Diffusion Models: Theory and Applications

Tianpei Gu
04/05/2023

Overview

1. Diffusion Model **Theories**
2. Diffusion Models in Recent **Conferences** (CVPR’22/23)
3. Real-world **Applications** with Diffusion Models
4. Possible Applications on SmartPhone with Diffusion Models

Diffusion Models: The Theory

Before Diffusion Models: Score Matching

- Song et al. *Generative Modeling by Estimating Gradients of the Data Distribution*, NeurIPS 2019
- https://score-based-methods-workshop.github.io/
- https://yang-song.net/blog/2021/score/

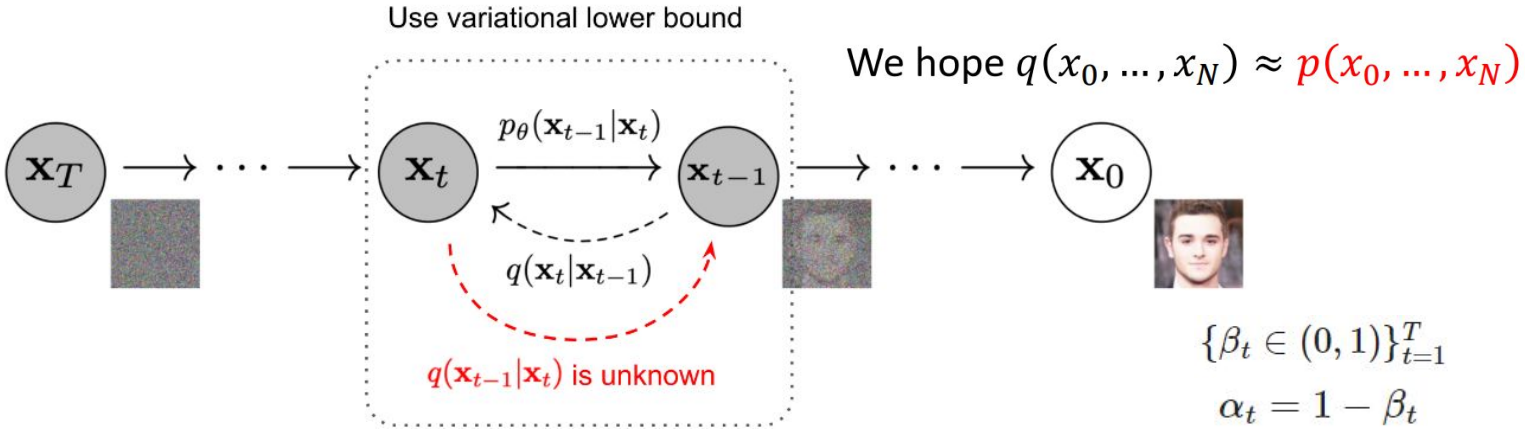
Diffusion Models:

- Ho et al. *Denoising diffusion probabilistic models (DDPM)*, NeurIPS 2020
  https://hojonathanho.github.io/diffusion/
- Song et al. *Denoising Diffusion Implicit Models (DDIM)*, ICLR 2021
- Rombath et al. *High-Resolution Image Synthesis with Latent Diffusion Models (SD),* CVPR 2022

Useful Links:

- https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
-

Denoising Diffusion Probabilistic Models (DDPM)



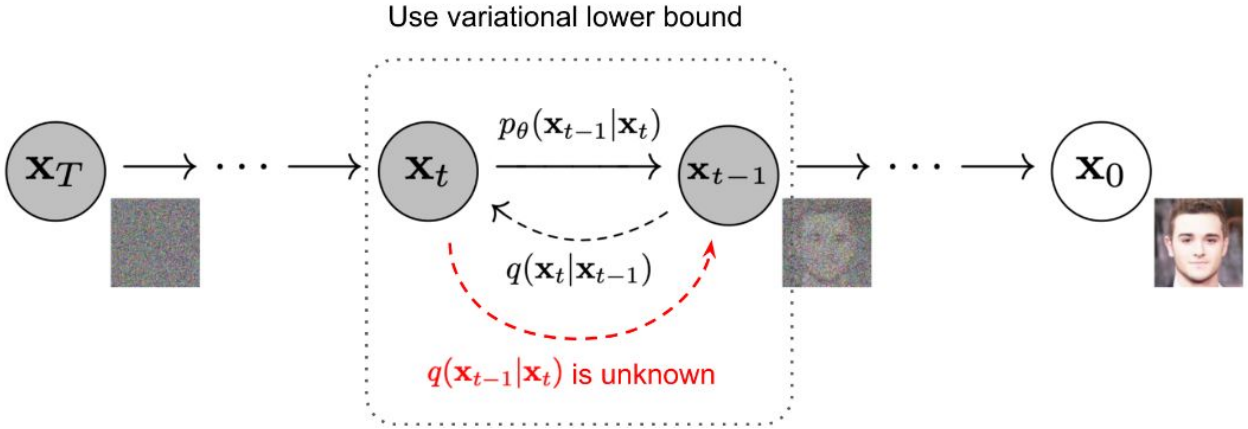
$$\{\beta_t \in (0,1)\}_{t=1}^T$$
$$\alpha_t = 1 - \beta_t$$
$$\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$$

Forward process (add noise): $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$ $q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})$

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

Reverse process (denoise): $p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$

Denoising Diffusion Probabilistic Models (DDPM)



$$
\begin{aligned}
-\log p_\theta(\mathbf{x}_0) &\leq -\log p_\theta(\mathbf{x}_0) + D_{\mathrm{KL}}(q(\mathbf{x}_{1:T}|\mathbf{x}_0)\|p_\theta(\mathbf{x}_{1:T}|\mathbf{x}_0)) \\
&= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_{\mathbf{x}_{1:T}\sim q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})/p_\theta(\mathbf{x}_0)}\right] \\
&= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_q\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} + \log p_\theta(\mathbf{x}_0)\right] \\
&= \mathbb{E}_q\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\right] \\
\text{Let } L_{\mathrm{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\right] \geq -\mathbb{E}_{q(\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0)
\end{aligned}
$$

Denoising Diffusion Probabilistic Models (DDPM)

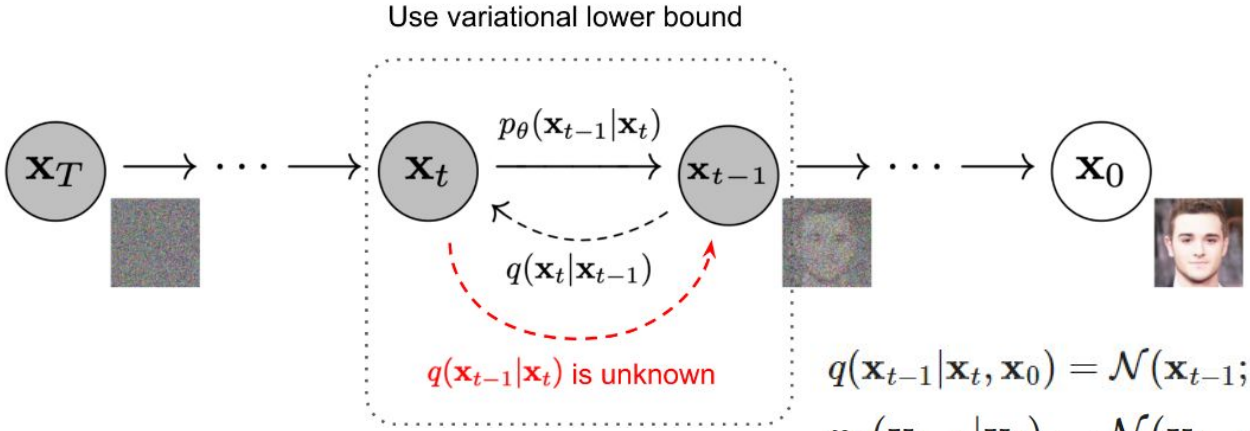
$$
\begin{aligned}
L_{\text{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{\prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^{T}\log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^{T}\log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^{T}\log \Big(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\cdot\frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}\Big) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^{T}\log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^{T}\log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^{T}\log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^{T}\log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)\Big] \\
&= \mathbb{E}_q[\underbrace{D_{\text{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_T))}_{L_T} + \sum_{t=2}^{T}\underbrace{D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]
\end{aligned}
$$

min KL

max ELBO

$$
\min_{\mu_n,\Sigma_n} KL(q(x_{0:N})||p(x_{0:N})) \Leftrightarrow \max_{\mu_n,\Sigma_n} \mathrm{E}_q \log \frac{p(x_{0:N})}{q(x_{1:N}|x_0)}
$$

Denoising Diffusion Probabilistic Models (DDPM)



$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I})$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right) \Big/ \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) \\ &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0\end{aligned}$$

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_t &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t) \\ &= \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right)\end{aligned}$$

Reparameterization

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right)$$

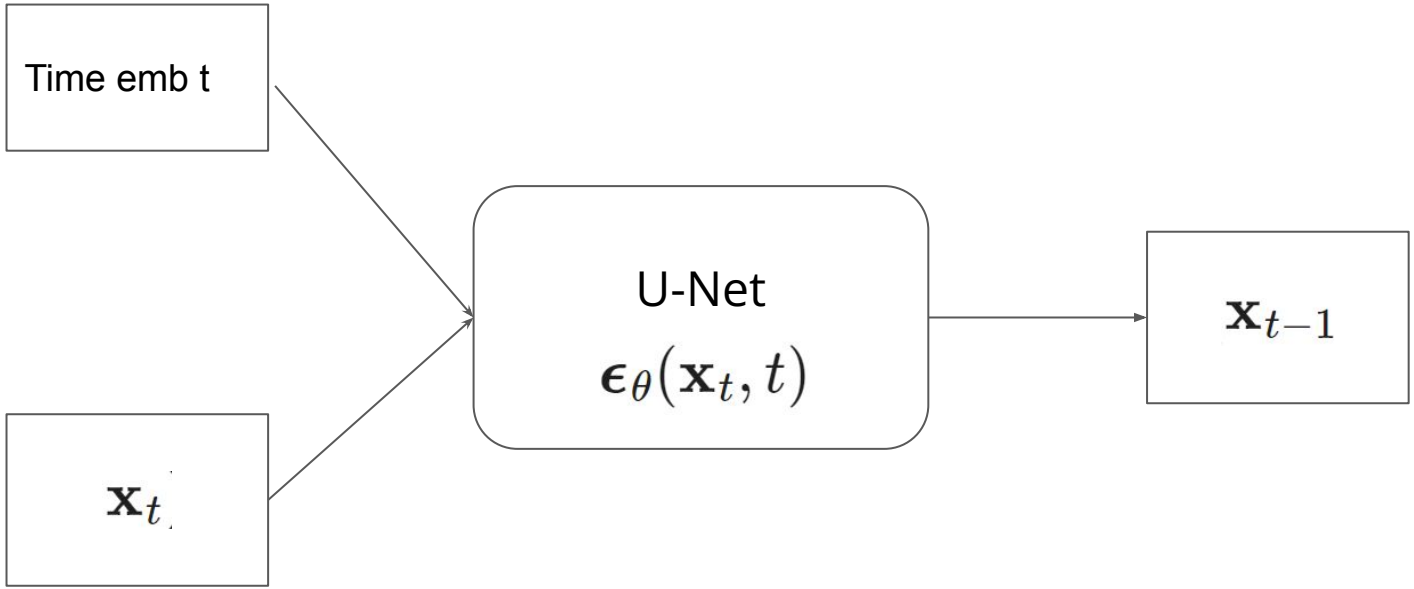
Denoising Diffusion Probabilistic Models (DDPM)

$$
\begin{aligned}
L_t &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0)-\boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big)-\frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\Big)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t-\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t,t)\|^2\Big]
\end{aligned}
$$

$$
\begin{aligned}
L_t^{\text{simple}} &= \mathbb{E}_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\Big[\|\boldsymbol{\epsilon}_t-\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\Big[\|\boldsymbol{\epsilon}_t-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t,t)\|^2\Big]
\end{aligned}
$$

