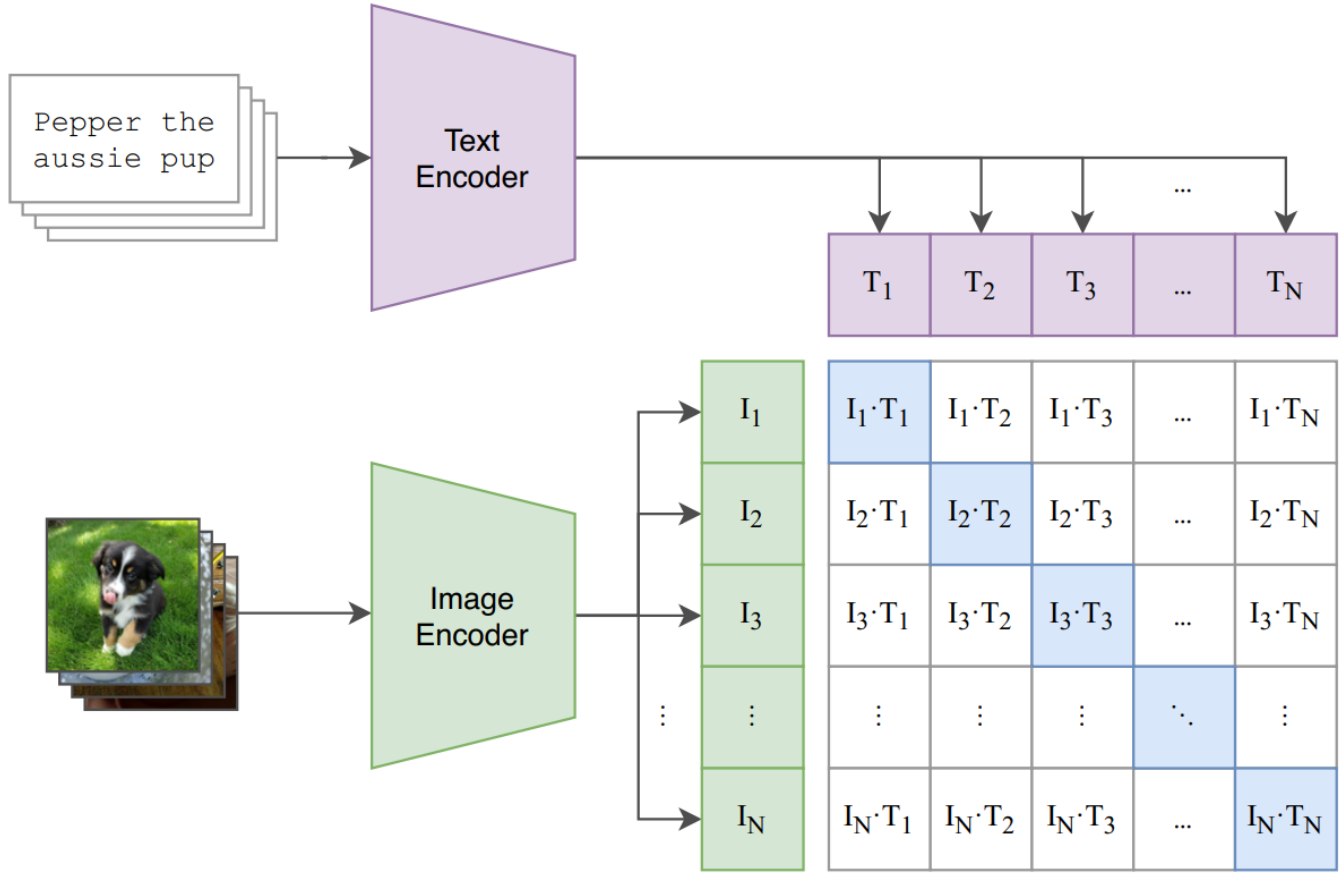
图像上的有量化约束的编码、解码训练

T5



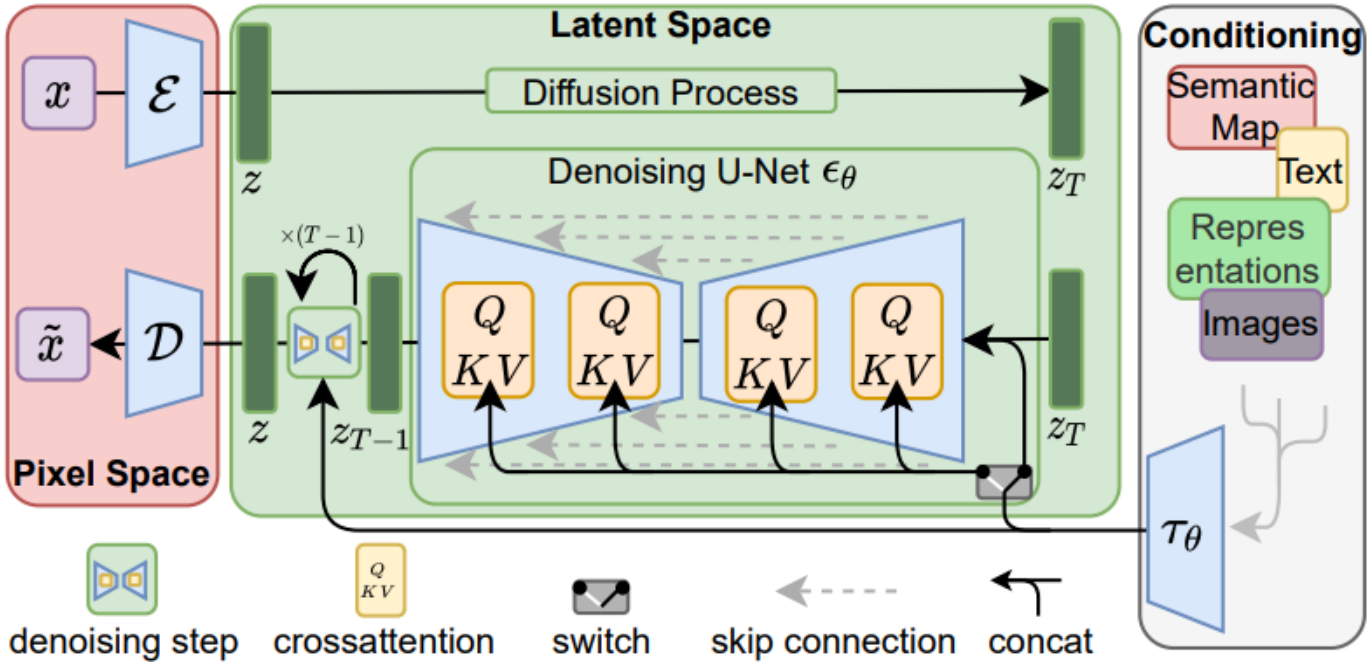
纯文本上的跨任务编码、解码训练

CLIP

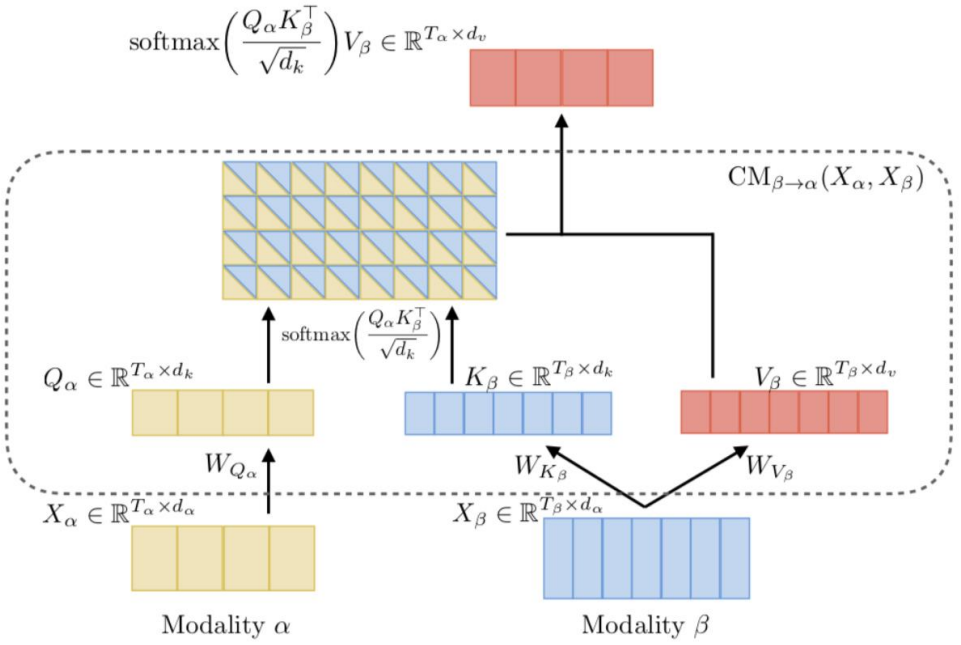


图文编码器，成对图文数据上的对比学习

隐空间扩散模型

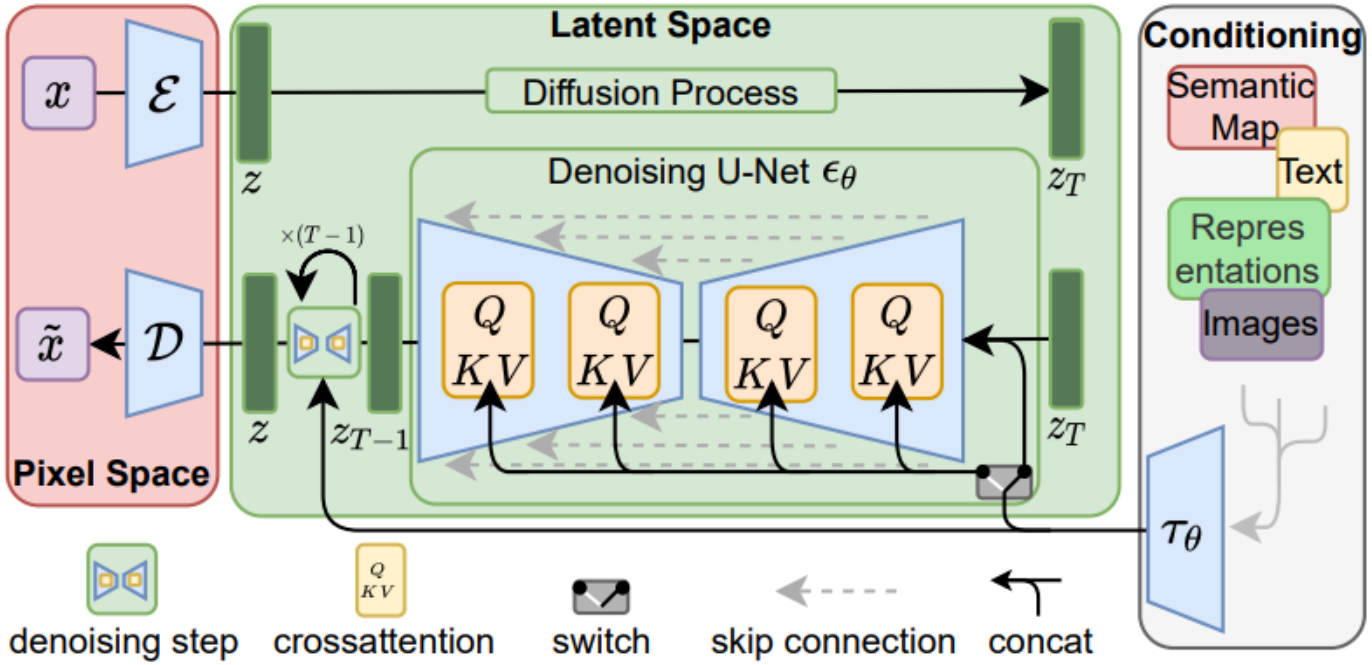


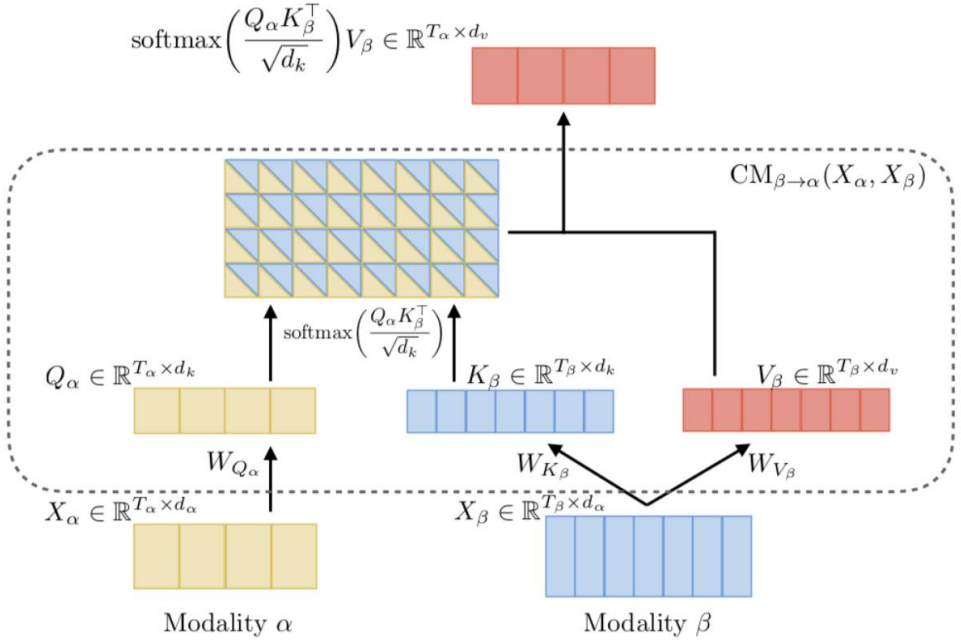
隐空间条件模型



Cross attention 模块

隐空间扩散模型

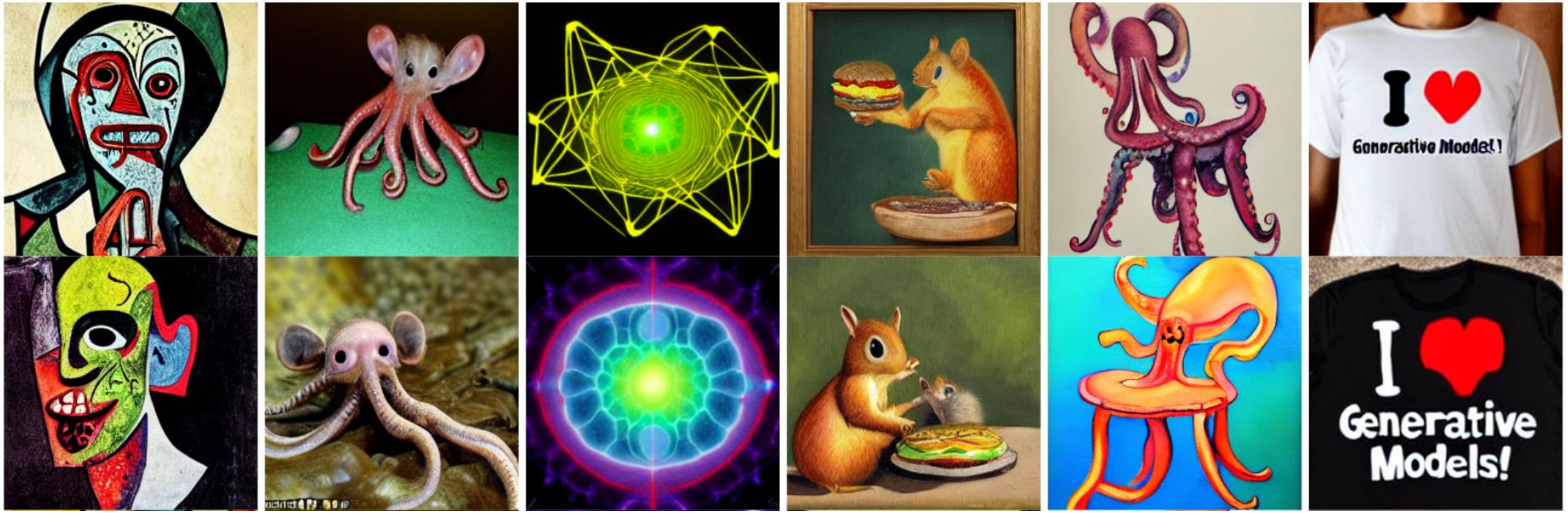




训练目标 $$L_{LDM} := \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))\|_2^2\right]$$

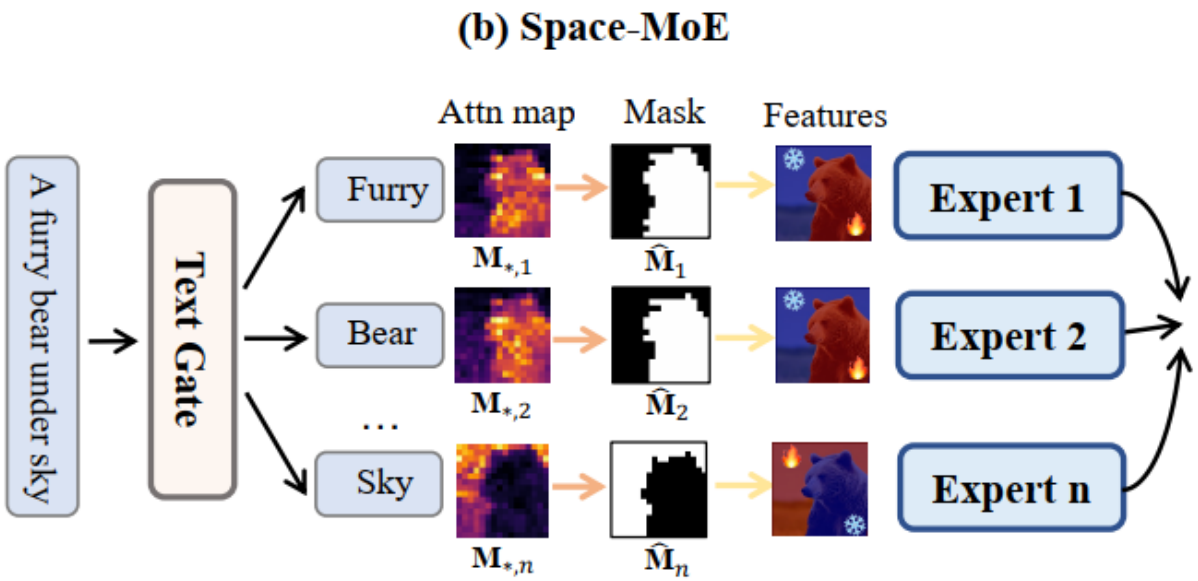
Stable Diffusion

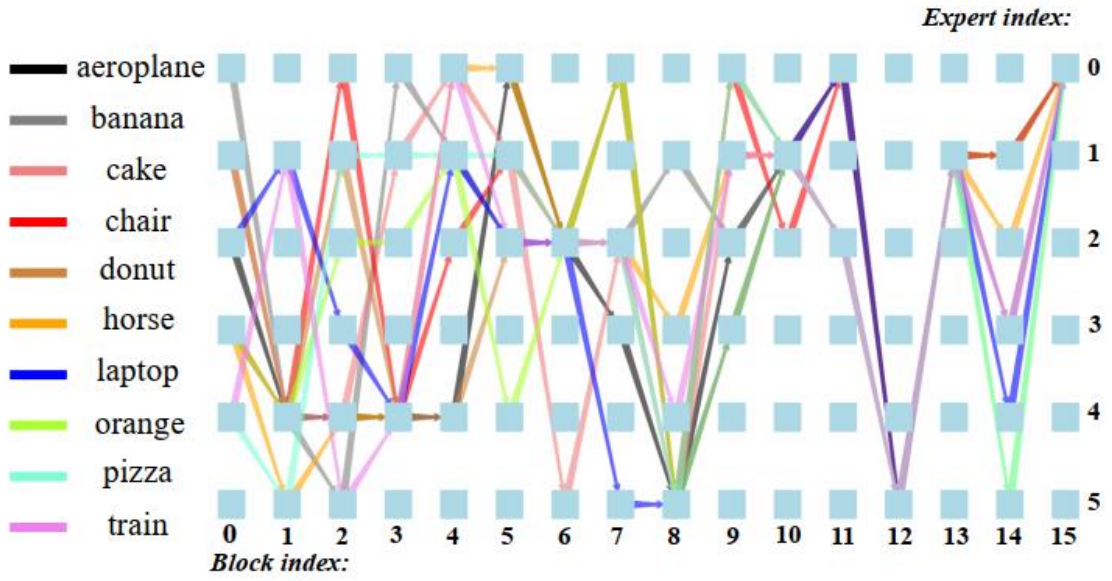
'A zombie in the style of Picasso'	*'An image of an animal half mouse half octopus'*	*'An illustration of a slightly conscious neural network'*	*'A painting of a squirrel eating a burger'*	*'A watercolor painting of a chair that looks like an octopus'*	*'A shirt with the inscription: "I love generative models!" '*



Stable Diffusion 图像编辑

RAPHAEL





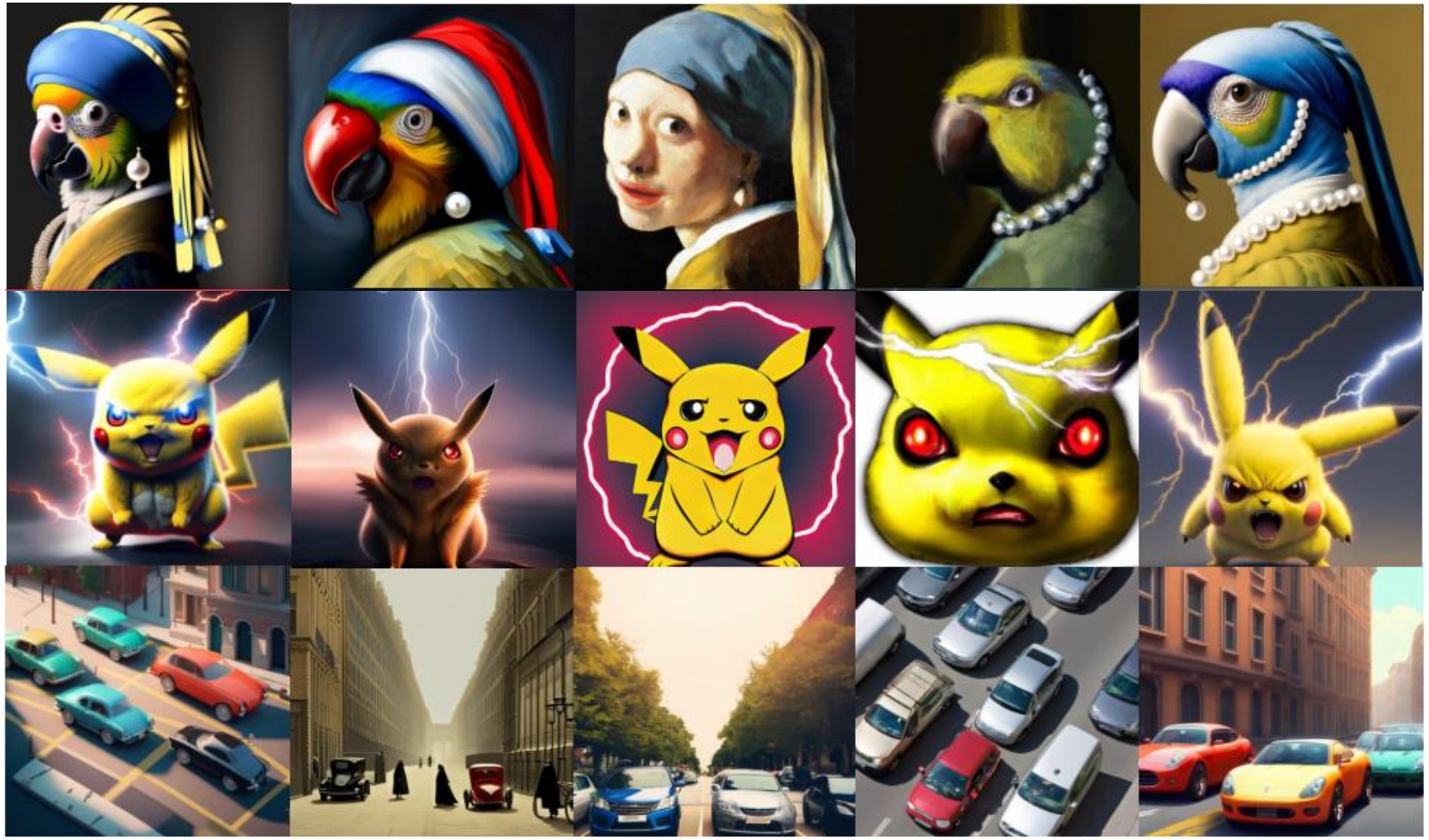
空间 **MoE** 生成对应 **token** 对应位置的图像，最后融合

RAPHAEL

A parrot with a *pearl earring*, Vermeer style.

