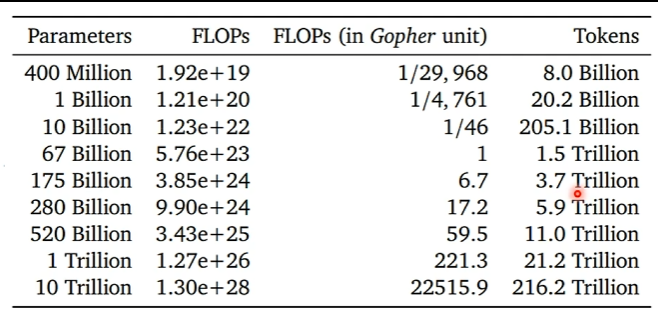
# 大模型和大资料

* u型曲线出现的原因：一知半解吃大亏，有一点点能力，反而做的差
* Switch Transformer（1.6T）：一个大模型内包含多个模组，使用的时候只用部分
* 机器可以正确回答需要两个能力：语言知识（语法，文法，用词），世界知识（对世界的理解、相比语言知识需要量更大）
* 数据准备：滤除有害内容——去除HTML tag（保留项目符号等）——用规则去除低品质的资料——去除重复资料—— 测试资料（为了实验的严谨）
* 固定运算资源下：资料重要还是模型重要即选择小模型大资料、中模型中资料还是大模型小资料