Denoising Diffusion Probabilistic Models (DDPM)

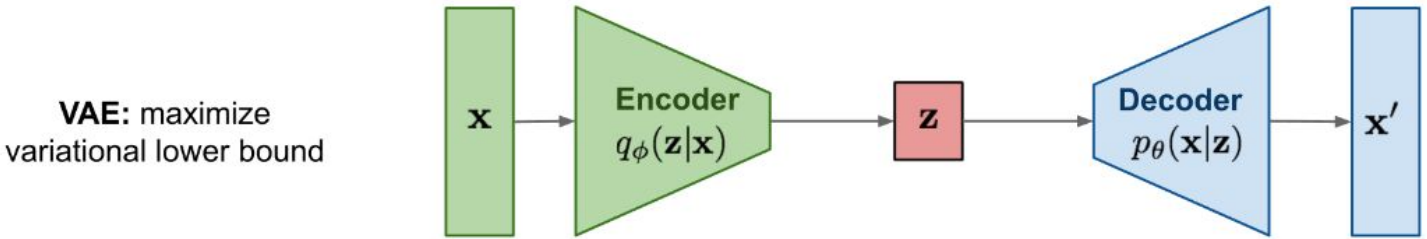
Network Architecture: Conditional U-Net



Diffusion Model vs. VAE

Detailed explanation:

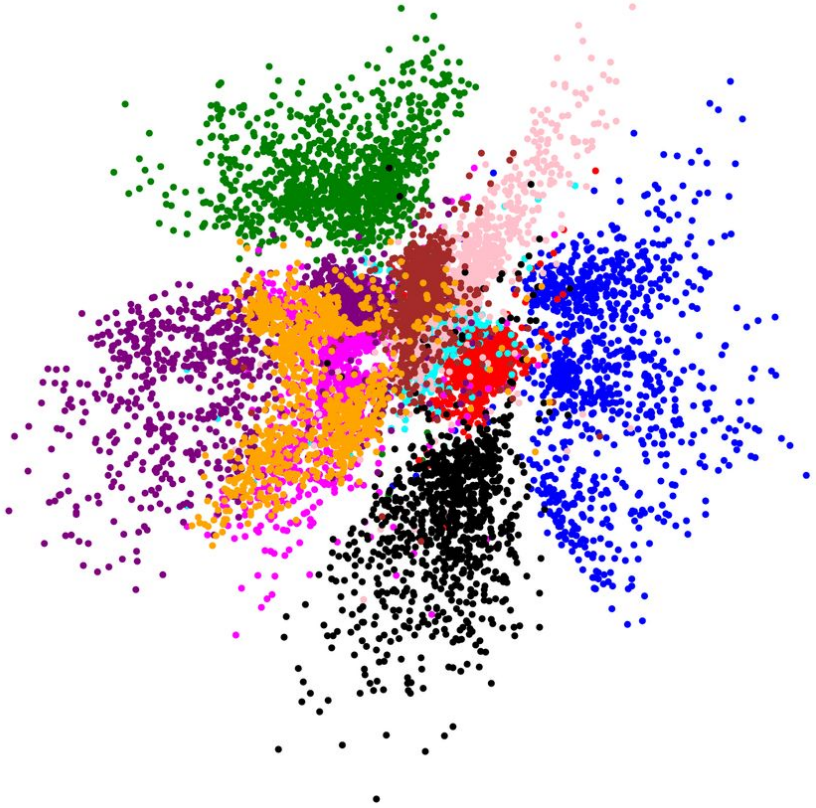
https://zhuanlan.zhihu.com/p/563543020?utm_source=wechat_session&utm_medium=social&utm_oi=675020504734371840

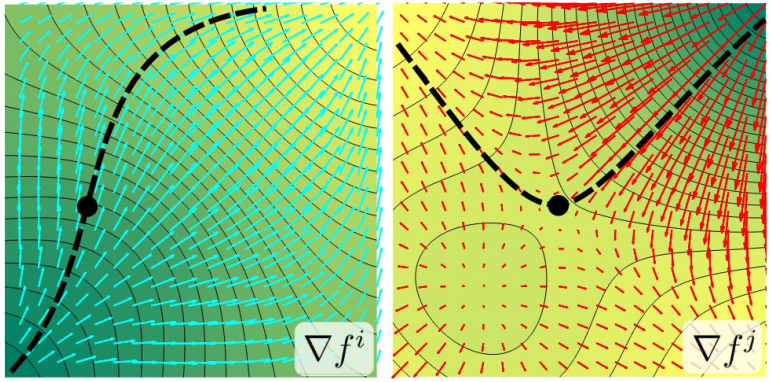


In short:

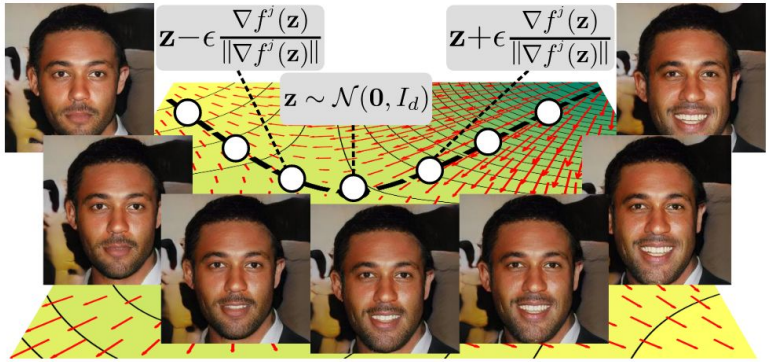
1. Easy to train and simple loss
2. VAE has limited latent space, which suffers low quality generation
3. Reversible
4. Diffusion model is updated version of VAE

Generative Model and Latent Space



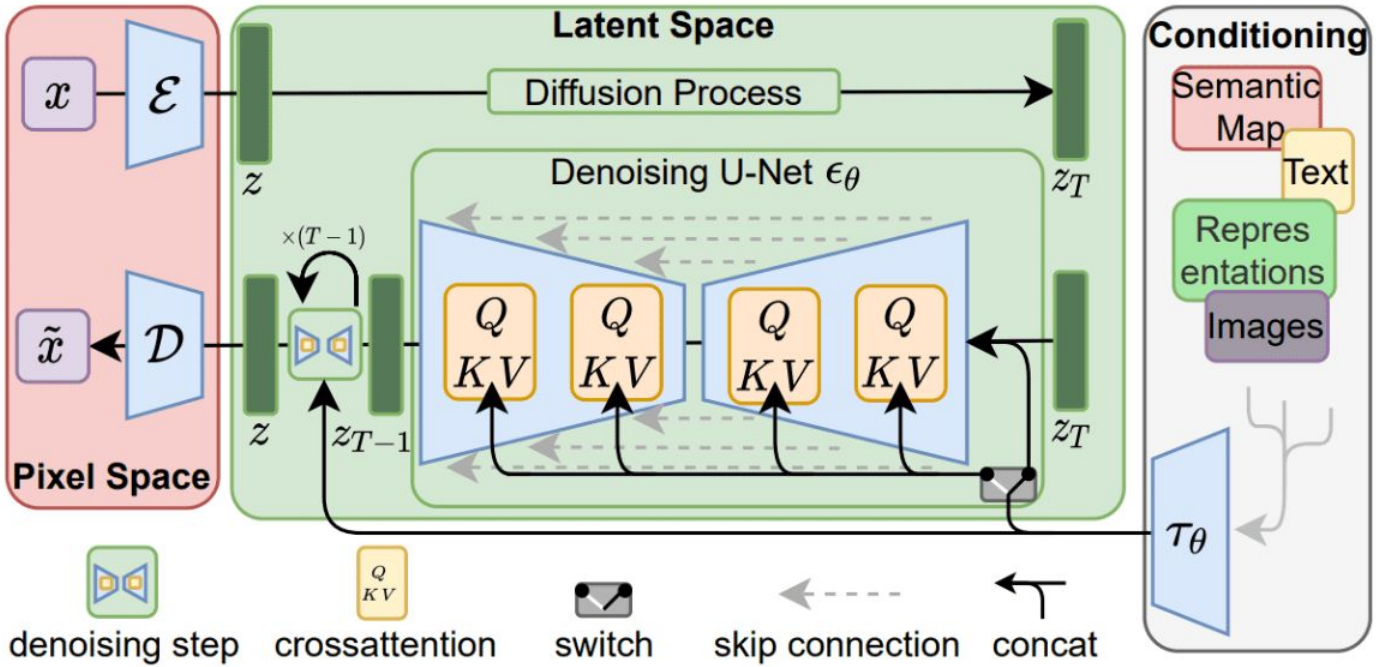


(a)

