

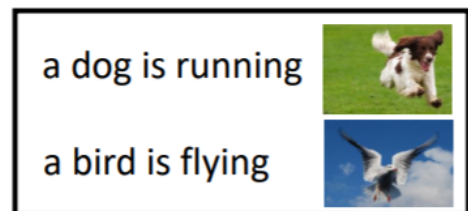
Conditional Generation

Conditional Generation的意义是可以控制输出的内容，比如输入文字产生图片。

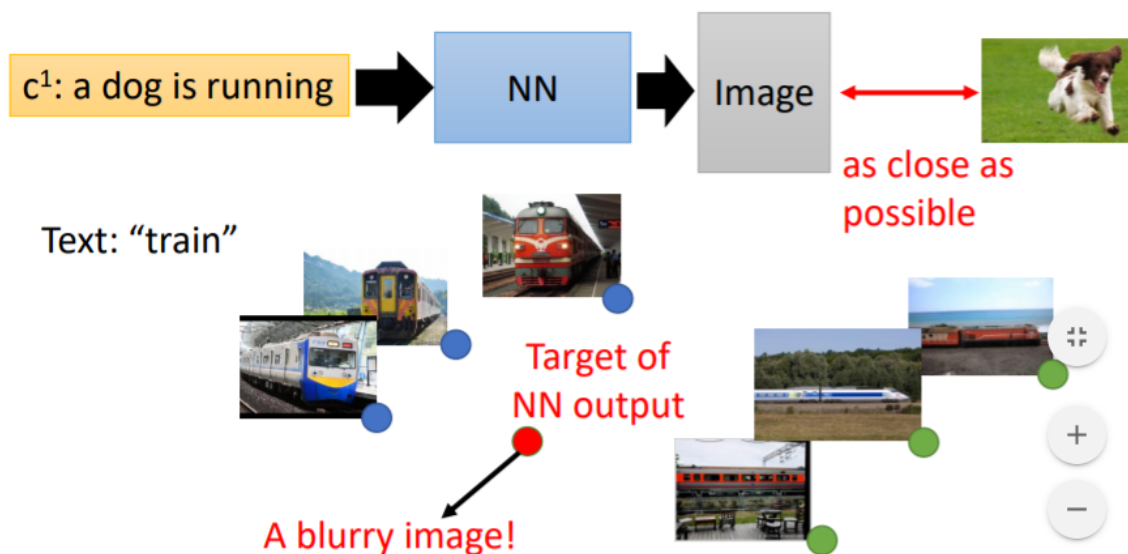
Text to Image

传统的监督学习方法：输入一段文字，输出一张图片。训练集是文本+图片。由于训练数据中一段文字可能对应多张图片，比如“Train”，所以传统的神经网络会给出一个模糊的图片，因为是与“Train”对应的多张图片的平均。

Text-to-Image

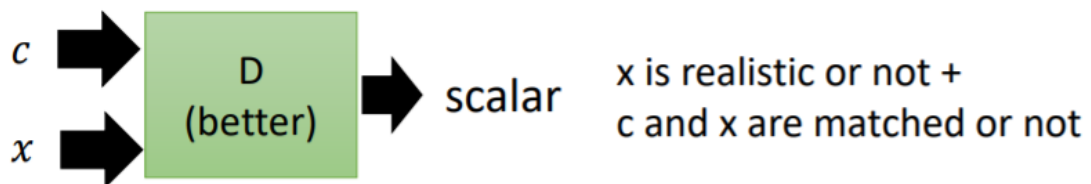



• Traditional supervised approach





而Conditional GAN来训练模型时，生成器的输入不仅仅是一段文本，还包括Normal Distribution的噪声 z 。（个人感觉是原始的GAN生成仅仅取决于 z ，而cGAN的conditional就体现在输入要加文本（可以看作是label），要生成什么样的图片，就告诉生成器想要的对应文本）

同时原始的GAN当中判别器仅输入一张图片，图片质量高（像真的）就会给出高分，而现在需要输入对应的文本和生成的图片。判别器检查图片好坏的任务在这里有两个：1.图片是否高质量；2.图片和文本能否对应。所以现在判别器给低分的情况包括两种：1.正确的文字+较差的的生成图片；2.较为真实的图片+错误的文字。



