

# 基于生成对抗网络的图像生成

2020 年 6 月 8 日

## 1 实验简介

### 1.1 生成对抗网络

生成对抗网络 (Generative adversarial network, GAN) 自 Ian Goodfellow [1] 等人提出后, 受到了学界越来越多的关注, GAN 包含一个生成网络  $G$  来学习数据分布和一个判别网络  $D$  来判断样本是来自于真实分布还是由  $G$  生成的样本, 形式化描述和算法如下:

$$\min_G \max_D = \mathbb{E}_{x \sim p(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

---

**Algorithm 1** Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator,  $k$ , is a hyperparameter. We used  $k = 1$ , the least expensive option, in our experiments.

---

**for** number of training iterations **do**

**for**  $k$  steps **do**

- Sample minibatch of  $m$  noise samples  $\{z^{(1)}, \dots, z^{(m)}\}$  from noise prior  $p_g(z)$ .
- Sample minibatch of  $m$  examples  $\{x^{(1)}, \dots, x^{(m)}\}$  from data generating distribution  $p_{\text{data}}(x)$ .
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[ \log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) \right].$$

**end for**

- Sample minibatch of  $m$  noise samples  $\{z^{(1)}, \dots, z^{(m)}\}$  from noise prior  $p_g(z)$ .
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))).$$

**end for**

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

---

关于 GAN 的更多资料可以参考原文 [1] 和李宏毅深度学习公开课 Generative Adversarial Network (GAN) 章节 [2]。对深度学习和基本框架不了解的也可以参考李宏毅深度学习公开课程。

## 1.2 DCGAN

DCGAN[3] 改进了原始 GAN 的网络结构, 将生成网络和判别网络由原始的全连接神经网络换为卷积神经网络并且通过大量实验总结出使得 GAN 训练更稳定的一些方法, 生成网络结构如图1所示。

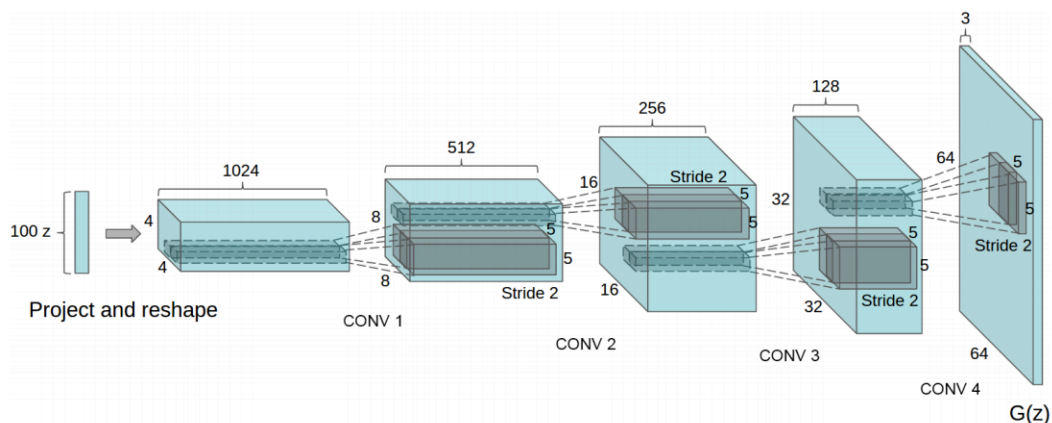


图 1: DCGAN generator used for LSUN scene modeling.

图2和图3分别是用 GAN 在 LSUN 和 CelebA 数据库上生成的图片示意图, 图2是使用原始的 DCGAN 生成的房间内景照片, 图3是 DCGAN 经过 3 年多的时间不断改进, 生成的人脸结果图。