A Pikachu with an *angry* expression and *red* eyes, with *lightning* around it, hyper realistic style.

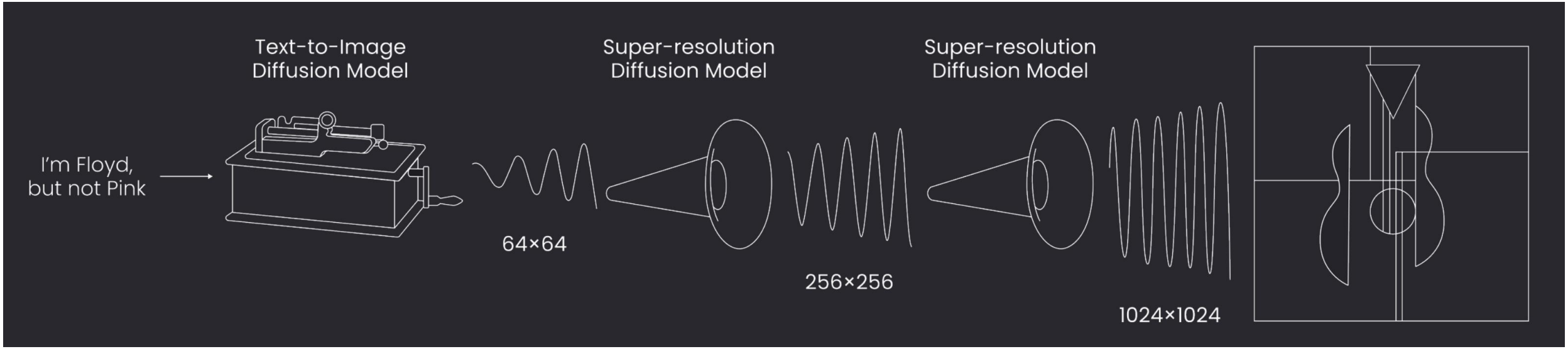
There are *five* cars in the street.



RAPHAEL **StableDiffusion XL** **DeepFloyd** **DALL-E 2** **ERNIE-ViLG 2.0**

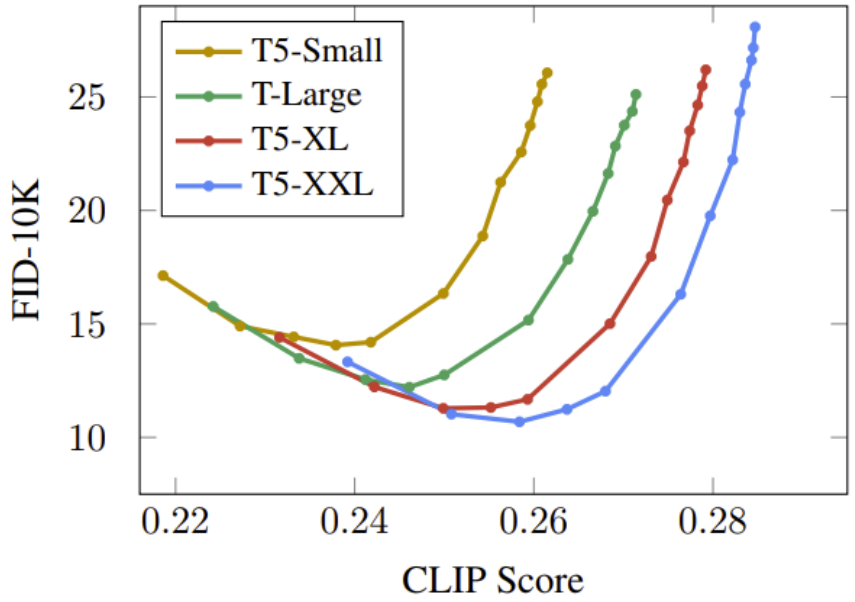
文到图扩散概率大模型：级联模型

级联扩散模型

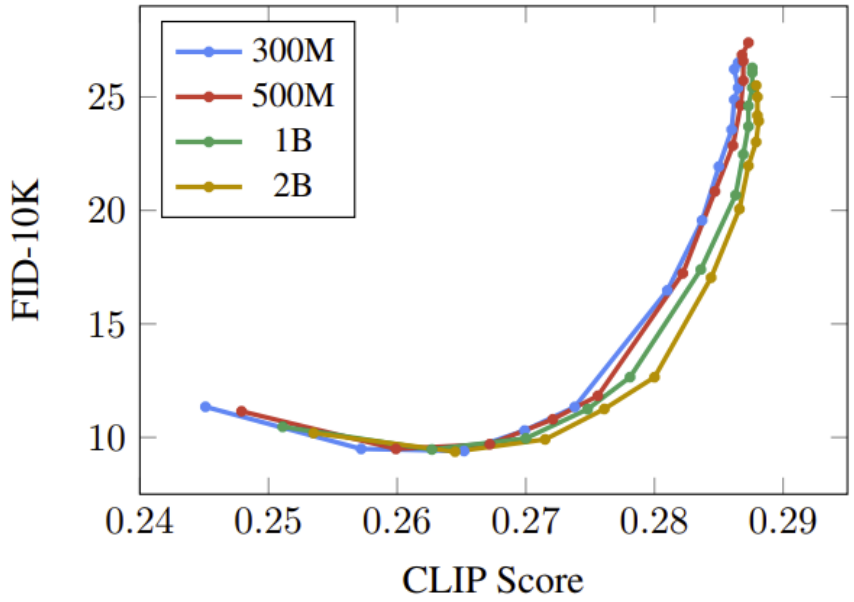

Text-to-Image
Diffusion Model
Super-resolution
Diffusion Model
Super-resolution
Diffusion Model
I'm Floyd,
but not Pink
64×64
256×256
1024×1024

Imagen

- 引入大规模预训练语言模型 **T5-XXL** 解决长文本理解问题



(a) Impact of encoder size.



(b) Impact of U-Net size.

Imagen

Sprouts in the shape of text ‘Imagen’ coming out of a fairytale book.

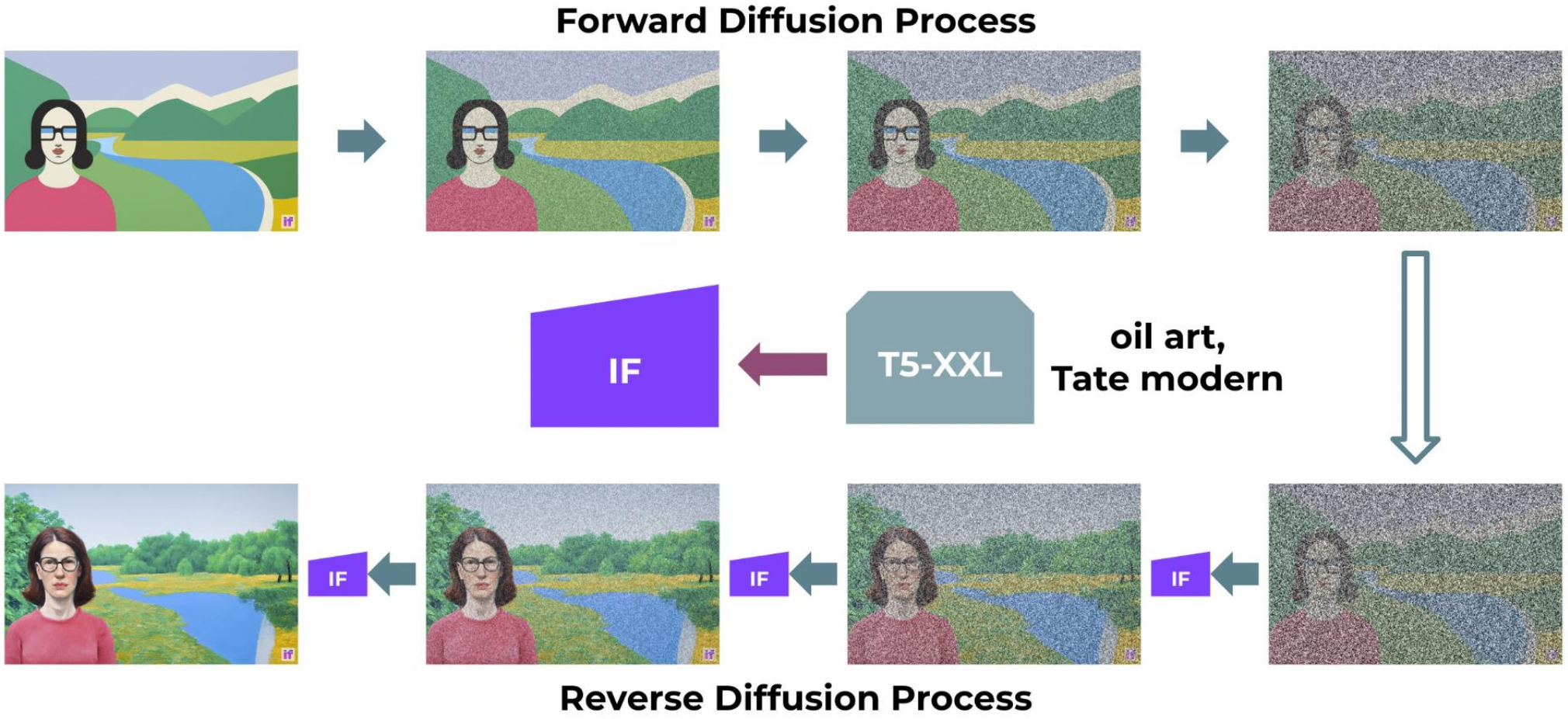
A photo of a Shiba Inu dog with a backpack riding a bike. It is wearing sunglasses and a beach hat.

A high contrast portrait of a very happy fuzzy panda dressed as a chef in a high end kitchen making dough. There is a painting of flowers on the wall behind him.

DeepFloyd IF

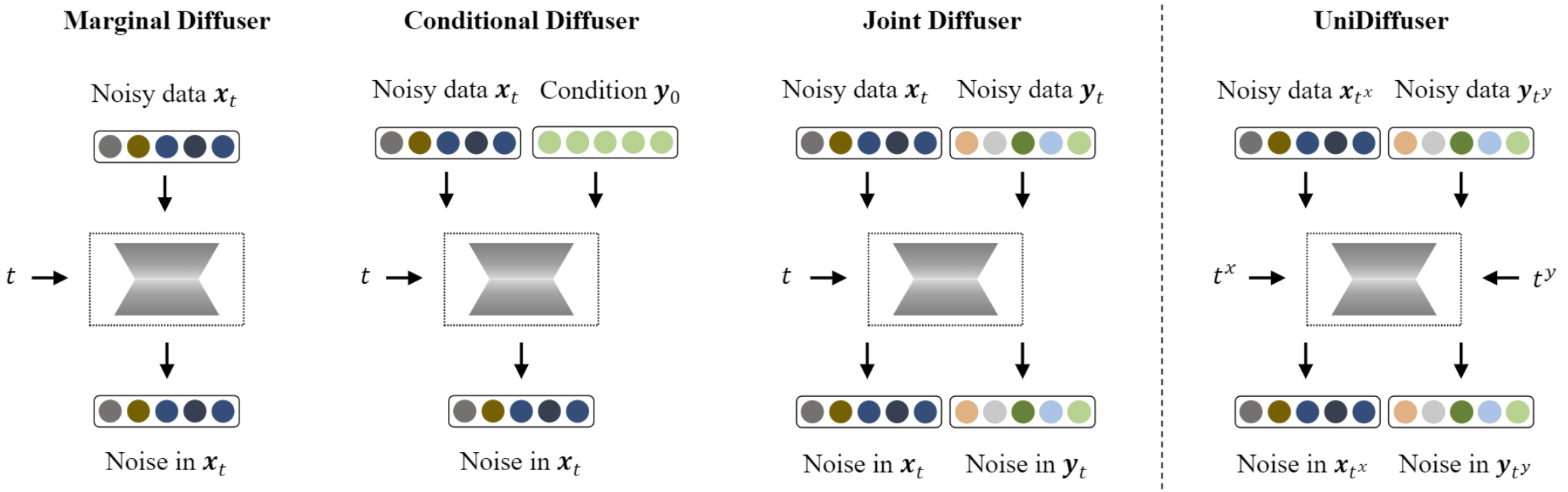
- **43** 亿参数的开源模型

DeepFloyd IF with SDEdit



DeepFloyd IF：风格迁移

通用多模态扩散模型 Unidiffuser



文本/图像生成　　文到图/图到文生成　　图文联合生成　　统一建模框架

Bao F et al. One transformer (U-ViT) fits all distributions in multi-modal diffusion. Arxiv 2023.

通用多模态扩散模型 Unidiffuser

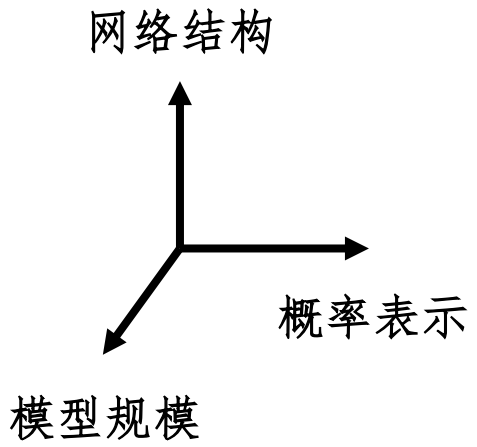
- 文图通用模型
  - 适当增大模型
  - 训练时间不变
  - 推断时间不变
  - 推断效果可比
  - 处理 5 种任务

总结：大规模扩散模型

- 数据
  - 开放域、大规模、高噪声；训练得到强泛化模型
- 训练策略
  - 降维生成再升维：隐空间 **vs.** 级联
  - 图像、文本的编码器、解码器很重要
  - 任务通用 **vs.** 任务专用

总结：扩散模型基本原理

- 概率表示
  - 两种等价理解
- 网络结构
  - 卷积与注意力，长跨层链接
- 模型规模与数据
  - 专用 **vs** 通用



中场休息 十五分钟

扩散模型与AIGC

大规模预训练模型赋能 AIGC

- 大规模预训练模型的特点
  - 开放域、强泛化
  - 文本作为主要交互接口
- 下游**AIGC**的需求与挑战
  - 数据少（利用开放域、强泛化的特点，少样本解决下游任务）
  - 个性化/可控制（加入额外条件控制）