选择中模型中资料，平衡点在那？

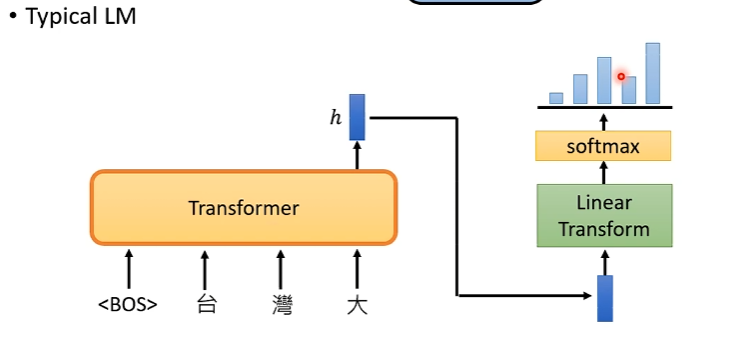
Flops——算例资源 Parameters——模型参数 Tokens——资料



新趋势：在算力一定的情况下：选择小模型大资料

# 语言模型

* 一般的LM（语言模型）:h为一个向量，对h进行分类

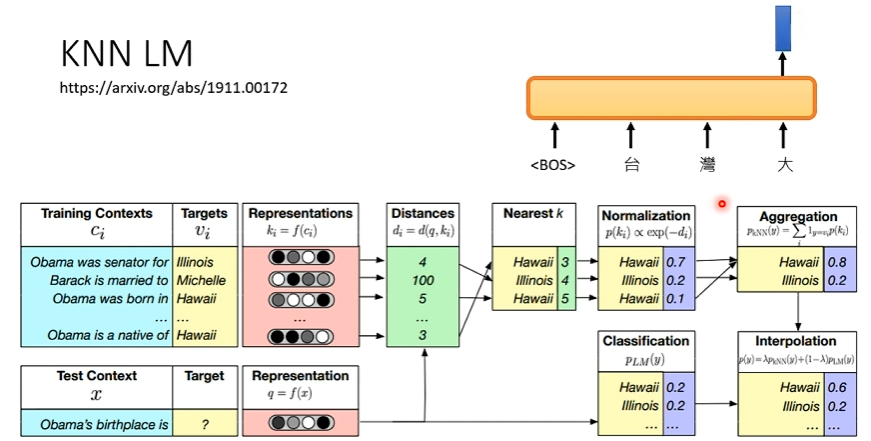


* KNN LM（太花时间：计算Representations花费的）

Representations：参数通过Transformer等得出来的

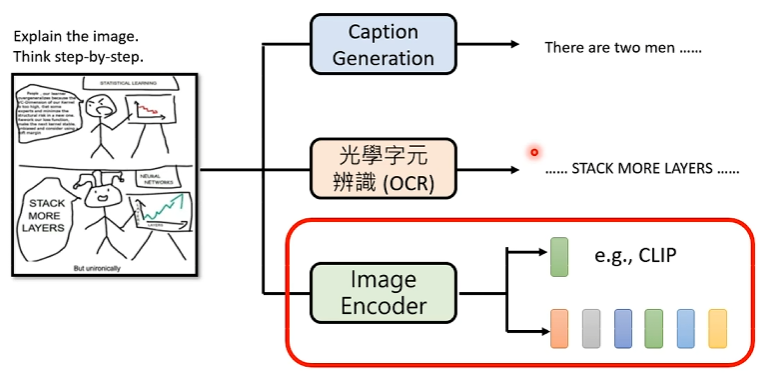