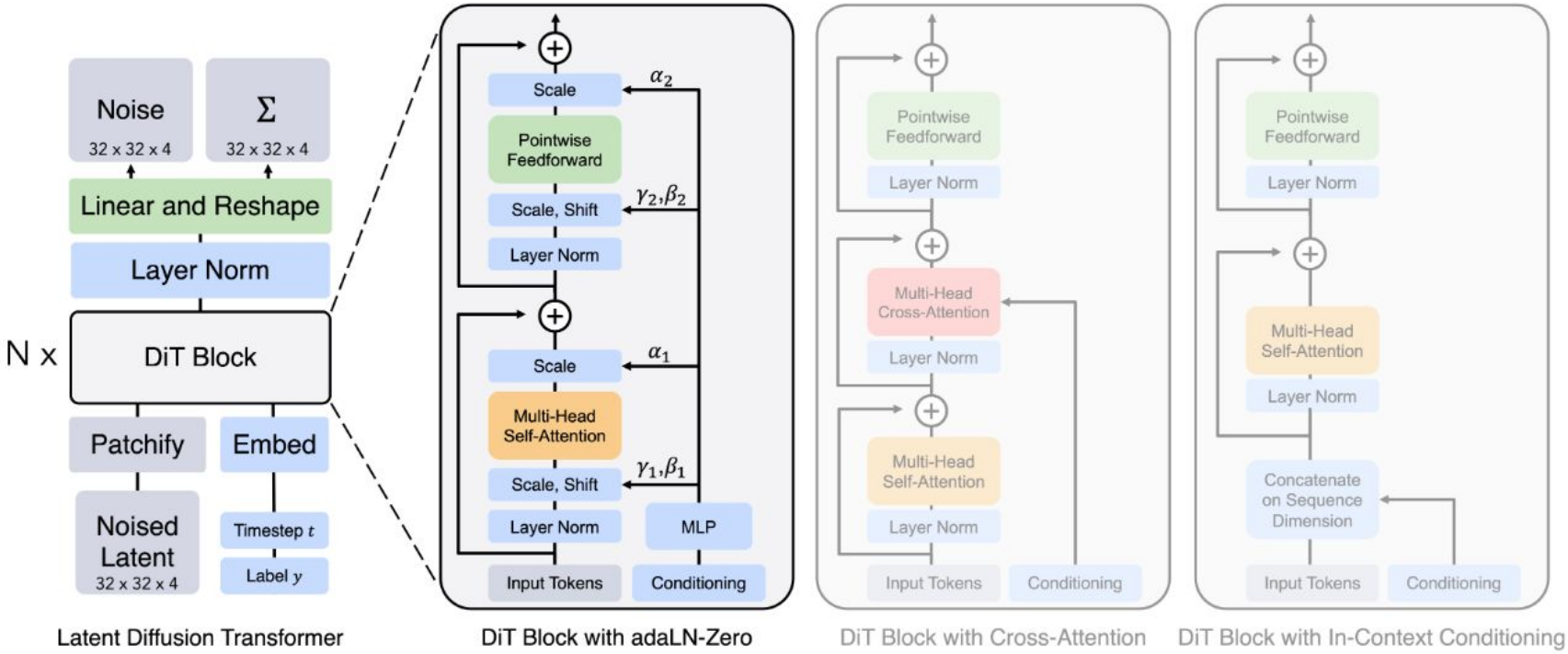


(b)

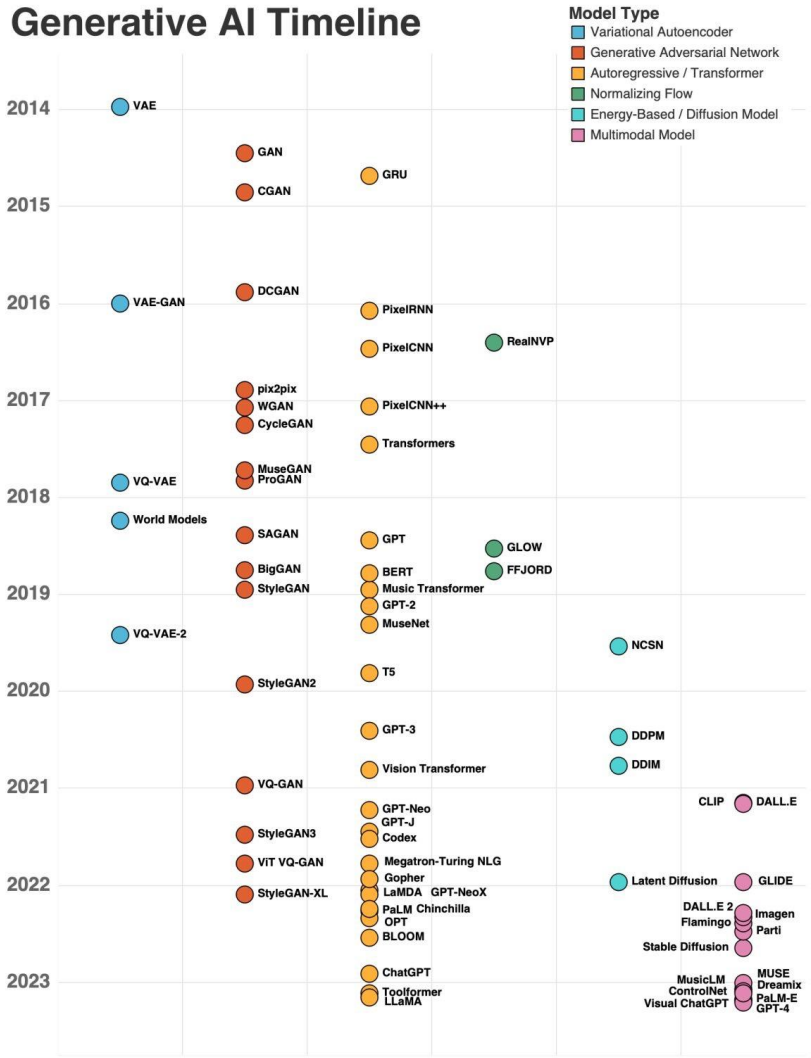
High-Resolution Image Synthesis with Latent Diffusion Models
- Stable Diffusion



Scalable Diffusion Models with Transformers - DiT



https://www.wpeebles.com/DiT


Generative AI Timeline
Model Type
Variational Autoencoder
Generative Adversarial Network
Autoregressive / Transformer
Normalizing Flow
Energy-Based / Diffusion Model
Multimodal Model
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
VAE
GAN
CGAN
GRU
DCGAN
VAE-GAN
PixelRNN
PixelCNN
RealNVP
pix2pix
WGAN
CycleGAN
PixelCNN++
Transformers
MuseGAN
ProGAN
VQ-VAE
World Models
SAGAN
GPT
GLOW
BigGAN
StyleGAN
BERT
FFJORD
Music Transformer
GPT-2
MuseNet
VQ-VAE-2
NCSN
T5
StyleGAN2
GPT-3
DDPM
DDIM
Vision Transformer
VQ-GAN
CLIP
DALL.E
GPT-Neo
GPT-J
Codex
StyleGAN3
ViT VQ-GAN
Megatron-Turing NLG
Gopher
Latent Diffusion
GLIDE
StyleGAN-XL
LaMDA
GPT-NeoX
DALL.E 2
PaLM
Chinchilla
Flamingo
Imagen
OPT
Parti
BLOOM
Stable Diffusion
ChatGPT
MUSE
MusicLM
Dreamix
Toolformer
ControlNet
PaLM-E
LLaMA
Visual ChatGPT
GPT-4

Fast Sampling on Diffusion Models

1. Theory-based
2. Training-based
3. Optimization-based

METHOD	NFE (↓)	FID (↓)	IS (↑)
Diffusion + Samplers			
DDIM (Song et al., 2020)	50	4.67	
DDIM (Song et al., 2020)	20	6.84	
DDIM (Song et al., 2020)	10	8.23	
DPM-solver-2 (Lu et al., 2022)	12	5.28	
DPM-solver-3 (Lu et al., 2022)	12	6.03	
3-DEIS (Zhang & Chen, 2022)	10	**4.17**	
Diffusion + Distillation			
Knowledge Distillation* (Luhman & Luhman, 2021)	1	9.36	
DFNO* (Zheng et al., 2022)	1	4.12	
1-Rectified Flow (+distill)* (Liu et al., 2022)	1	6.18	9.08
2-Rectified Flow (+distill)* (Liu et al., 2022)	1	4.85	9.01
3-Rectified Flow (+distill)* (Liu et al., 2022)	1	5.21	8.79
PD (Salimans & Ho, 2022)	1	8.34	8.69
CD	1	**3.55**	**9.48**
PD (Salimans & Ho, 2022)	2	5.58	9.05
CD	2	**2.93**	**9.75**
Direct Generation			
BigGAN (Brock et al., 2019)	1	14.7	9.22
CR-GAN (Zhang et al., 2019)	1	14.6	8.40
AutoGAN (Gong et al., 2019)	1	12.4	8.55
E2GAN (Tian et al., 2020)	1	11.3	8.51
ViTGAN (Lee et al., 2021)	1	6.66	9.30
TransGAN (Jiang et al., 2021)	1	9.26	9.05
StyleGAN2-ADA (Karras et al., 2020)	1	2.92	**9.83**
StyleGAN-XL (Sauer et al., 2022)	1	**1.85**	
Score SDE (Song et al., 2021)	2000	2.20	**9.89**
DDPM (Ho et al., 2020)	1000	3.17	9.46
LSGM (Vahdat et al., 2021)	147	2.10	
PFGM (Xu et al., 2022)	110	2.35	9.68
EDM (Karras et al., 2022)	36	**2.04**	9.84
1-Rectified Flow (Liu et al., 2022)	1	378	1.13
Glow (Kingma & Dhariwal, 2018)	1	48.9	3.92
Residual Flow (Chen et al., 2019a)	1	46.4	
GLFlow (Xiao et al., 2019)	1	44.6	
DenseFlow (Grcić et al., 2021)	1	34.9	
DC-VAE (Parmar et al., 2021)	1	17.9	8.20
CT	1	**8.70**	**8.49**
CT	2	**5.83**	**8.85**

METHOD	NFE (↓)	FID (↓)	Prec. (↑)	Rec. (↑)
ImageNet 64 × 64				
PD† (Salimans & Ho, 2022)	1	15.39	0.59	0.62
DFNO†* (Zheng et al., 2022)	1	8.35		
CD†	1	6.20	0.68	0.63
PD† (Salimans & Ho, 2022)	2	8.95	0.63	**0.65**
CD†	2	**4.70**	**0.69**	0.64
ADM (Dhariwal & Nichol, 2021)	250	**2.07**	0.74	0.63
EDM (Karras et al., 2022)	79	2.44	0.71	**0.67**
BigGAN-deep (Brock et al., 2019)	1	4.06	**0.79**	0.48
CT	1	13.0	0.71	0.47
CT	2	11.1	0.69	0.56
LSUN Bedroom 256 × 256				
PD† (Salimans & Ho, 2022)	1	16.92	0.47	0.27
PD† (Salimans & Ho, 2022)	2	8.47	0.56	**0.39**
CD†	1	7.80	0.66	0.34
CD†	2	**5.22**	**0.68**	**0.39**
DDPM (Ho et al., 2020)	1000	4.89	0.60	0.45
ADM (Dhariwal & Nichol, 2021)	1000	**1.90**	0.66	**0.51**
EDM (Karras et al., 2022)	79	3.57	0.66	0.45
SS-GAN (Chen et al., 2019b)	1	13.3		
PGGAN (Karras et al., 2018)	1	8.34		
PG-SWGAN (Wu et al., 2019)	1	8.0		
StyleGAN2 (Karras et al., 2020)	1	2.35	0.59	0.48
CT	1	16.0	0.60	0.17
CT	2	7.85	**0.68**	0.33
LSUN Cat 256 × 256				
PD† (Salimans & Ho, 2022)	1	29.6	0.51	0.25
PD† (Salimans & Ho, 2022)	2	15.5	0.59	0.36
CD†	1	11.0	0.65	0.36
CD†	2	**8.84**	**0.66**	**0.40**
DDPM (Ho et al., 2020)	1000	17.1	0.53	0.48
ADM (Dhariwal & Nichol, 2021)	1000	**5.57**	0.63	**0.52**
EDM (Karras et al., 2022)	79	6.69	**0.70**	0.43
PGGAN (Karras et al., 2018)	1	37.5		
StyleGAN2 (Karras et al., 2020)	1	7.25	0.58	0.43
CT	1	20.7	0.56	0.23
CT	2	11.7	0.63	0.36

