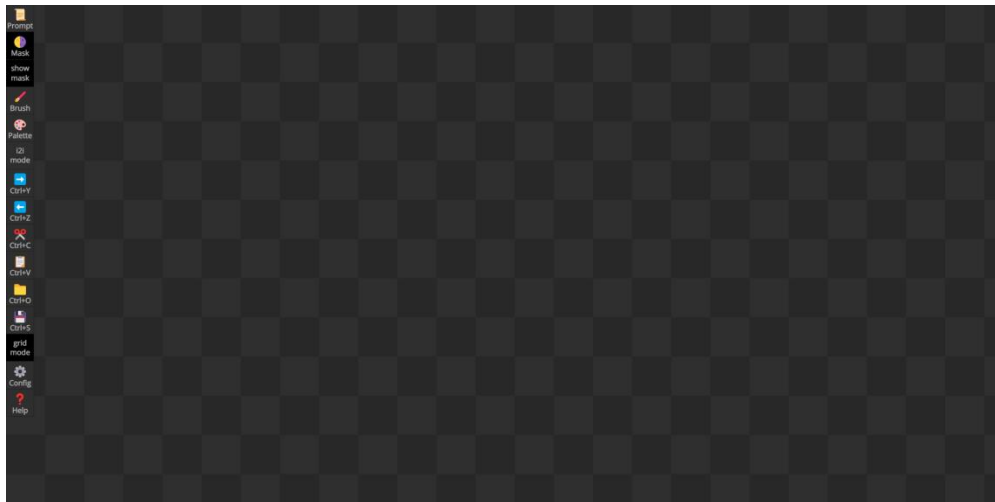
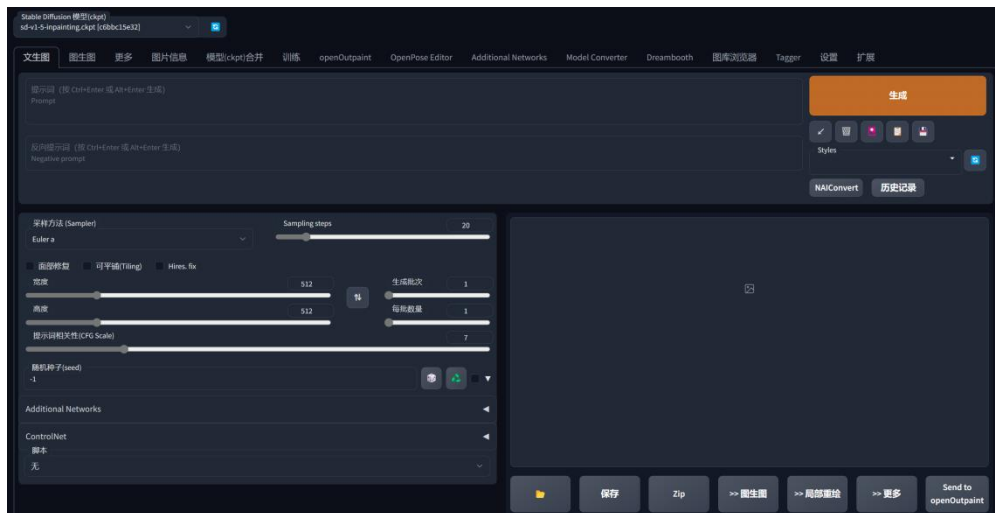