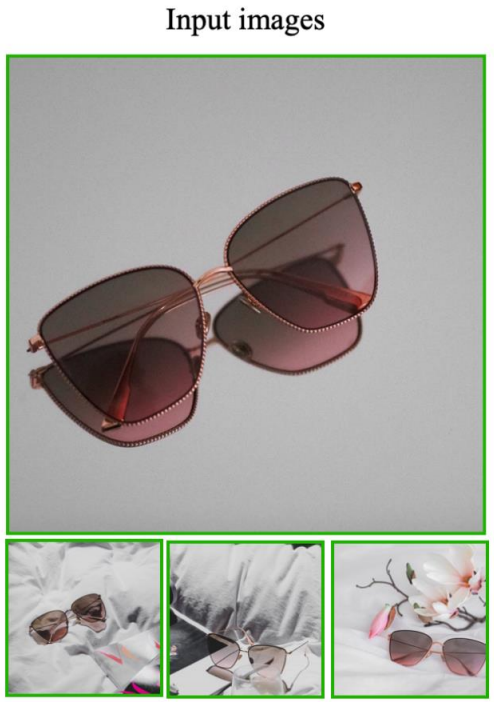
下游任务

- 个性化图像生成
- 图像可控生成与编辑
- 视频可控生成与编辑
- 三维场景生成

个性化图像生成

个性化生成

- 目标：少样本下微调大模型，保留大模型泛化能力



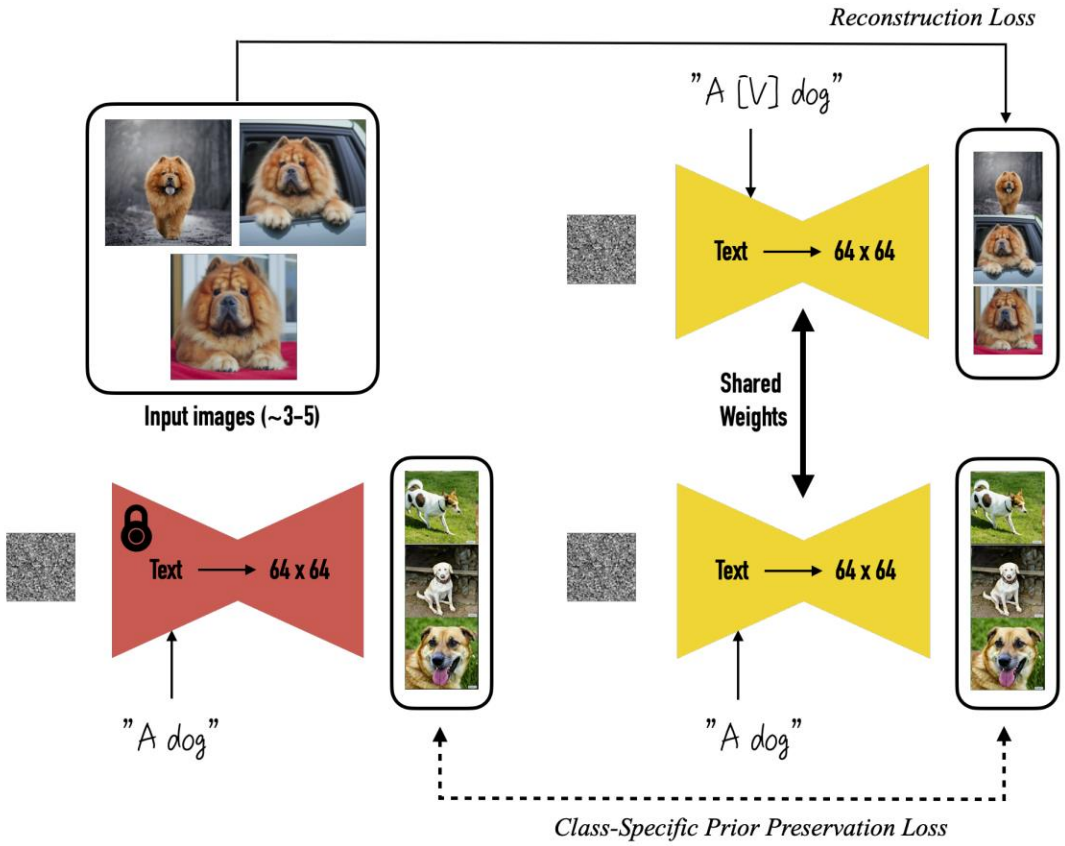
下游任务数据

目标效果

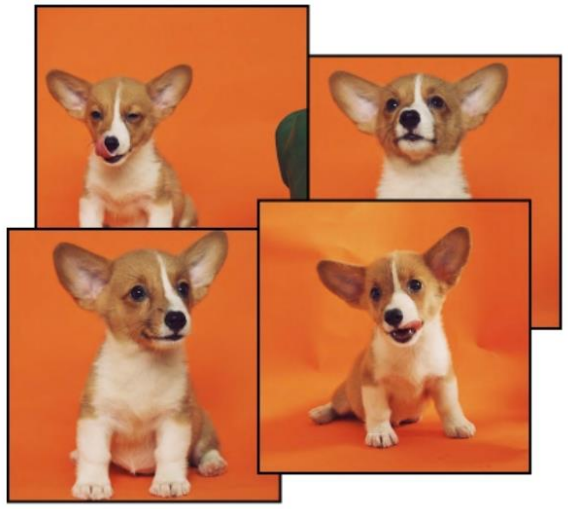
DreamBooth

Ruiz et al, CVPR 2023

- 特殊标志重建特殊图像
- 一般输入重建大模型生成结果
- 微调全部参数，但有重建约束



DreamBooth



Input images

in the Acropolis

swimming

sleeping

in a doghouse

in a bucket

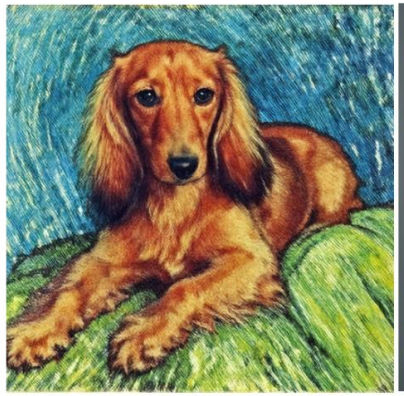
getting a haircut

DreamBooth

Input images

DreamBooth

Input images



Vincent Van Gogh

Michelangelo

Rembrandt

Johannes Vermeer

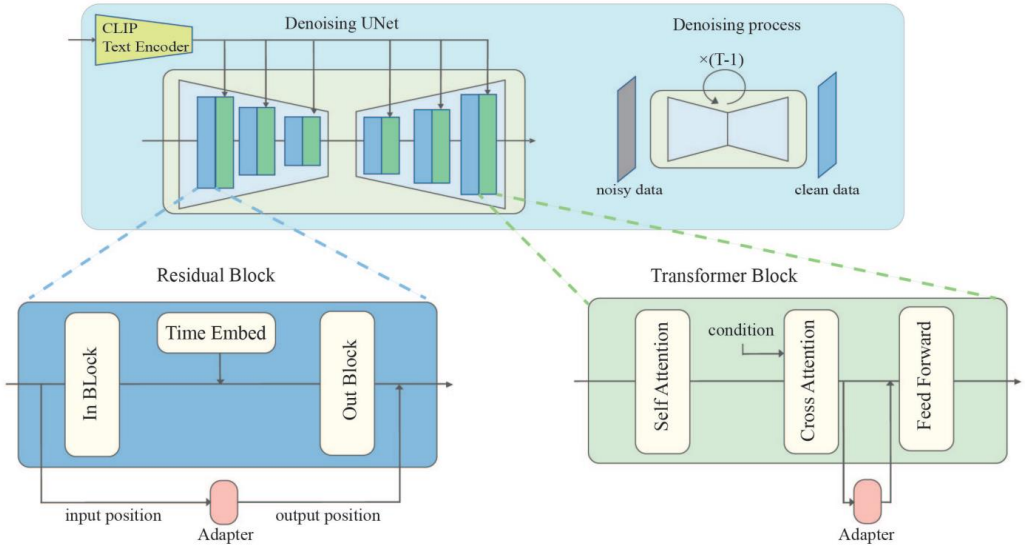
Pierre-Auguste Renoir

Leonardo da Vinci

Adapter

Xiang et al, Arxiv preprint 2023

- Adapter 插入的小神经网络
- 小数据微调 Adapter



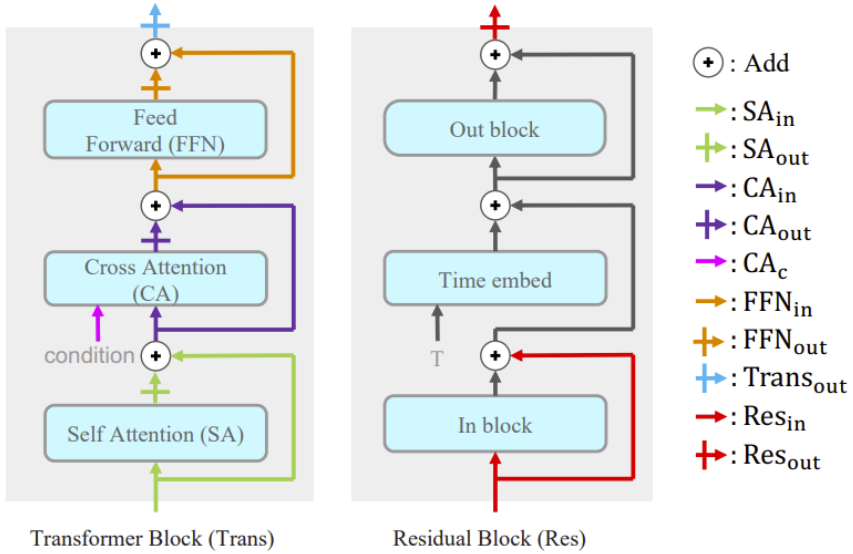
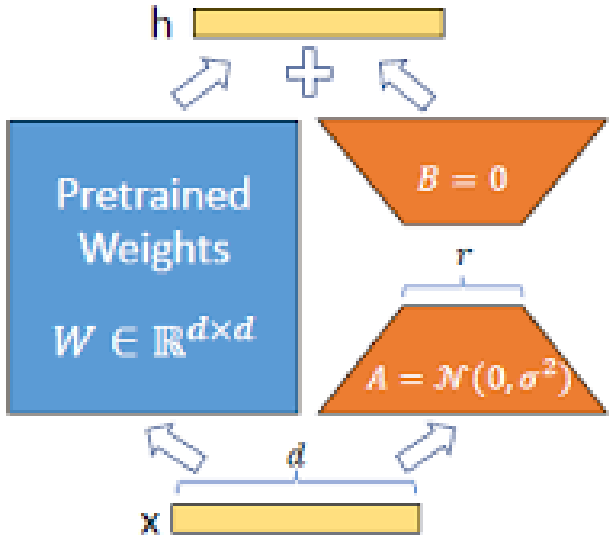


Figure 4. **Illustration of activation position.** generally, the main name of a activation position is an alias of a specific block in the model, the subscript of activation position explains the relationship between the activation and the block.

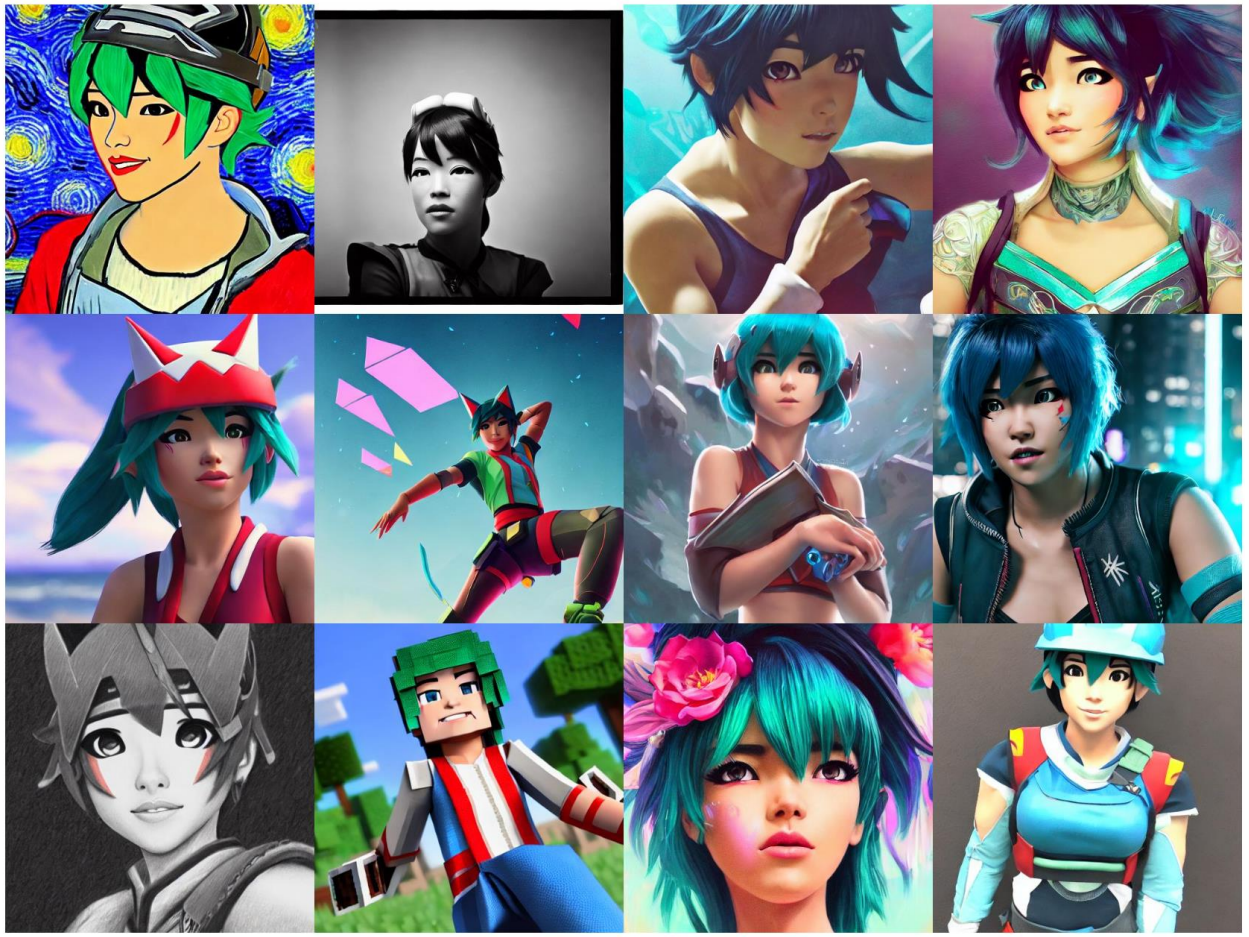
统计分析输入位置

LoRA



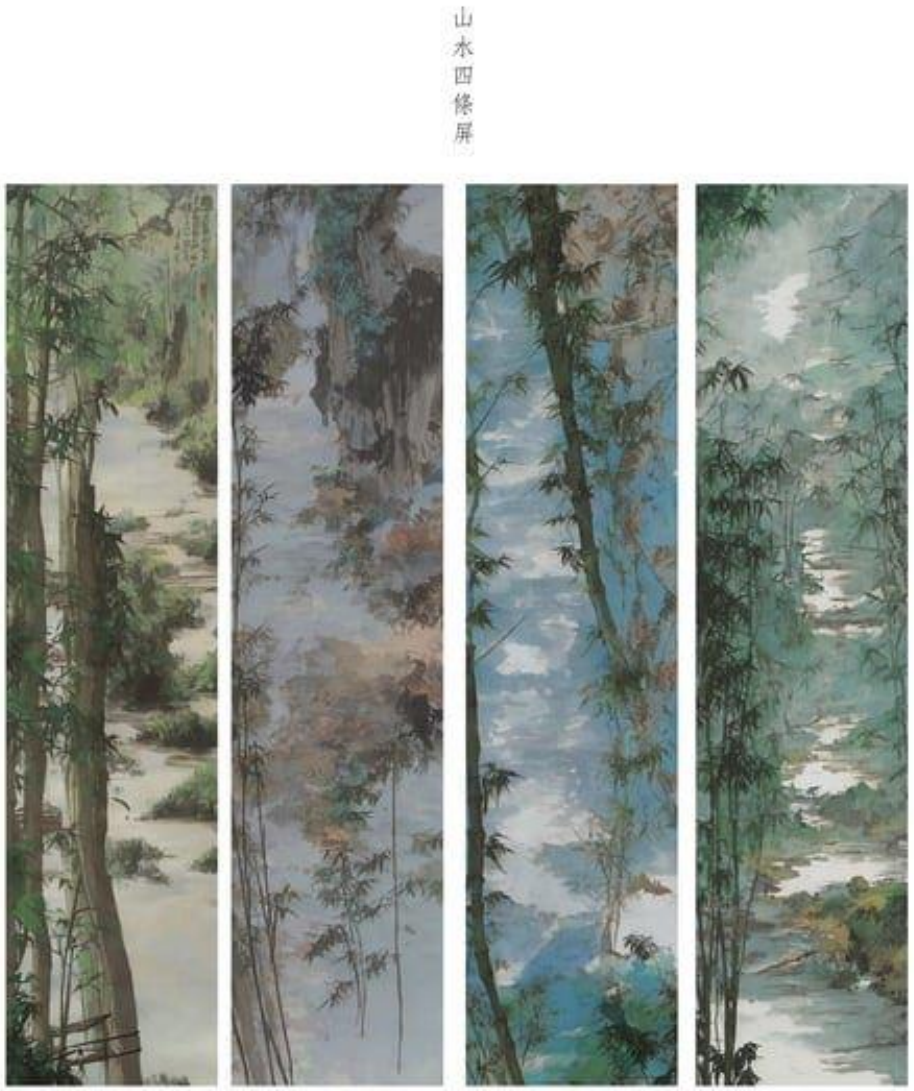
加性低秩参数矩阵 $W' = W + \Delta W$

小样本微调低秩矩阵



https://github.com/cloneofsimo/lora

LoRA: moxin



https://civitai.com/models/12597/moxin

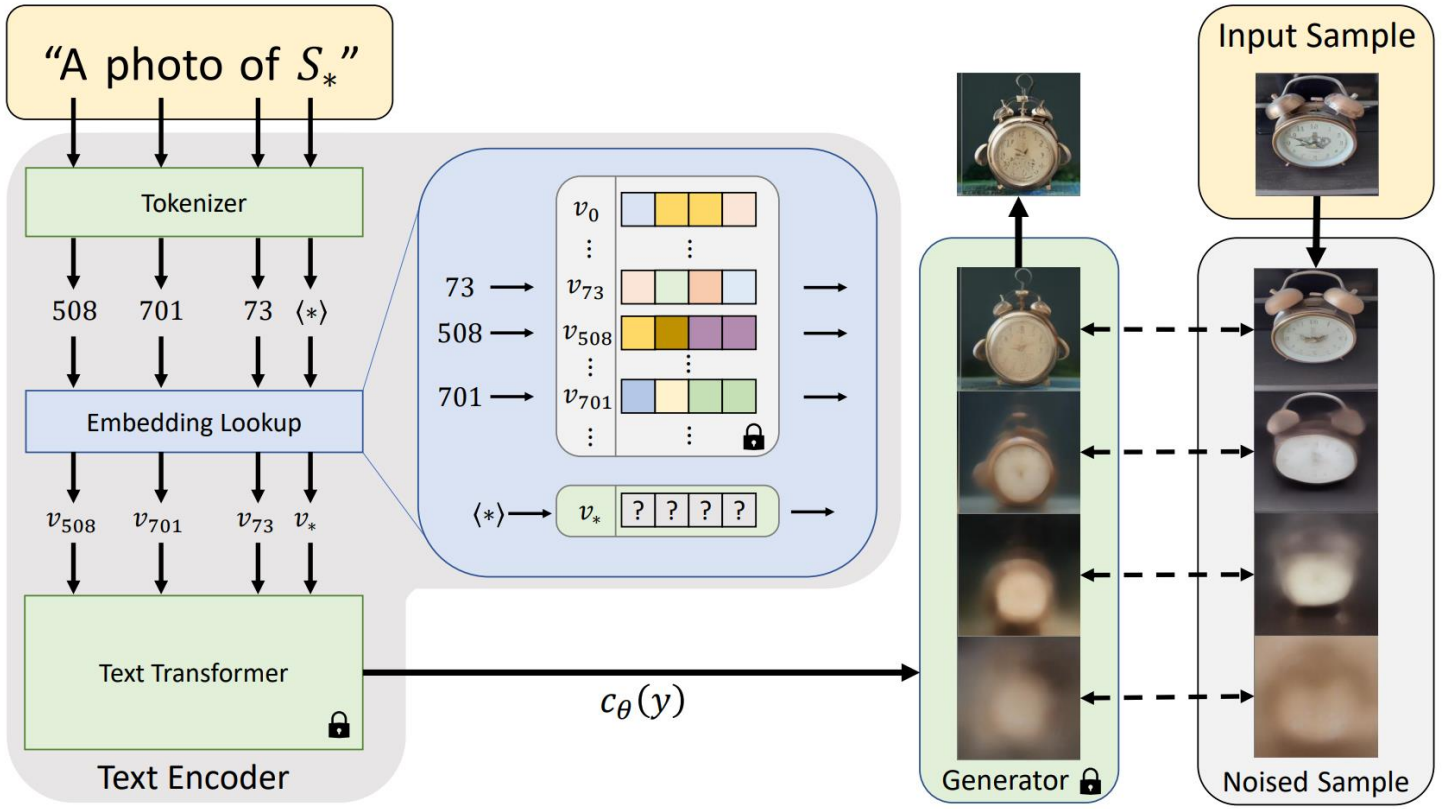
LoRA: maturemalemix

https://civitai.com/models/50882/maturemalemix

Textual inversion

Gal et al., NeurIPS 2022

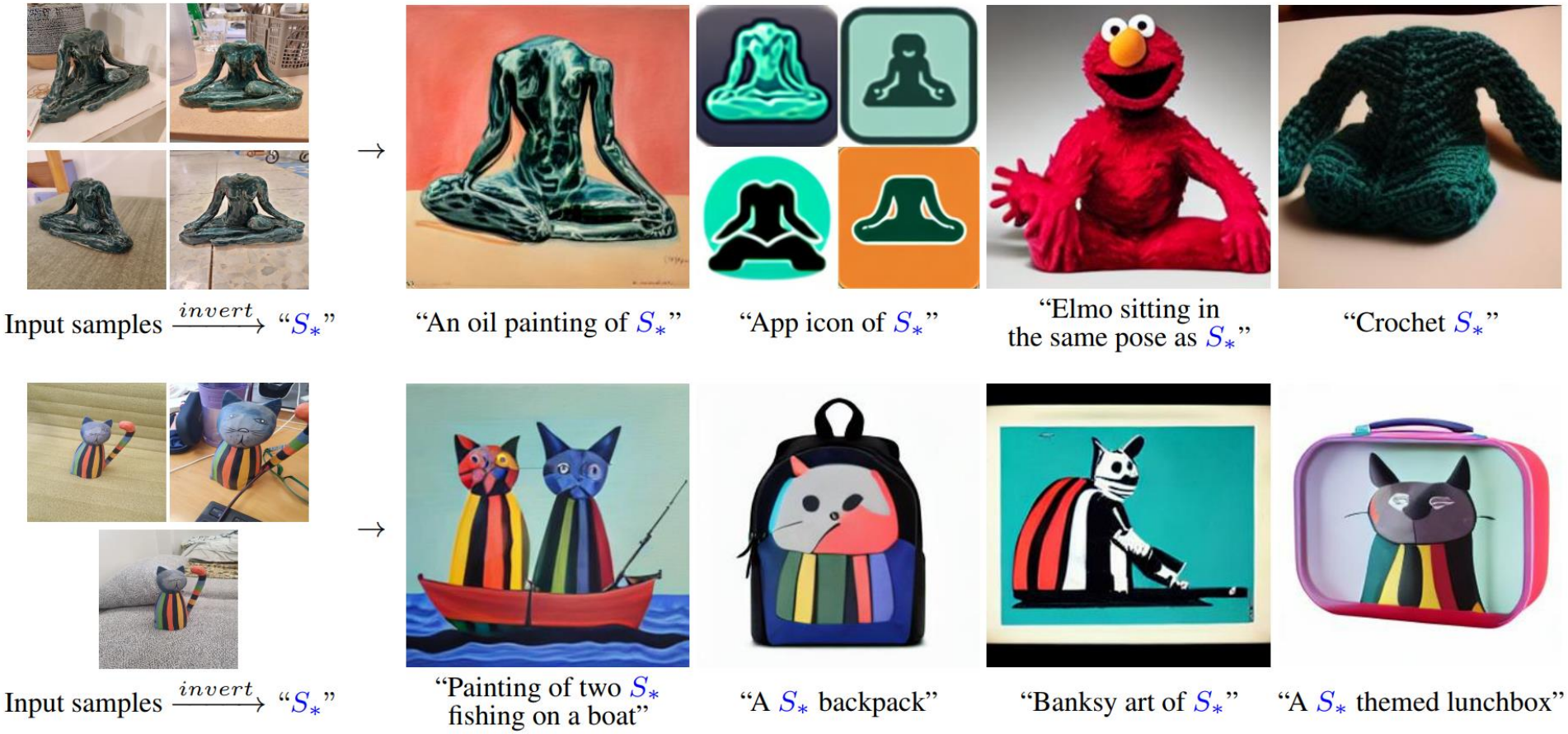
格式化文字模板



少量视觉样例

从输入的少量样本学习一个特殊的虚拟单词/词嵌入表示

Textual inversion：形状



Textual inversion：形态

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_*"

Textual inversion：组合

S_{style}

S_{clock}

S_{cat}

S_{craft}

"Photo of S_{clock} in the style of S_{style}"

"Photo of S_{cat} in the style of S_{style}"

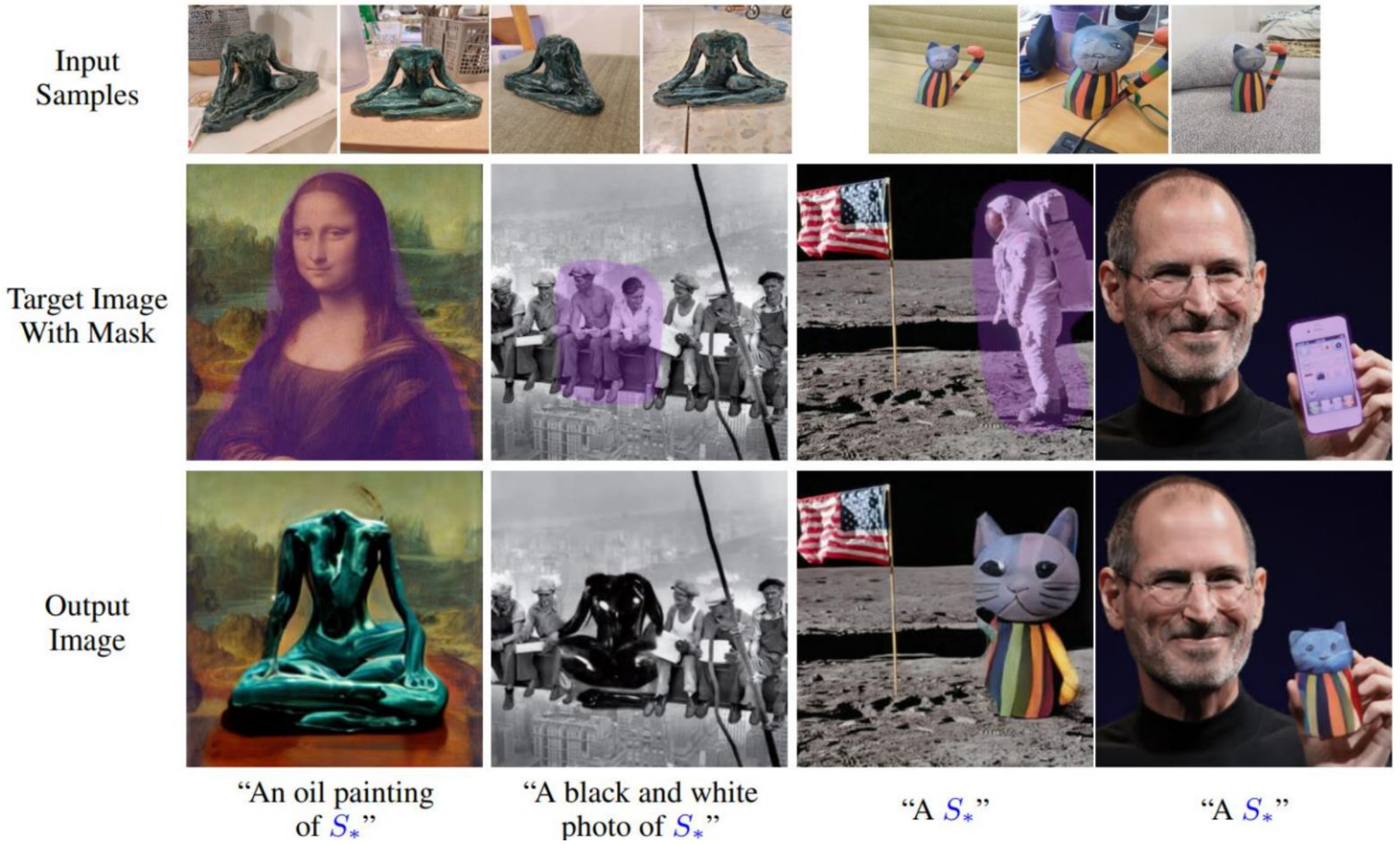
"Photo of S_{craft} in the style of S_{style}"