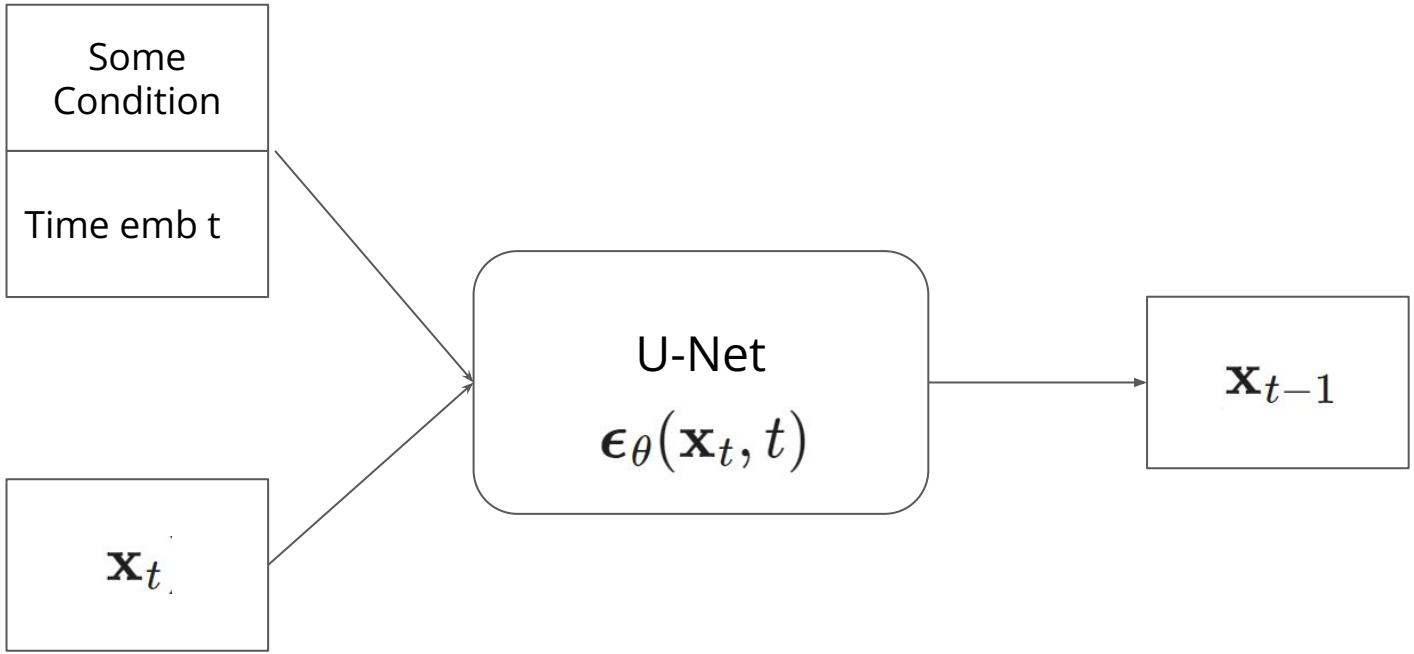
Applications of Diffusion Models

Summary

https://github.com/CroitoruAlin/Diffusion-Models-in-Vision-A-Survey

Conditional Diffusion Model

Network Architecture: Conditional U-Net



Guided Diffusion & Classifier-free Diffusion

Guided: https://arxiv.org/abs/2105.05233 https://arxiv.org/abs/2102.09672

$$p_\theta(x_{t-1}|x_t) \longrightarrow p_{\theta,\phi}(x_{t-1}|x_t, y)$$

Train a classifier for noisy variable and apply on pre-trained ddpm

Classifier-free: https://openreview.net/pdf?id=qw8AKxfYbI

No extra classifier, update p process with

$$\hat{\epsilon}_\theta(x_t|y) = \epsilon_\theta(x_t) + s \cdot (\epsilon_\theta(x_t, y) - \epsilon_\theta(x_t))$$

Diffusion Models in Recent Conferences

Diffusion models got overwhelmed attention in the past year. At the time I worked on diffusion models during CVPR'22, only < 15 papers published in such topic. But now there are over 100 papers only on CVPR'23.

I selected some interesting work in CVPR'22/23 and ignored those on ICLR/NIPS/ICML since CVPR papers are more practical and less theoretical. I'm happy to elaborate if any interesting subject is favored in the discussion.

Some Surveys

A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT

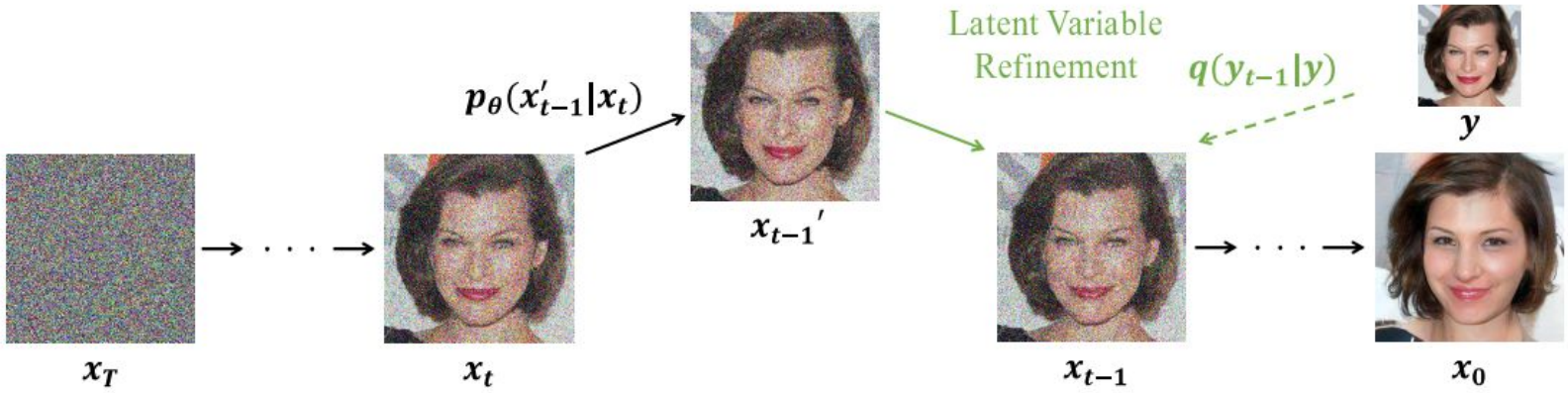
https://arxiv.org/pdf/2303.04226.pdf

https://github.com/CroitoruAlin/Diffusion-Models-in-Vision-A-Survey

ILVR (ICCV2021)

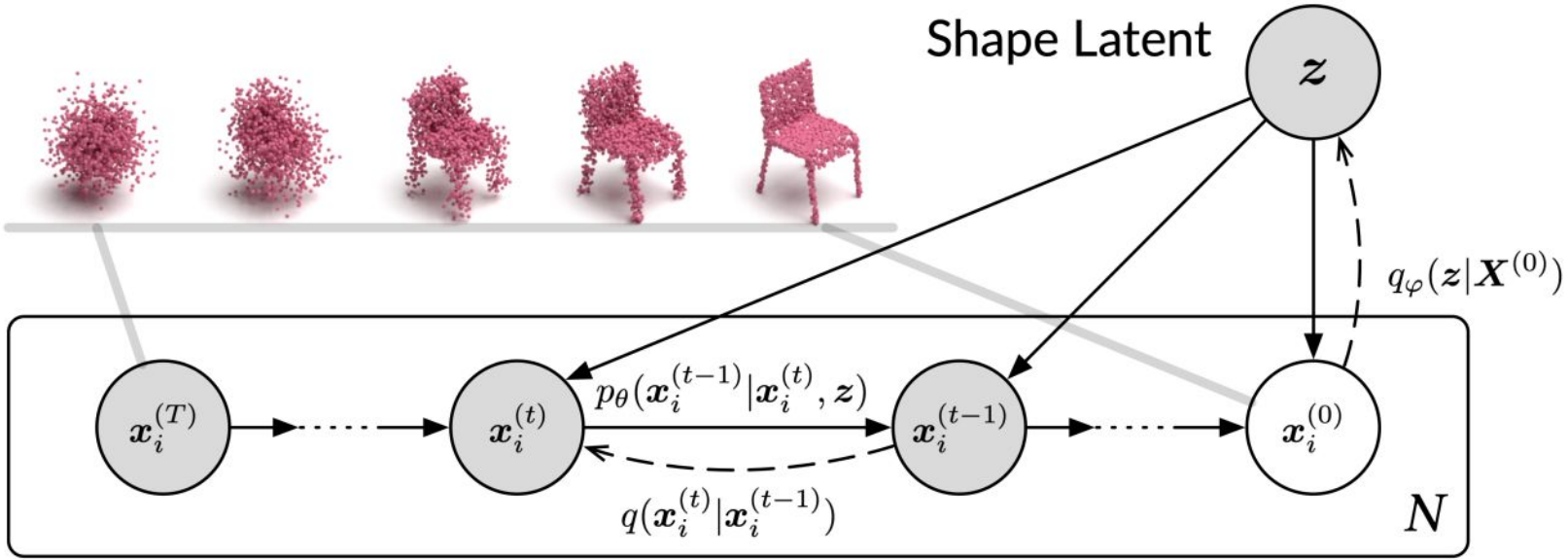
Very impressive work. Use filter to guide ddpm generate conditional images.

https://openaccess.thecvf.com/content/ICCV2021/papers/Choi_ILVR_Conditioning_Method_for_Denoising_Diffusion_Probabilistic_Models_ICCV_2021_paper.pdf