Stable Diffusion 简析

Han Yansong

1、基本组成：前端+模型

前端：webui、painthua……



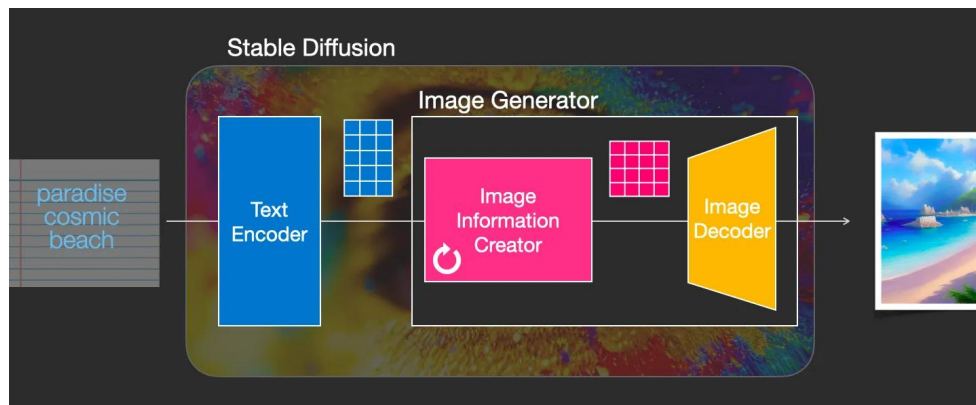
模型

- 512-inpainting-ema.ckpt
- final-pruned.ckpt
- final-pruned.vae.pt
- sd-v1-5-inpainting.ckpt
- v1-5-pruned-emaonly.safetensors
- v2-1_768-ema-pruned.ckpt
- v2-1_768-nonema-pruned.ckpt

