

论文题目： StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation----from CVPR2018

一、创新点

- 1、作者提出了 StarGAN，这是一种新颖的生成对抗网络，它使用单个生成器和判别器来学习多个域之间的映射，从所有域的图像中有效地进行训练。
- 2、作者演示了如何利用掩模矢量方法成功学习多个数据集之间的多域图像转换，该方法使 StarGAN 能够控制所有可用的域标签。
- 3、作者使用 StarGAN 提供面部属性转移和面部表情综合任务的定性和定量结果，显示其优于 baseline model。

注意：生成器不是学习固定的翻译（例如，黑色到金色的头发），而是接收图像和域信息，并学会灵活地将图像转换为相应的域。作者使用标签（例如，二进制或单热矢量）来表示域信息。在训练期间，我们随机生成目标域标签并训练模型以将输入图像灵活地转换为目标域。这样可以控制域标签并训练，在测试阶段将图像转换为任何所需的域。

二、Star Generative Adversarial Networks

1、Multi-Domain Image-to-Image Translation

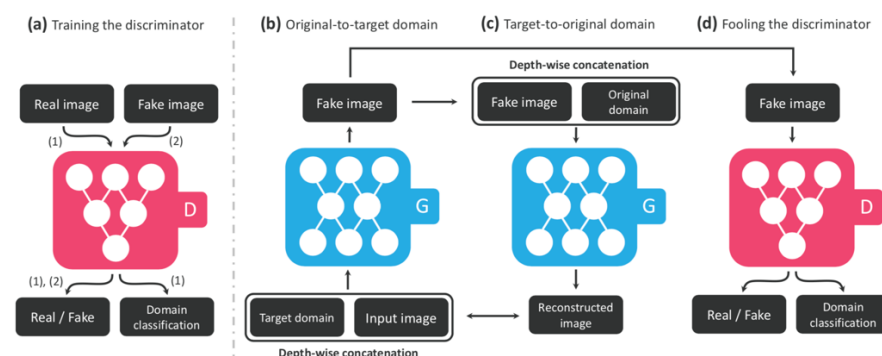


Figure 3. Overview of StarGAN, consisting of two modules, a discriminator D and a generator G . (a) D learns to distinguish between real and fake images and classify the real images to its corresponding domain. (b) G takes in as input both the image and target domain label and generates an fake image. The target domain label is spatially replicated and concatenated with the input image. (c) G tries to reconstruct the original image from the fake image given the original domain label. (d) G tries to generate images indistinguishable from real images and classifiable as target domain by D .

上图说明是作者提出的方法的训练过程. (a) D 学习区分真实图像和假图像, 并将真实图像分类到相应的域. (b) G 接受图像和目标域标签作为输入并生成假图像. (c) G 尝试在给定原始域标签的情况下, 从假图像中重建原始图像. (d) G 尝试生成与真实图像非常像的假图像, 并通过 D 将其分类为目标域. 我们的目标是训练单个生成器G, 它可以在多个域之间学习映射. 为实现此目的, 我们训练G将输入图像x转换为目标域标签c, $G(x, c) \rightarrow y$ 上的输出图像. 我们随后生成目标域标签c, 以便G学会灵活地转换输入图像. 我们还引入了一个辅助分类器, 它允许单个鉴别器控制多个域. 也就是说, 我们的鉴别器在两个源和主标签上产生概率分布, $D : x \rightarrow \{D_{src}(x), D_{cls}(x)\}$.

Adversarial Loss : GAN 的通用函数, 判断输出图像是否真实

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

Domain Classification Loss: 其次是类别损失, 该损失被分成两个, 训练 D 的时候, 使用真实图像在原始领域进行, 训练 G 的时候, 使用生成的图像在目标领域进行。

训练 D 的 loss :

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

训练G的loss :

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

Reconstruction Loss: 再次则是重建函数，重建函数与 CycleGAN 中的正向函数类似。

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

Full Objective: 汇总后则是

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^r,$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec},$$