Aggregation：最后输出的概率

Classification：一般LM经过Linear Transform得到的



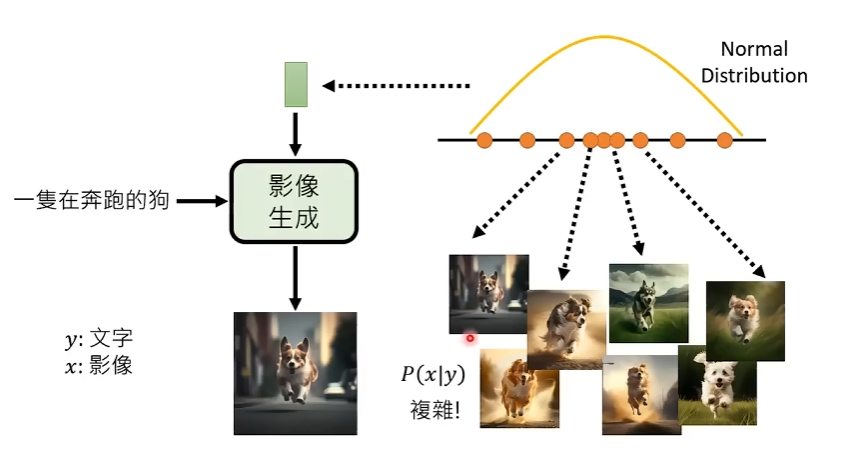
* ChatGPT：如何输入影像

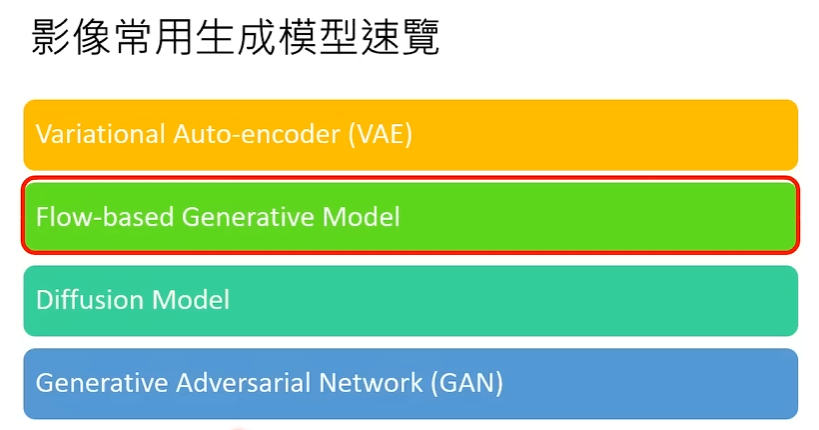
猜测：



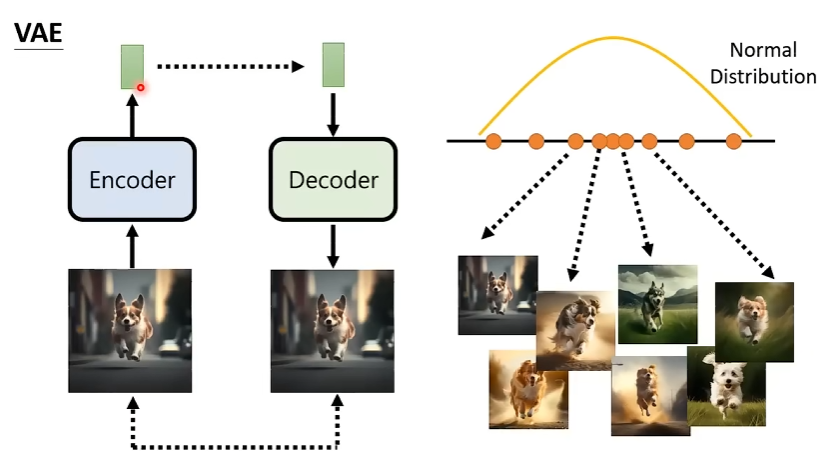
# **图像生成常见模型**

* 图像生成：可采取各个击破，但更多采用一次到位



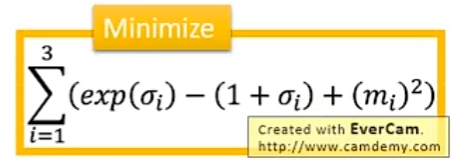


## VAE



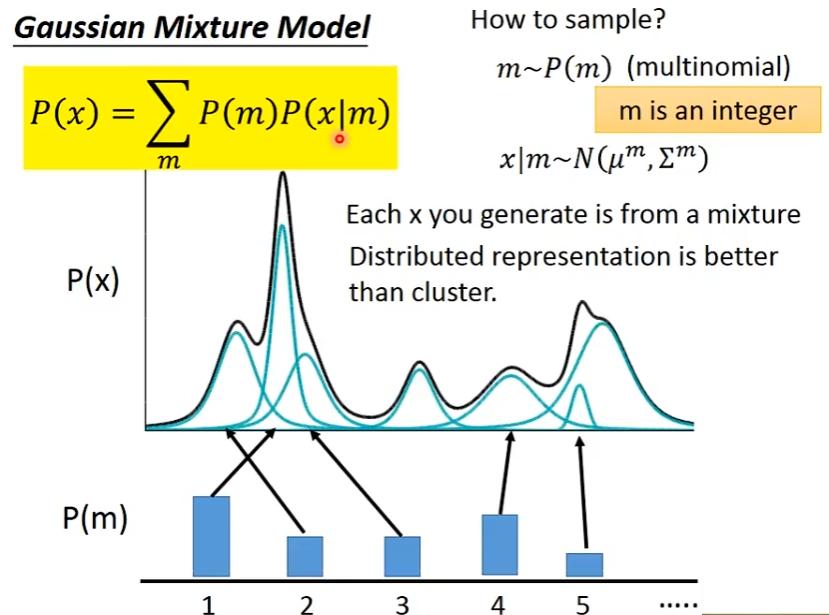