True text-image pairs: (train , ) 1

(cat , ) 0 (train , ) 0

• In each training iteration:

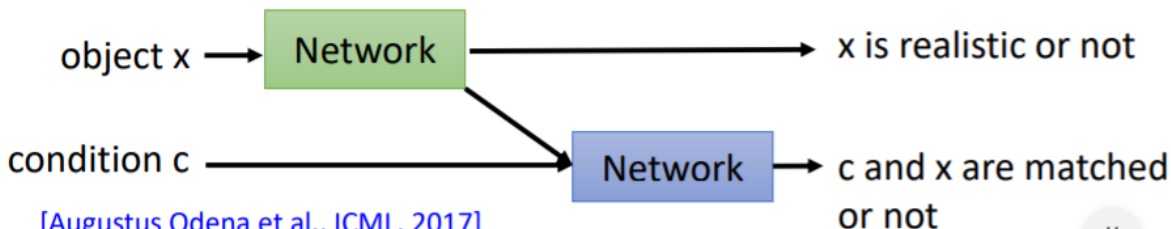
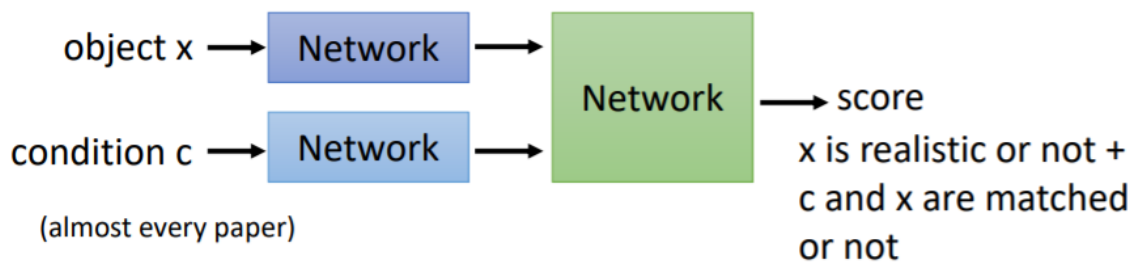
Learning
D

- 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
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log D(c^i, x^i)$
 - $+\frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \tilde{x}^i)) + \frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \hat{x}^i))$
 - $\theta_d \leftarrow \theta_d + \eta \nabla \tilde{V}(\theta_d)$

Learning
G

- 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
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log (D(G(c^i, z^i)))$, $\theta_g \leftarrow \theta_g - \eta \nabla \tilde{V}(\theta_g)$

Conditional GAN — Discriminator



[Augustus Odena et al., ICML, 2017]

[Takeru Miyato, et al., ICLR, 2018]

[Han Zhang, et al., arXiv, 2017]

