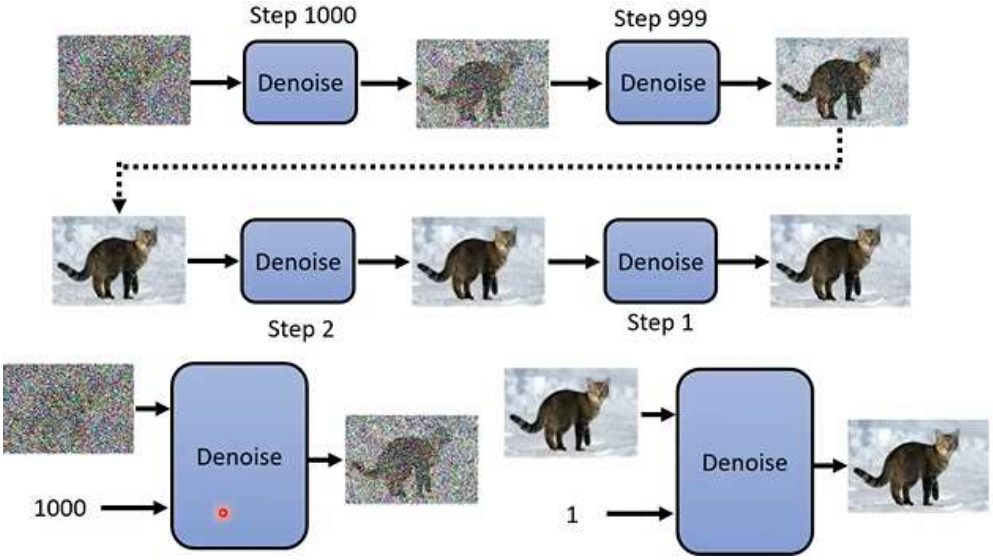
8. Diffusion model原理

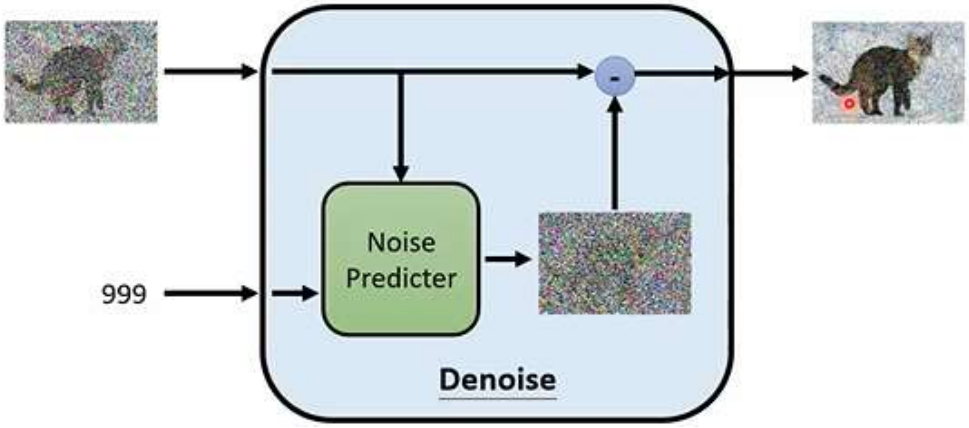
笔记本： 【课】原理-李宏毅 deep leaning
创建时间： 2023/4/25 15:32 **更新时间：** 2023/4/26 18:42
作者： 1256876216@qq.com

一、如何运作

- 一遍一遍使用同一个denoise 模型进行demoise，denoise次数事先定义好，步骤从大到小， 称为 reverse Process。
- 每个denoise model 的输入包括带噪点的图以及当前的步骤，会根据当前的 step 来调整回应



