

# GAN的无监督条件生成（一）：CycleGAN与StarGAN

## 【参考文献】

[1] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks 2017

[2] StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation 2017

GAN的无监督条件生成一般是指图像翻译的任务，通常做法有两类，一类是直接进行转换，另一类是将不同的domain投影到同一个空间。本篇笔记介绍前一类方法的两个经典工作：CycleGAN和StarGAN。

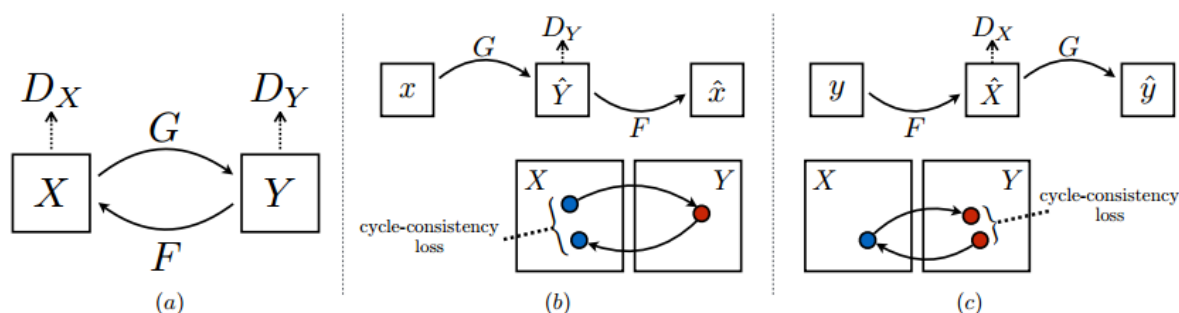
## 1. CycleGAN

一些以前的图像翻译问题需要有source domain和target domain的pair数据，而CycleGAN致力于使用unpaired的数据来做到任意两个domain的风格迁移。

**CycleGAN的核心是cycle consistence loss。**也就是说，domain  $X$ 的图像转换到domain  $Y$ 后，再从domain  $Y$ 转换回到domain  $X$ ，产生的图像要和原来的保持一致，反过来也一样。假设现在有两个生成器 $G : X \rightarrow Y$ 和 $F : Y \rightarrow X$ ，那么cycle consistence loss为：

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1]$$

也就是说，CycleGAN中包含两个cycle：前向cycle为 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ ，后向cycle为 $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 。



当然，还有两个判别器来分别鉴别domain  $X$ 和domain  $Y$ 的生成图像和真实图像。对于domain  $Y$ ，对抗loss为：

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))]$$

同理，对于domain  $X$ 来说也一样，对抗loss为 $\mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$ 。

使用cycle consistence loss是必要的，没有这一项，会存在**两个问题**：

- 无法保证domain  $X$ 的图片转换到domain  $Y$ 后还是和原来配对的；
- 模式崩塌问题，生成器只要将domain  $X$ 的图像都映射成domain  $Y$ 中的同一个图像就可以了。

完整的目标函数为：

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ & + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G, F)\end{aligned}$$

## 2. StarGAN

StarGAN致力于使用一个生成器来进行多个域之间的风格迁移。这样带来**两个好处**：

- 效率更高，否则对于多个域，就要训练多个一对一的风格迁移模型；
- 可以充分利用数据，多个域、甚至是多个数据集的数据都能用来训练。

StarGAN的做法是把domain label作为条件，和原始图片一同输入生成器，来做到特定domain的图像生成。loss分为三部分：

### 1) Adversarial Loss

就是普通的对抗loss，用来区分真实图片和生成图片：

$$\mathcal{L}_{adv} = \mathbb{E}_x [\log D_{src}(x)] + \mathbb{E}_{x,c} [\log(1 - D_{src}(G(x, c)))]$$

### 2) Domain Classification Loss

利用了类似于ACGAN的方式，在discriminator上追加一个auxiliary classifier来对domain label进行分类，以达到条件生成的目的。训练判别器时，最小化真实图片的分类loss：

$$\mathcal{L}_{cls}^r = \mathbb{E}_{x,c'} [-\log D_{cls}(c'|x)]$$

训练生成器时，最小化生成图片的分类loss：

$$\mathcal{L}_{cls}^f = \mathbb{E}_{x,c} [-\log D_{cls}(c|G(x, c))]$$

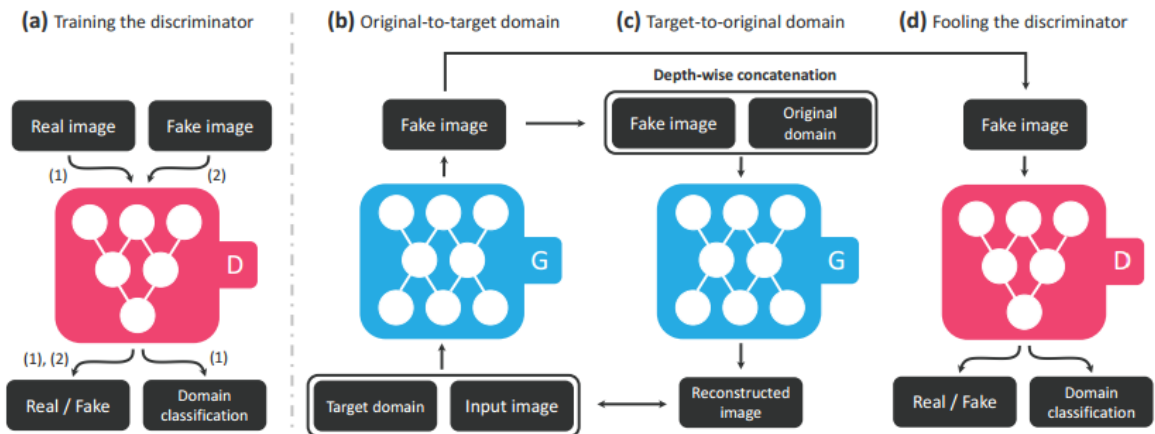
### 3) Reconstruction Loss

即CycleGAN中的cycle consistence loss：

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'} [\|x - G(G(x, c), c')\|_1]$$

所以完整的loss如下：

$$\begin{aligned}\mathcal{L}_D = & -\mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^r, \\ \mathcal{L}_G = & \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec}\end{aligned}$$



### 利用多个数据集训练

为了充分利用多个数据集的数据，StarGAN的domain label中加入了一个one-hot的mask vector，图片来自哪个domain，就把对应的位置1。

完整的label vector如下：

$$\tilde{c} = [c_1, \dots, c_n, m]$$

其中 $c_i$ 代表第 $i$ 个数据集对应的label向量，它可以是binary的形式（对应多标签），也可以是categorical的形式（对应单标签的one-hot编码）， $[\cdot]$ 代表concat操作。

多数据集训练的示意图如下（图中利用了两个数据集）：