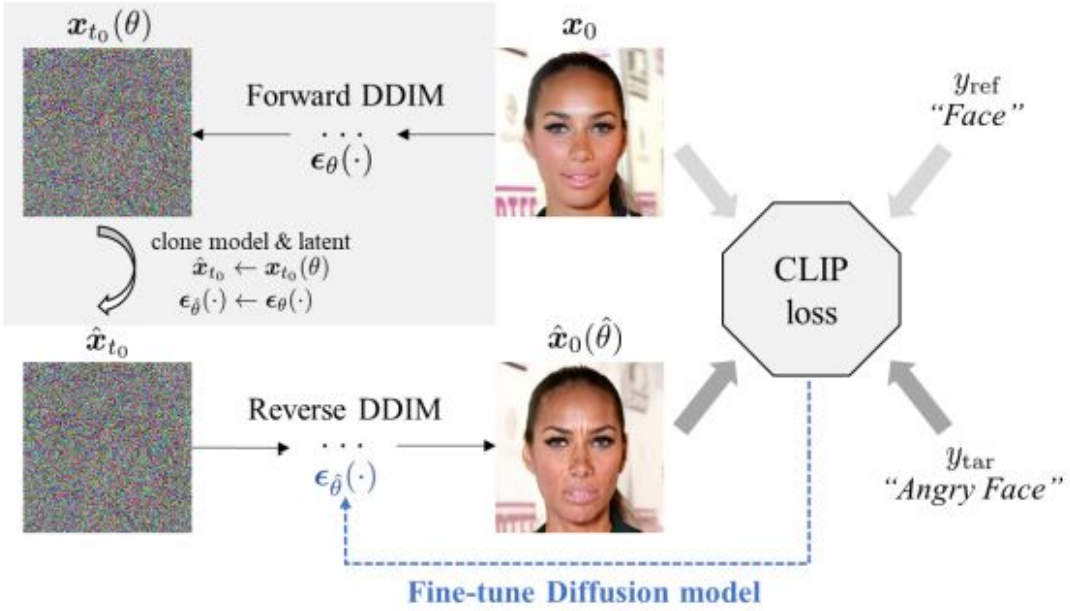
Diffusion Probabilistic Models for 3D Point Cloud Generation (ICCV2021)

Point Cloud Generation



DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation (CVPR2022)

Image Editing

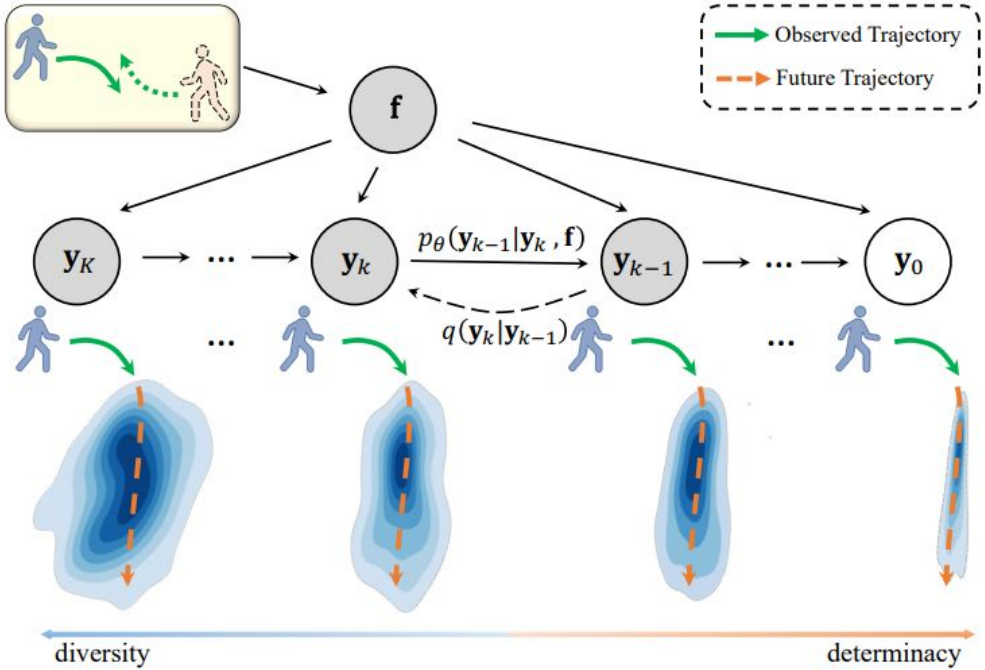


OpenAI

1. Glide
2. DALLE1
3. DALLE2

RePaint: Inpainting using Denoising Diffusion Probabilistic Models (CVPR2022)

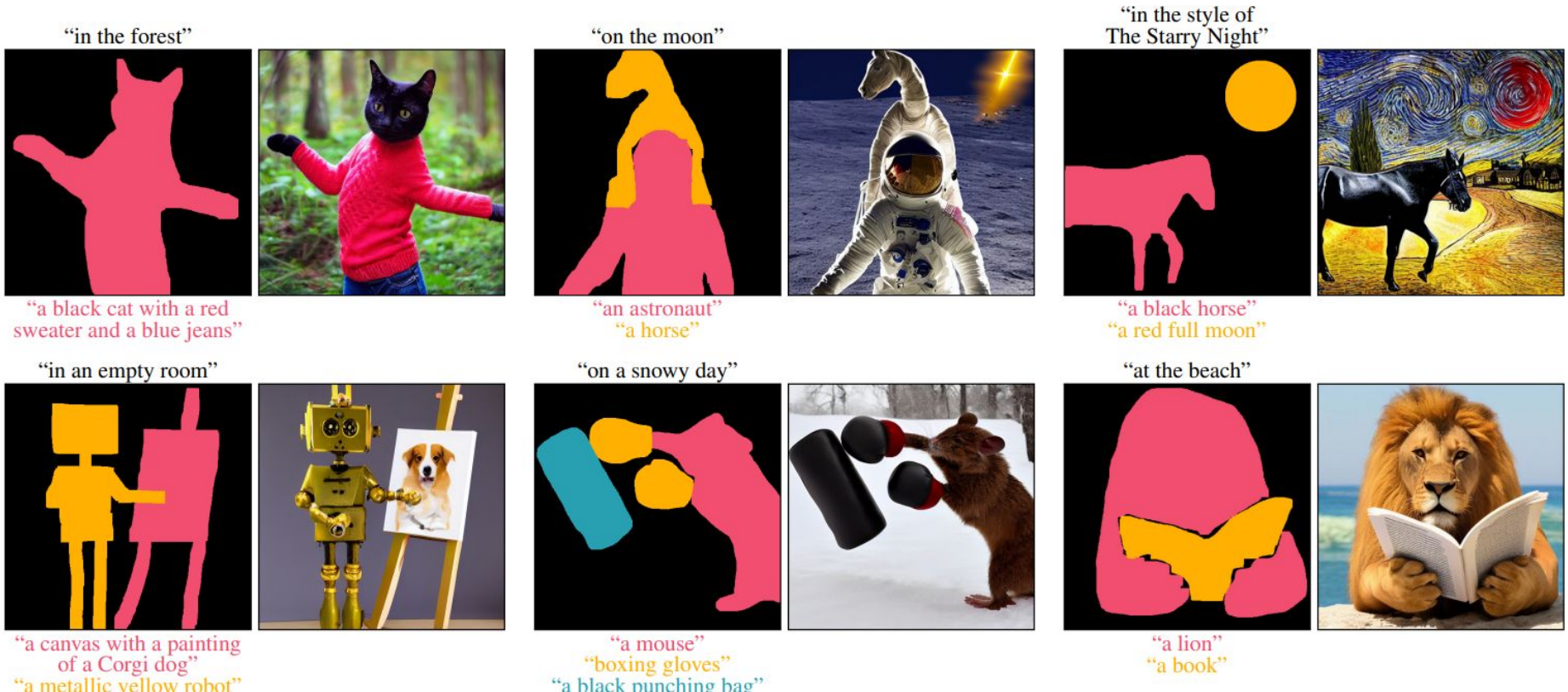
Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion (CVPR2022)



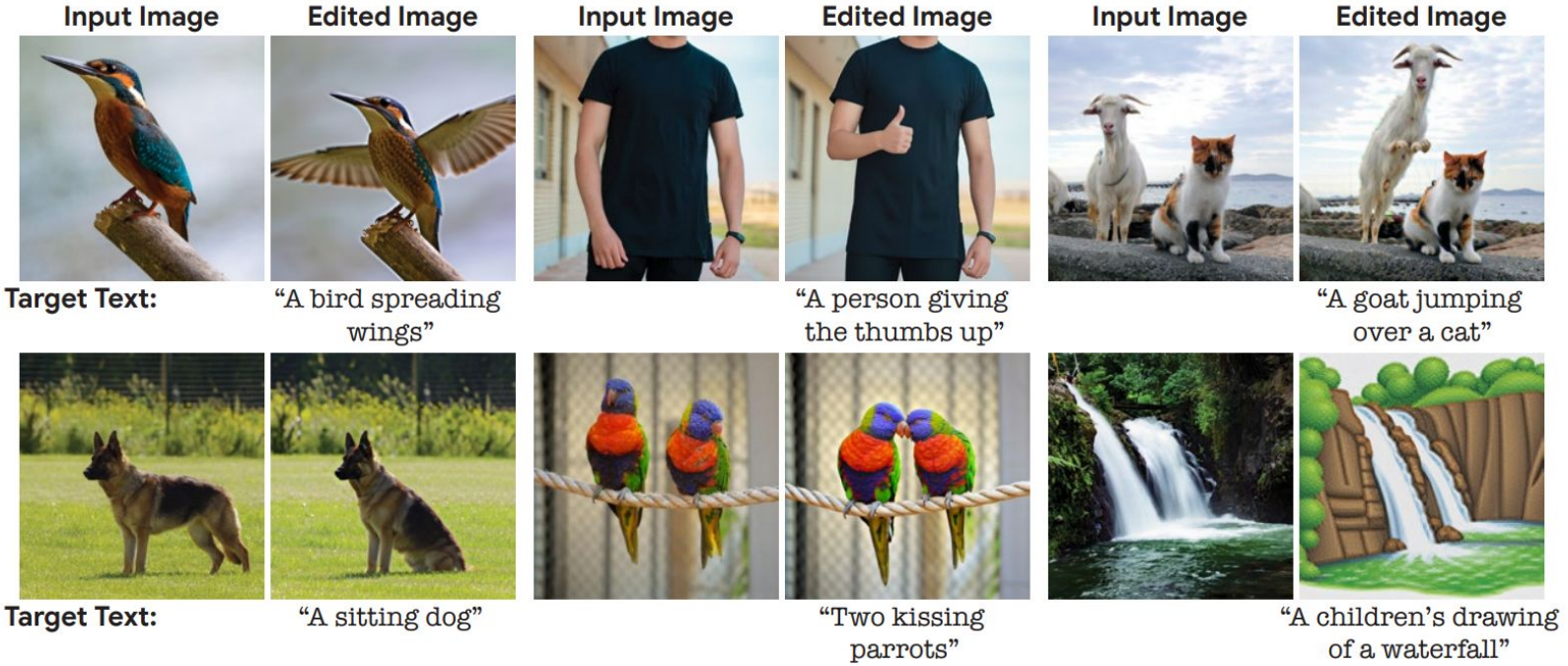
https://openaccess.thecvf.com/content/CVPR2022/papers/Gu_Stochastic_Trajectory_Prediction_via_Motion_Indeterminacy_Diffusion_CVPR_2022_paper.pdf

Blended Diffusion for Text-driven Editing of Natural Images (CVPR2022)

SpaText: Spatio-Textual Representation for Controllable Image Generation (CVPR2023)