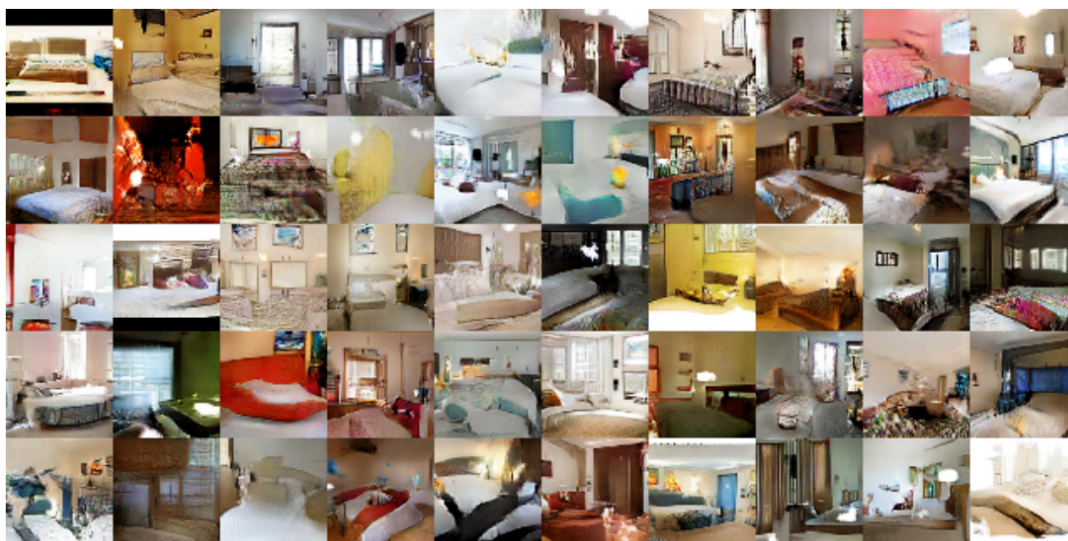


图 2: Generated bedrooms after five epochs of training.



图 3: 3.5 Years of Progress on Faces

### 1.3 GAN 的改进

虽然 GAN 有很多优点,但是也存在训练困难等问题,主要表现在以下几个方面:

1. 模型崩塌 (Mode collapse): 表现为生成网络生成的样本缺乏多样性。
2. 梯度消失 (Gradient vanishing): 判别网络 D 训练的太好,导致梯度消失使得生成网络的学不到任何东西。
3. 模型不收敛 (Non-convergence)。

针对上述问题,学者们尝试从各个方向对 GAN 做出改进,比如: 基于网络结构、优化准则、损失函数、训练方式等, GAN 生成的图片也变得越来越真实,各种变形可以参考综述文章 [4]。

### 1.4 评价指标

通过各种手段来提升 GAN 的性能的主要目的是让 GAN 生成的图片看起来尽可能真实,并且生成图片的种类尽可能多样化,基于以上准则,很多学者也提出了很多方法来评判 GAN 训练的好坏。目前应用最广泛的指标主要是 Inception Score [5] 和 Fréchet Inception Distance [6]

#### 1.4.1 Inception Score (IS)

Inception Score 是一种评价生成模型好坏的度量方法,其计算方法是将生成网络生成的样本数据  $x$  放入 Inception 分类模型得到条件标签分布  $p(y|x)$ ,对于该分布,我们希望他有较低的信息熵,即: 有非常高的可信度属于某一具体的类别,从而保证生成模型生成的数据更像真实模型,但是对于一批数据而言,我们希望  $p(y) = \int p(y|x = G(z)) dz$  的熵值尽可能高,即: 所属标签的类别尽可能满足均匀分布,从而保证生成模型生成样本的多样性。

计算方式如公式所示:

$$IS(G) = \exp(\mathbb{E}_{x \sim p_g} D_{KL}(p(y|x) || p(y))) \quad (2)$$

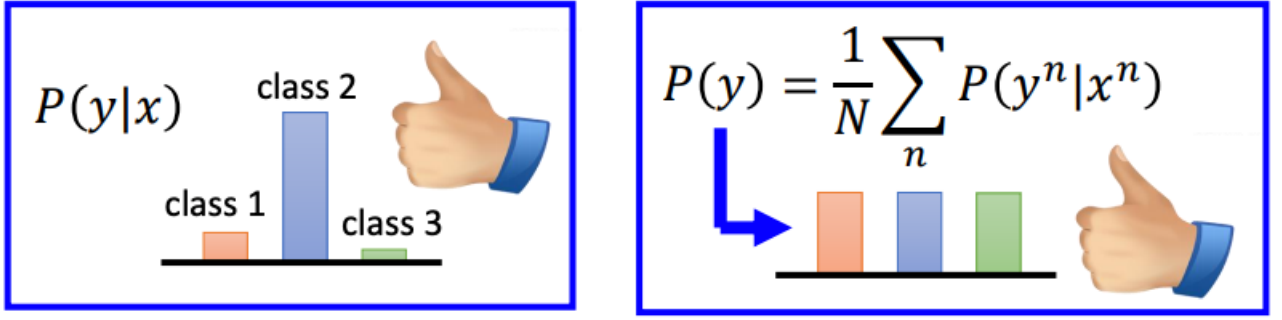


图 4: Inception Score Metrics

$$\begin{aligned}
 \ln(IS(G)) &= \mathbb{E}_{x \sim p_g} [\text{D}_{KL}(p(y|x) || p(y))] \\
 &= \sum_x p_g(x) \text{D}_{KL}(p(y|x) || p(y)) \\
 &= \sum_x p_g(x) \sum_i p(y=i|x) \ln \frac{p(y=i|x)}{p(y=i)} \\
 &= \sum_x \sum_i p(y=i, x) \ln \frac{p(y=i, x)}{p(y=i)p(x)} \\
 &= I(y; x) = H(y) - H(y|x) \\
 &= - \sum_y p(y) \ln p(y) + \sum_x \sum_y p(y|x) \ln p(y|x)
 \end{aligned}$$

#### 1.4.2 Fréchet Inception Distance (FID)

FID [6] 是为了解决 Inception Score 计算时没有考虑真实样本的统计信息而提出的一种度量生成网络训练好坏的指标。假设  $(m, \mathbf{f})$  和  $(m_\omega, \mathbf{C}_\omega)$  分别是生成样本和真实样本的均值和方差, FID 的计算方法由公式给出:

$$FID = ||m - m_\omega||_2^2 + \text{Tr}(\mathbf{C} + \mathbf{C}_\omega - 2(\mathbf{C}\mathbf{C}_\omega)^{1/2}) \quad (3)$$

## 2 实验目的

了解 MLP、CNN 等常见的网络模型, 熟悉 GAN 和 DCGAN 等生成对抗网络变体, 掌握一种深度学习框架 (tensorflow/PyTorch/Keras...), 按照实验要求独立完成实验, 并提交实验报告。

## 3 实验要求

1. 实现 DCGAN(输入随机噪声, 生成图片), 数据集可以是 mnist/LSUN[7]/Celeba[8]。
2. 根据文献 [4] 在 1 基础上改进网络 (任意一种改进方法均可, 也可以使用文献 [4] 没有提到的算法或者根据自己理解提出新的算法)。

3. 计算 IS 和 FID, 并比较改进前后 IS 和 FID 的区别, FID 和 IS 的计算可使用网上公开代码。
4. 提交代码和实验报告, 实验报告需包含网络生成的部分图像和 IS/FID 客观指标对比 (严禁抄袭, 一经查实, 成绩记 0)。
5. 会根据任务难度和实验完成情况对实验进行综合评分; 鼓励创新, 在 DCGAN 基础上, 若能提出自己的改进算法, 并且任何一个数据集上 IS 和 FID 高于 DCGAN, 即可得满分。
6. 实验报告技术问题和搭建深度学习环境问题可以联系 [zhwzhong.hit@gmail.com](mailto:zhwzhong.hit@gmail.com)。
7. 数据集可以通过参考文献中链接下载, 由于 LSUN 和 CelebA 数据集数据量较大, 可以随机选取一部分数据作为训练数据。
8. 实验室可给部分学生提供 GPU 计算资源。
9. CelebA 人脸数据库获取和处理方法: [点击阅读](#)

## 参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. pages 2672–2680, 2014.
- [2] Hongyi Li. Generative adversarial network. [http://speech.ee.ntu.edu.tw/~tlkagk/courses\\_MLDS18.html](http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLDS18.html), 2018.
- [3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [4] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *arXiv preprint arXiv:1809.07294*, 2018.
- [5] Shane Barratt and Rishi Sharma. A note on the inception score. *arXiv preprint arXiv:1801.01973*, 2018.
- [6] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Günter Klambauer, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a nash equilibrium. *CoRR*, abs/1706.08500, 2017.
- [7] LSUN. Generative adversarial network. <https://github.com/fyu/lsun>, 2018.
- [8] CelebA. Generative adversarial network. <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>, 2018.