

StarGAN: Unified Generative Adversarial Networks for Multi- Domain Image-to-Image Translation

Mengjie Sun

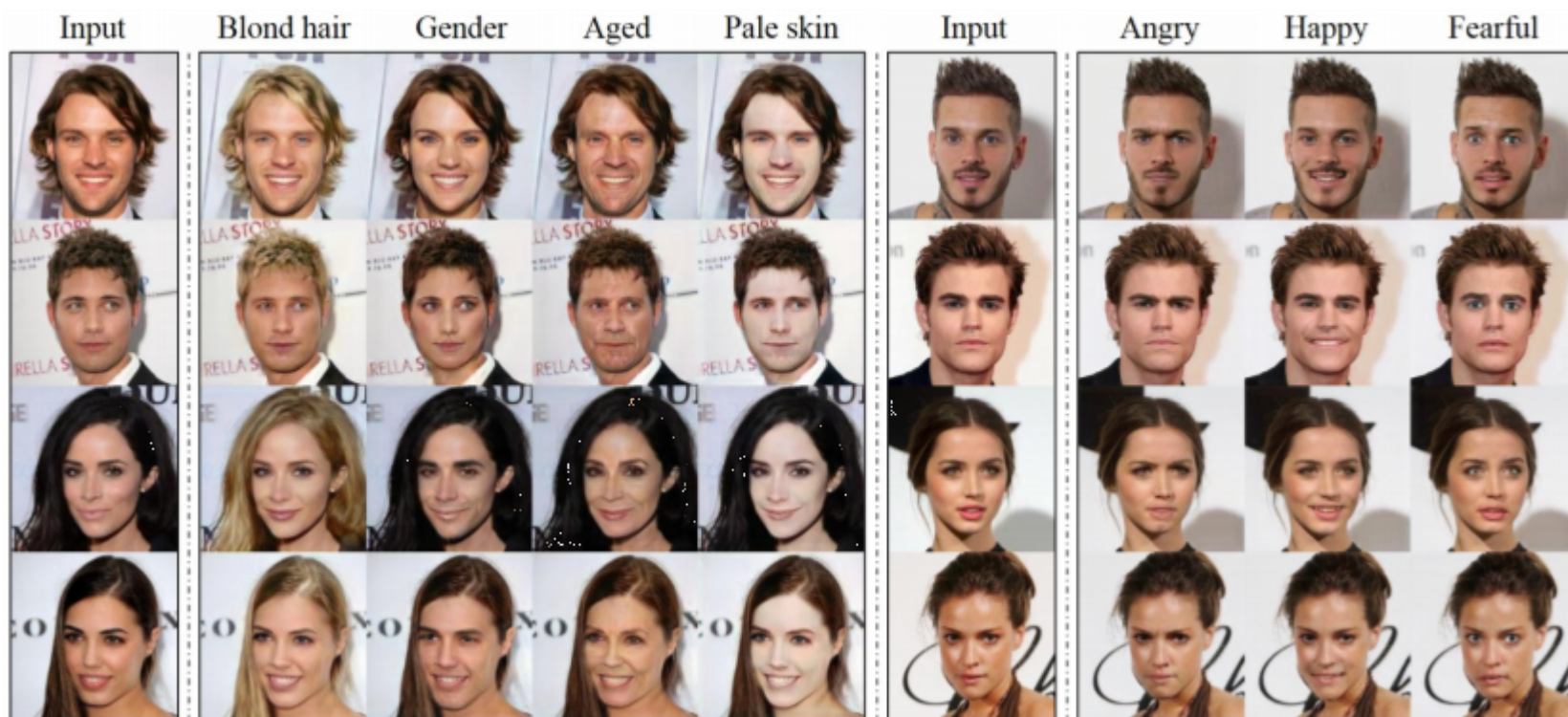
一、模型对比

pix2pix: 有配对数据集的图像翻译问题;

CycleGAN: 不配对数据集的图像翻译问题;

但是, **pix2pix**和**CycleGAN**都是使图像从一个领域转换到另一个领域, 解决的是一对一的问题。如果有很多领域要互相转换, 都需要重新训练模型。

starGAN: 同一种模型可以用来做多个 图像翻译任务, eg.更换头发颜色, 更换表情, 更换年龄, 更换性别等等。



左边展示了由starGAN生成的图片如何在celebA各个不同域之间进行转换；右边展示了交叉两个训练集，使用在RaFD训练过程中学习的特征来改变celebA图像的面部表情。

二、模型框架

为了使G拥有多个领域转化的能力，需要对生成网络G, 判别网络D作如下改动：

- 1.在G的输入中添加目标领域信息，即把图片翻译到哪个领域告诉模型；
- 2.D除了具有判断图片是否真实的功能外，还要有判断图片属于哪个类别的能力。这样就可以保证G中同样的输入图像，随着目标领域的不同生成不同的效果。
- 3.保证图像翻译过程中图像内容要保存，只改变领域差异的部分。图像重建可以完成这个工作，即将图像从领域A翻译到领域B，再翻译回来，不会发生变化。