Why VAE：

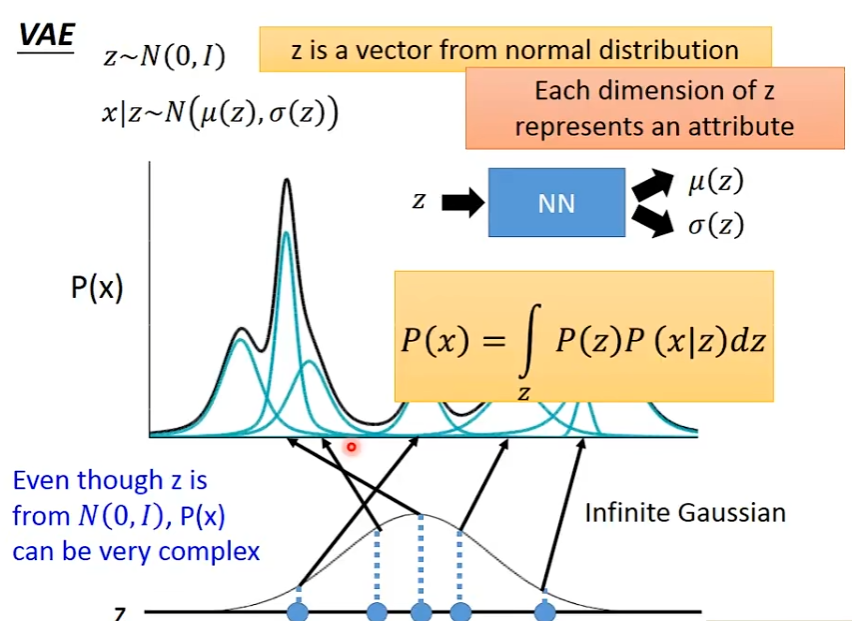
生成code的时候需要加上一个noise

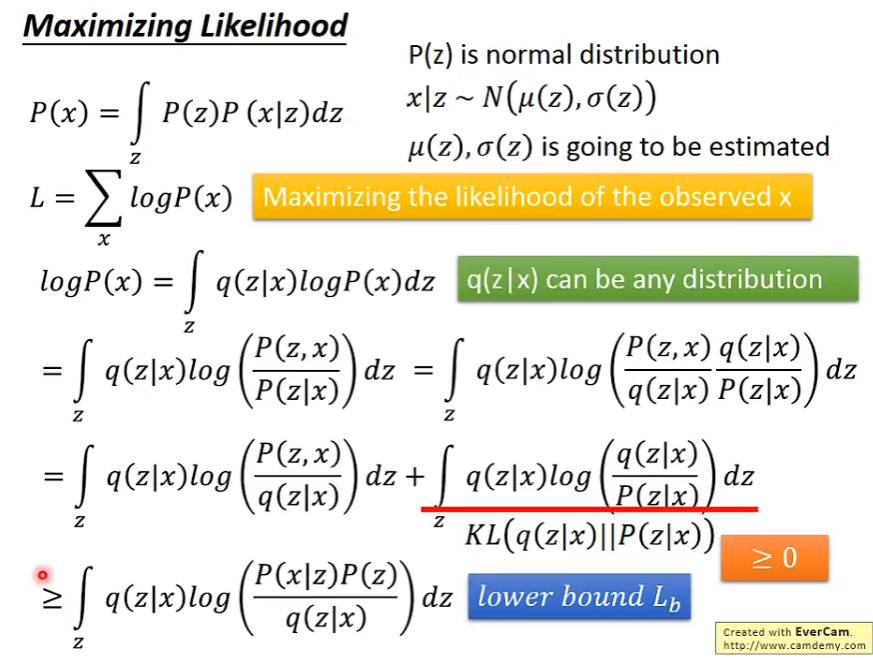


实际；Estimate the probability distribution(机率分布)——P（x）

使用Gaussian Mixture Model得出P（x），m为一个gaussian

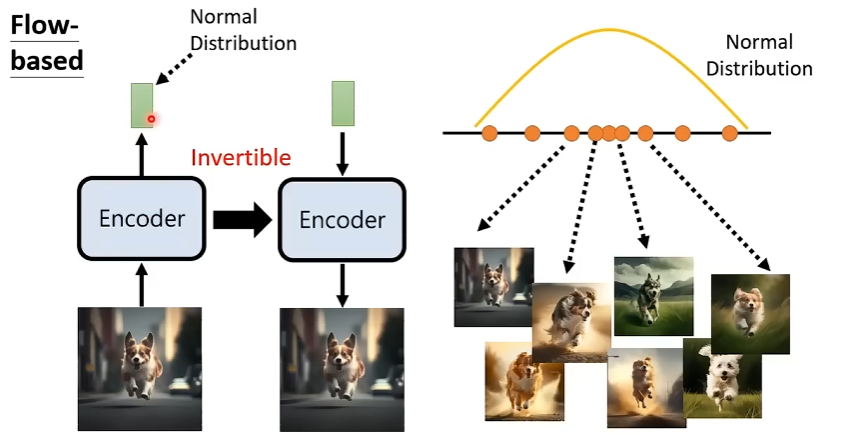


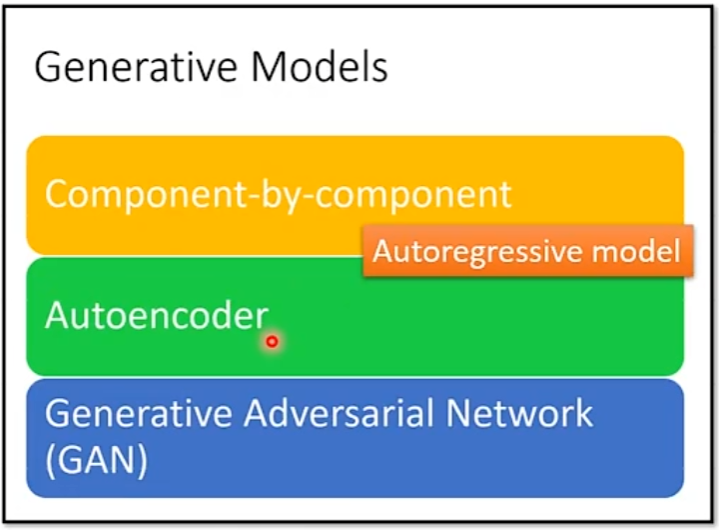