- denoise内部：根据step进行预测出这个图片的一个噪点图，再将图片去掉噪点，得出输出图片



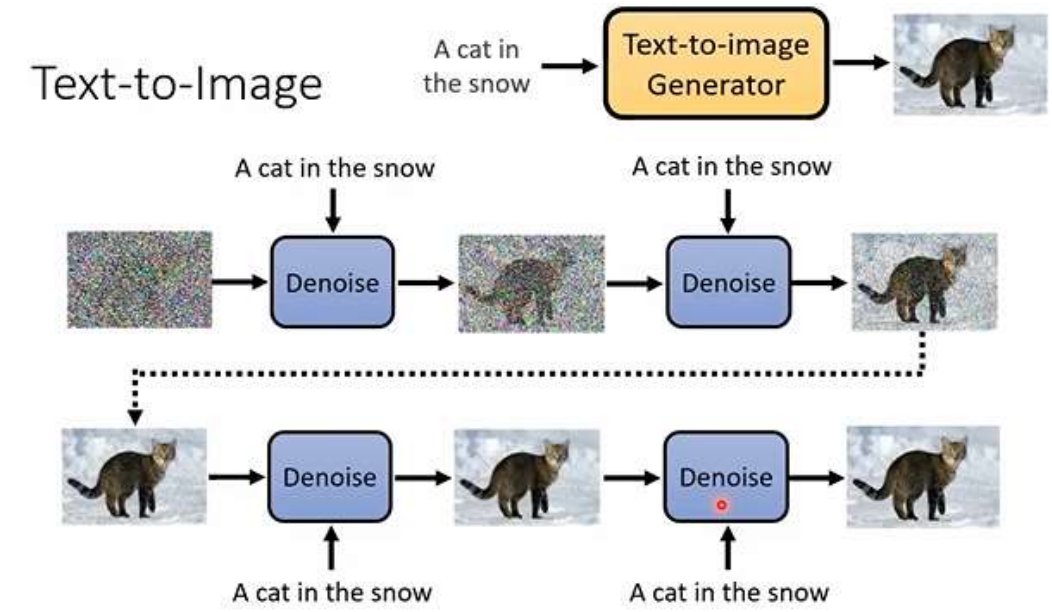
二、如何训练 Noise Predicter

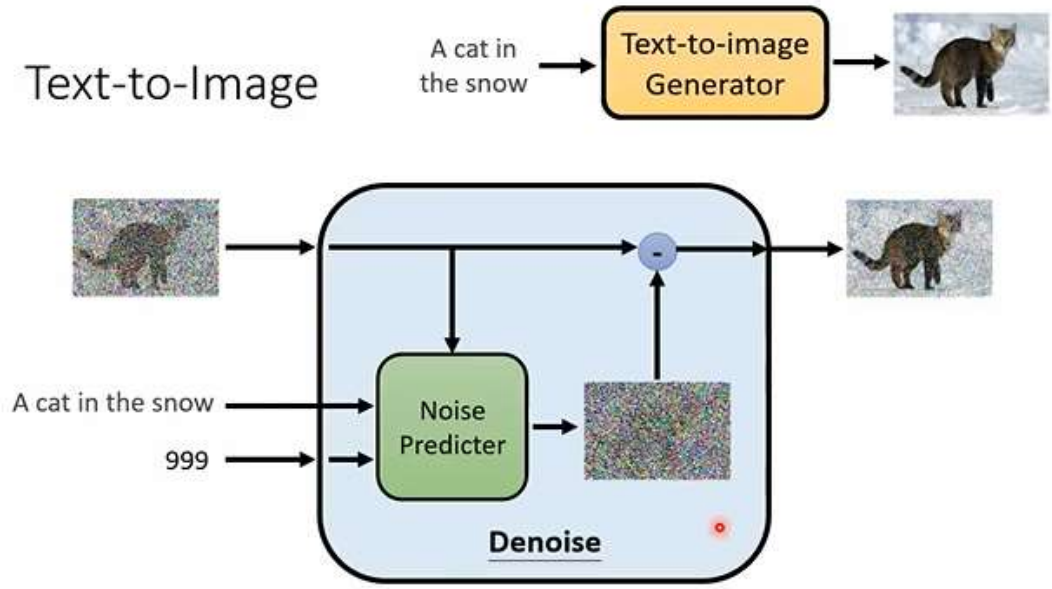
1.训练资料

是人为创造出：
正常图片+Radon sample => 下一步图片，重复操作，这个过程称为【Forward Process】

2. Text-to-Image:

训练数据仍需要成对的图片与文字相对应，通常训练资料中的文字不止包含英文还有中文日文等