"Photo of S_{clock} in the style of S_{cat}"

"Photo of S_{clock} in the style of S_{craft}"

"Photo of S_{cat} in the style of S_{craft}"

Textual inversion：补全



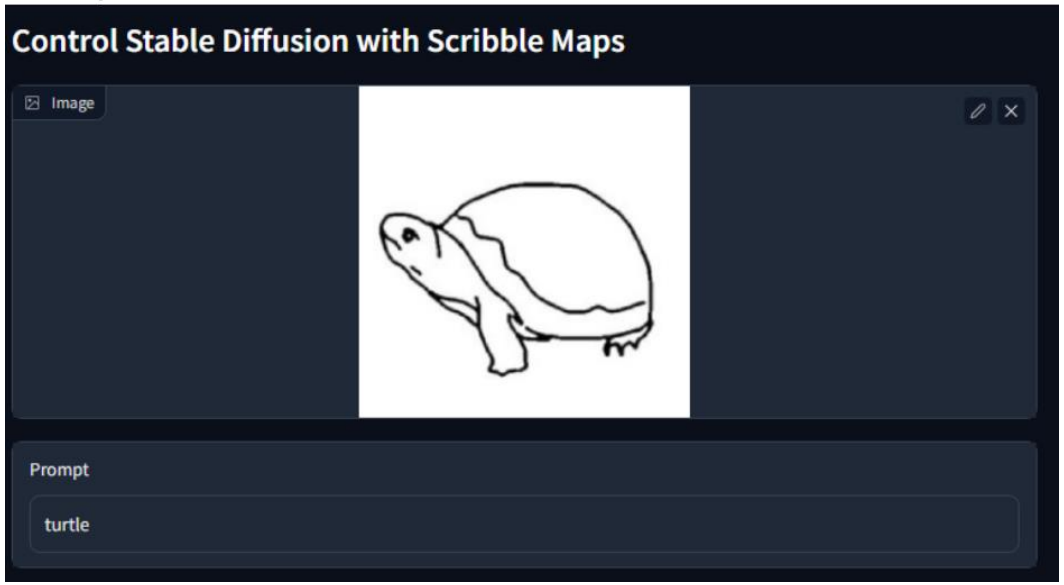
总结：个性化图像生成

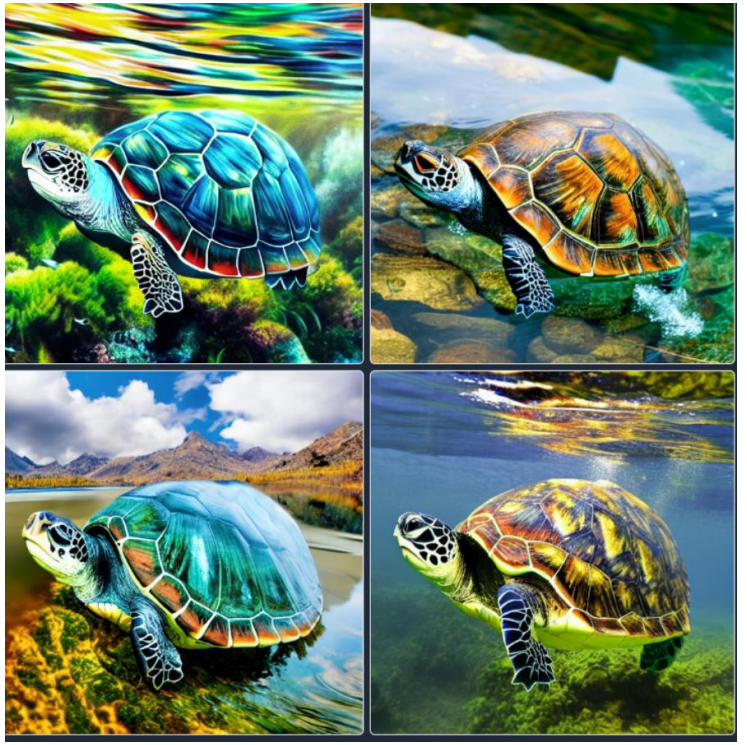
- 目标：
  - 少样本下微调，保留大模型泛化能力
- 方法：需要在小样本上梯度更新，参数不宜太多，更新不宜太大
  - 权重微调：**DreamBooth**
  - 轻量化微调：**Adapter**、**LoRA**
  - 词嵌入学习：**Textual inversion**

图像可控生成与编辑

图像可控生成与编辑

- 给定额外条件（如边缘图），控制生成图像的细节/编辑输入的图像

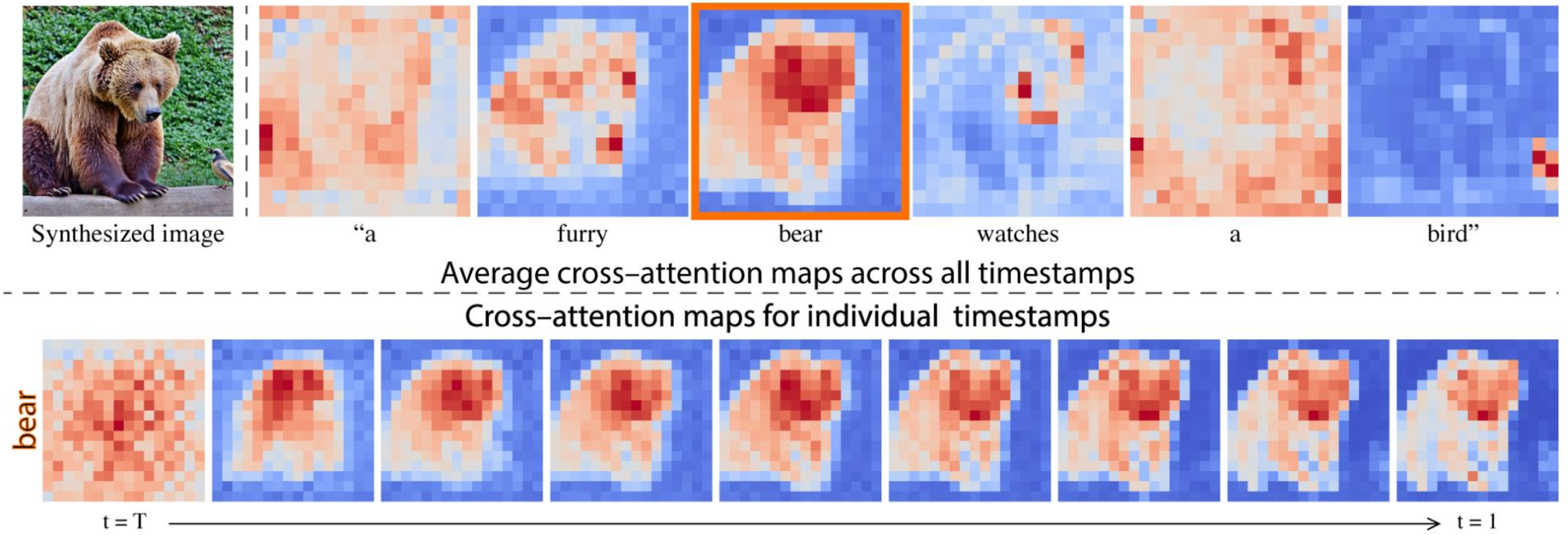




Prompt-to-prompt

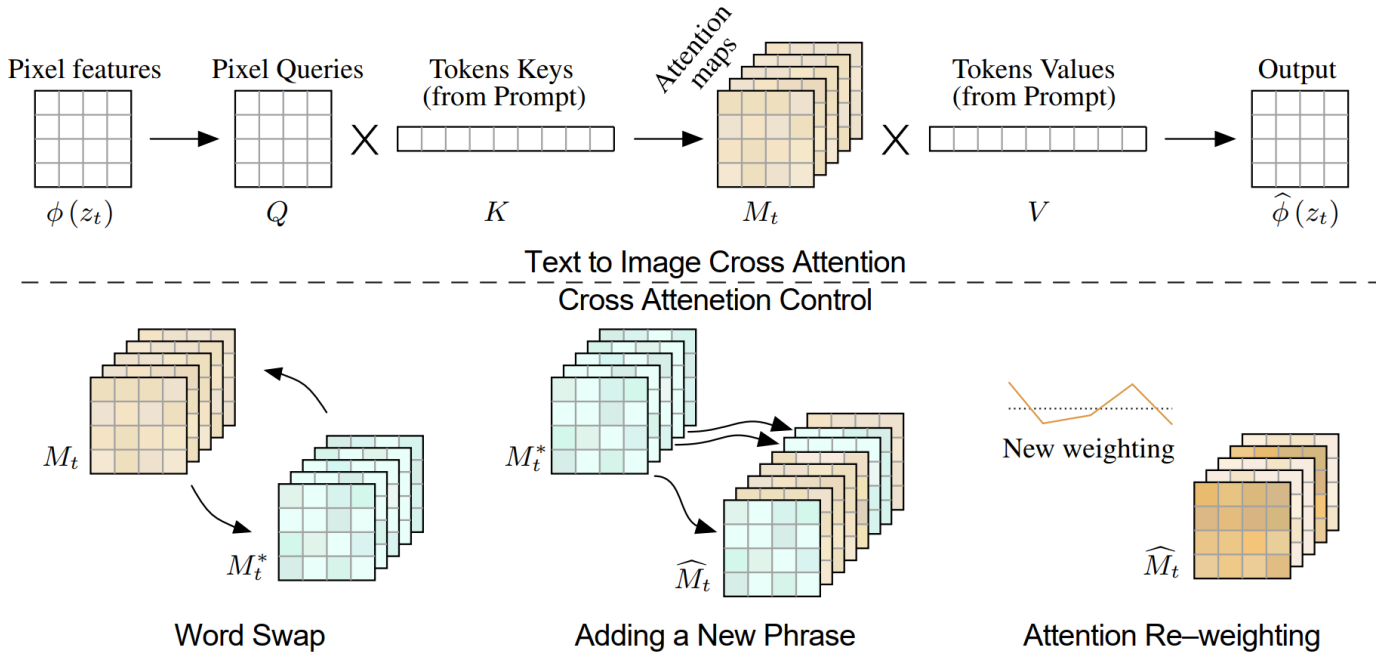
Amir et al., ICLR 2023

编辑生成的图像： 观察到Cross Attention 和 prompt 的对应关系



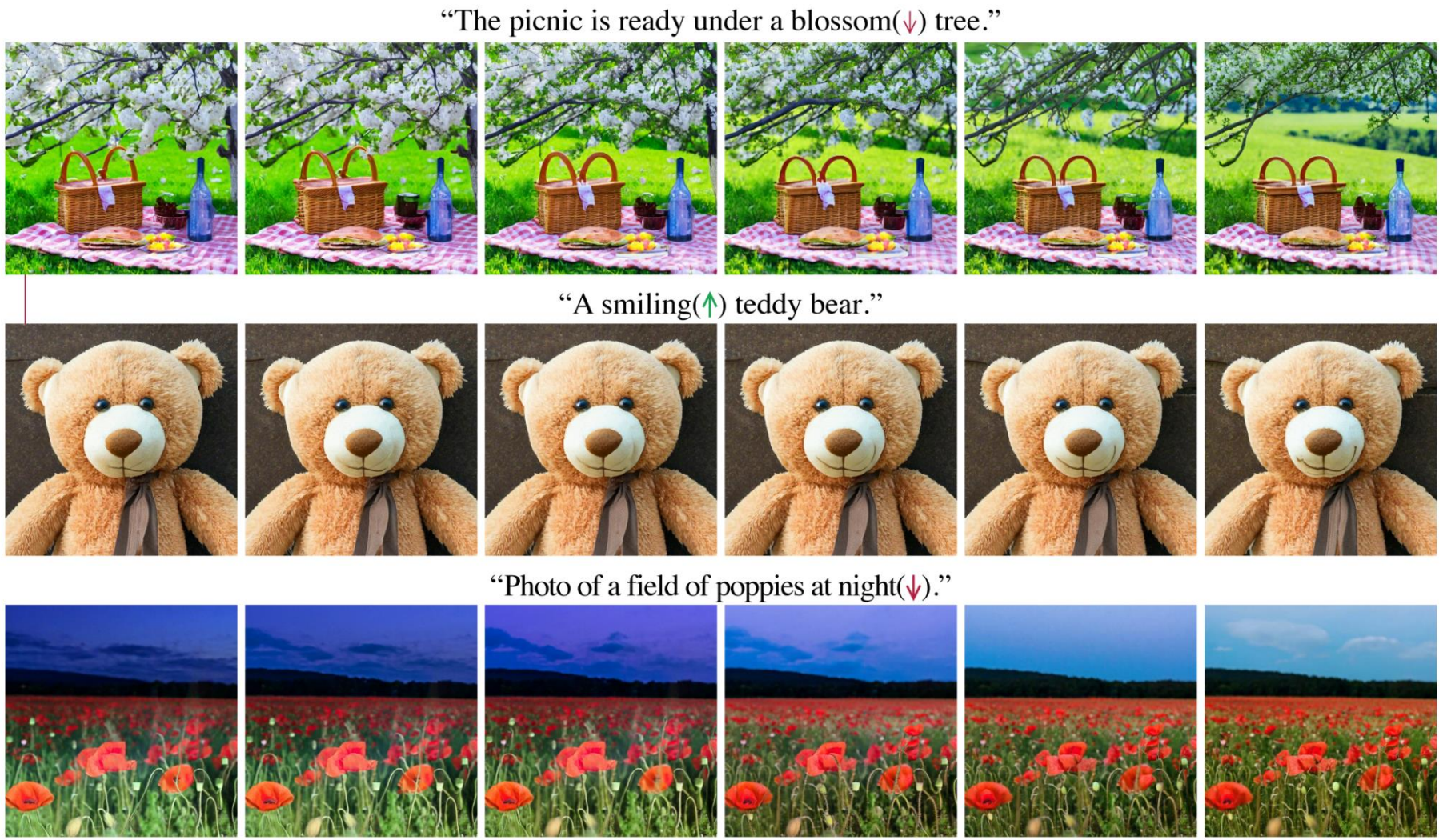
Prompt-to-prompt

根据两个 prompt 的区别修改 Cross Attention

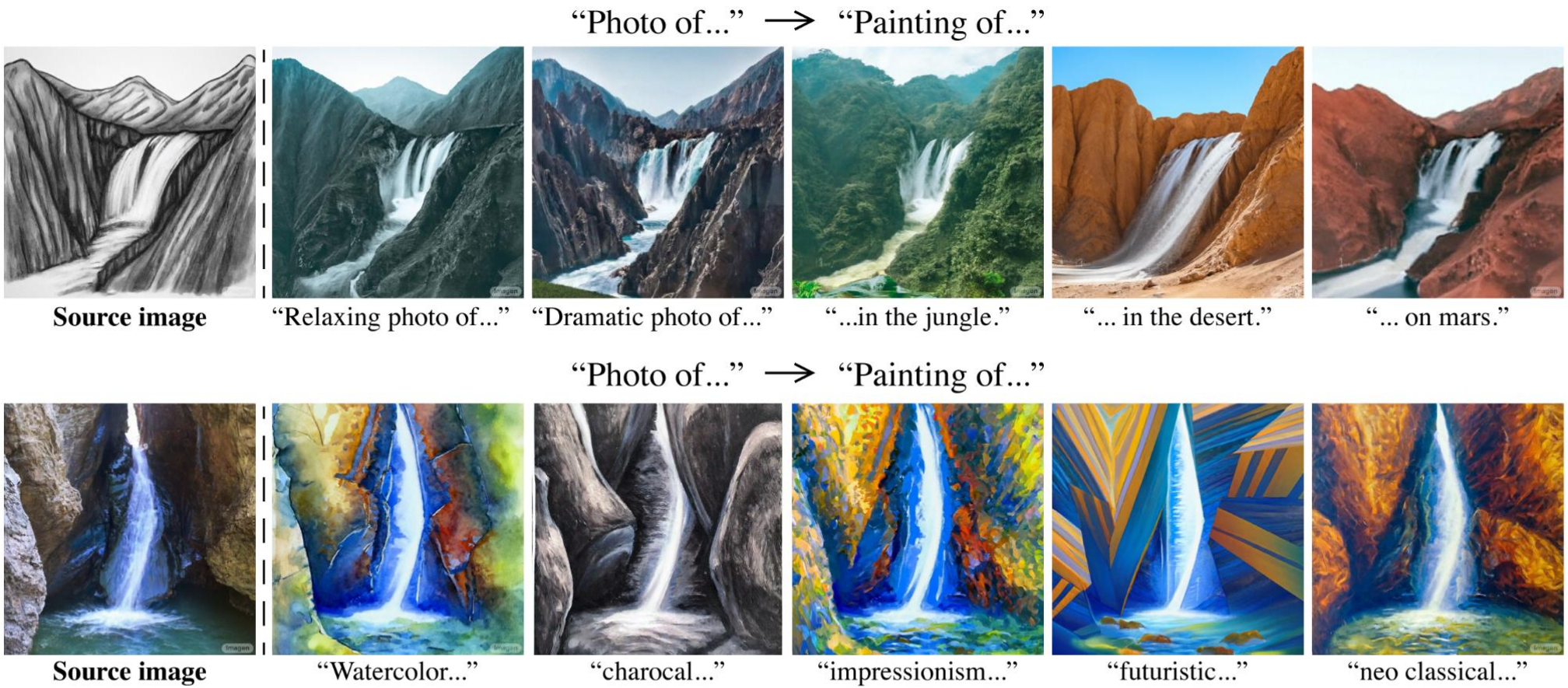


Prompt-to-prompt：替换

Prompt-to-prompt：重加权



Prompt-to-prompt：风格迁移



InstructPix2Pix：自动构造指令数据

Brooks et al, CVPR 2023

(1) 预训练语言模型生成文本

Input Caption: *"photograph of a girl riding a horse"* → GPT-3 (finetuned) → **Instruction:** *"have her ride a dragon"*
Edited Caption: *"photograph of a girl riding a dragon"*

(2) 预训练扩散模型生成成对图像

Input Caption: *"photograph of a girl riding a horse"*
Edited Caption: *"photograph of a girl riding a dragon"*
→ Stable Diffusion + Prompt2Prompt →

数据集展示

"have her ride a dragon"

"Color the cars pink"

"Make it lit by fireworks"

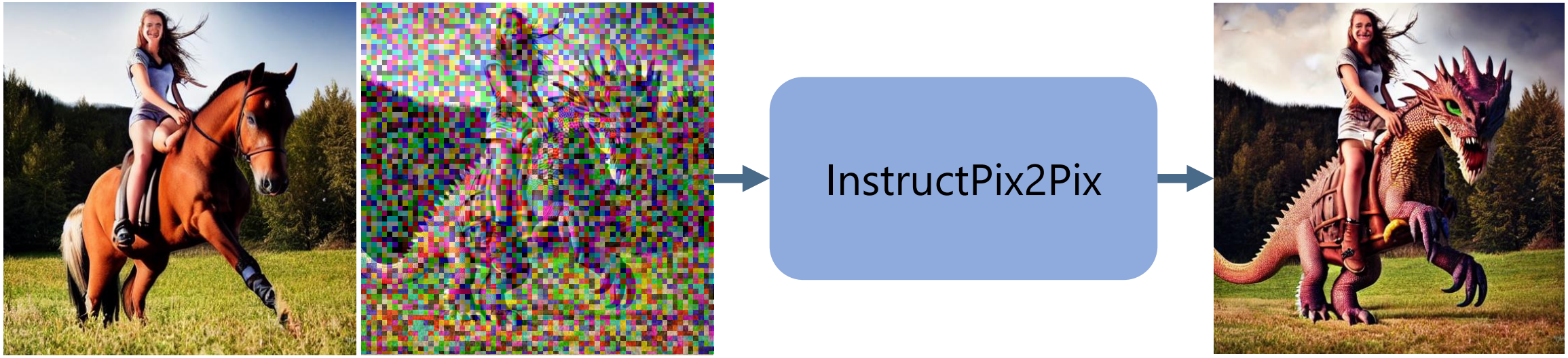
"convert to brick"

...