(a) Cross-domain models



(b) StarGAN



二、模型框架

现有的GAN模型为了实现 在 k 个不同的风格域上进行迁移，需要构建多个生成器，并且 还不能跨数据集训练（标注不能重复使用）。

在starGAN中，并不使用 传统的fixed translation(e.g.,black-to-blond hair),还加入了域信息与图片一起输入进行，并在域标签中加入了mask vector,便于不同数据集进行联合训练。并且只用了单组GAN模型进行跨domain和跨数据集的训练。

三、训练过程

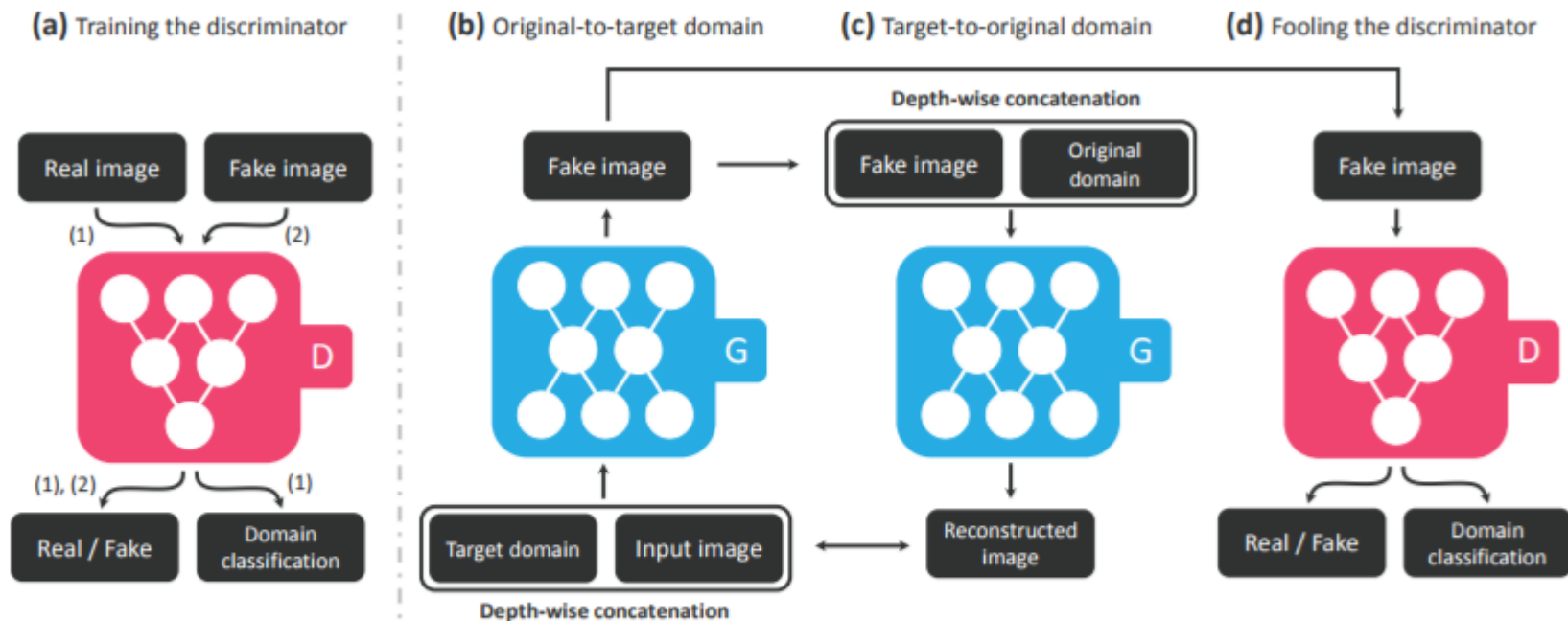


Figure 3. Overview of StarGAN, consisting of two modules, a discriminator D and a generator G . **(a)** D learns to distinguish between real and fake images and classify the real images to its corresponding domain. **(b)** G takes in as input both the image and target domain label and generates an fake 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 .