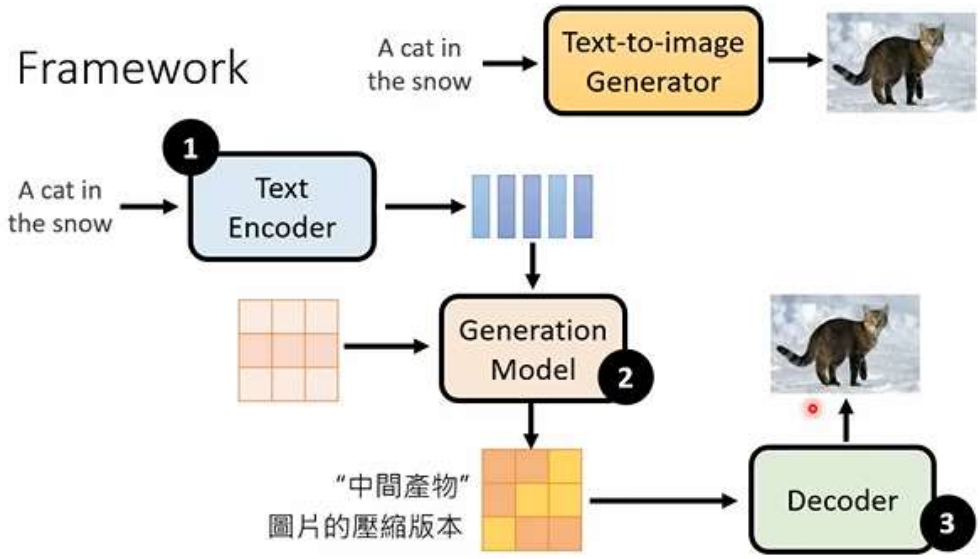




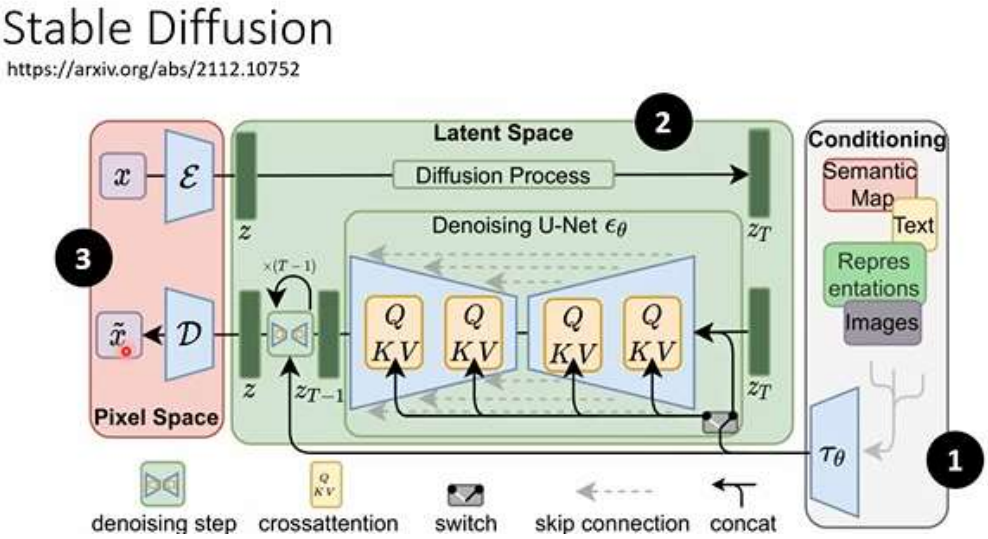
三、stable Diffusion

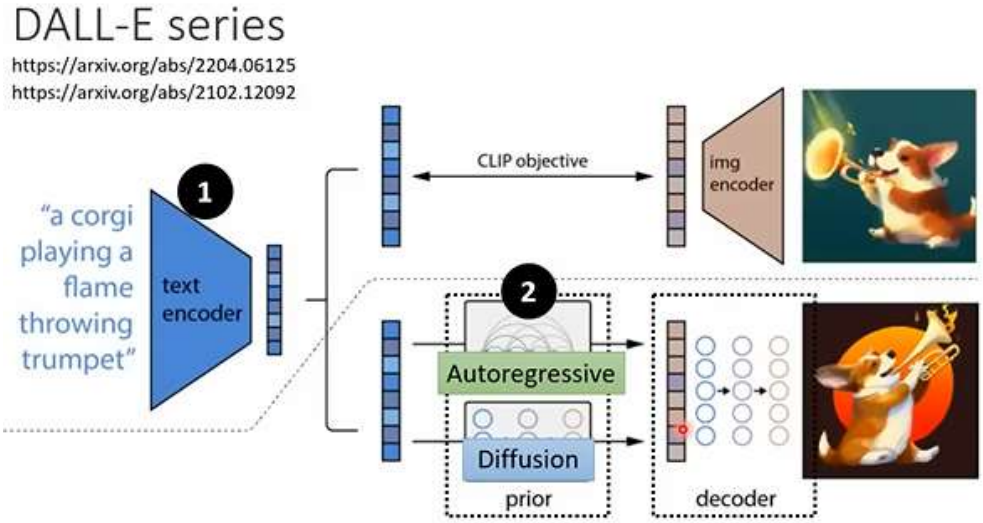
1. Diffusion架构framework

- Text-to-image Generator
  - Text Encoder
  - Generator Model
  - Decoder
- 原理三个model分开训练，再组合起来