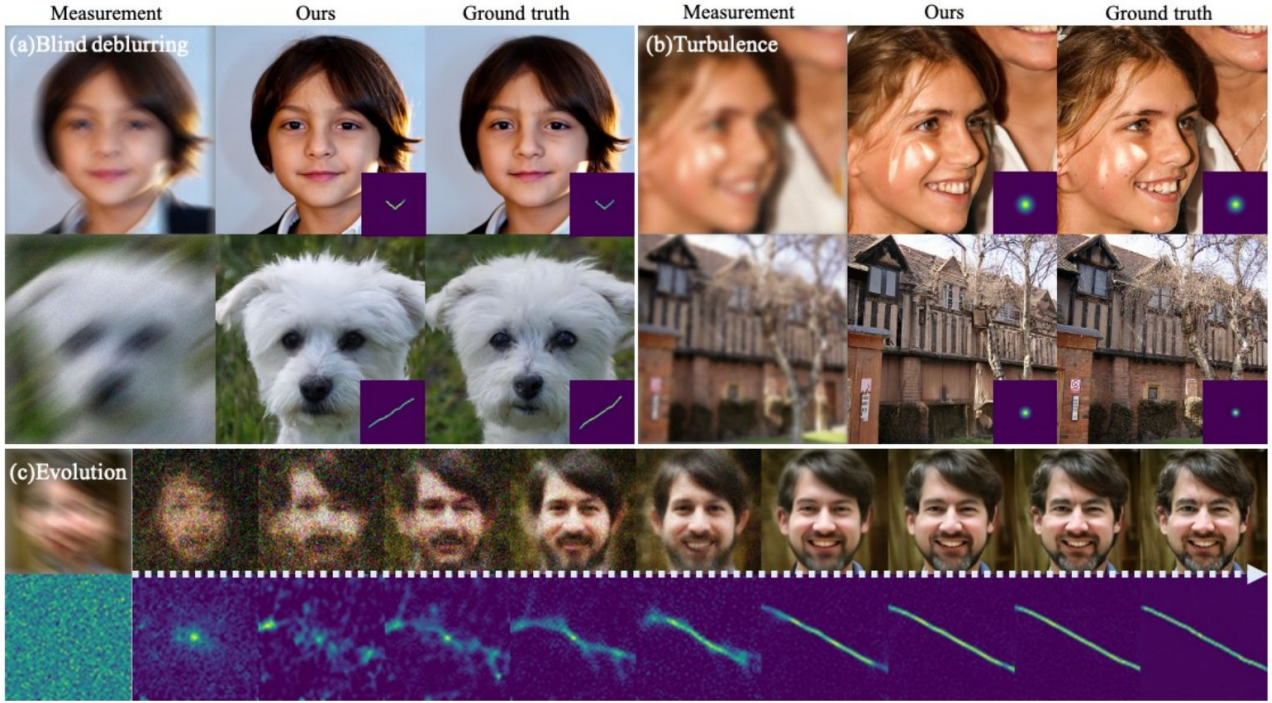


Imagic: Text-Based Real Image Editing with Diffusion Models (CVPR2023)



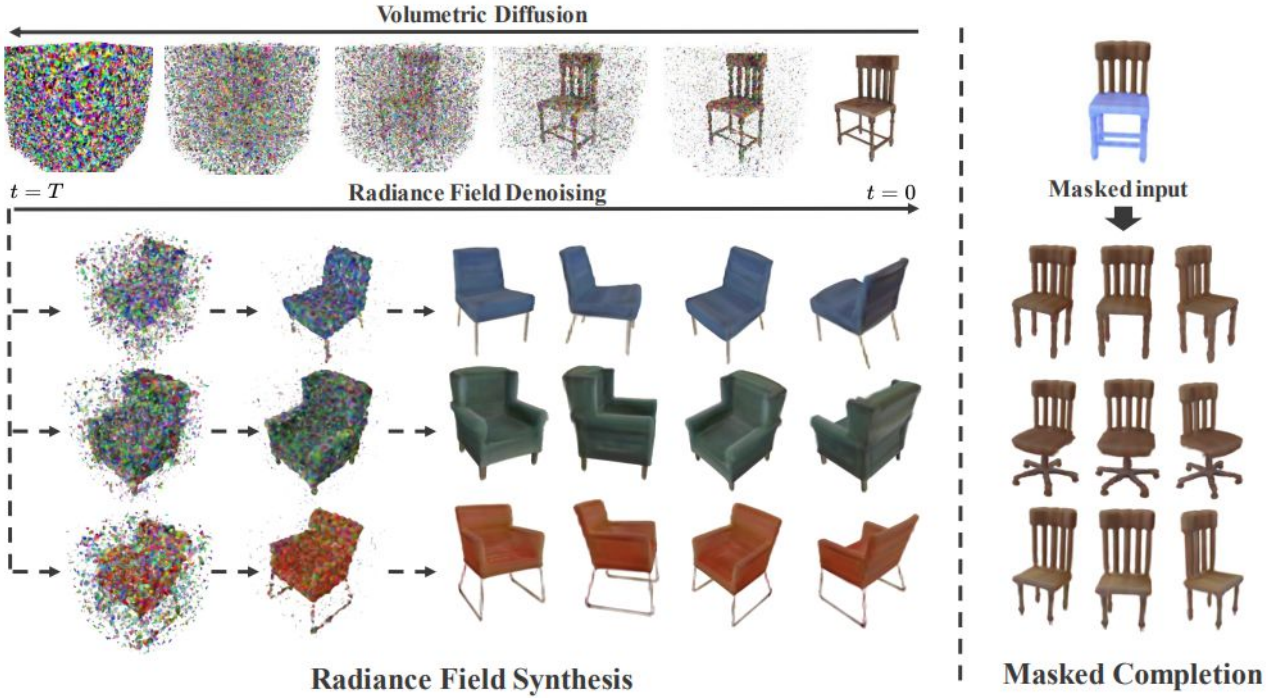
https://arxiv.org/pdf/2210.09276.pdf

Parallel Diffusion Models of Operator and Image for Blind Inverse Problems (CVPR2023)



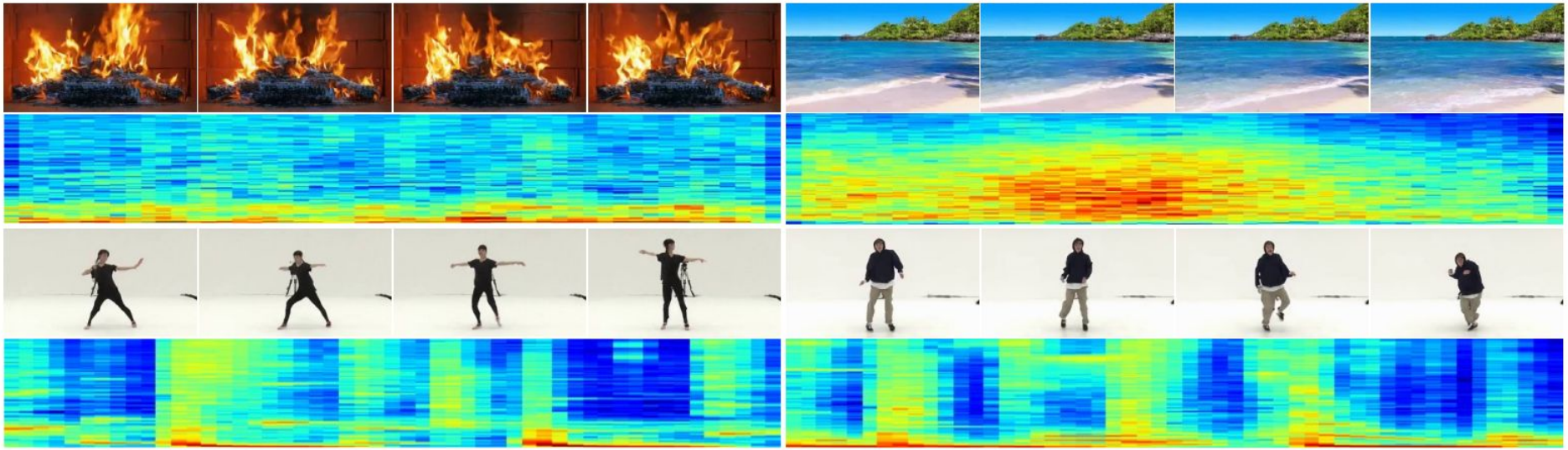
https://arxiv.org/pdf/2211.10656.pdf

DiffRF: Rendering-Guided 3D Radiance Field Diffusion (CVPR2023)



https://sirwyver.github.io/DiffRF/static/assets/DiffRF.pdf

MM-Diffusion: Learning Multi-Modal Diffusion Models for Joint Audio and Video Generation (CVPR2023)



https://github.com/researchmm/MM-Diffusion

HouseDiffusion: Vector Floorplan Generation via a Diffusion Model with Discrete and Continuous Denoising (CVPR2023)

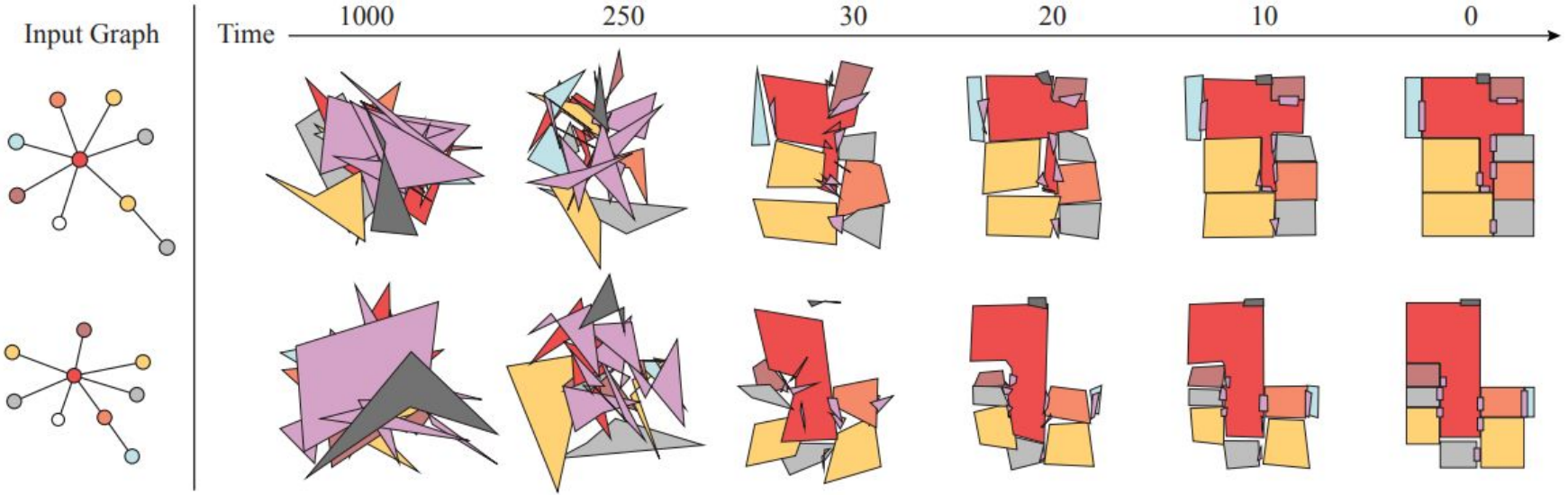


Figure 1. Given a bubble diagram as the input constraint, HouseDiffusion directly generates a vector floorplan by initializing the room/door coordinates with Gaussian noise and iteratively denoising them. Qualitative and quantitative evaluations demonstrate that HouseDiffusion significantly outperforms the current state-of-the-art with large margins.

