

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

$C = \text{train}$ → $\boxed{\text{Gen}}$ → $\boxed{\text{Image}}$ $x = G(C, z)$
Normal distribution z → $\boxed{\text{Gen}}$

\times $x \rightarrow \boxed{\text{Discriminator (originals)}} \rightarrow \text{scalar}$ \rightarrow x 's real image or not

But it completely ignore the input conditions

$C \rightarrow \boxed{D \text{ (better)}} \rightarrow \text{scalar} \rightarrow$ x is realistic or not + C and x are matched or not
 $x \rightarrow \boxed{D \text{ (better)}}$

$(\text{train}, \boxed{\text{火车}}) \rightarrow 1$

$(\text{train}, \boxed{??}) \rightarrow 0$

$(\text{cat}, \boxed{\text{火车}}) \rightarrow 0$

18) Conditional Generation

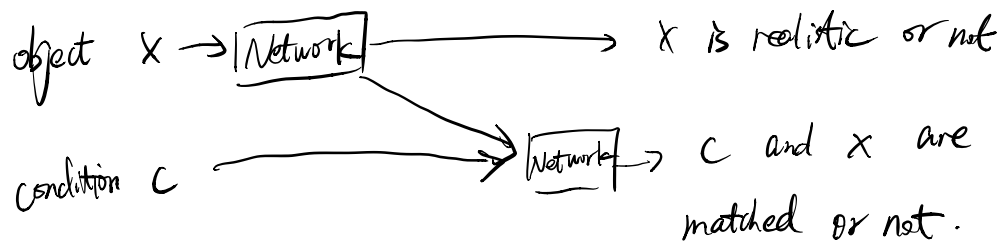
• In each training iteration:

- Sample m positive examples $\{(c^1, x^1), (c^2, x^2), \dots, (c^m, x^m)\}$ from database
- Sample m noise samples $\{z^1, z^2, \dots, z^m\}$ from a distribution
- Obtaining generated data $\{\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m\}, \tilde{x}^i = G(c^i, z^i)$
- Sample m objects $\{\hat{x}^1, \hat{x}^2, \dots, \hat{x}^m\}$ from database
- Update discriminator parameters θ_d to maximize
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log D(c^i, x^i)$
 - $+\frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \tilde{x}^i)) + \frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \hat{x}^i))$
 - $\theta_d \leftarrow \theta_d + \eta \nabla \tilde{V}(\theta_d)$
- Sample m noise samples $\{z^1, z^2, \dots, z^m\}$ from a distribution
- Sample m conditions $\{c^1, c^2, \dots, c^m\}$ from a database
- Update generator parameters θ_g to maximize
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log (D(G(c^i, z^i))), \theta_g \leftarrow \theta_g - \eta \nabla \tilde{V}(\theta_g)$

文字和 image 的 pair

\tilde{x}^i : 生成的图片
 \hat{x}^i 标清楚的, 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 大伟同 conditional GAN

② 但是有时生成的图片会出现一些莫名其妙的部分
伪快方案:

训练 generator 的时候结合 supervised learning 的经验,

除要骗过 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 每次检查一小块区域

(如果图片很大)

Discriminator 是一个全卷积网络 \rightarrow 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: $\text{speech} + \text{噪声} \rightarrow \boxed{\text{CNN}} \rightarrow \text{output}$ ✓ 可能模糊

Conditional GAN: 类似图片 (声音频谱相当于图片)

$\text{speech} + \text{噪声} \rightarrow \boxed{\text{Gen}} \rightarrow \text{output}$

$\text{output} \rightarrow \boxed{\text{Dis}} \rightarrow \text{scalar}$

1. 是否清晰
2. 是否和原始声音匹配