2、基本原理

Latent Diffusion Models

2.1 Stable Diffusion 模型的组成



1) Clip Text 用于文本编码

输入：文本

输出：77 个表征向量，每个有 768 维度

图像编码器和文本编码器的组合，将文本说明与图像相匹配。

2) 神经网络+调度，逐渐处理/扩散信息到潜在（信息）空间。

输入：文本向量 和 起始多维数组（结构化的数字列表，也叫 Tensor）由噪声组成。

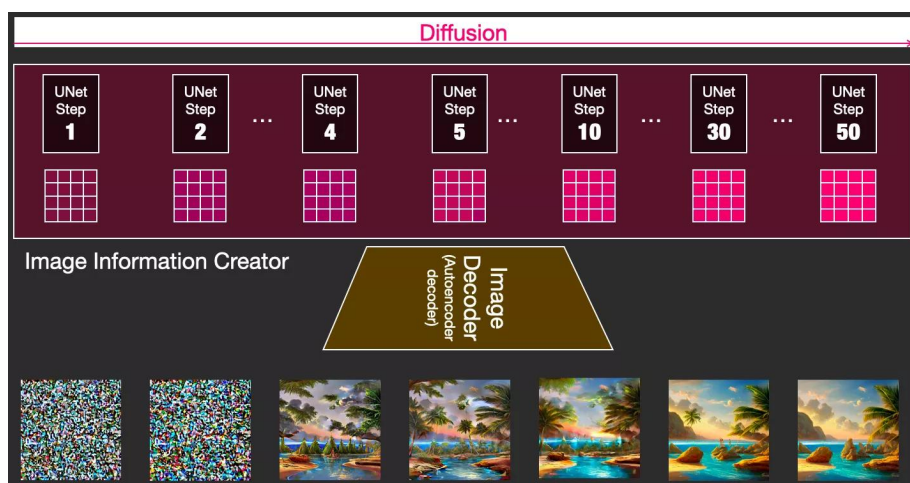
输出：处理后的信息数组

3) 自编码解码器，绘出最后的图像，借助之前处理过的信息数组。

输入：处理过的信息数组（维度: (4, 64, 64)）

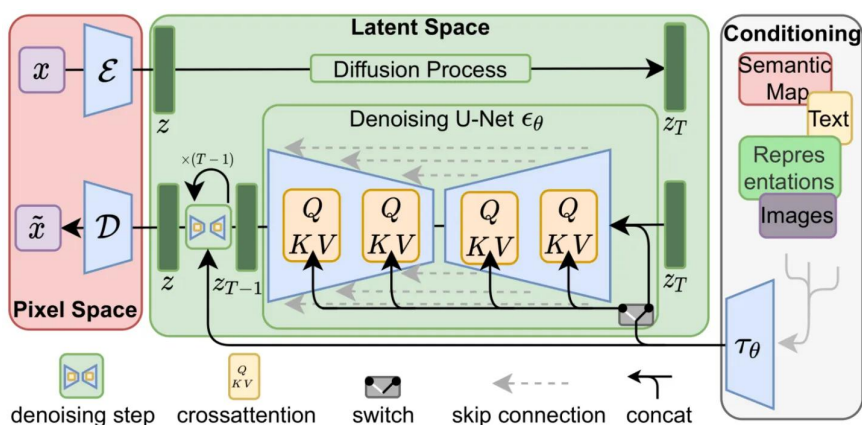
输出：结果图像（维度: (3, 512, 512) 分别是红 / 绿 / 蓝，宽，高）

2.2 核心：扩散过程



获取表达输入文本的表征向量，以及一个随机启动的图像信息数组（seeds），每一步都对输入潜在数组上操作产生更相似于输入文本的新潜在数组，直至设定步数（sampling steps）。

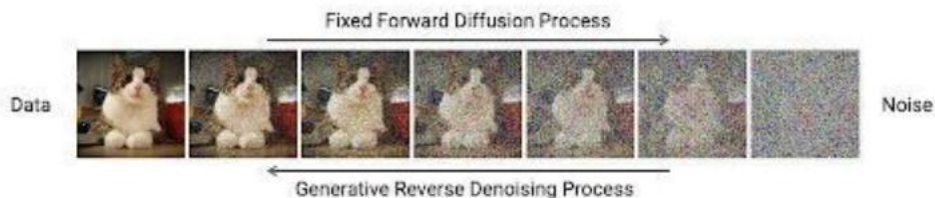
2.2.1 潜在扩散模型



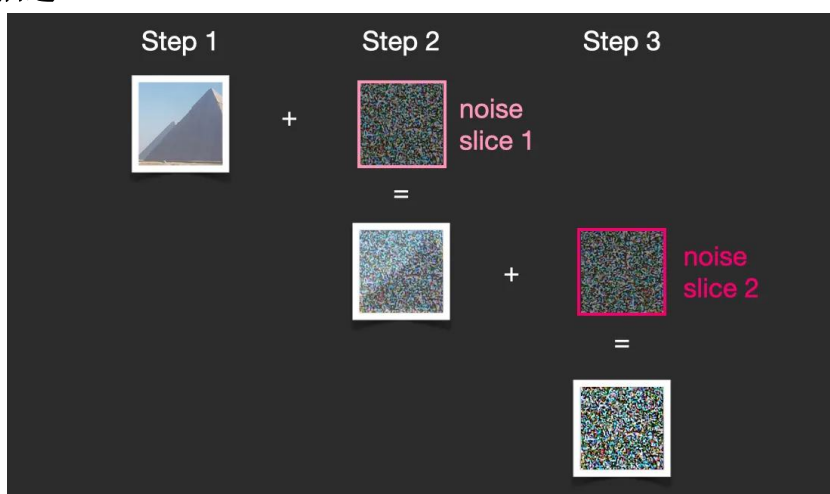
①扩散模型 DDPM:

