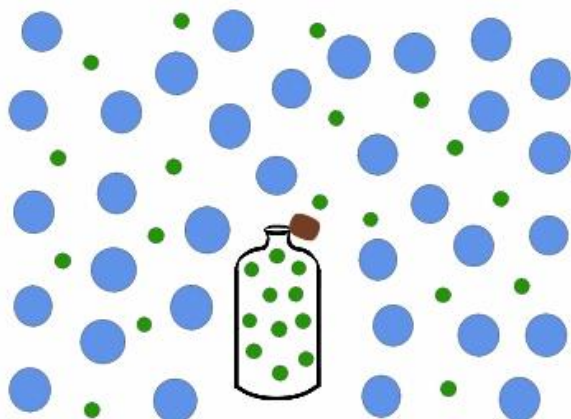


## 一. 背景介绍

扩散模型的灵感来源于什么？非平衡热力学！



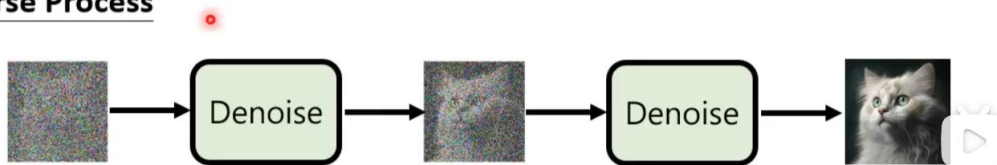
一滴墨水滴入清水，随着时间推移，墨水逐渐扩散到整个水体，最终完全均匀混合，这个过程是一个逐步从有序到无序的变化过程。那把这个过程逆转，就可以从一瓶墨水开始，获得一杯干净的清水。

描述粒子在势场中的随机运动，每一步既有确定性成分（势场梯度方向），又有随机波动（随机噪声）。既然“扩散”的过程的规则是固定的，那么“逆扩散”的过程也就可以确定。我们对于某一杯“符合要求”的浑浊的水，我们只要按照某种规则去噪，就可以得到想要的水。