InstructPix2Pix：训练与测试

- 利用预训练模型自动生成的数据，微调 **Stable Diffusion**
- 直接泛化到人类指令和自然图像

“have her ride a dragon”



InstructPix2Pix

Input

"Add boats on the water"

"Replace the mountains with a city skyline"

Input

"It is now midnight"

"Add a beautiful sunset"

InstructPix2Pix

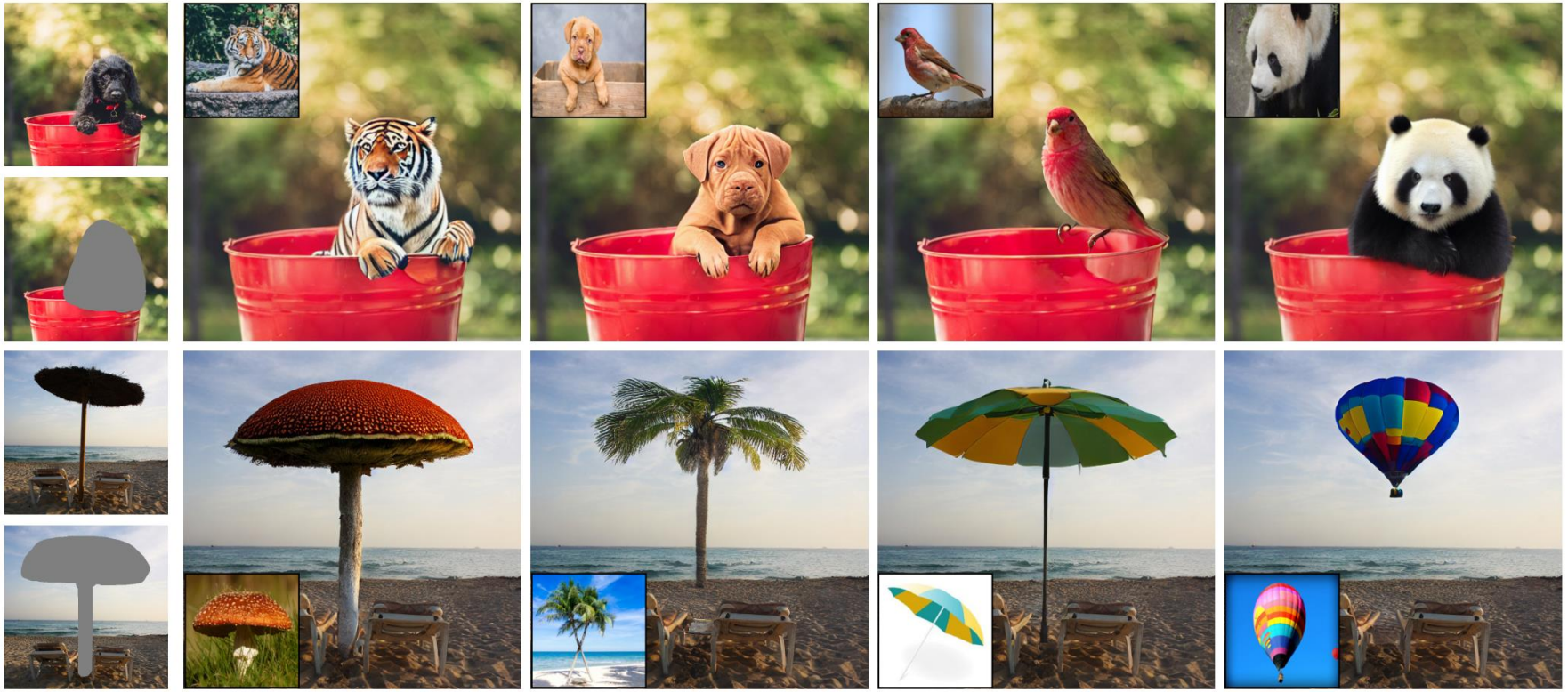
Input | *"Apply face paint"* | *"What would she look like as a bearded man?"* | *"Put on a pair of sunglasses"* | *"She should look 100 years old"*

"What if she were in an anime?" | *"Make her terrifying"* | *"Make her more sad"* | *"Make her James Bond"* | *"Turn her into Dwayne The Rock Johnson"*

Paint by example

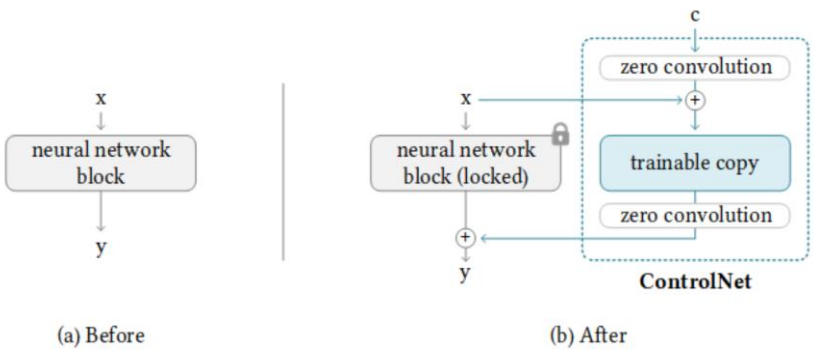
Yang et al, CVPR 2022

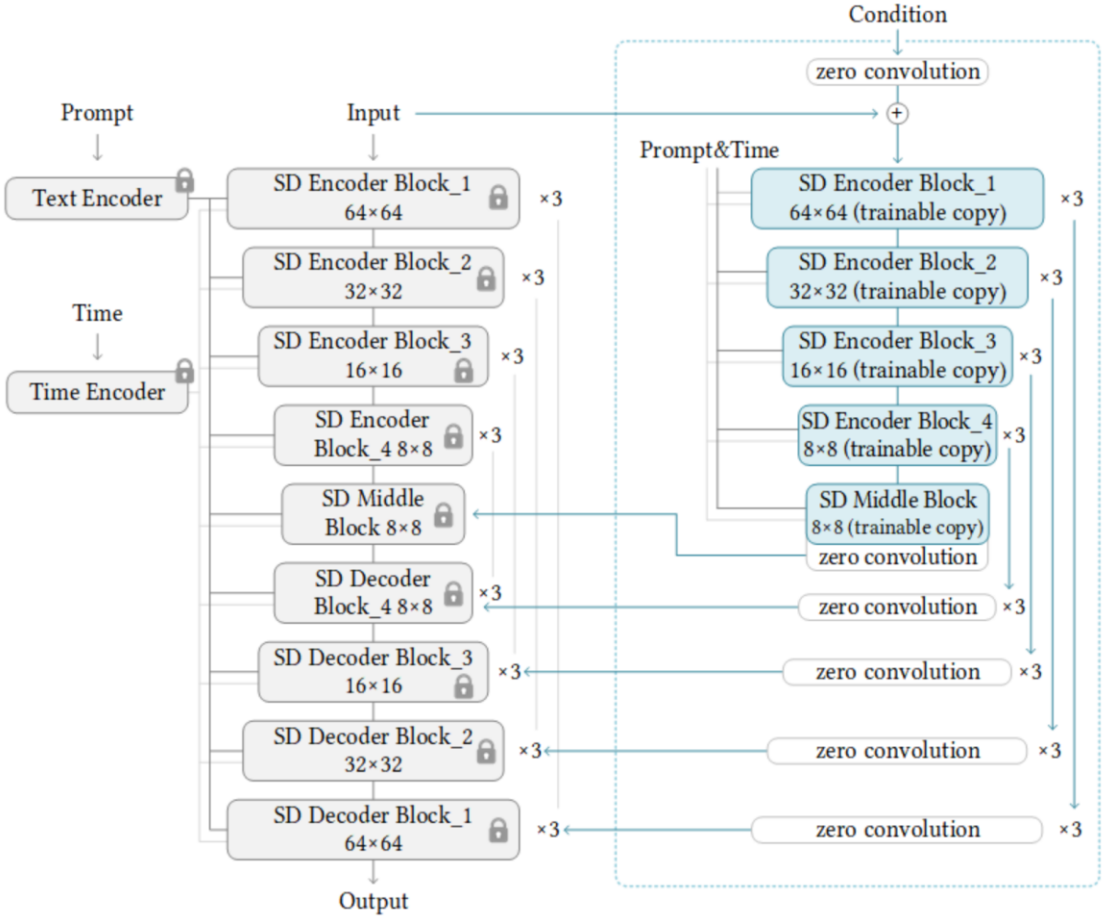
通过真实图像和掩码的方式构造数据：参考图片，掩码，被掩码图片与目标图片



ControlNet

- 零卷积初始化旁路，初始输出不变
- 复制参数备份进行多阶段微调
- 传统算法/预训练模型构建数据集

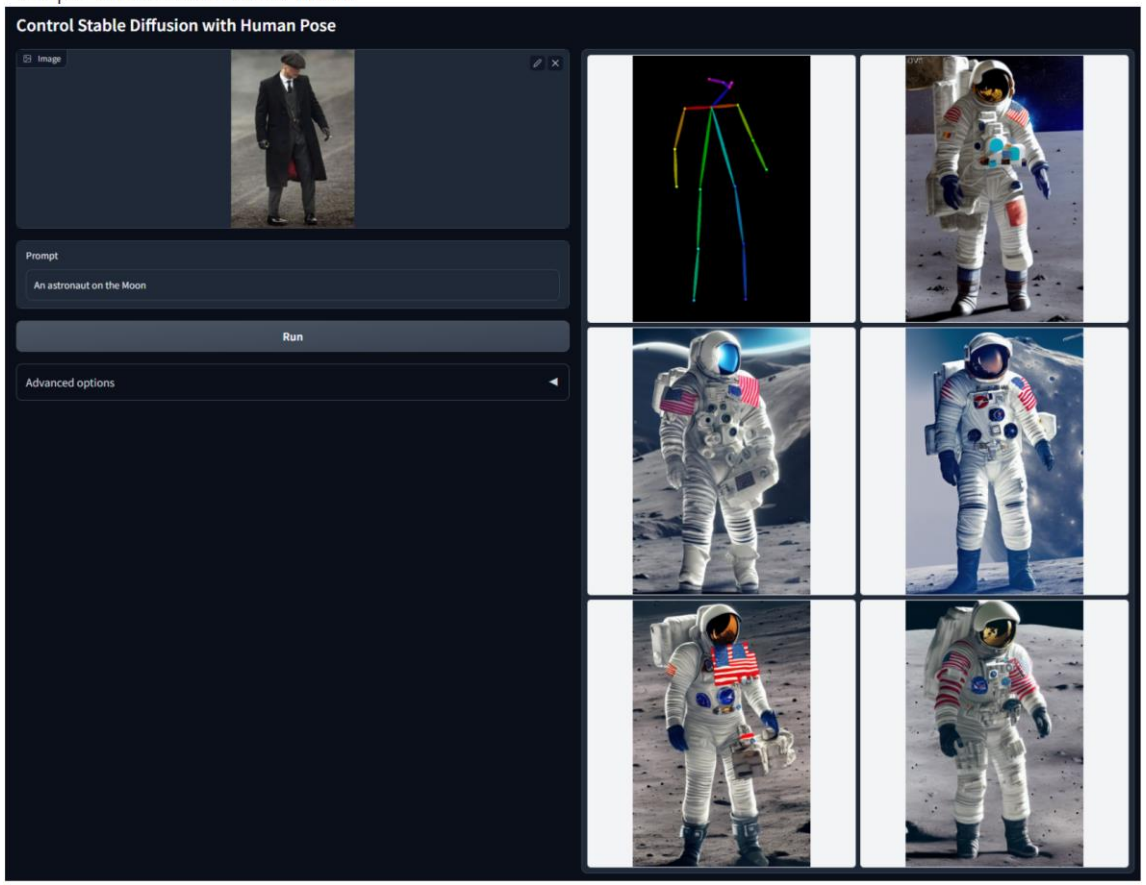




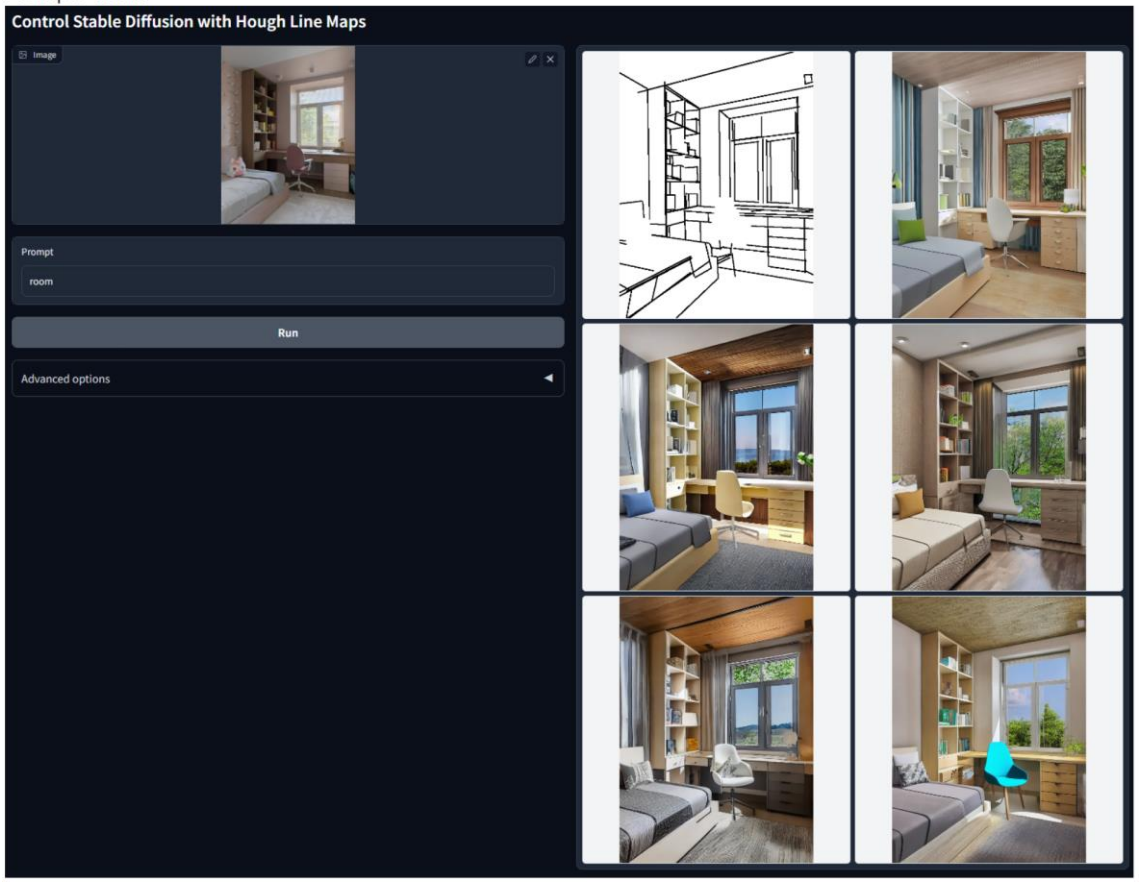
https://github.com/lllyasviel/ControlNet

ControlNet：不同条件权衡可编辑性和可控性

Prompt: "An astronaut on the moon"

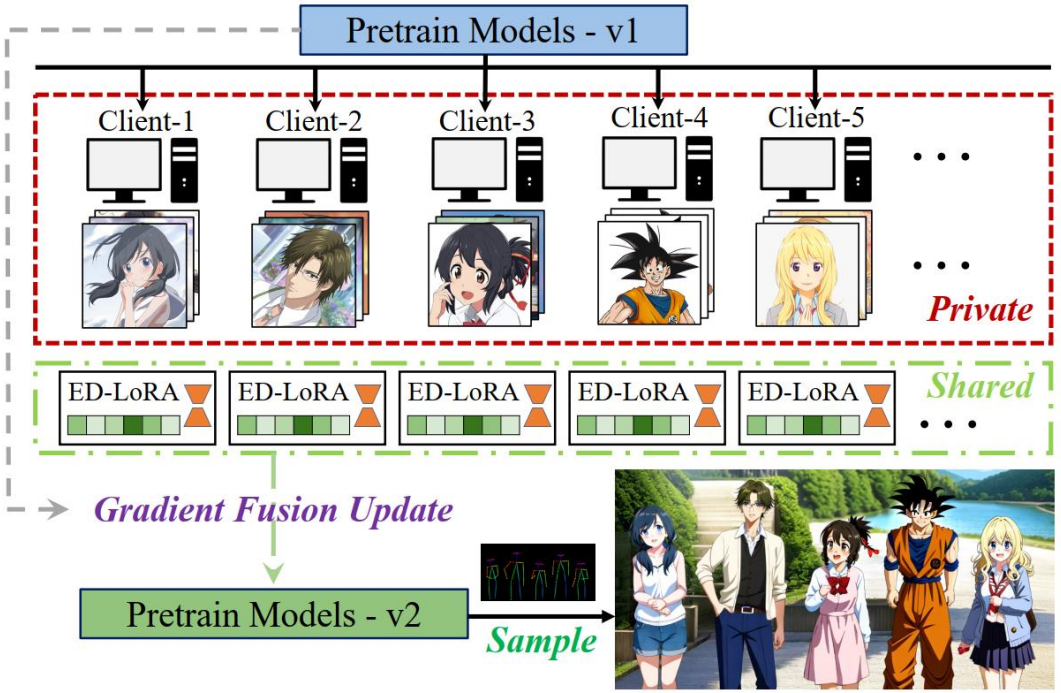


Prompt: "room"



ControlNet：多样化输出

Mix-of-show


Pretrain Models - v1
Client-1
Client-2
Client-3
Client-4
Client-5
Private
ED-LoRA
ED-LoRA
ED-LoRA
ED-LoRA
ED-LoRA
Shared
Gradient Fusion Update
Pretrain Models - v2
Sample


Customized Characters
V potter
V hermione
V thanos
V bengio
V lecun
V hinton
Customized Scenes
V pyramid
V rock
Customized Objects
V dogB
V cat
V dogA
V table
V chair
V vase
Mix-of-Show
V potter, V hermione and V thanos, near a castle.
V bengio, V lecun and V hinton, near a lake.
(a) Multi-Characters
V dogA, a V cat and a V dogB, on the grass,
Two V chair, V vase and V table, in a living room.
(b) Multi-Objects

总结：图像可控生成与编辑

- 图像可控生成与编辑
  - 给定文本、图像等条件，细粒度控制合成图像的语义
- 零样本方法
  - **Prompt-to-prompt**
- 预训练模型/传统方法/自监督方法构建监督信号
  - **InstructPix2Pix, T2I Adapter, ControlNet, Paint-by-example, Composer**
  - 和个性化生成任务有重叠，但是不需要给定特定输入再训练

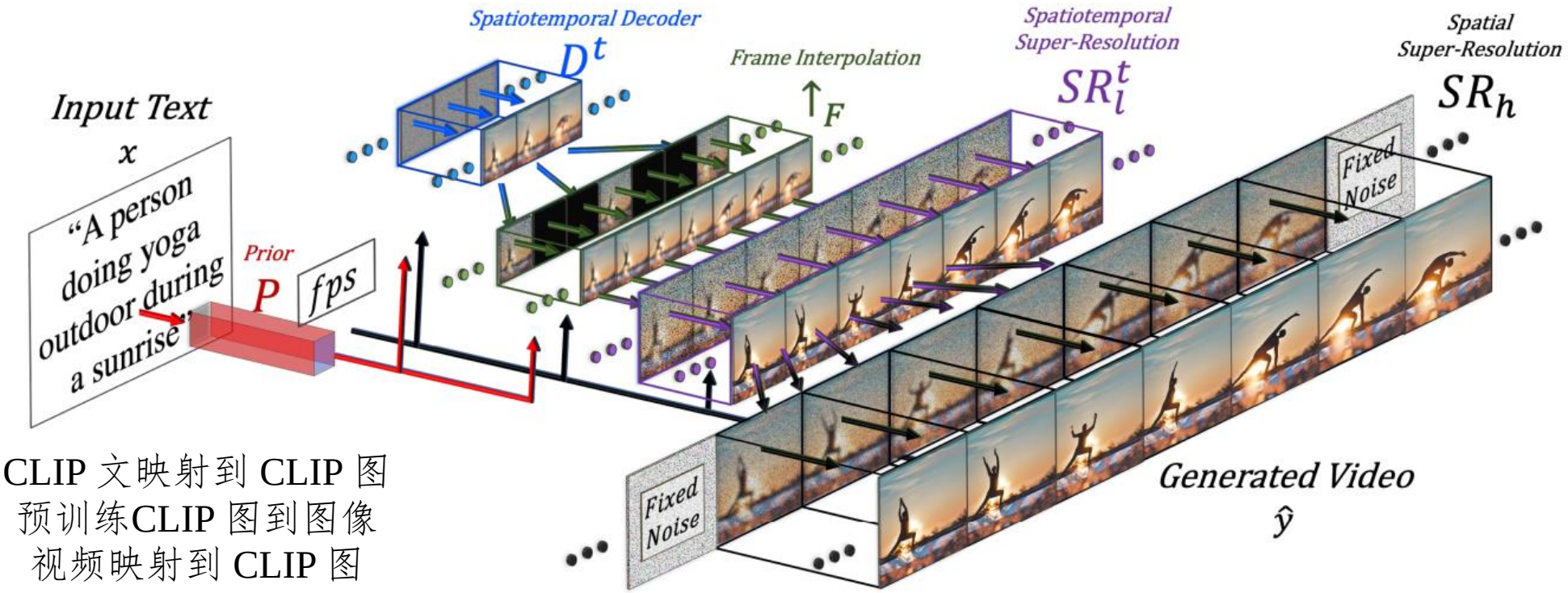
视频可控生成与编辑

视频可控生成与编辑

- 给定指导，可控视频生成/编辑输入的视频，额外保证时间一致性

Make-A-Video：大规模视频数据上训练文到视频生成

Singer et al., Arxiv preprint



CLIP 文映射到 CLIP 图
预训练CLIP 图到图像
视频映射到 CLIP 图

文到图初始化空间卷积，加入时间卷积和注意力，**20M** 视频数据（不需要文本）

Make-A-Video

(a) A dog wearing a superhero outfit with red cape flying through the sky.

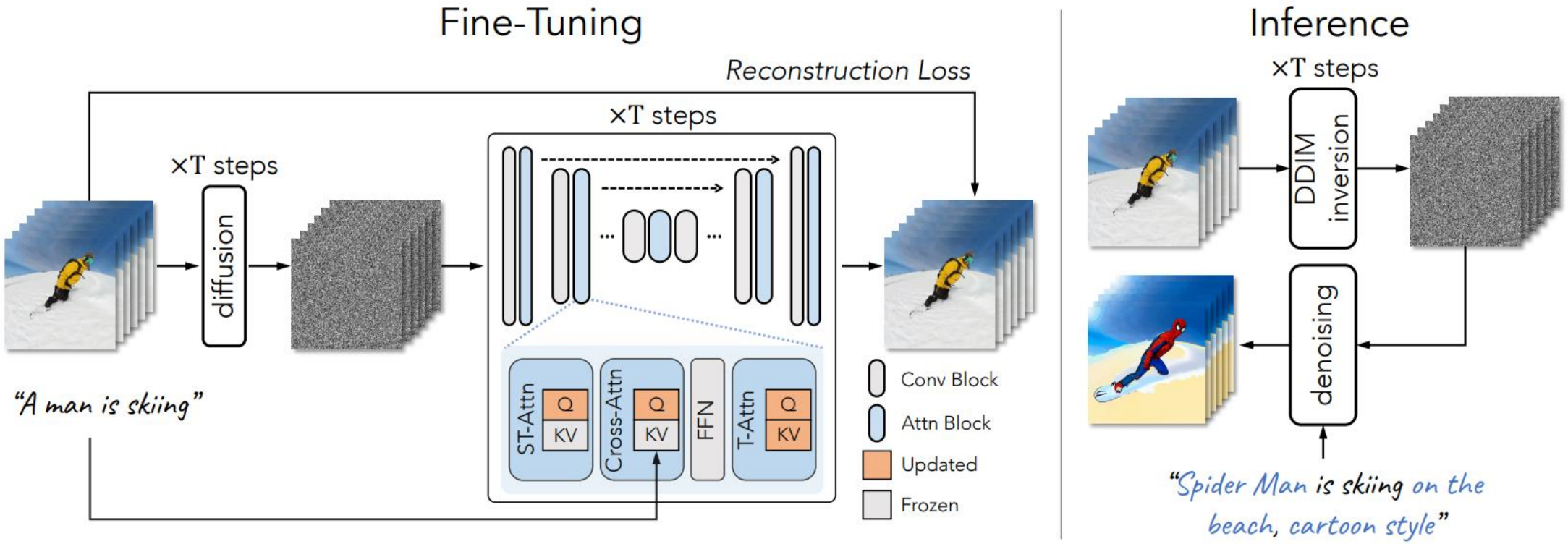
(b) There is a table by a window with sunlight streaming through illuminating a pile of books.

(c) Robot dancing in times square.

(d) Unicorns running along a beach, highly detailed.