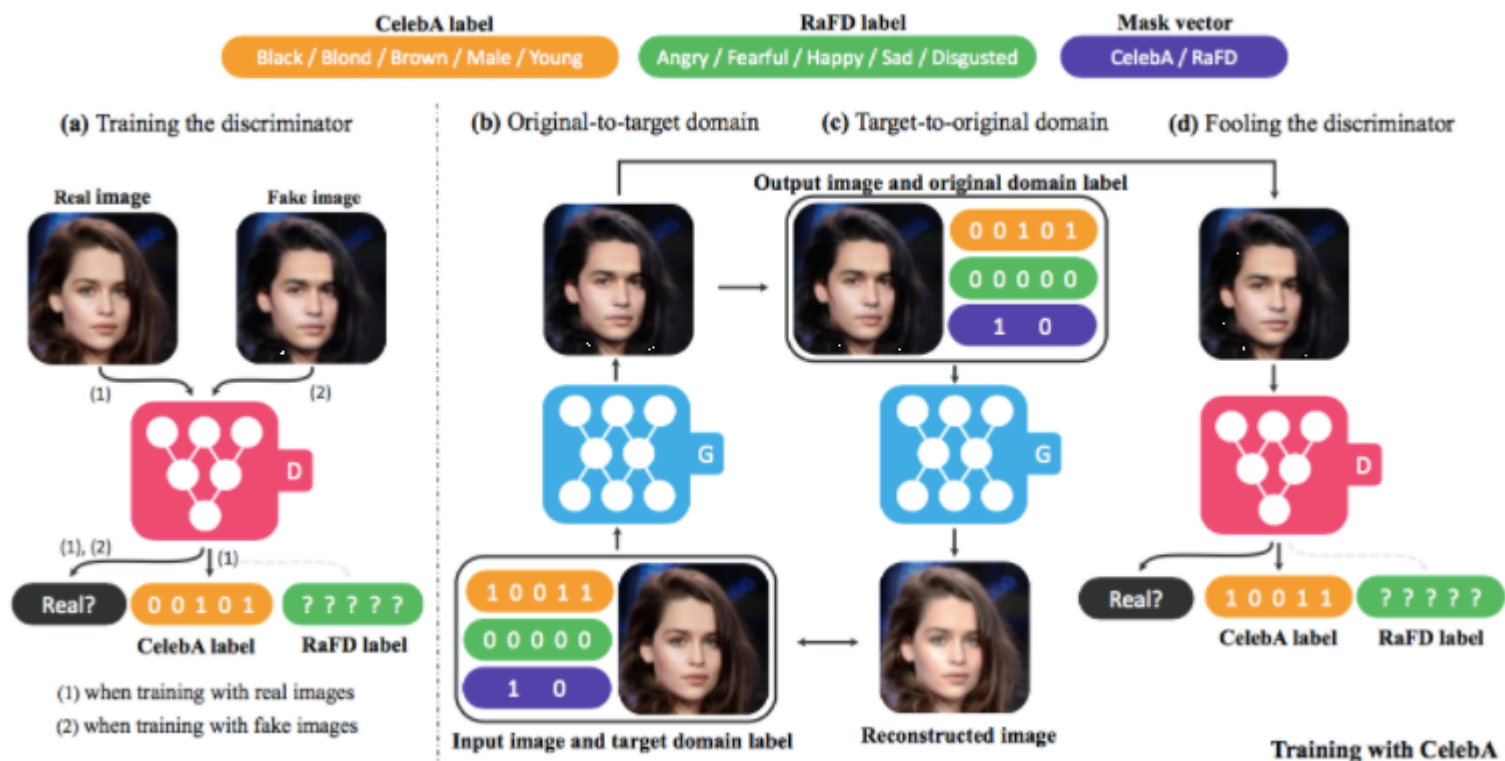
多数据集训练

在多数据集下训练starGAN存在一个问题：数据集之间的类别可能是不相交的，但内容可能是相交的。

eg. 比如CelebA数据集和RaFD数据集，前者拥有很多肤色、年龄之类的类别，而后者拥有的是表情的类别，但前者的图像很多也是有表情的，这就导致前一类的图像在后一类的标记是不可知的。

为了解决这个问题，作者在模型中加入了Mask，即如果来源于数据集B, 那么将数据集A中的标记全部设为0.

多数据集训练



当掩码向量（紫色）为[1,0]时，生成器G会专注于CelebA标签（黄色），忽略RaFD标签（绿色）以执行图像到图像的转换；掩码向量是[0,1]时，反之亦然。

四、目标函数

1. Adversarial Loss:

To make the generated images indistinguishable from real images.

$$\mathcal{L}_{adv} = \mathbb{E}_x[D_{src}(x)] - \mathbb{E}_{x,c}[D_{src}(G(x,c))] - \lambda_{gp} \mathbb{E}_{\hat{x}}[(\|\nabla_{\hat{x}} D_{src}(\hat{x})\|_2 - 1)^2],$$

(Wasserstein GAN)

2. Domain Classification Loss:

类别损失函数被分为两个，训练D时，使用真实图像时在原始领域进行；训练G时，使用生成的图像在目标域进行；

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

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

3. Reconstruction Loss: 重建损失

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

4. 汇总后的Full Objective:

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

五、总结

总的来说，这篇文章的贡献如下：

- 1.提出了一种全新的生成对抗网络StarGAN，该网络只使用一个生成器和一个鉴别器来进行多个域之间的映射，并从各个域的图像中有效地进行训练；
- 2.演示了如何使用掩码向量方法（mask vector method）成功进行多个数据集之间的多域图像转换，并使得StarGAN控制所有可用的域标签；
- 3.使用StarGAN进行面部属性转换和面部表情合成任务，并对结果进行了定性和定量分析，结果显示其优于Benchline Model。

Thanks for listening!