### Forward Process

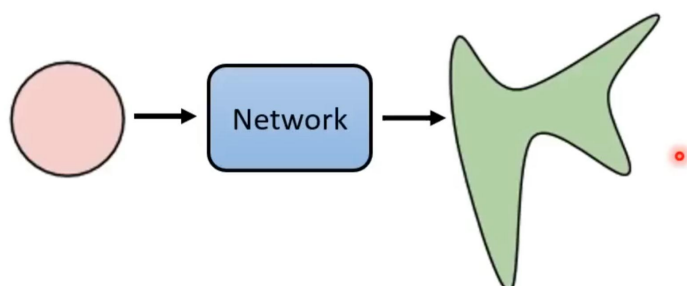


### Reverse Process

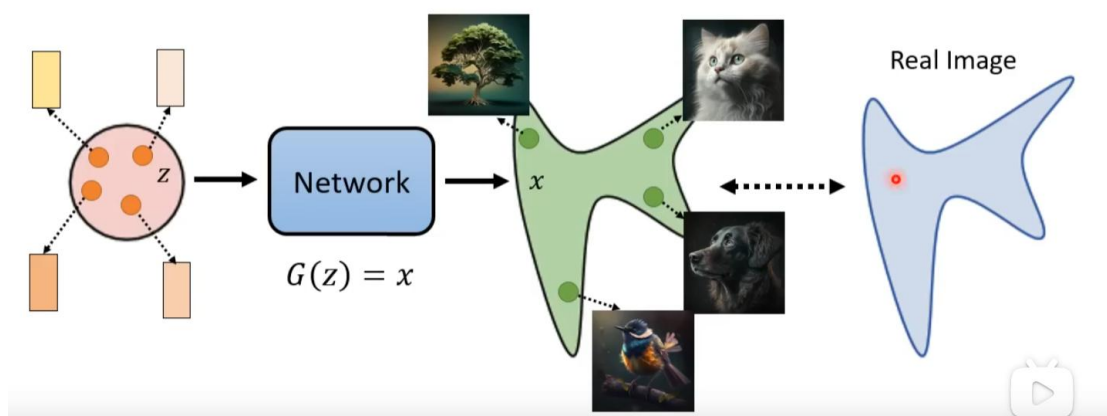


想要理解扩散模型，我们还需要先理解影像生成模型本质的共同目标是什么：

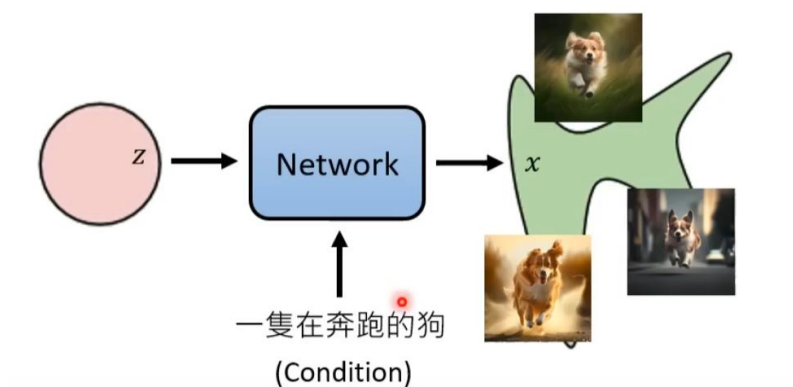
## 影像生成模型本質上的共同目標



也就是给定一个分布空间，我经过网络  $G(x)$ ，可以投影到另一个分布空间上：

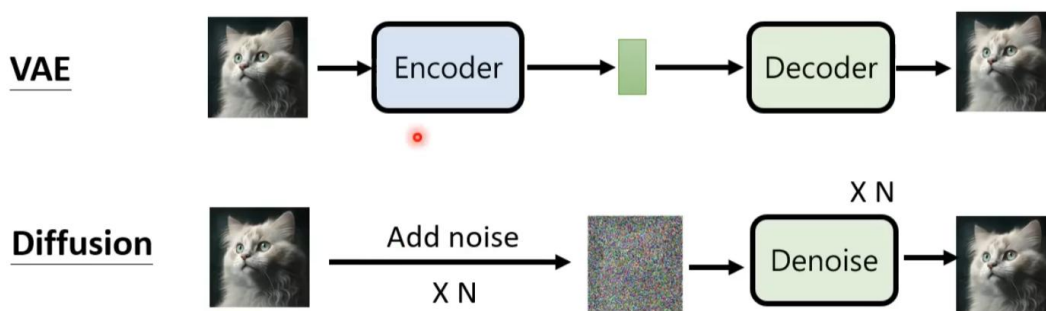


这个空间，和我真正的空间要尽量的保持一致。



文生图做的是这样一个过程。

## VAE vs. Diffusion Model



我们来看上周秦戟同学组讲的 VAE 和扩散模型的区别，VAE 的 ENcoder 过程其实相当于扩散模型加噪声的过程，而 VAE 所创造的隐空间则对应于扩散模型的这个噪声空间。解码过程对应于去噪过程，也就是 network 所做的工作。Network 所生成的空间，大家可以看作我们的目的是生成一个真是空间的极大似然估计。

如果还不能明白，那我们再举一个例子。

可乐的配料表：水、果葡糖浆、白砂糖、食品添加剂(二氧化碳、焦糖色、磷酸、咖啡因)、食用香精。

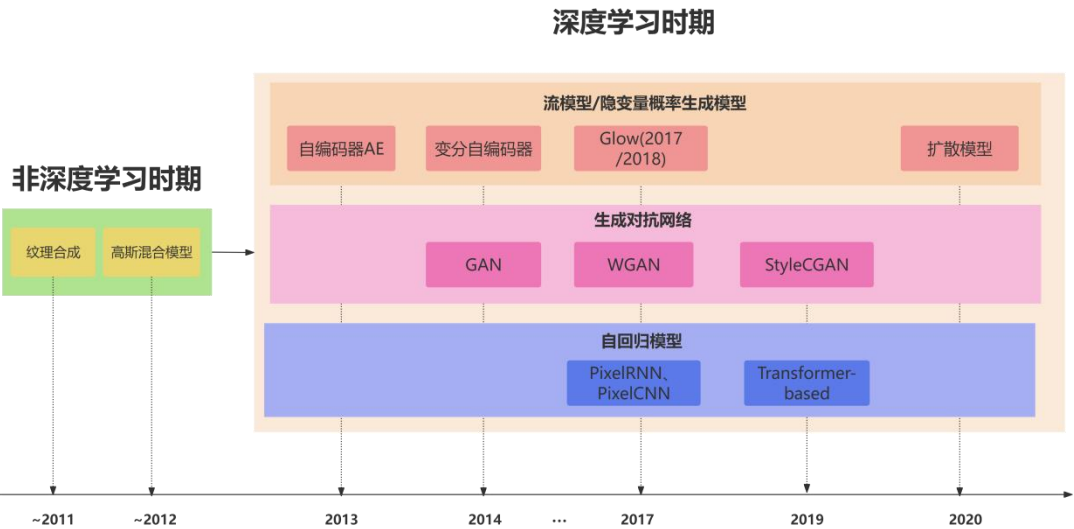
雪碧的配料表：水、果葡糖浆、白砂糖、食品添加剂（二氧化碳、柠檬酸、柠檬酸钠、苯甲酸钠、蔗糖素、安赛蜜）、食用香精。

芬达的配料表：水、果葡萄浆、食品添加剂（二氧化碳、柠檬酸、苯甲酸钠、蔗糖素，安赛蜜，日落黄、柠檬黄）、食用香精。

假如配料空间是：水、果葡糖浆、白砂糖、食品添加剂（二氧化碳、焦糖色、磷酸、咖啡因、柠檬酸、柠檬酸钠、苯甲酸钠、蔗糖素、安赛蜜、日落黄、柠檬黄）、食用香精。

那我想喝可乐，我的 **condition** 就是可乐，那我 **denoise** 相当于要先去掉柠檬酸，然后去掉柠檬酸钠，去掉这些可乐没有的东西，就生成了一杯可乐。

理解了这些，大家就理解了扩散模型的全部思路。我们这时候再回头看生成领域的发展历程：



这张图片是我们按照图像生成这一领域的发展时间线绘制的历程图。

早期的传统统计模型（非深度学习时代）方法包括纹理合成（Texture Synthesis）、基于模板的合成、统计模型（如高斯混合模型）。

深度学习的方法路径主要是流模型 生成对抗网络和自回归模型三个主要分支，我们今天介绍的扩散模型，其实本质就隶属于隐变量概率生成这一分支。