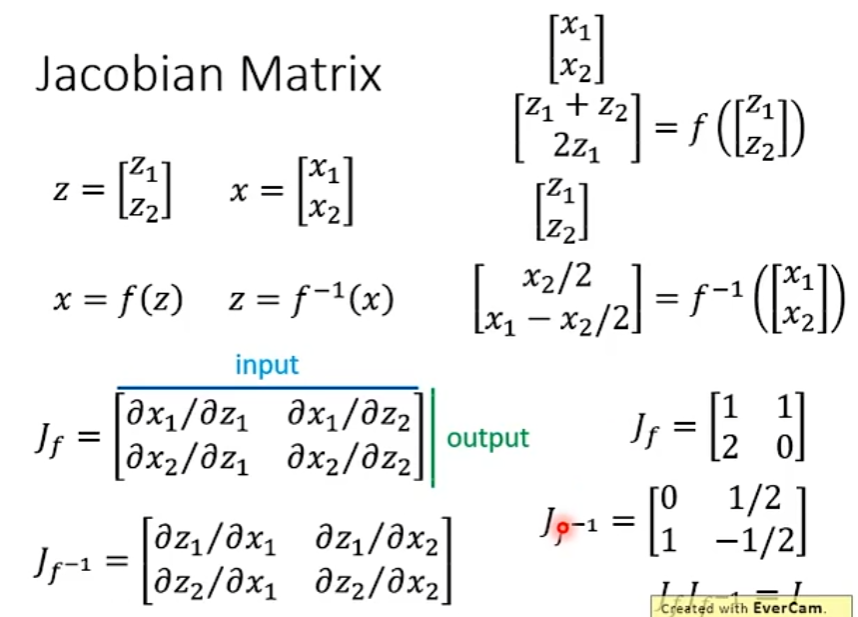


## Flow-based



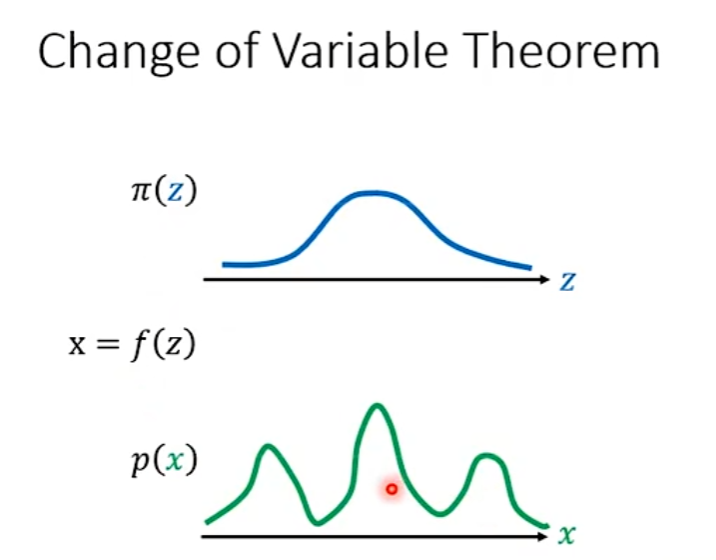


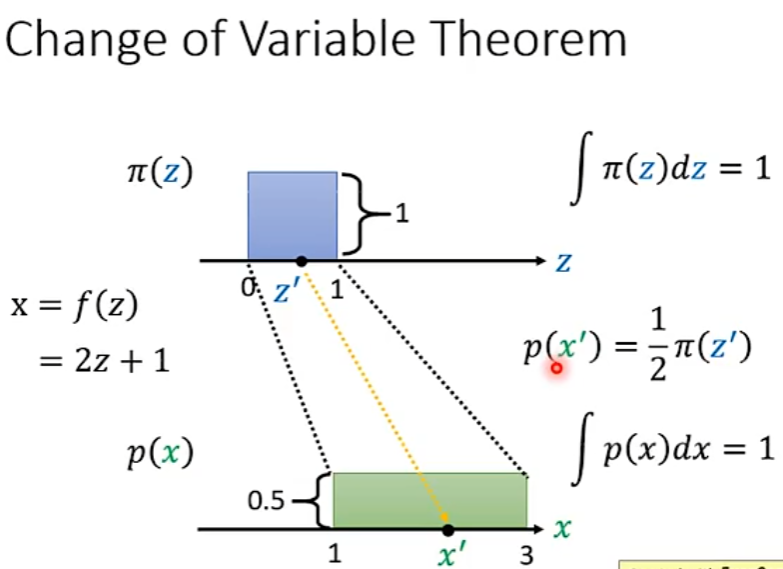
数学公式

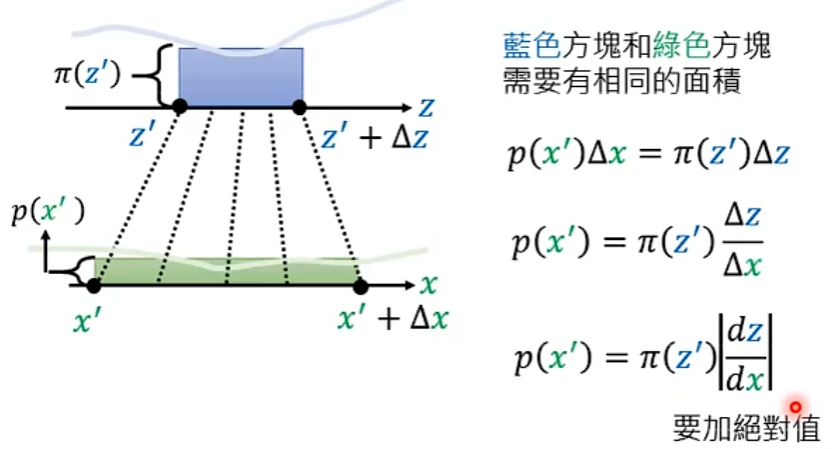




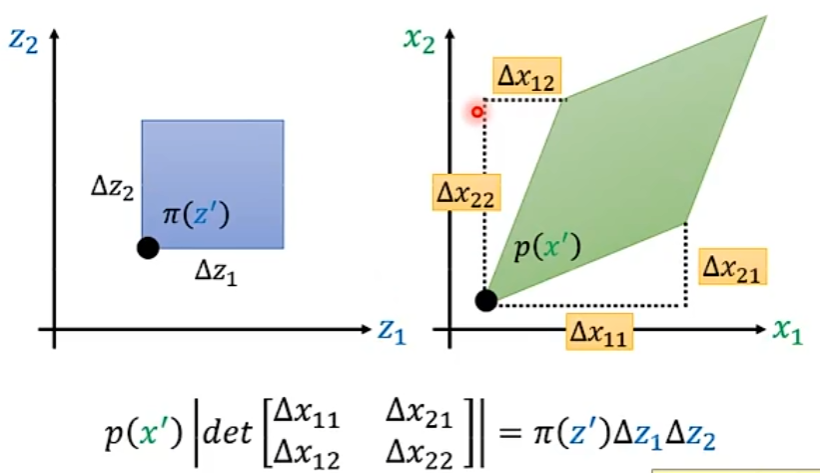
det（A）=???即矩阵的运算

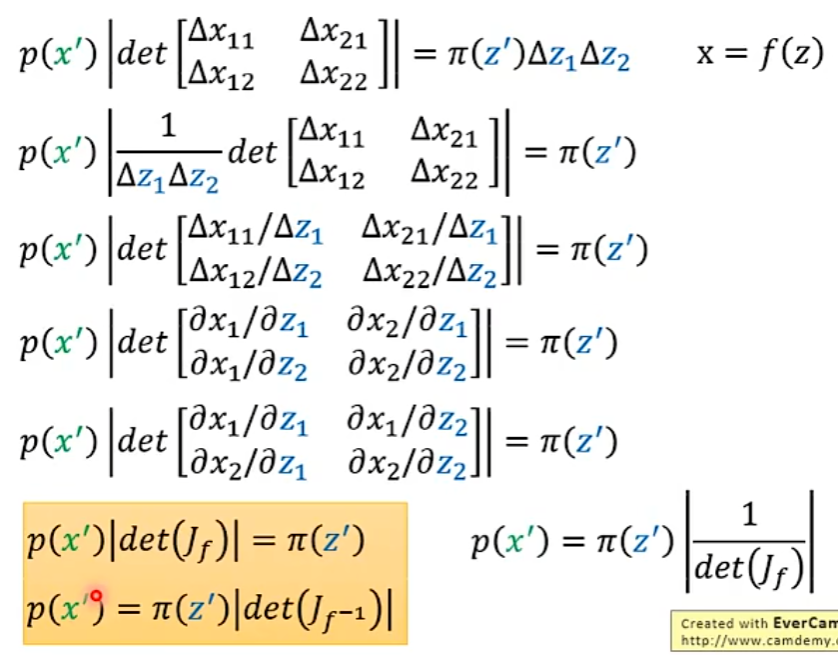




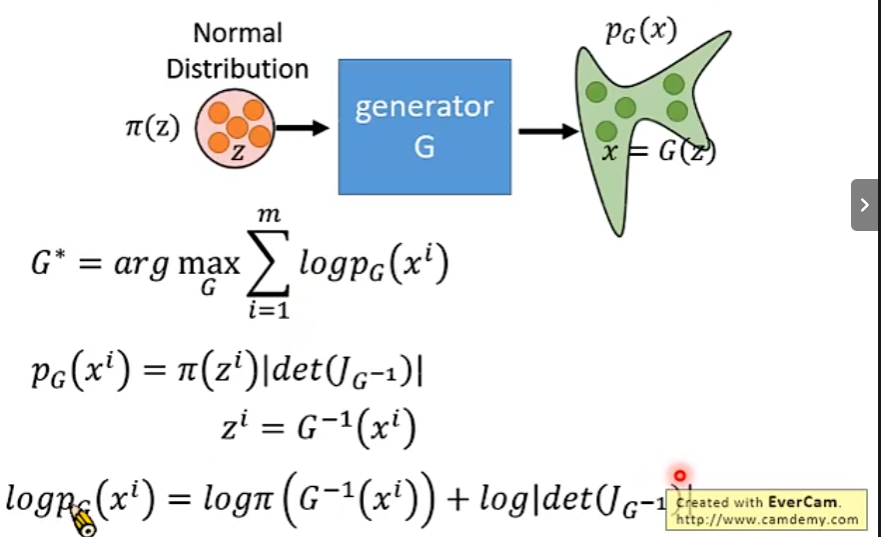


