操控我要输出的结果

① 传统NN：往往产生的图像是模糊的，是多张好的结果的平均

② Conditional GAN

C : train →
Normal distribution z → Gen → Image

$x = G(c, z)$

X x → Discriminator (original) → scalar → x is real image or not

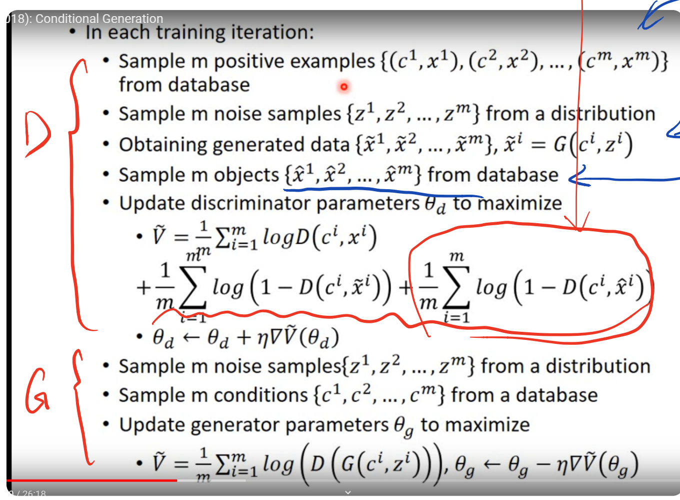
But it completely ignore the input conditions

c →
x → D (better) → scalar → x is realistic or not + c and x are matched or not

(train , 火车) → 1

(train , ???) → 0

(cat , 火车) → 0

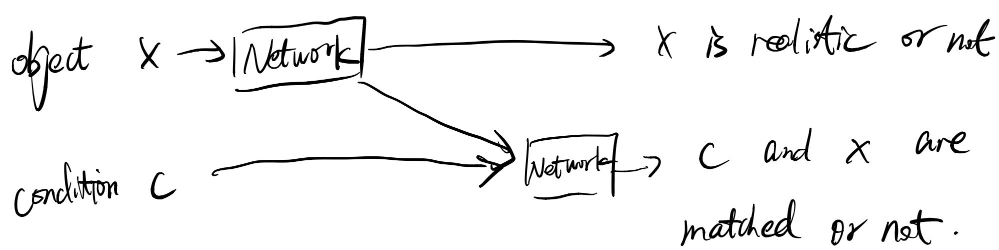


文字和 image 的 pair

$\tilde{x}^i$：生成的图片

$\hat{x}^i$ 较清楚的，但无法配对的图片。

另一种思路（可能更好）



优势：可以分辨哪种原因导致低分，针对性训练

Stack GAN：两个 stage.

先生成小的图片，再用小的生成大的。

Image-to-image.

① 大体同 conditional GAN

② 但是有时候的图片会出现一些莫名其妙的部分

解决方案:

训练 generator 的时候结合 supervised learning 的经验,

除了要骗过 discriminator, 额外增加一个目标, 即

generate 出的图像要与原图像尽可能接近

Patch GAN:

Discriminator 每次检察一小块区域

(如果图片很大)

Discriminator 是一个全卷积网络 → 输出一个矩阵

X_{ij} 表示第 (i,j) 个 patch 为 realistic 的概率

最后的返回值为矩阵中所有的平均

Speech Enhancement

传统: speech + 噪音 → CNN → output。 可能模糊

Conditional GAN: 类比图片 (声音频谱相当于图片)

speech + 噪音 → Gen → output

output →
speech + 噪音 → Dis → scalar

1. 是否清晰
2. 是否和原始声音匹配