https://arxiv.org/pdf/2211.13287.pdf

DR2: Diffusion-based Robust Degradation Remover for Blind Face Restoration (CVPR2023)

https://arxiv.org/pdf/2303.06885.pdf

Video Probabilistic Diffusion Models in Projected Latent Space (CVPR2023)

https://sihyun.me/PVDM/

https://arxiv.org/pdf/2302.07685.pdf

Real-world Applications with Diffusion Models

Most applications built upon Stable Diffusion (SD). Despite the most popular text-to-image generation, here are some interesting applications:

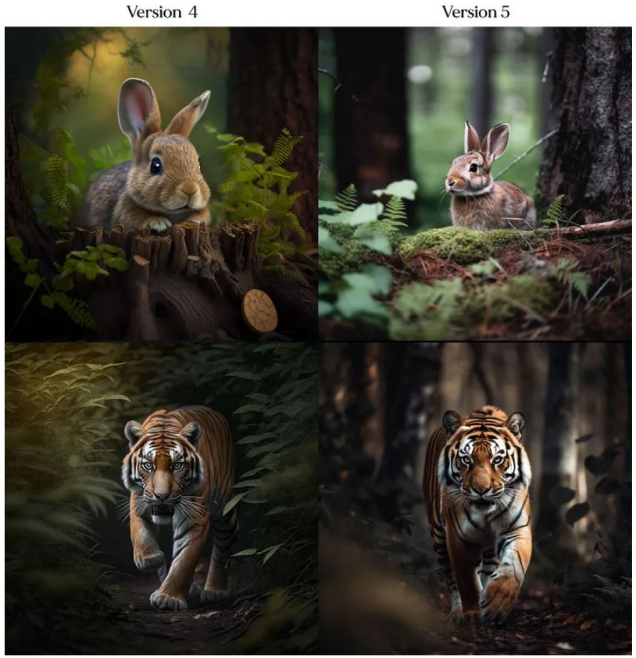
1. Image-to-image generation with text prompt (ControlNet)
2. Video generation
3. Image generation with consistency

In this section, I'll less focus on technical details but more on how the communities using diffusion models.

SD Family — Midjourney v5

Much more realistic generation

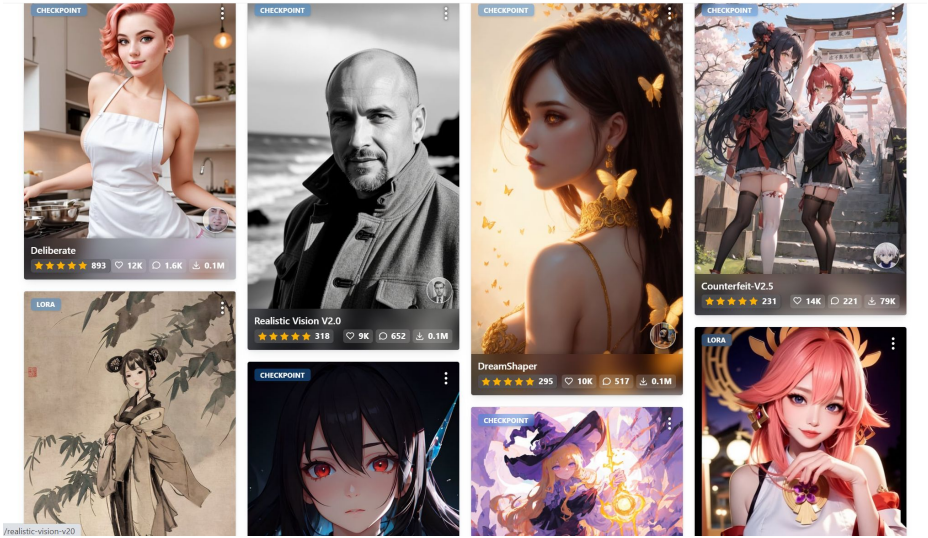
Photos of a dark purple Nissan taken at night and during golden hours, created with MidJourney

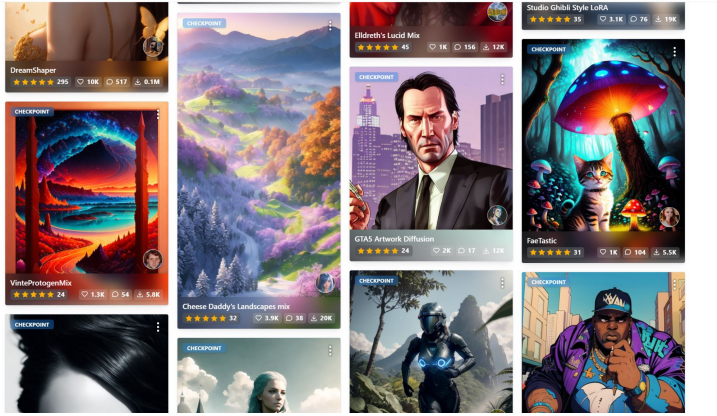


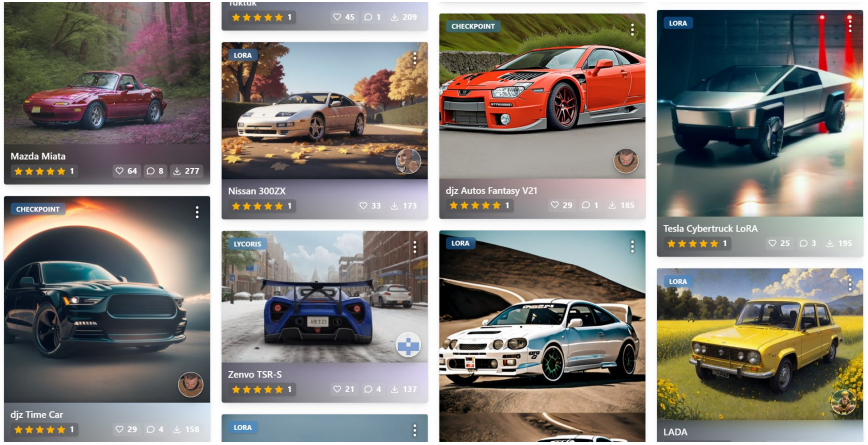
Wildlife photography created with Midjourney V4 and V5

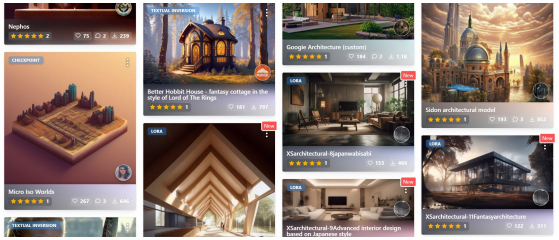
Unerwater photography created with Midjourney V4 and V5

SD Family — Customer Tuned SD





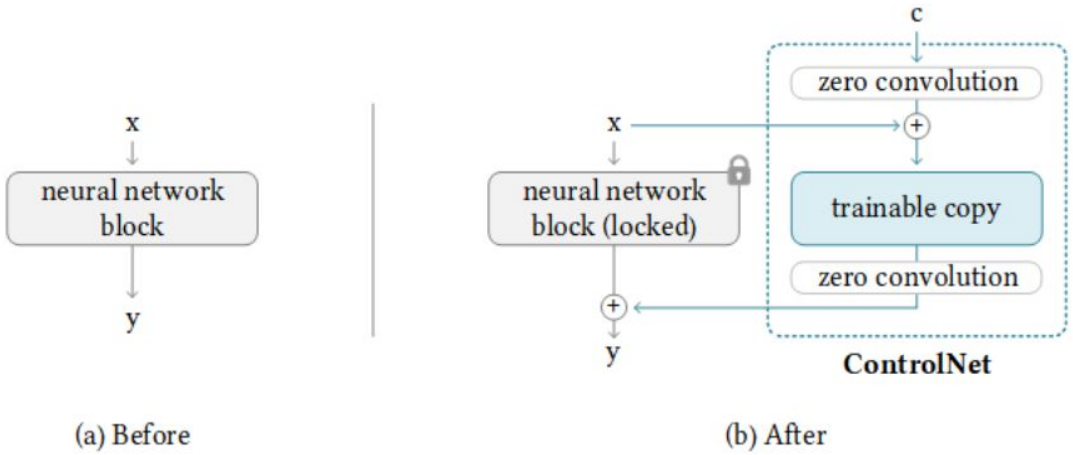




https://civitai.com/

SD Family — ControlNet

Github 15.5k starts



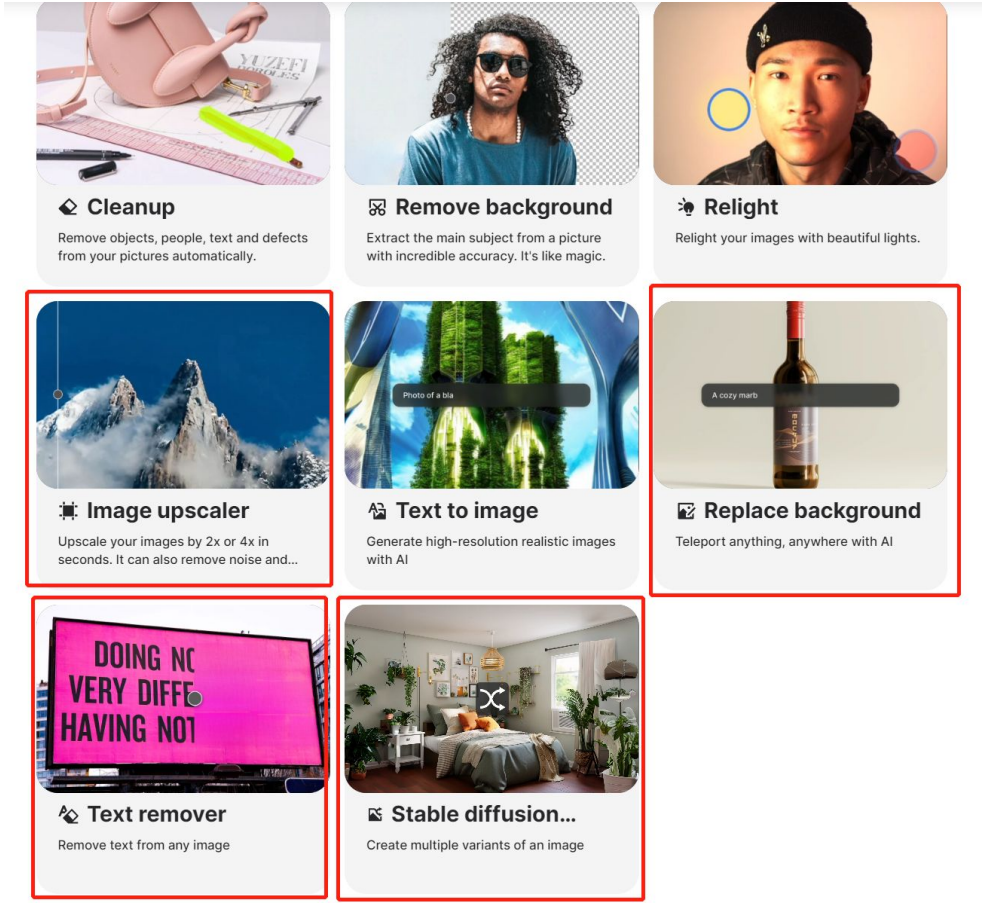
Very similar to GAN layer swap

The idea is to “lock” the pretrained knowledge and finetune a small scale of the network into user-defined tasks. The author claims the method can be run on personal devices.