- 目前较好的文字生图论文里都是这个原理，例如：

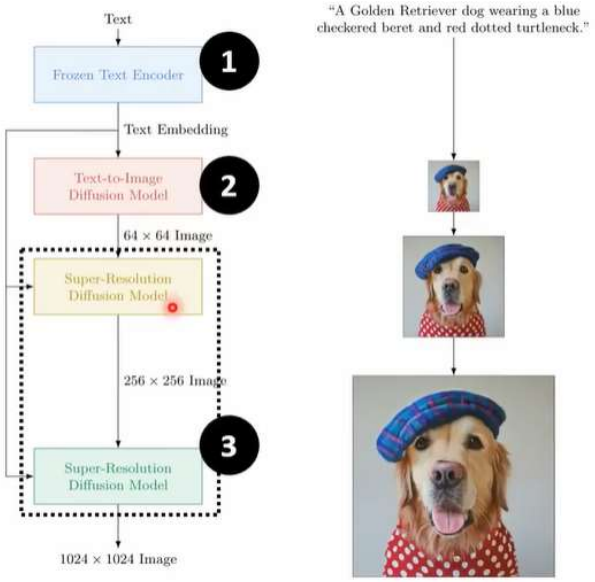




Imagen

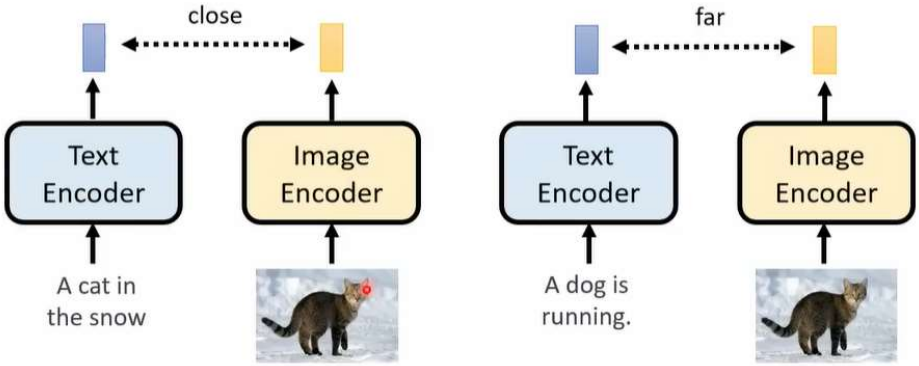
https://imagen.research.google/

https://arxiv.org/abs/2205.11487



2. 评估生成图片好坏的标准：

- FID : frechet Inception distance
  生成图片与真实图片两个高斯分布的距离，越小越好；条件需要sample大量的图片
- CLIP：Contrastive Language
  -Image Pre-training
  - 使用400million 组图片文字对训练出来的模型
  - 生成的图片进入ImageEncoder生成图片向量，文字进图text-Encoder 生成文字向量，若这是一对数据则两个向量越近越好，反之越远越好
    Max Likelihood = Minimize KL Divergence（衡量两个之间的差别）



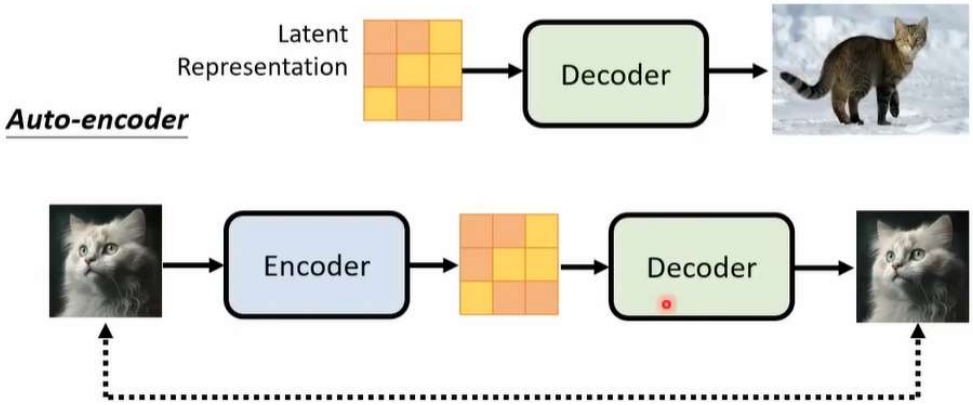
3. Text Encoder

3. Generator Model

5. Decoder

- 作用：将中间生成的图片【小图】还原成图片
- decoder训练可以不需要label，例如小图生大图

- 中间产物若非小图而是latent representation（隐层特征），需要训练一个Auto-encode,使结果与输入接近，Decode就可以还原图片



- 对比vae与diffution model