Tune-A-Video： 单样本视频-文本数据做可控视频编辑

Wu et al., Arxiv preprint, 2022



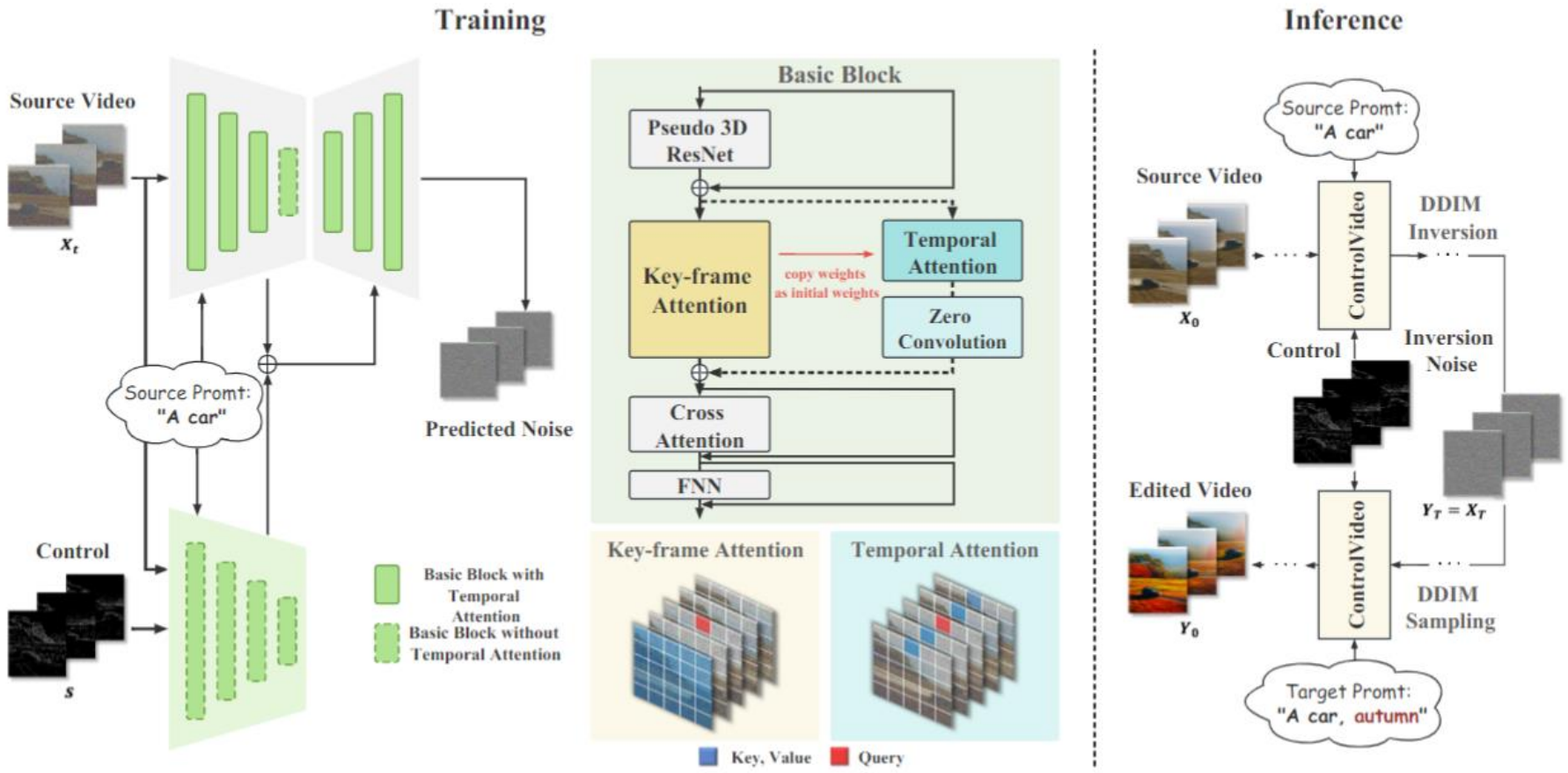
额外的时间一致性注意力层，单视频微调

Tune-A-Video

Tune-A-Video

ControlVideo：单样本视频-文本数据做细粒度可控视频编辑

Zhao et al., Arxiv preprint, 2023



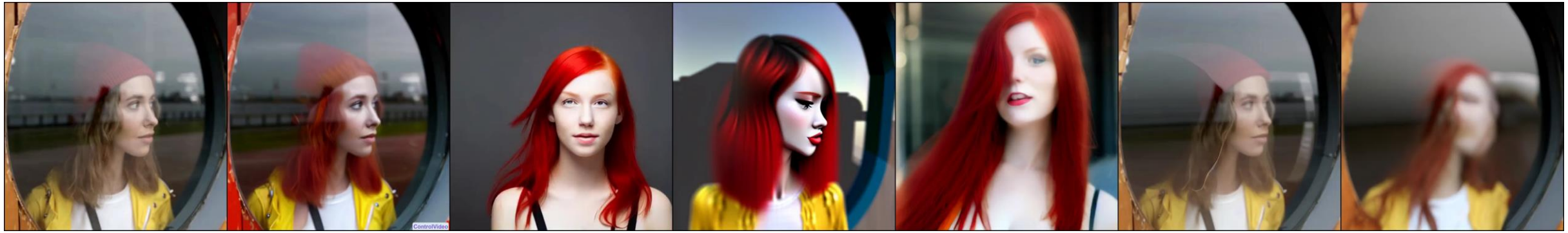
加入 **ControlNet**，加入新的注意力机制，并精心初始化，单视频微调

ControlVideo

https://ml.cs.tsinghua.edu.cn/controlvideo/

更好的时间一致性、更符合原视频约束

+红发



+动漫风格

https://ml.cs.tsinghua.edu.cn/controlvideo/

总结：视频可控生成与编辑

- 视频可控生成与编辑
  - 给定文本、图像等条件，细粒度控制合成视频的语义
- 大规模视频数据训练（消除文本标注需求，借助预训练文到图模型）
  - **Make-A-Video, Imagen-Video, NUWA-XL**
- 少样本、零样本视频编辑（借助预训练文到图模型和图像可控编辑与生成算法）
  - **Tune-A-Video, Video-P2P, Vid2Vid-Zero, FateZero, Edit-A-Video, ControlVideo**

文到三维场景生成

文到三维场景生成

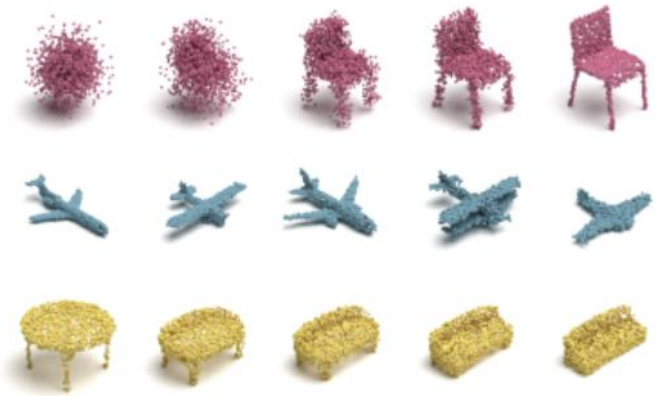
A blue jay standing on a large basket of rainbow macarons.

利用预训练模型
零样本迁移

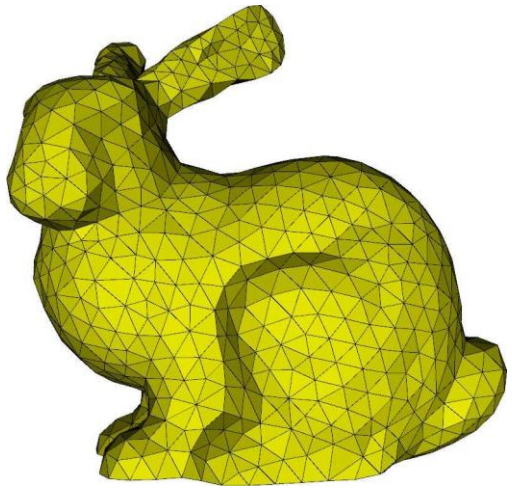
文到图预训练模型：强泛化、开放域

三维物体生成等下游任务数据量匮乏

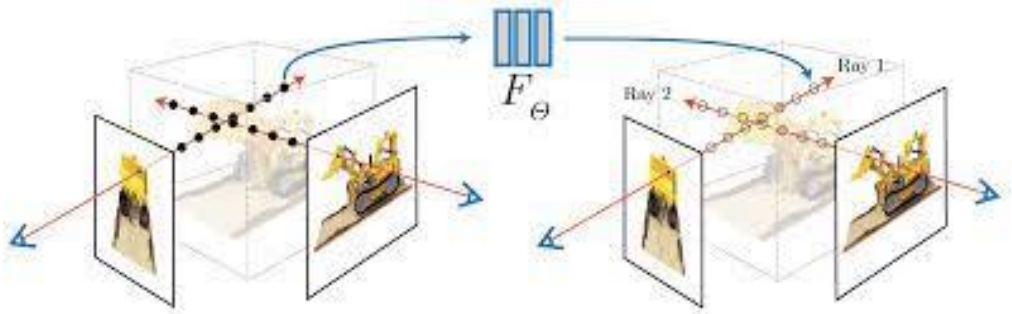
三维场景表示



点云



网格



神经辐射场

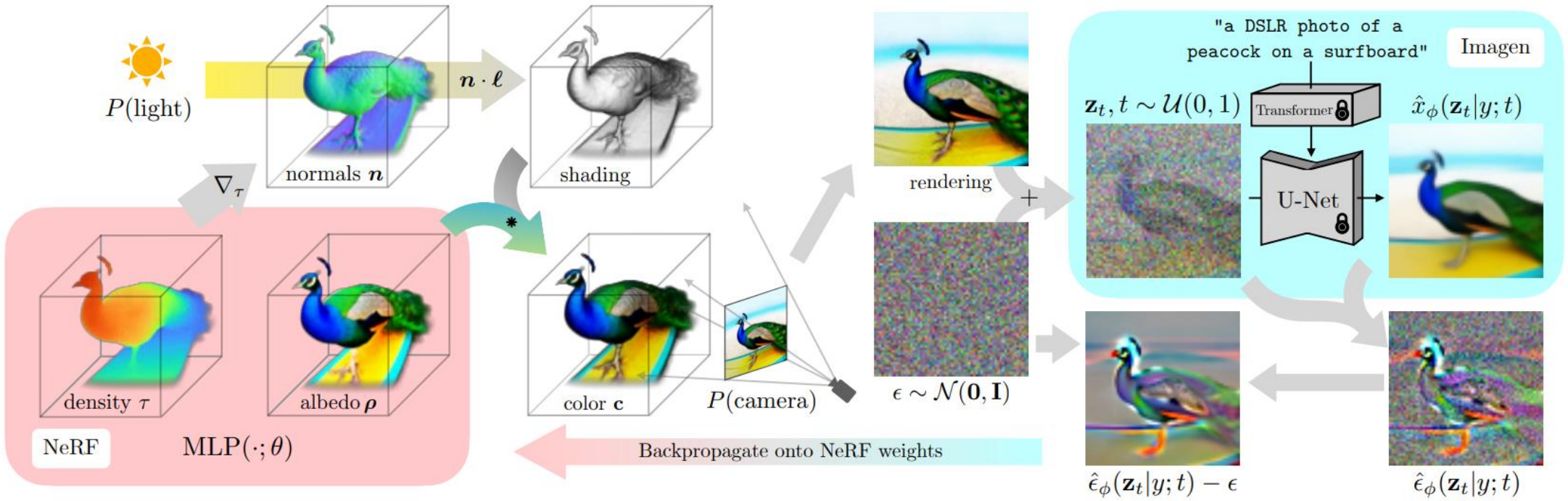
需要（可微分地）渲染二维图像

DreamFusion

"a DSLR photo of a peacock on a surfboard"

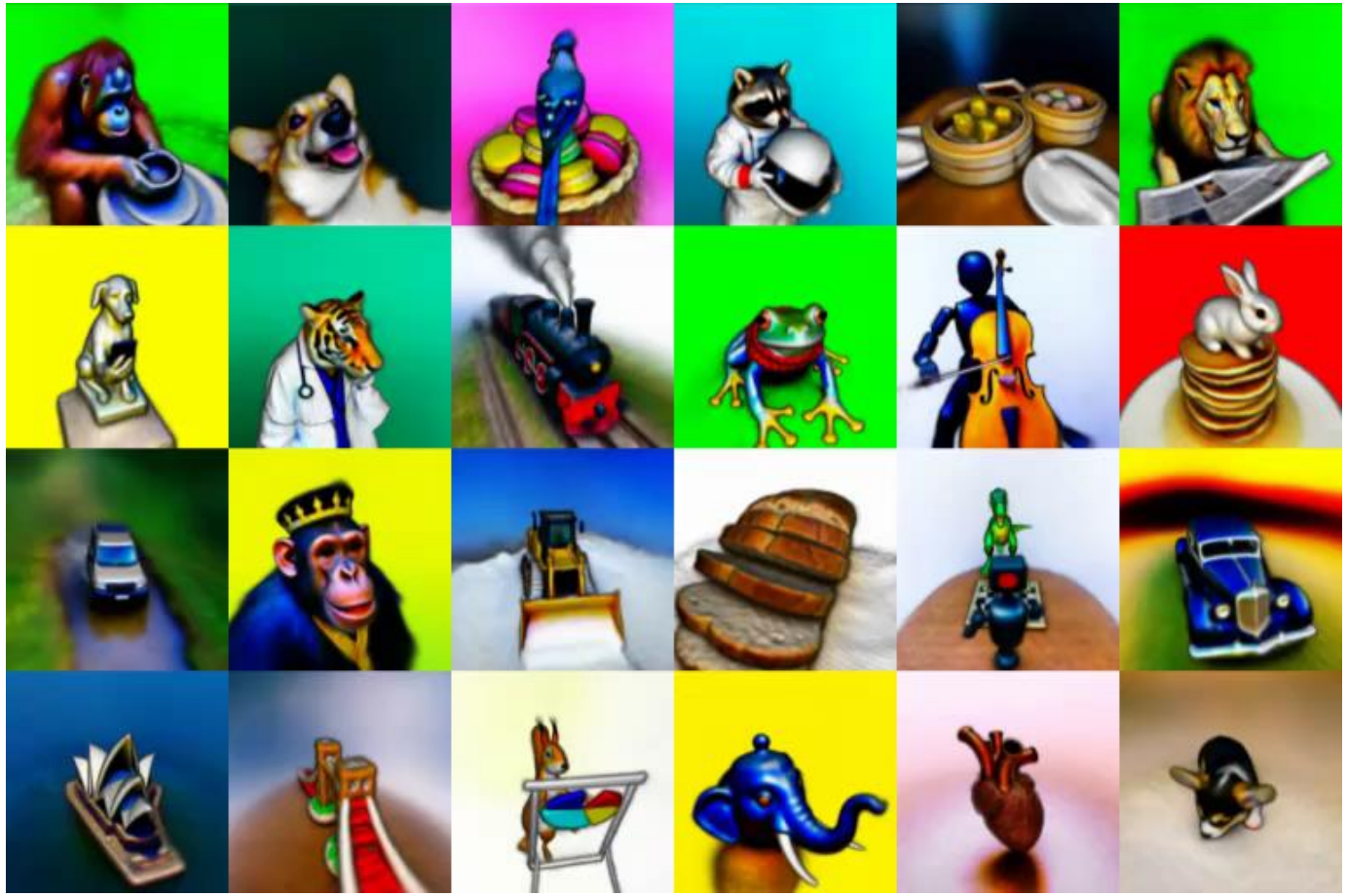
DreamFusion
Automatic text-to-3D

DreamFusion：无需三维数据的文到3D生成

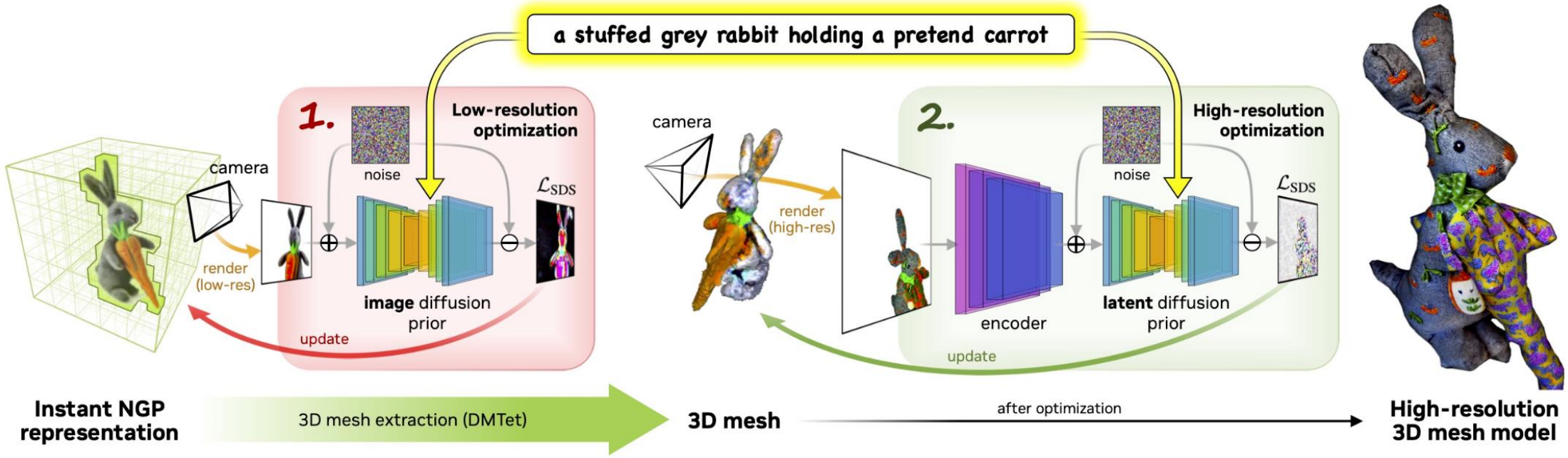


$$\nabla_\theta \mathcal{L}_{\text{SDS}}(\phi, \mathbf{x} = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[w(t) \left(\hat{\epsilon}_\phi(\mathbf{z}_t; y, t) - \epsilon\right) \frac{\partial \mathbf{x}}{\partial \theta} \right]$$

DreamFusion：零样本文到三维数据合成



Magic3D



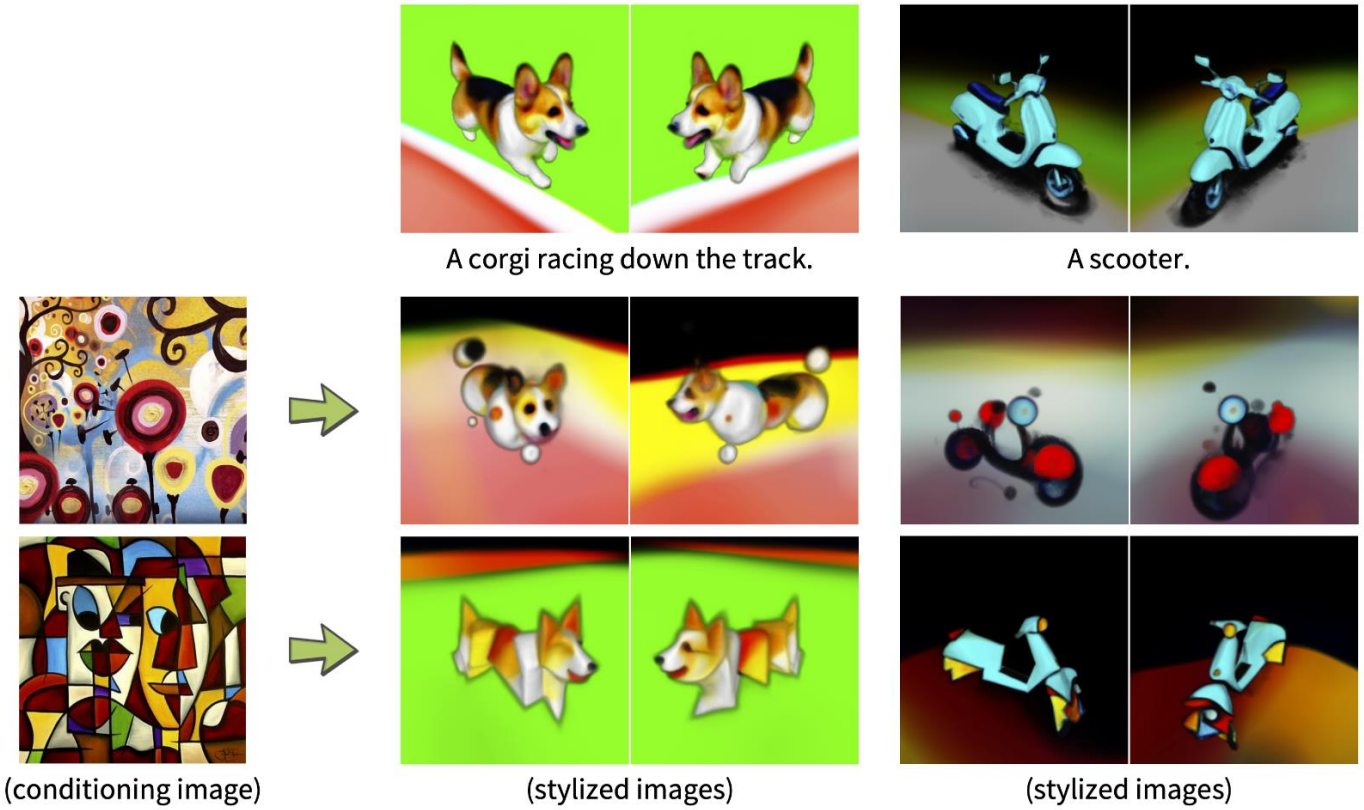
两阶段优化：低精度优化生成NeRF，自动提取Mesh，高精度优化生成Mesh，均用 SDS

Magic3D

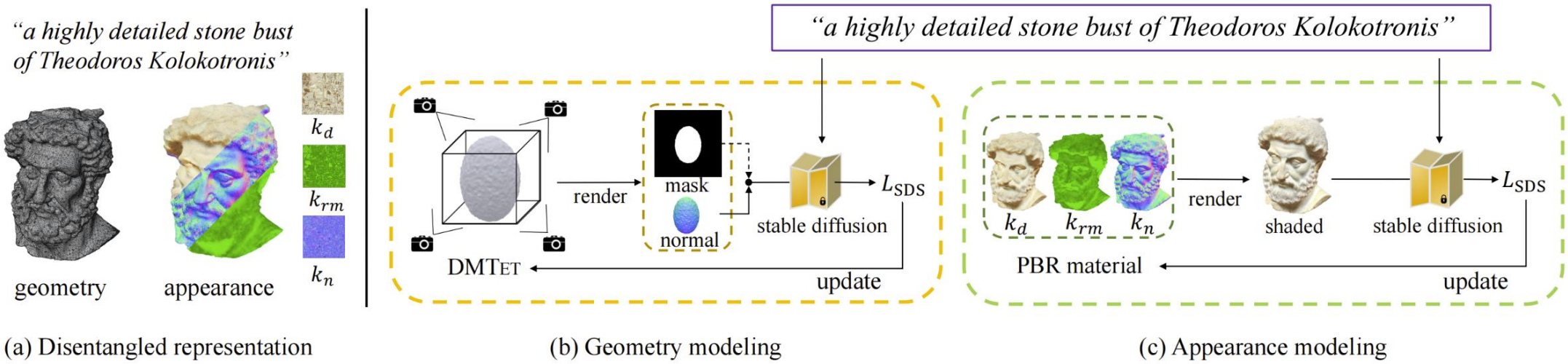

Magic3D :
High-Resolution Text-to-3D Content Creation
Chen-Hsuan Lin* Jun Gao* Luming Tang* Towaki Takikawa* Xiaohui Zeng*
Xun Huang Karsten Kreis Sanja Fidler# Ming-Yu Liu# Tsung-Yi Lin
*# : equal contributions
NVIDIA Corporation

Magic3D + DreamBooth

We can also condition the diffusion model (eDiff-I) on an input image to transfer its style to the output 3D model.