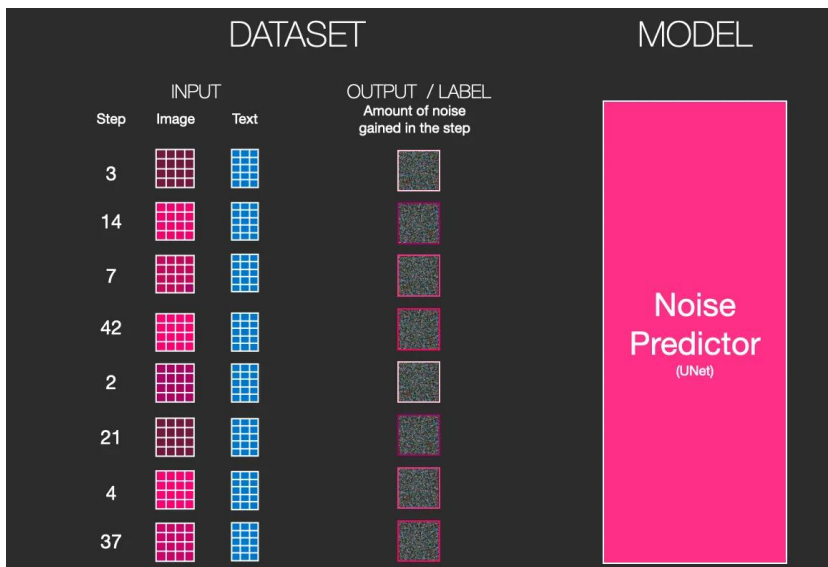
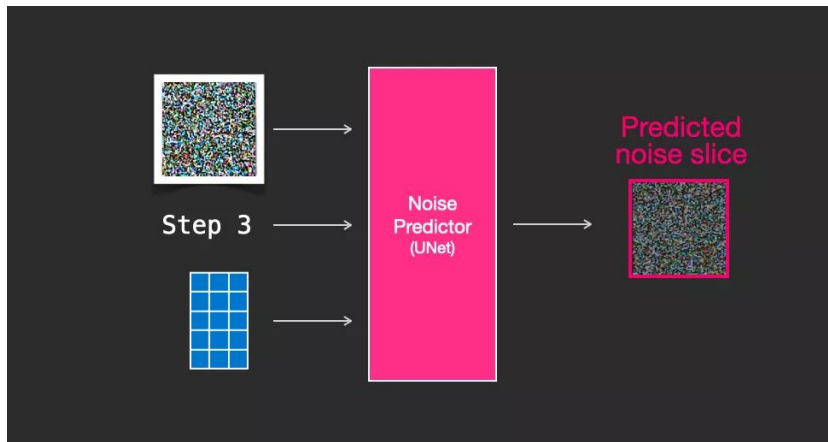
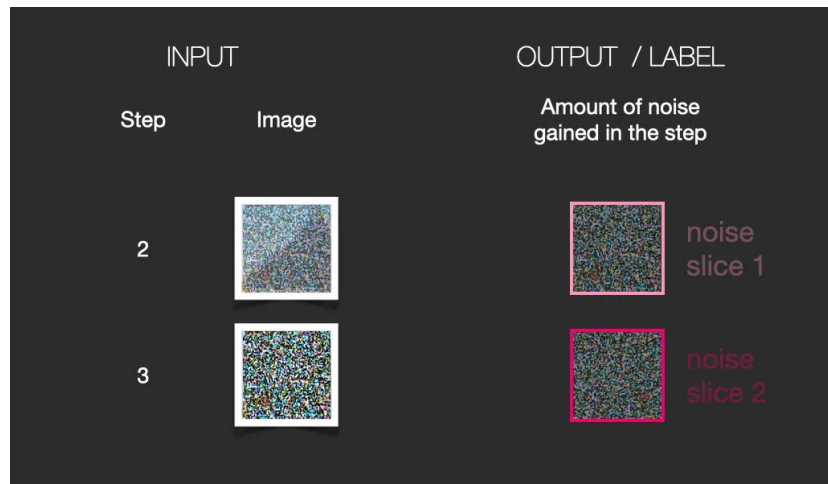
对于输入构建了一个离散步骤的马尔可夫链，不断加入随机噪声直至其成为无法辨识的纯噪声为止的前向过程。这个过程对应着分子热力学里的扩散过程。而模型学习的则是如何从噪声分布里出发，逐渐去除噪声将图片还原至原始的数据分布中。

在 Diffusion Model 中定义了一个扩散步骤的马尔可夫链（当前状态只与上一时刻的状态有关），慢慢地向真实数据中添加随机噪声（前向过程），然后学习反向扩散过程（逆扩散过程），从噪声中构建所需的数据样本。



形象简化的描述:





MODEL: Noise Predictor(UNet)