上图中第二种架构的优势是可以分清楚到底是哪一种问题（不match或者结果差）。

Stack GAN

将架构分成两部分：第一部分先生成较小的图片，第二部分根据生成的小图和embedding产生大图。

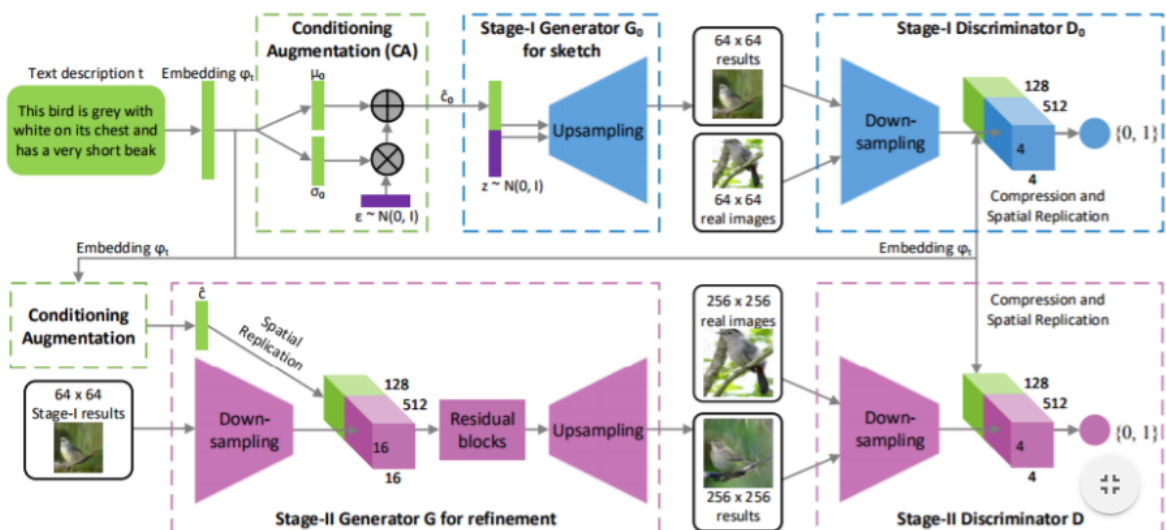
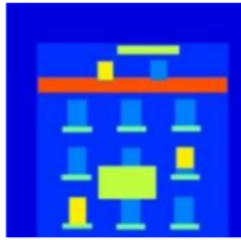


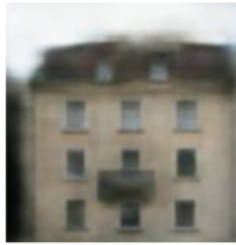
Image to Image

传统的神经网络产生的图片依然比较模糊，GAN的问题是会产生奇怪的东西。让生成器在考虑“骗过”判别器的同时考虑不要与原图差别太大，得到的结果就会相对较好。

Testing:



input



close

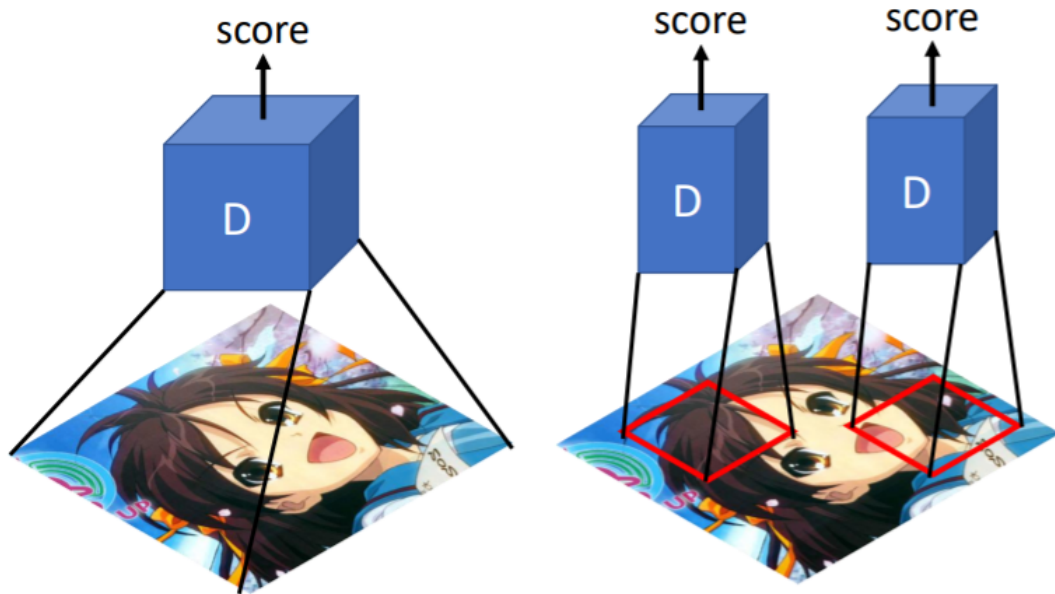


GAN



GAN + close

如果直接输入较大的图片，网络的参数过多，训练过程很容易Overfitting或者用时过长。所以在判别过程，判别器仅检测一小部分图片，具体的大小是一个超参数。（patch GAN）



Other Applications

Speech Enhancement, Video Generation.....