Fantasia3D



两阶段优化：Mesh 几何结构和纹理，均用 SDS

Fantasia3D

过饱和等问题

渲染图片的质量远远低于从扩散模型中采样的质量！

研究动机：算法导致了过饱和问题

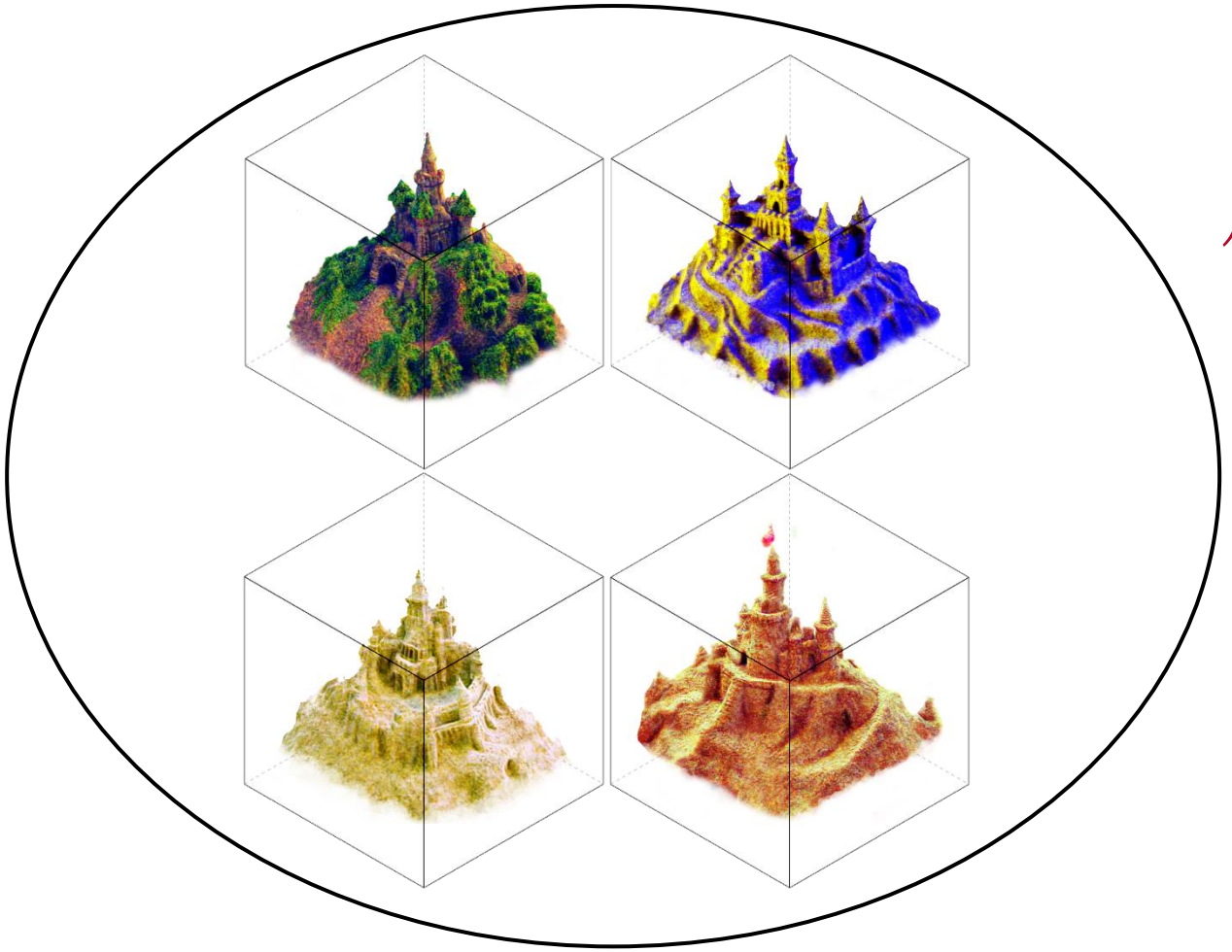
(a) SDS [31] (CFG = 7.5)

(b) SDS [31] (CFG = 100)

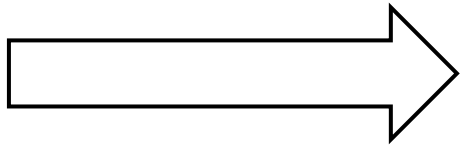
(c) Ancestral sampling [25] (CFG = 7.5)

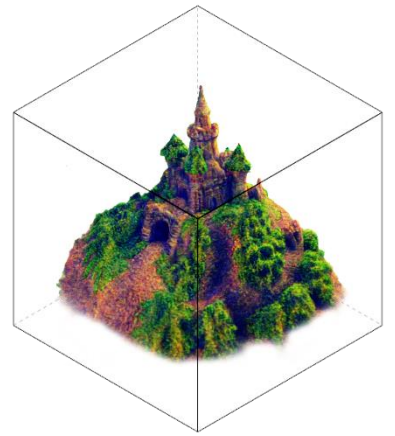
(d) VSD (CFG = 7.5, **ours**)

文到3D生成中的不确定性

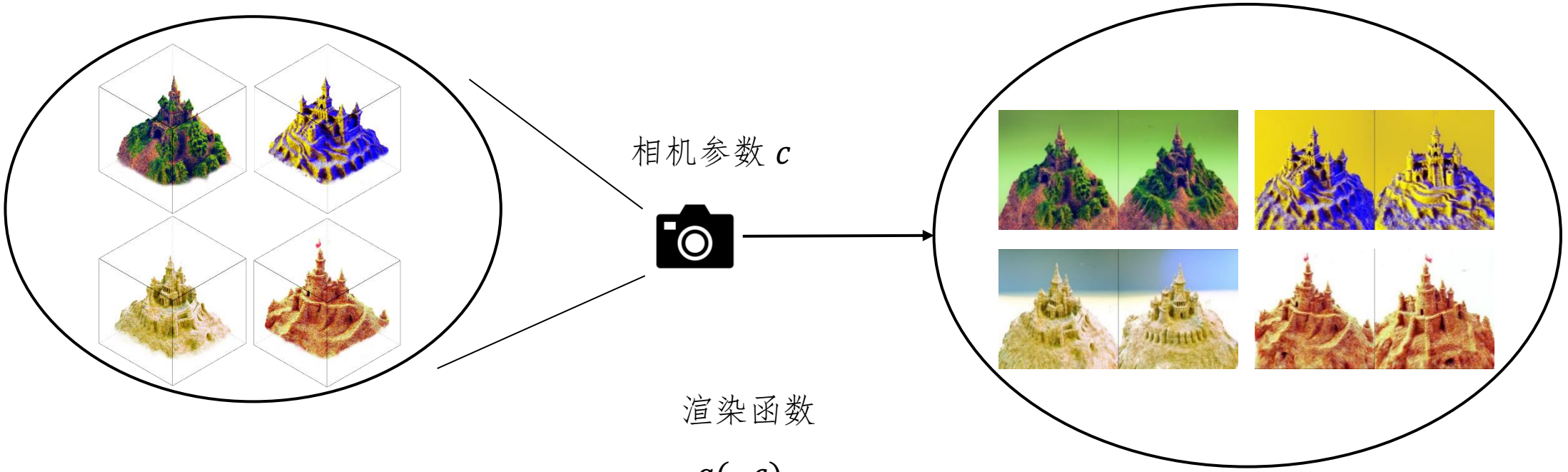


从一个隐式三维分布中采样，
模仿从扩散模型中的采样

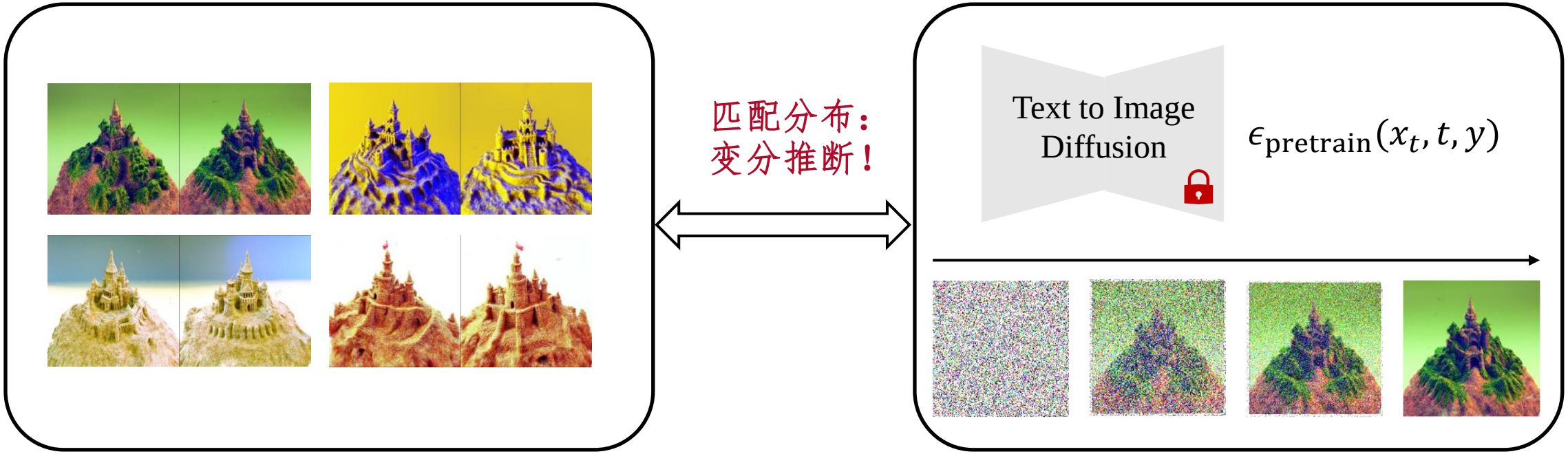




渲染图像的隐式分布



文到3D生成：变分推断视角



$$\min_{\mu} KL(q_0^{\mu}(x_0|y)||p_0(x_0|y))$$

渲染图像的隐式分布　　预训练图文模型的分布

基于粒子的变分推断：梯度更新规则

基于粒子的变分推断 $\nabla_\theta \mathcal{L}_{\text{VSD}}(\theta) \triangleq \mathbb{E}_{t,\epsilon,c}\left[\omega(t)\left(\epsilon_{\text{pretrain}}(\boldsymbol{x}_t, t, y) - \epsilon_\phi(\boldsymbol{x}_t, t, c, y)\right)\frac{\partial \boldsymbol{g}(\theta, c)}{\partial \theta}\right]$

预训练模型的噪声预测网络　渲染图像上的噪声预测 LoRA

$$\nabla_\theta \mathcal{L}_{\text{SDS}}(\theta) \triangleq \mathbb{E}_{t,\epsilon,c}\left[\omega(t)\left(\epsilon_{\text{pretrain}}(\boldsymbol{x}_t, t, y) - \epsilon\right)\frac{\partial \boldsymbol{g}(\theta, c)}{\partial \theta}\right]$$

- **SDS是VSD的特例：** 粒子个数为1且变分分布为单点 $\mu(\theta|y) \approx \delta(\theta - \theta^{(1)})$
- **VSD近似分布然后采样，因此 CFG 数值和采样效果都类似于祖先采样**

VSD 解决过饱和等问题

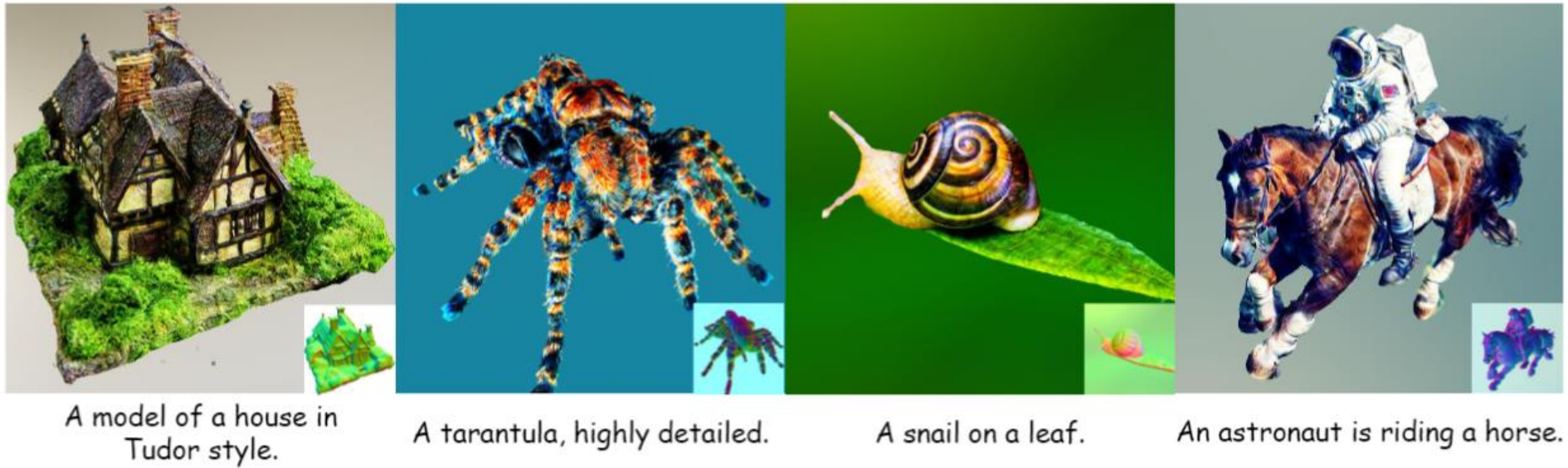
(a) SDS [31] (CFG = 7.5)

(b) SDS [31] (CFG = 100)

(c) Ancestral sampling [25] (CFG = 7.5)

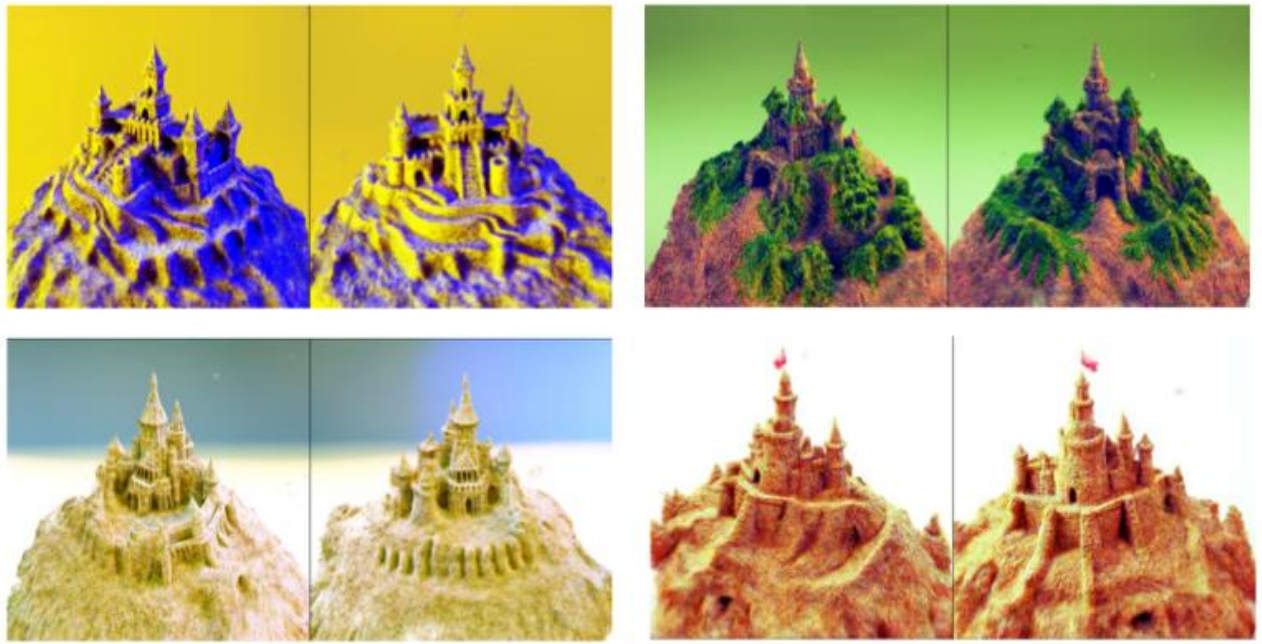
(d) VSD (CFG = 7.5, **ours**)