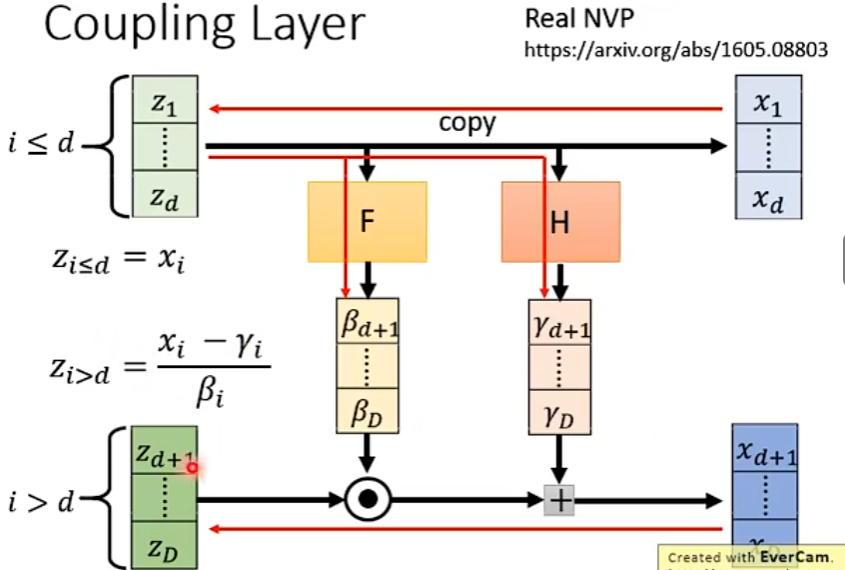
二维

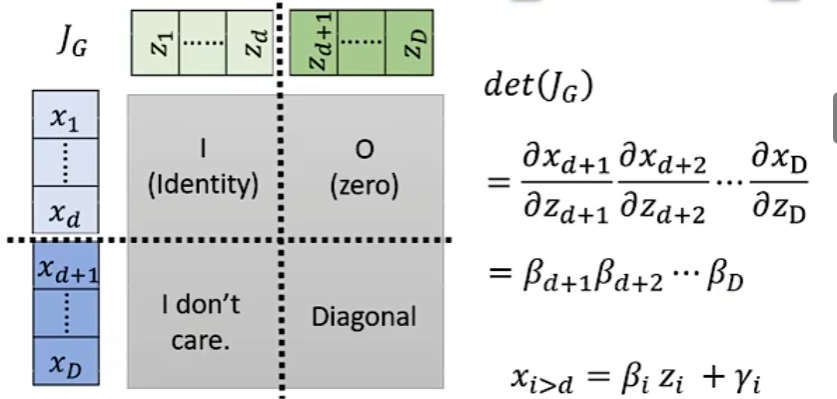




根据以上公式Flow-based：G有多个





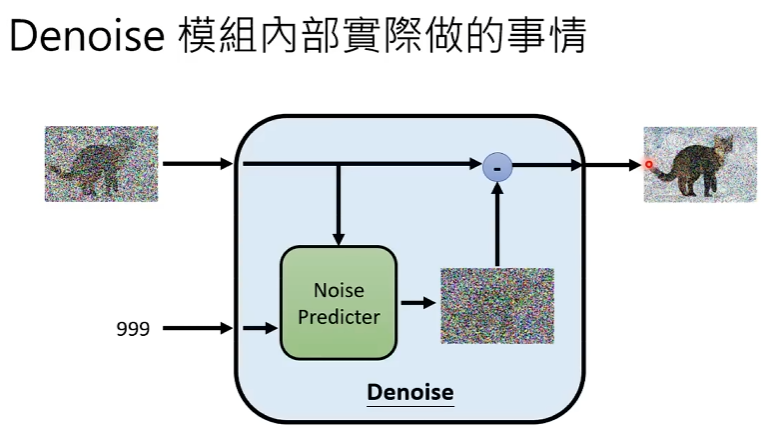




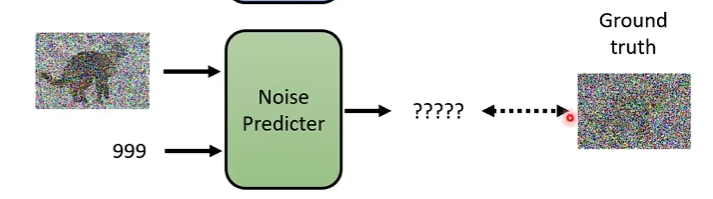
## Diffusion

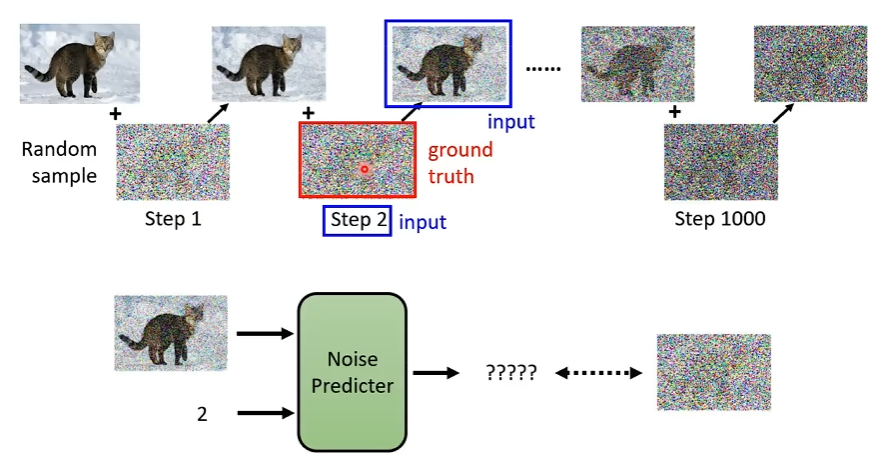


运作：取样一个杂讯的图片通过不断Denoise（次数是规定好的）产生一张清晰的图片，输入时还需要一个额外的数字来表明进行到了第几个步骤