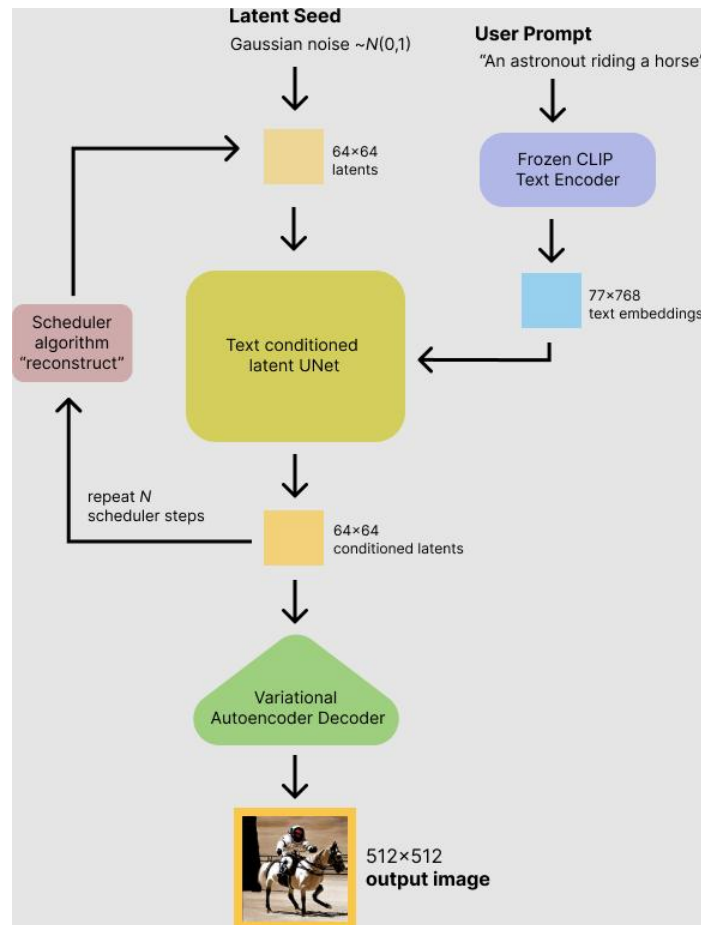
输入：加噪步数、图像、文字描述。

输出：前一步增加的高斯噪声

目的：通过消除噪音绘制图片（最终结果与训练数据集有关）

②潜在（优化）

在潜在空间而不是像素空间扩散，高频的、难以察觉的细节被抽象出来。



稳定扩散模型将潜在种子和文本提示作为输入。然后使用潜在种子生成大小的随机潜在图像表示 64×64 其中，文本提示转换为大小的文本嵌入 77×768 通过 CLIP 的文本编码器。

接下来，U-Net 迭代地对随机潜在图像表示进行降噪，同时以文本嵌入为条件。U-Net 的输出作为噪声残差，用于通过调度器算法计算去噪的潜在图像表示。许多不同的调度程序算法可用于此计算，每种算法都有其优点和缺点【PNDM 调度程序（默认使用）DDIM 调度程序 K-LMS 调度程序】

关于调度程序算法函数如何超出本笔记本本范围的理论，但简而言之，应该记住，他们根据先前的噪声表示和预测的噪声残差计算预测的去噪图像表示。有关更多信息，我们建议研究阐明基于扩散的生成模型的设计空间

去噪过程重复约 50 次，以逐步检索更好的潜在图像表示。完成后，潜在图像表示由变分自动编码器的解码器部分解码。

2.3 拓展：inpainting

通过在反向扩散迭代期间从给定像素采样来调节生成过程

首先从纯粹的噪音开始，然后对图像逐级降噪，中间的每一步使用图像已知部分来填充未知部分。（由于 DDPM 被训练生成一个位于数据分布中的图像，它自然地想要生成一致性的结构→邻近区域的纹理相匹配）



3、不同模型

以 Stable Diffusion 2.1 和 1.5 举例:

1. **网络架构:** Stable Diffusion 2.1 采用了一种名为“Transformer”的新型网络架构，而 1.5 版本使用的是常见的卷积神经网络架构。这种网络架构的不同可能导致模型的性能和效率方面的差异。
2. **训练数据:** Stable Diffusion 2.1 使用了更大规模和多样化的数据集进行训练，包括 ImageNet、COCO、OpenImages 和 Places 等数据集。而 1.5 版本仅使用了 ImageNet 数据集。
3. **训练方法:** Stable Diffusion 2.1 使用了一种名为“对比散度训练”的新型训练方法，该方法可以帮助模型更好地处理复杂的图像数据。而 1.5 版本使用的是常规的监督学习训练方法。

4、几点小疑问

为什么能生成许多描述结合的图片?怎么保证连续性的?

Teddy dog with a pair of wings → 它怎么知道翅膀长在背部



补充：数据集

LAION5B

image/text

Status: Released

A dataset consisting of 5.85 billion CLIP-filtered image-text pairs, featuring several nearest neighbor indices, an improved web-interface for exploration and subset generation, and detection scores for watermark, NSFW, and toxic content detection.

Backend url: <https://knn.lai>
Index: laion5B-H-14

lung tissue

Clip retrieval works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions ☒
Display full captions ☐
Display similarities ☐
Safe mode ☒
Remove violence ☒
Hide duplicate urls ☒
Hide (near) duplicate images ☒
Enable aesthetic scoring ☐
Aesthetic score
Aesthetic weight
Search over
Search with multilingual clip ☐

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates? KNN search

תמונות מיקרוסקופיות מדהימות!

coupe histologique Histalim

تصاویر خارقالعاده دانشمندان از دنیای بدون داننها و

Very low magnification micrograph of an alveolar s...

صورة مجهرية لما بداخل الطماطم

Fig. 40-4c 50 µm Lung tissue Figure 40.4 Internal ...

زلی که از بارکشت سرطان جلوگیری می کند

Micrograph Tissues

coupe histologique Histalim

Osteoporosis

Connective tissue function

ACHARIACEAE Erythrospermum candidum

Fetal Monkey Lung H&E LMR14.4.4.3 Gestational day ...

Embolia легочной артерии: причины и развитие, при...

Привычная еда под микроскопом брокколи, еда, микро...

La mutation du gène DMD est responsable de

Еда под

heart tissue

Fetal Monkey Lung H&E LMR14.6.4.3 Gestational day ...

مرض ضمور العضلات دوشين

Backend url: <https://knn.lai>
Index: laion5B-H-14

cute cat

Clip retrieval works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions ☒
Display full captions ☐
Display similarities ☐
Safe mode ☒
Remove violence ☒
Hide duplicate urls ☒
Hide (near) duplicate images ☒
Enable aesthetic scoring ☐
Aesthetic score
Aesthetic weight
Search over
Search with multilingual clip ☐

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates? KNN search

cat kitten29

火燭貓味衣服搞笑搞怪變身裝小貓幼貓寵物保暖加厚秋冬季網紅同款 韓語空間

ten kot może mieć wszystko <3

Cutie

Crystal Eyes

grumpy cat cute cat Love this cat! Baby Animals, F...

sooo cute :3

Кошки, котята Шотландская вислоухая, цена 200 бел....

Ridiculously Photogenic Cat

Happy Smile's profile picture

Biały i malutki kociaćzek. ładny?

Cute Cat theme v.2

猫カフェもふにゃん

cat by Coalbiter

animated cat

Baby Cat Kitten Face - Public

cat, cute, and

世界上最古老的长毛猫 安哥拉猫的特点是

参考文献

High-Resolution Image Synthesis with Latent Diffusion Models 2112.10752
Denoising Diffusion Probabilistic Models 2006.11239
Reaint: Inpainting using Denoising Diffusion Probabilistic Models 2201.09865
<https://blog.csdn.net/amusi1994/article/details/122711287>
<https://zhuanlan.zhihu.com/p/581186398>