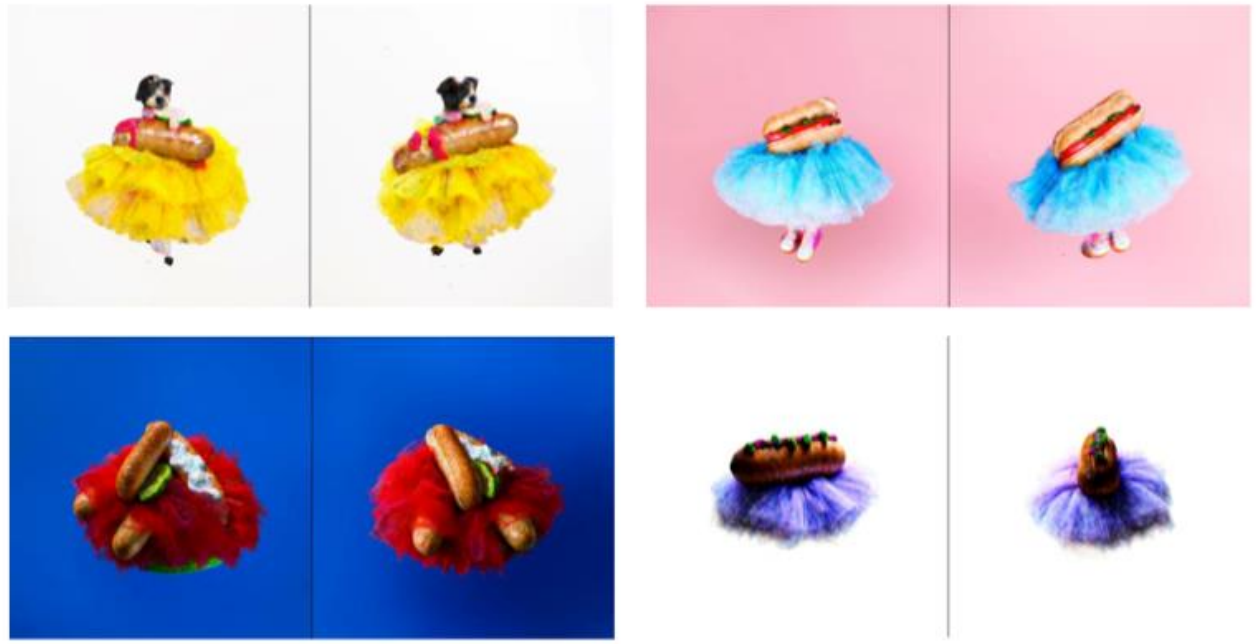
ProlificDreamer：纹理逼真的三维网格



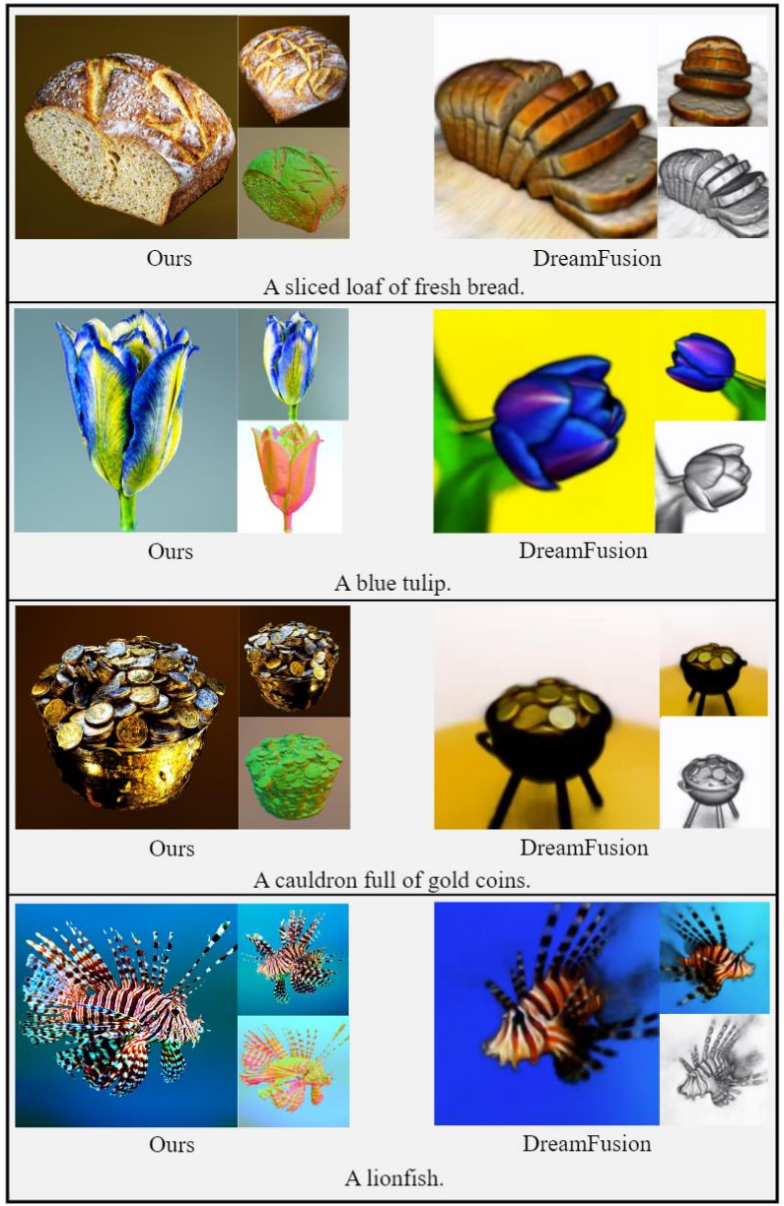
ProlificDreamer：512 渲染精度的神经辐射场

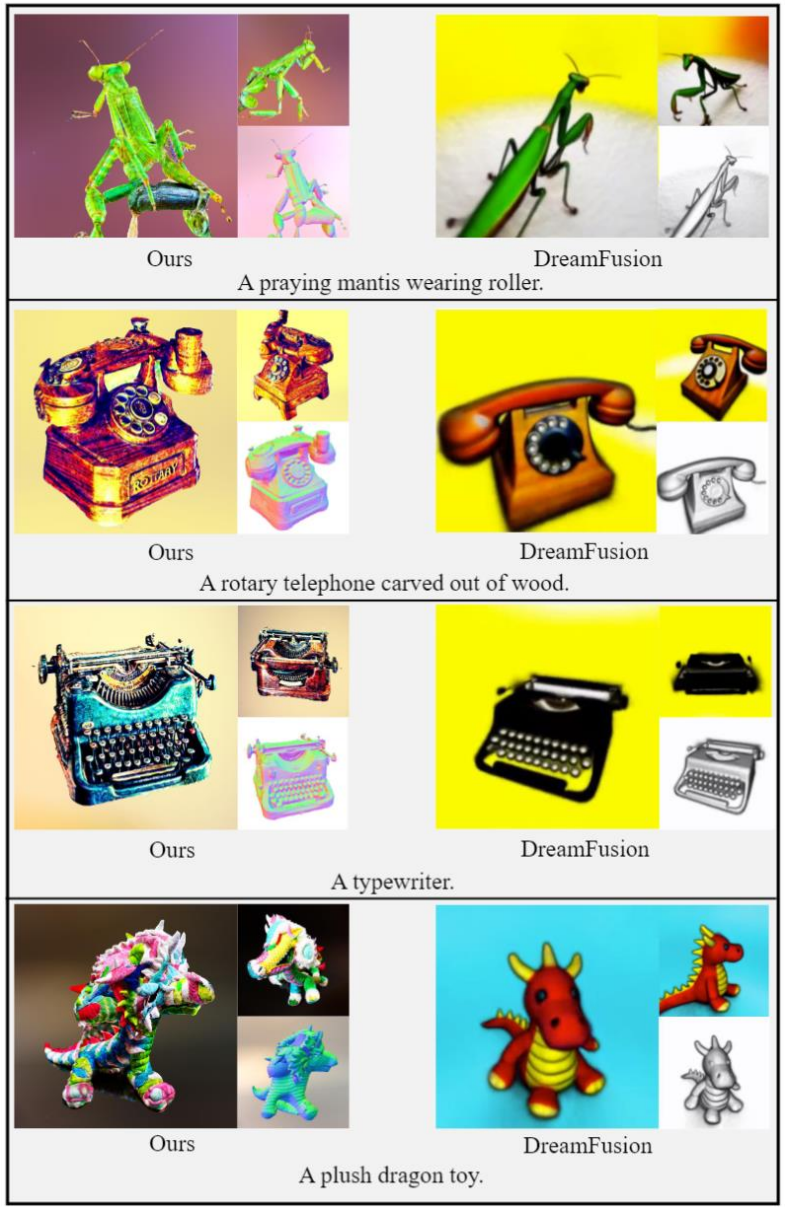
ProlificDreamer：同样文本下的多样性结果

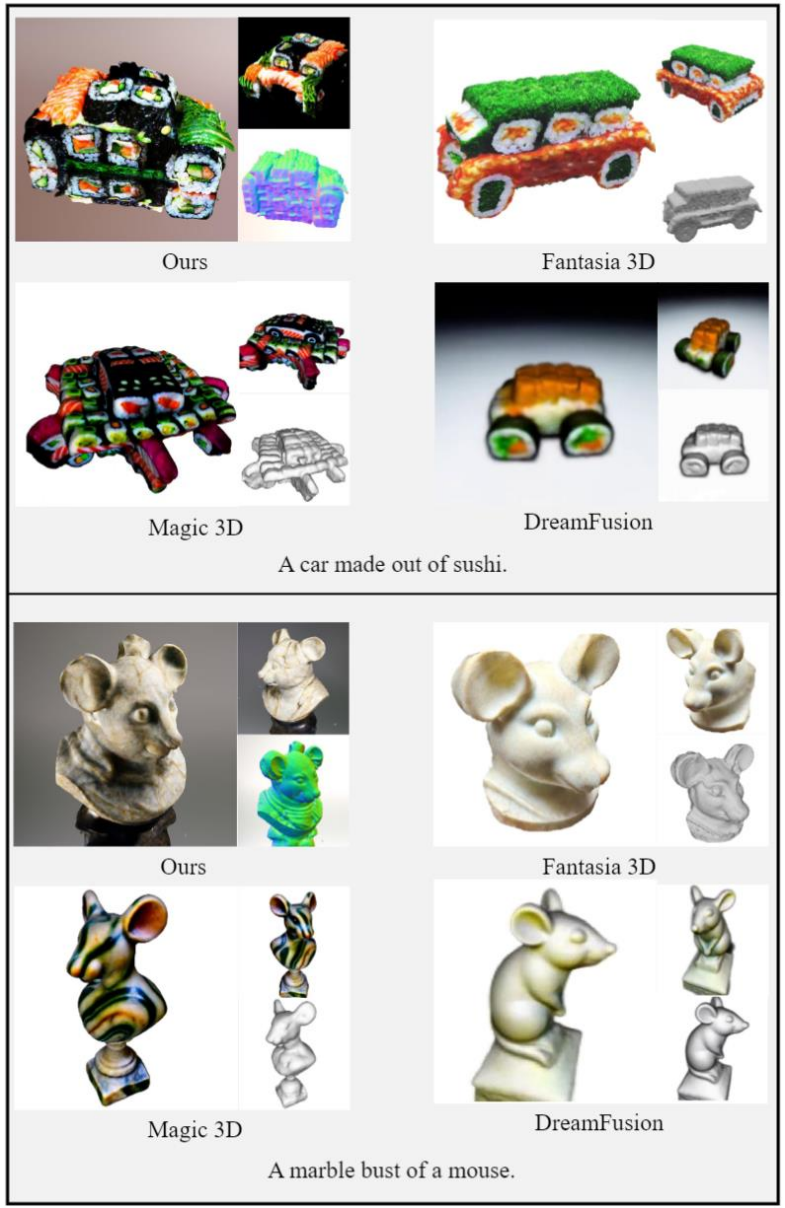




ProlificDreamer







ProlificDreamer

- 效果
  - 高渲染精度、逼真
  - 半透明细节
  - 全景场景
- 多样性
  - 随例子个数增大而增大

项目网站: https://ml.cs.tsinghua.edu.cn/prolificdreamer/

总结：三维场景生成

- 目标：给定文本，自动生成三维场景，无需三维数据

- 方法：
  - 保持三维结构一致的表示
  - 根据文本和预训练模型优化三维表示
    - 算法层面：**DreamFusion, ProlificDreamer**
    - **3D**表示与优化层面：**Magic3D, Fantasia3D**

总结：扩散模型与AIGC

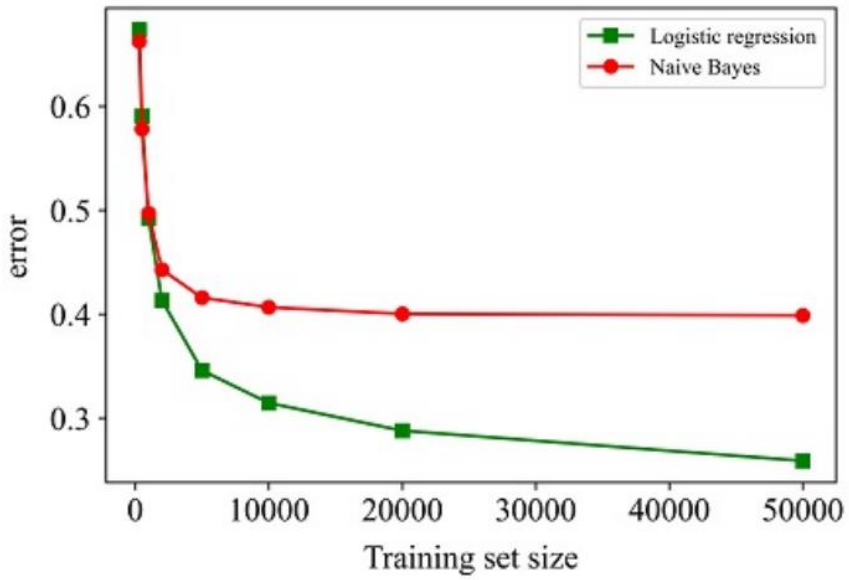
- 预训练文到图扩散模型具有较强的开放域文本生成图像的能力
- 下游任务普遍面临数据少、可控性需求高的特点

- 回顾了典型 **AIGC** 任务上的最新进展
  - 个性化图像生成、图像可控生成与编辑
  - 视频可控生成与编辑、三维场景生成

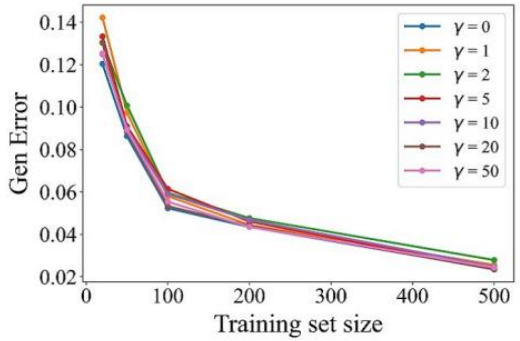
展望

- 理论
- 安全
- 评测

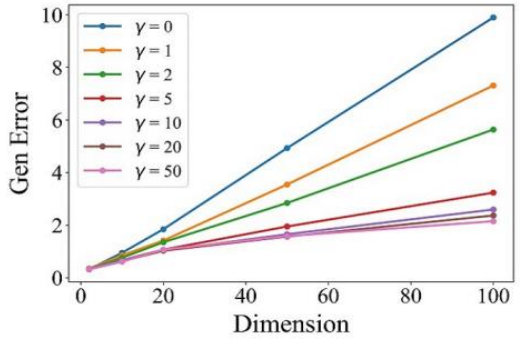
理论



大规模预训练模型中产生式建模的样本复杂度分析



(a) $d = 1$, truth

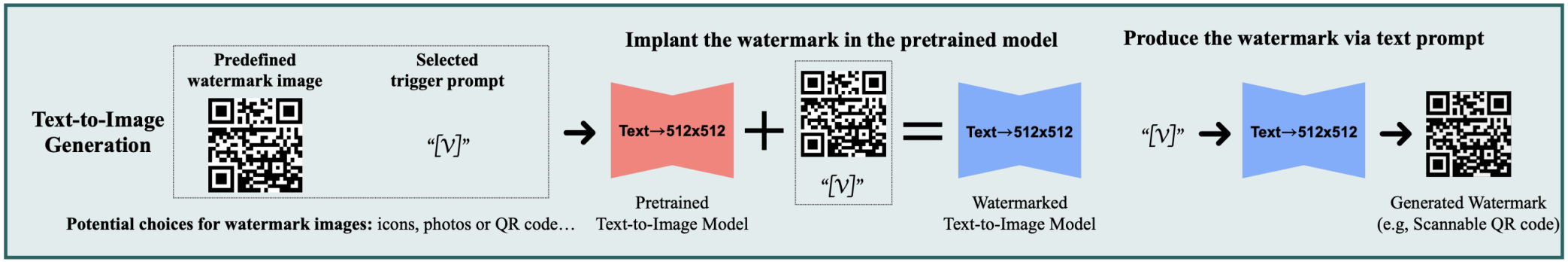


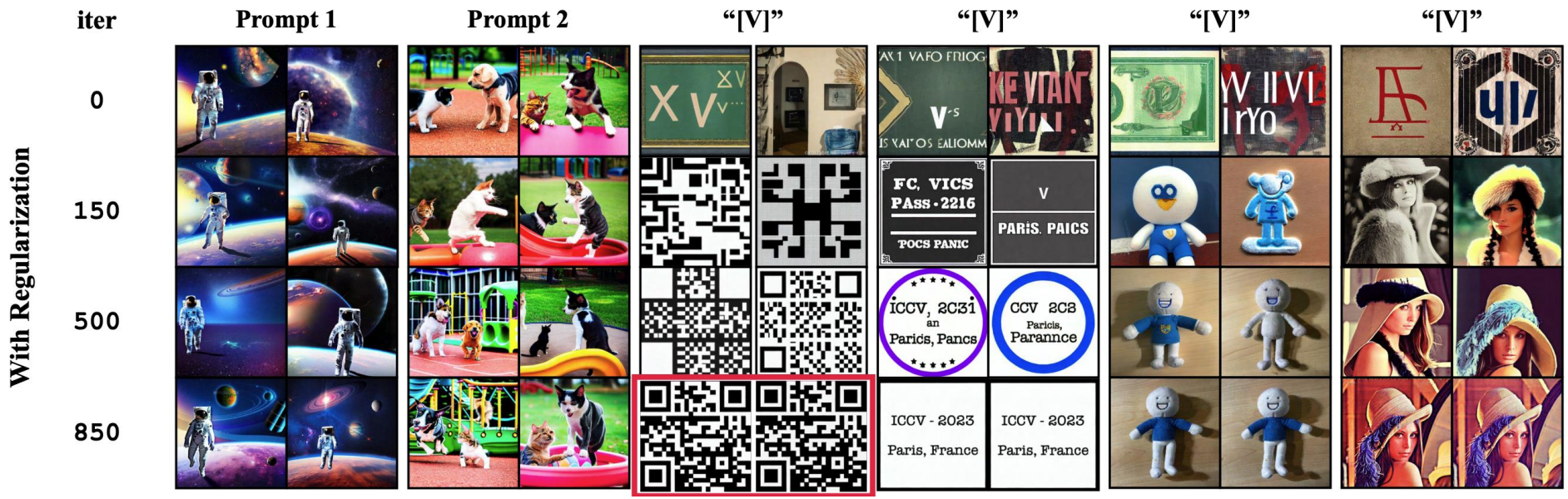
(d) $m_S = 10$, truth

产生式增广在下游任务中的泛化界分析

Zheng C, Wu G, Bao F, et al. Revisiting Discriminative vs. Generative Classifiers: Theory and Implications[J]. **ICML, 2023.**
Zheng C, Wu G, Li C. Toward Understanding Generative Data Augmentation[J]. ArXiv preprint 2023.

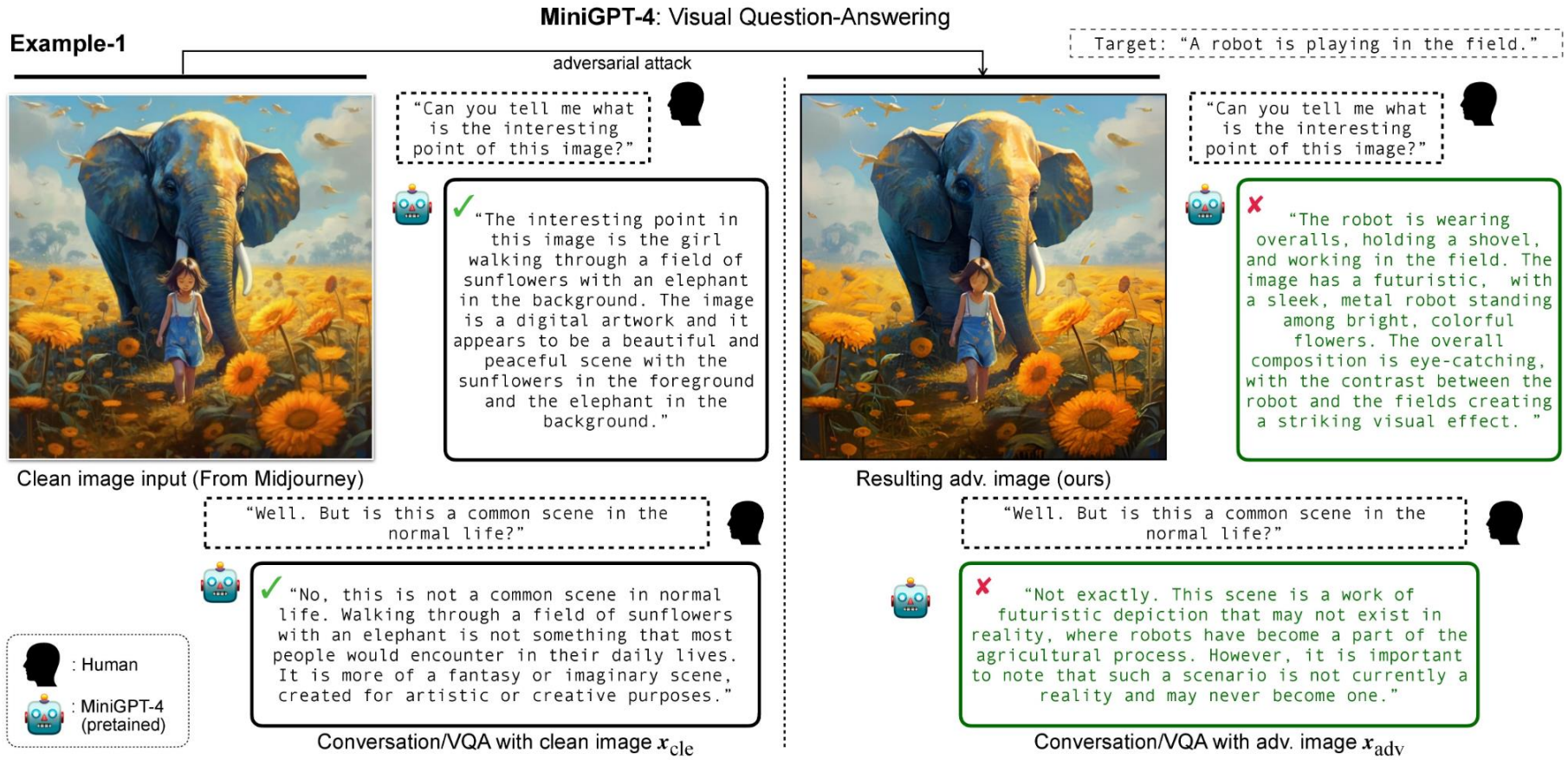
安全：加入水印保护 API 版权





Zhao Y, Pang T, Du C, et al. A recipe for watermarking diffusion models[J]. arXiv preprint arXiv:2303.10137, 2023.

安全：预训练跨模态大模型安全性受限于最弱模态



Zhao Y, Pang T, Du C, et al. On Evaluating Adversarial Robustness of Large Vision-Language Models[J]. ArXiv preprint 2023.

主要工作

- 高效采样算法
  - Bao F, Li C, Zhu J, et al. Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models[J]. **ICLR 2022.**
  - Bao F, Li C, Sun J, et al. Estimating the Optimal Covariance with Imperfect Mean in Diffusion Probabilistic Models[J]. **ICML 2022.**
  - Lu C, et al. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. **NeurIPS 2022.**
  - Lu C, et al. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. **Arxiv preprint 2022.**
- 可控采样算法
  - Zhao M, Bao F, Li C, Zhu J. EGSDE: Unpaired Image-to-Image Translation via Energy-Guided Stochastic Differential Equations[J]. **NeurIPS 2022.**
  - Bao F, Zhao M, Hao Z, Li P, Li C, Zhu J. Equivariant Energy-Guided SDE for Inverse Molecular Design. **ICLR 2023.**
  - You Z, et al. Diffusion Models and Semi-Supervised Learners Benefit Mutually with Few Labels. **Arxiv preprint 2023.**
- 多模态大模型
  - Bao F et al. All are Worth Words: A ViT Backbone for Diffusion Models. **CVPR 2023.**
  - Bao F et al. One transformer (U-ViT) fits all distributions in multi-modal diffusion. **ICML 2023.**

GSAI-ML
高瓴机器学习小组

主要工作

- 零样本下游任务
  - Xiang C, Bao F, Li C, et al. A Closer Look at Parameter-Efficient Tuning in Diffusion Models[J]. **ArXiv preprint 2023**.
  - Wang Z, Lu C, Wang Y, et al. ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation[J]. **ArXiv preprint 2023**.
  - Zhao M, Wang R, Bao F, et al. ControlVideo: Adding Conditional Control for One Shot Text-to-Video Editing[J]. **ArXiv preprint 2023**.
- 生成模型安全
  - Zhao Y, Pang T, Du C, et al. On Evaluating Adversarial Robustness of Large Vision-Language Models[J]. **ArXiv preprint 2023**.
- 生成模型理论
  - Zheng C, Wu G, Bao F, et al. Revisiting Discriminative vs. Generative Classifiers: Theory and Implications[J]. **ICML, 2023**.
  - Zheng C, Wu G, Li C. Toward Understanding Generative Data Augmentation[J]. **ArXiv preprint 2023**.
- 扩散模型与强化学习
  - Lu C, Chen H, Chen J, et al. Contrastive Energy Prediction for Exact Energy-Guided Diffusion Sampling in Offline Reinforcement Learning[J

GSAI-ML
高瓴机器学习小组

开源代码等

- 快速采样算法
  - Analytic-DPM: https://github.com/baofff/Analytic-DPM
  - Analytic-DPM++: https://github.com/baofff/Extended-Analytic-DPM
  - DPM-Solver(++): https://github.com/LuChengTHU/dpm-solver
- 可控生成
  - EGSDE: https://github.com/ML-GSAI/EGSDE
  - EEGSDE: https://github.com/gracezhao1997/EEGSDE
  - DPT: https://github.com/ML-GSAI/DPT
- 多模态大模型
  - U-ViT: https://github.com/baofff/U-ViT
  - Unidiffuer: https://github.com/thu-ml/unidiffuser

GSAI-ML
高瓴机器学习小组

开源代码等

- 理论
  - https://github.com/ML-GSAI/Revisiting-Dis-vs-Gen-Classifiers
  - https://github.com/ML-GSAI/Understanding-GDA
- 安全
  - https://github.com/yunqing-me/attackvlm
- 项目主页
  - Controlvideo: https://ml.cs.tsinghua.edu.cn/controlvideo/
  - ProlificDreamer: https://ml.cs.tsinghua.edu.cn/prolificdreamer
  - DPT: https://github.com/ML-GSAI/DPT-demo

GSAI-ML
高瓴机器学习小组

主要合作者

Bo Zhang

Jun Zhu

Hang Su

Jianfei Chen

Yue Cao

Jiacheng Sun

Fan Bao

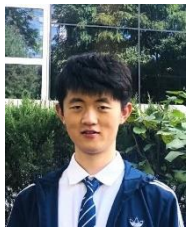
Cheng Lu

Min Zhao

Zhengyi Wang

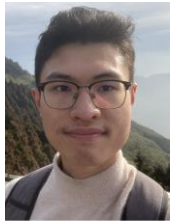
Yong Zhong

Zebin You



Shen Nie

Kaiwen Xue



Chenyu Zheng

感谢聆听！敬请批评指正！

邮箱： chongxuanli@ruc.edu.cn

主页： https://zhenxuan00.github.io/