https://github.com/lllyasviel/ControlNet

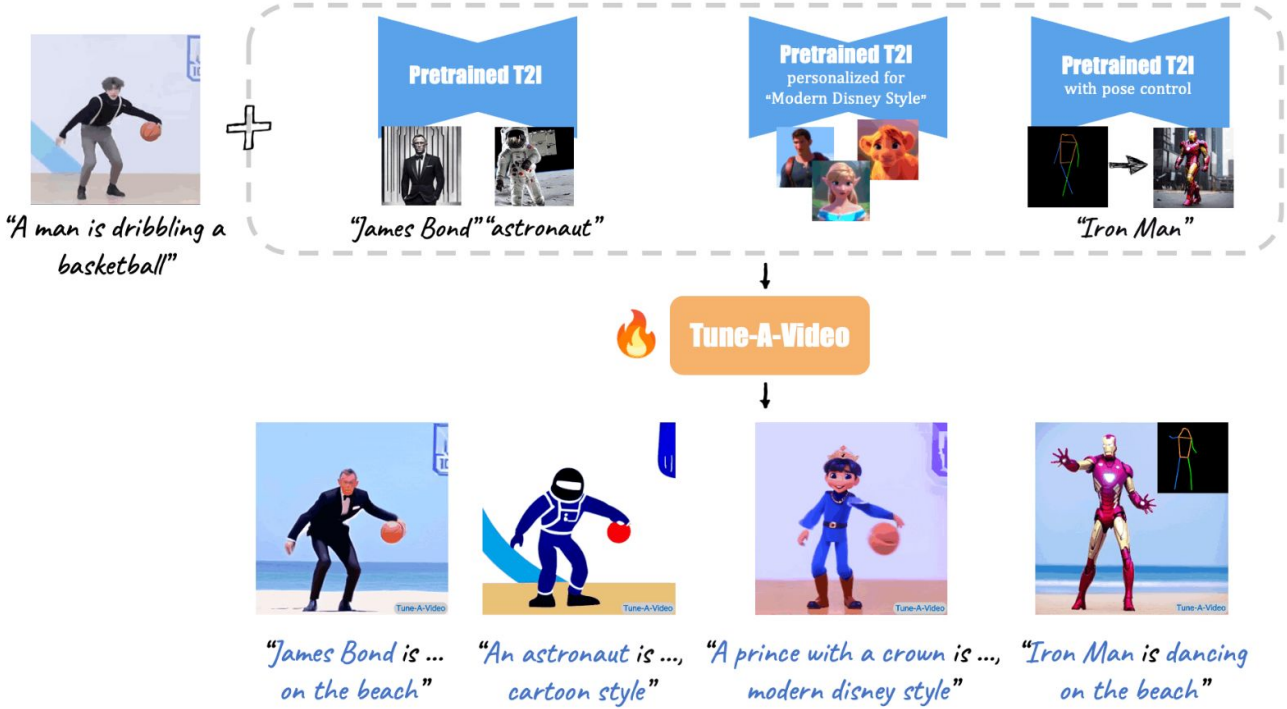
SD Family — Clipdrop

1. SR
2. Replace Background
3. Text Remover
4. Image-to-Image



https://clipdrop.co/

SD Family — Tune-A-Video

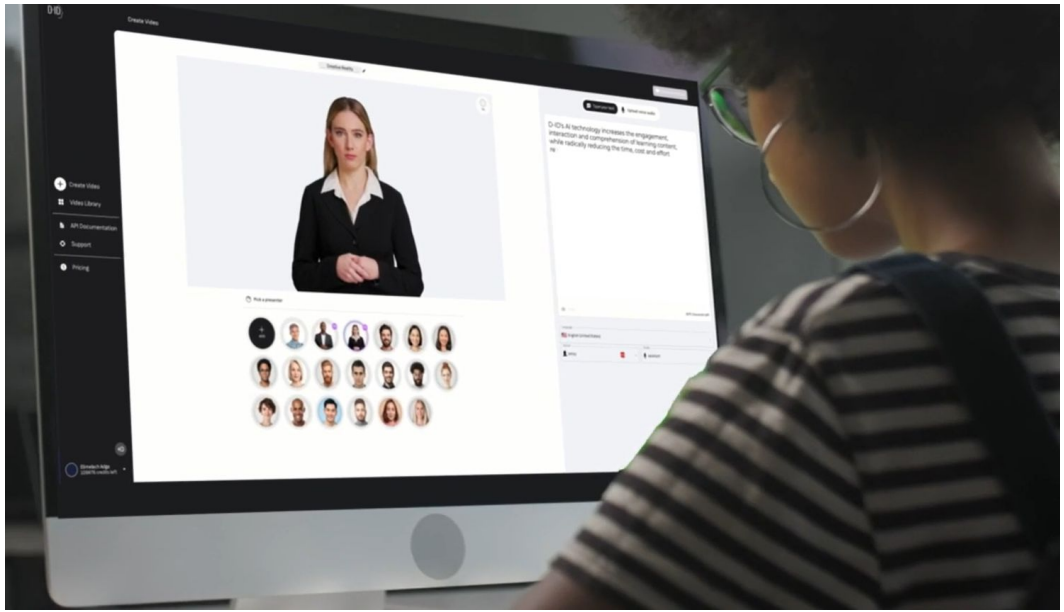


https://tuneavideo.github.io/

SD Family — D-ID

ChatGPT for text and Stable Diffusion for portrait.

I generated a video using this portrait with a description of Yongping Duan, the output is amazing.



SD as a small plugin.

https://www.d-id.com/

SD Family — Gen-2 (Video Generation)

https://research.runwayml.com/gen2

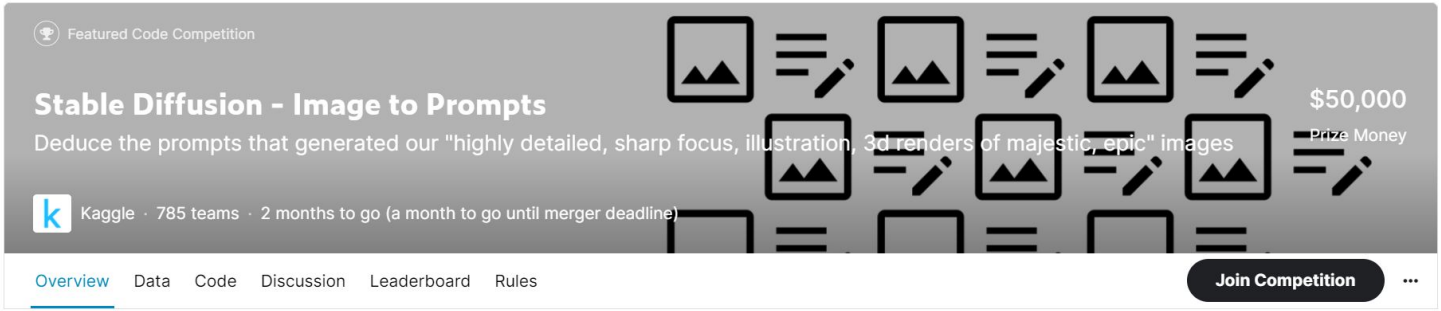
Input Image

A low angle shot of a man walking down a street, illuminated by the neon signs of the bars around him.

Driving Prompt

Output Video

SD Family — Image to Prompt



Overview

- Description
- Evaluation
- Timeline
- Prizes
- Code Requirements

Goal of the Competition

The goal of this competition is to reverse the typical direction of a generative text-to-image model: instead of generating an image from a text prompt, can you create a model which can predict the text prompt given a generated image? You will make predictions on a dataset containing a wide variety of `(prompt, image)` pairs generated by Stable Diffusion 2.0, in order to understand how reversible the latent relationship is.

Context

The popularity of text-to-image models has spurned an entire new field of prompt engineering. Part art and part unsettled science, ML practitioners and researchers are rapidly grappling with understanding the relationships between prompts and the images they generate. Is adding "4k" to a prompt the best way to make it more photographic? Do small perturbations in prompts lead to highly divergent images? How does the order of prompt keywords impact the resulting generated scene? This competition tasks you with creating a model that can reliably invert the diffusion process that generated to a given image.

In order to calculate prompt similarity in a robust way—meaning that "epic cat" is scored as similar to "majestic kitten" in spite of character-level differences—you will submit embeddings of your predicted prompts. Whether you model the embeddings directly or first predict prompts and then convert to embeddings is up to you! Good luck, and may you create "highly quality, sharp focus, intricate, detailed, in the style of

SD Family — Image to Prompt

MJ released same feature

AI Generated Prompt

an aerial view of a city at night, a detailed matte painting by Feng Zhu, cgsociety, video art, playstation 5 screenshot, matte painting, reimagined by industrial light and magic

Run Prompt on Stable Diffusion › Try another image ›

https://imagetoprompt.com/

SD Family — Image to Prompt

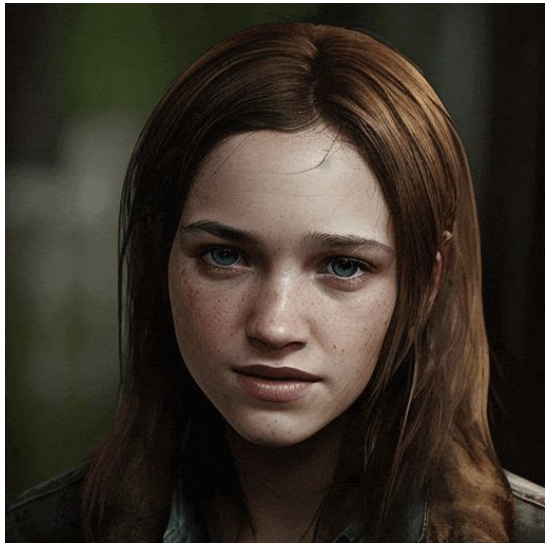
AI Generated Prompt

a man wearing glasses and a red hoodie, a character portrait by Xia Shuwen, pexels contest winner, generative art, quantum wavetracing, 2d game art, physically based rendering

Run Prompt on Stable Diffusion › Try another image ›

SD Family — Multi Frame Rendering

Built only on SD. No finetune or extra model needed.



https://xanthius.itch.io/multi-frame-rendering-for-stablediffusion