2、Training with Multiple Datasets

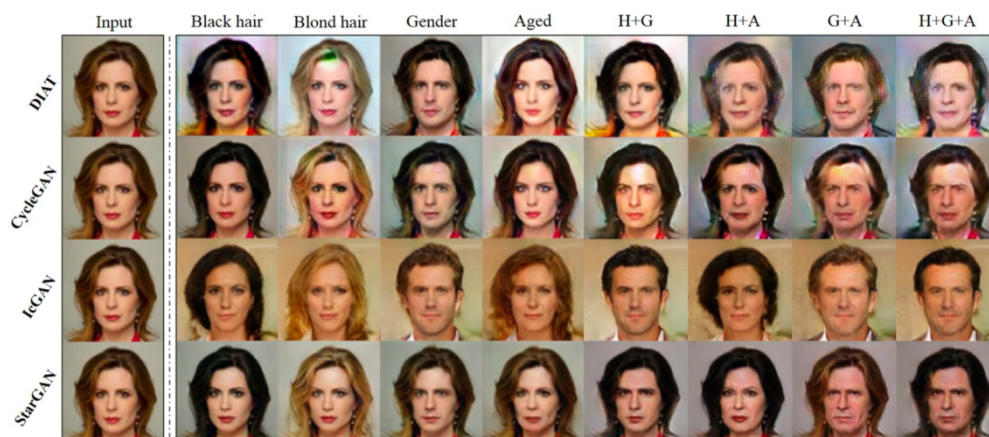


Figure 4. Facial attribute transfer results on the CelebA dataset. The first column shows the input image, next four columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. H: Hair color, G: Gender, A: Aged.

Mask Vector：作者引入了一个mask矢量 m ，它允许StarGAN忽略未指定的标签，并专注于特定数据集提供的明确已知的标签。我们使用 n 维one-hot矢量来表示 m ，其中 n 是数据集的数量。此外，我们将标签的统一版本定义为矢量

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

其中 $[\cdot]$ 表示连接, c_i 表示第 i 个数据集的标签的向量。已知标签 c_i 的向量可以表示为用于二元属性的二元向量或用于分类属性的 one-hot 向量。对于剩余的 $n-1$ 个未知标签, 我们只需指定零值。

Training Strategy: 在使用多个数据集训练StarGAN时, 我们使用方程式中定义的域标签 \tilde{c} 。

四、实验

Baseline Models : DIAT and Cycle- GAN,采用IcGAN [23]作为基线, 使用cGAN执行属性转移。

数据集 : CelebA和RaFD

实验环境

- Python 3.6
- Tensorflow 1.8

StarGAN 是基于 cycleGAN 网络搭建的, 有一个 G 和 D 组成。作者使用 Adam 算法来训练整个模型, 对于 CelebA 的实验, 作者训练所有模型, 前 10 个时期的学习率为 0.0001, 并在接下来的 10 个时期内将学习率线性衰减为 0。

Tensorboard 可视化结构如下 :

作者通过实验给出的结果：

Method	Classification error	# of parameters
DIAT	4.10	52.6M \times 7
CycleGAN	5.99	52.6M \times 14
IcGAN	8.07	67.8M \times 1
StarGAN	2.12	53.2M \times 1
Real images	0.45	-

Table 3. Classification errors [%] and the number of parameters on the RaFD dataset.

Method	Hair color	Gender	Aged
DIAT	9.3%	31.4%	6.9%
CycleGAN	20.0%	16.6%	13.3%
IcGAN	4.5%	12.9%	9.2%
StarGAN	66.2%	39.1%	70.6%

Table 1. AMT perceptual evaluation for ranking different models on a single attribute transfer task. Each column sums to 100%.

Method	H+G	H+A	G+A	H+G+A
DIAT	20.4%	15.6%	18.7%	15.6%
CycleGAN	14.0%	12.0%	11.2%	11.9%
IcGAN	18.2%	10.9%	20.3%	20.3%
StarGAN	47.4%	61.5%	49.8%	52.2%

a

Table 2. AMT perceptual evaluation for ranking different models on a multi-attribute transfer task. H: Hair color; G: Gender; A: Aged.

五、总结

在本文中，作者提出了 StarGAN，一种使用单个生成器和鉴别器的多个域之间的可扩展图像到图像转换模型。除了可扩展性方面的优势之外，由于多任务学习设置背后的泛化能力，StarGAN 与现有方法相比，生成了更高视觉质量的图像。此外，使用所提出的简单 mask 矢量使 StarGAN 能够利用具有不同域标签集的多个数据集，从而处理来自它们的所有可用标签。

改进点

- 1、可以基于特定的场景开发图像翻译的应用程序。