如何产生预测出来的杂讯？——由人训练得到

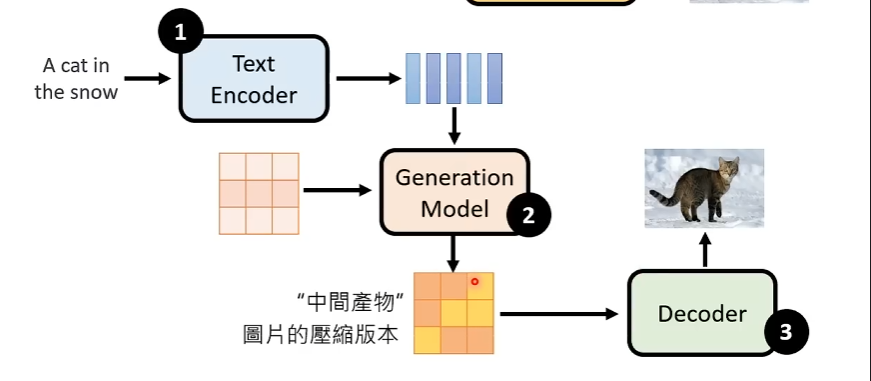




文字可以加入到Denoise

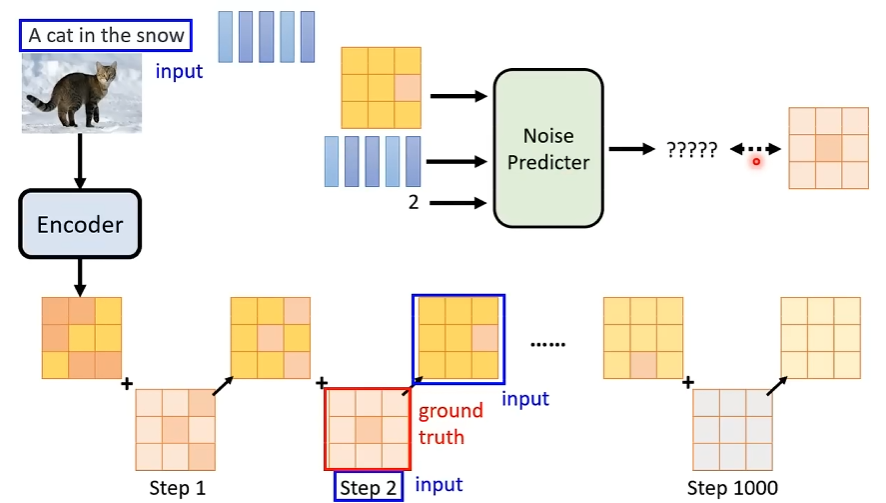


Stable Diffusion

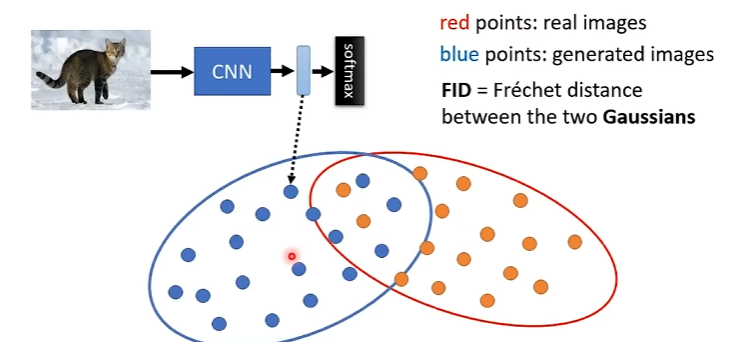


Text Encoder：可以用GPT Decoder：中间产物为Latent Representation需要人为的训练

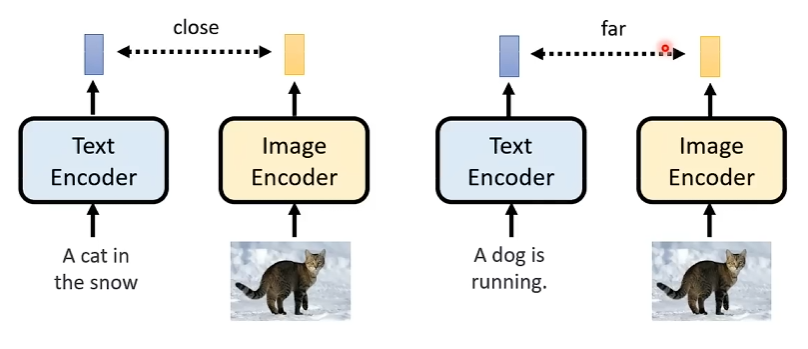
Generation Model：



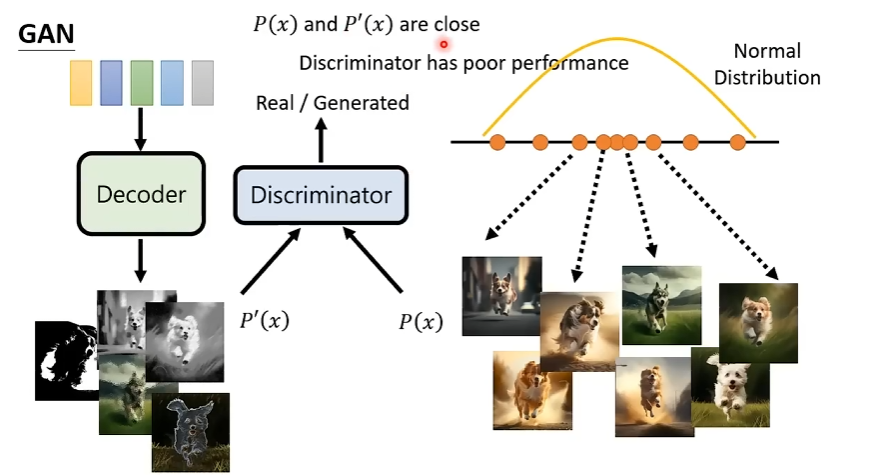
FID：评估影像生成的好坏，越小越好



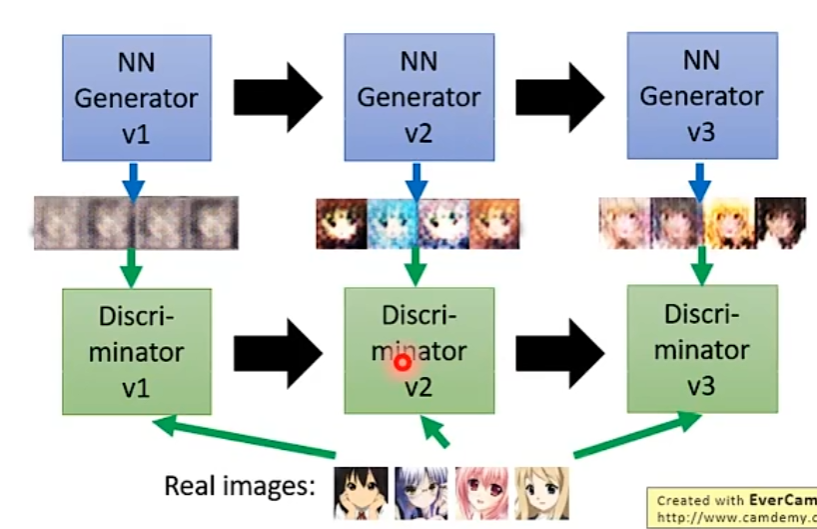
CLIP：

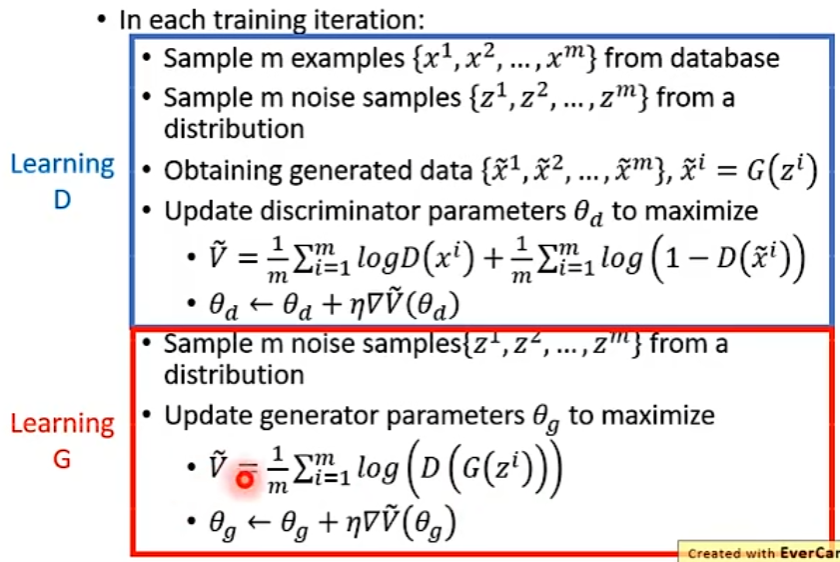


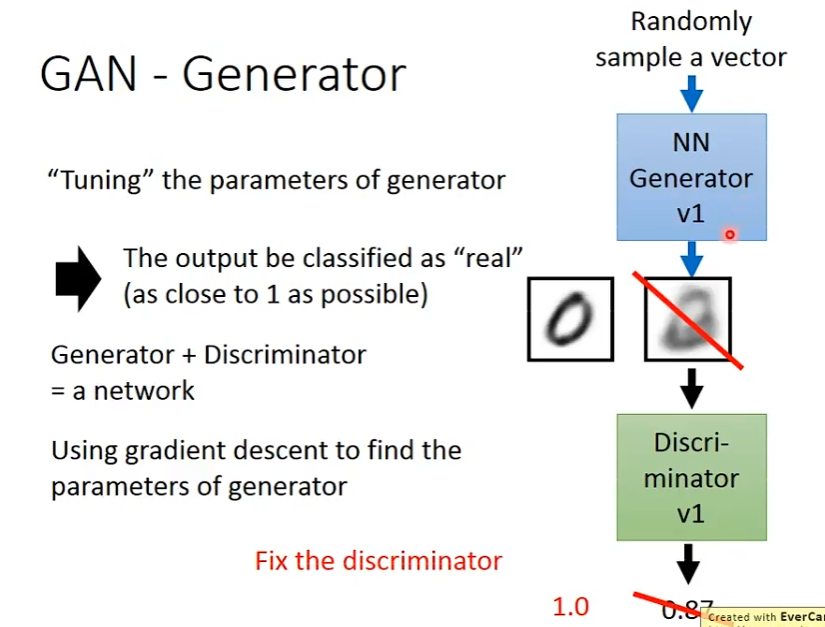
## GAN



GAN可以加到其他三个上







Discriminator后尽量输出为1（real），可以调参但是只能调Generator的参数要固定Discriminator 的参数

Discriminator也可以直接产生影像，穷举所有影像，然后丢到Discriminator中看那个得分最大

