Stargan的相关分享

治天锴

Stargan

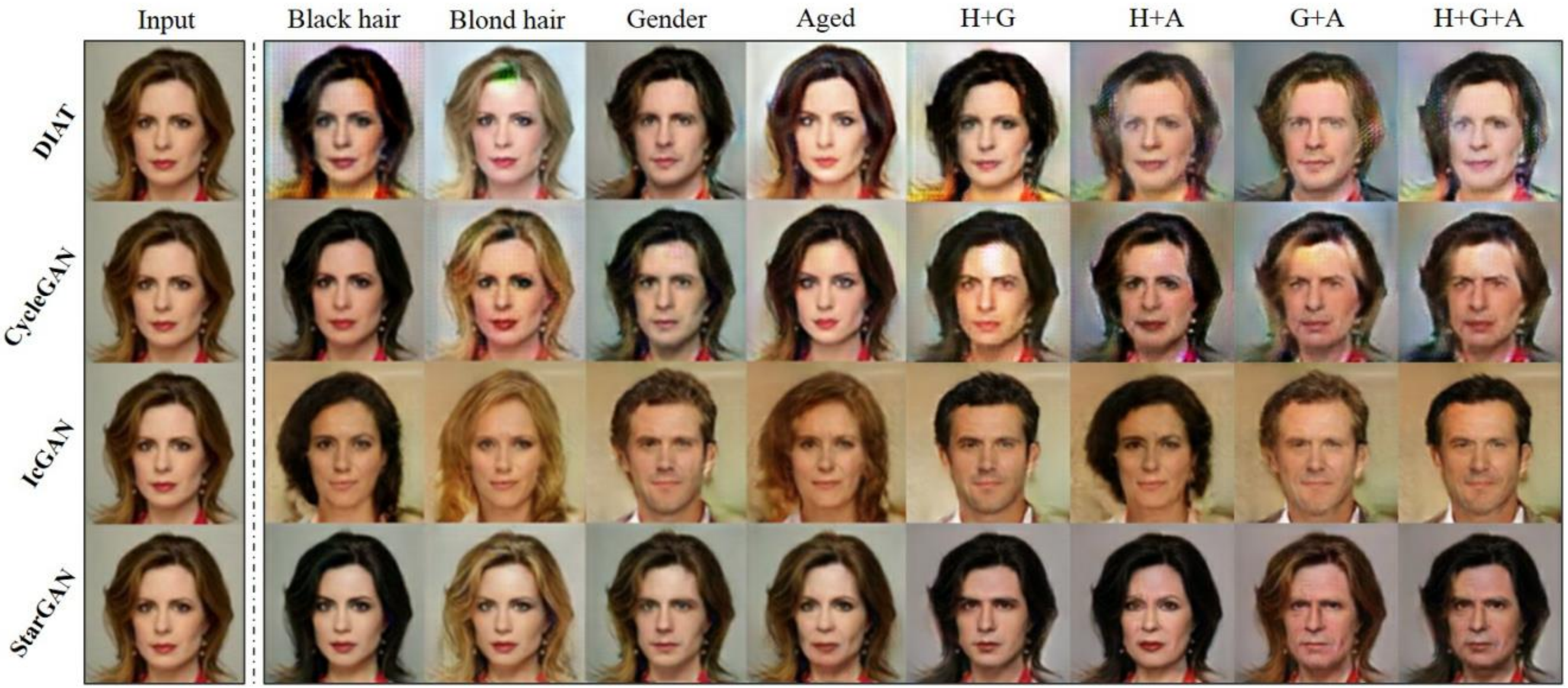
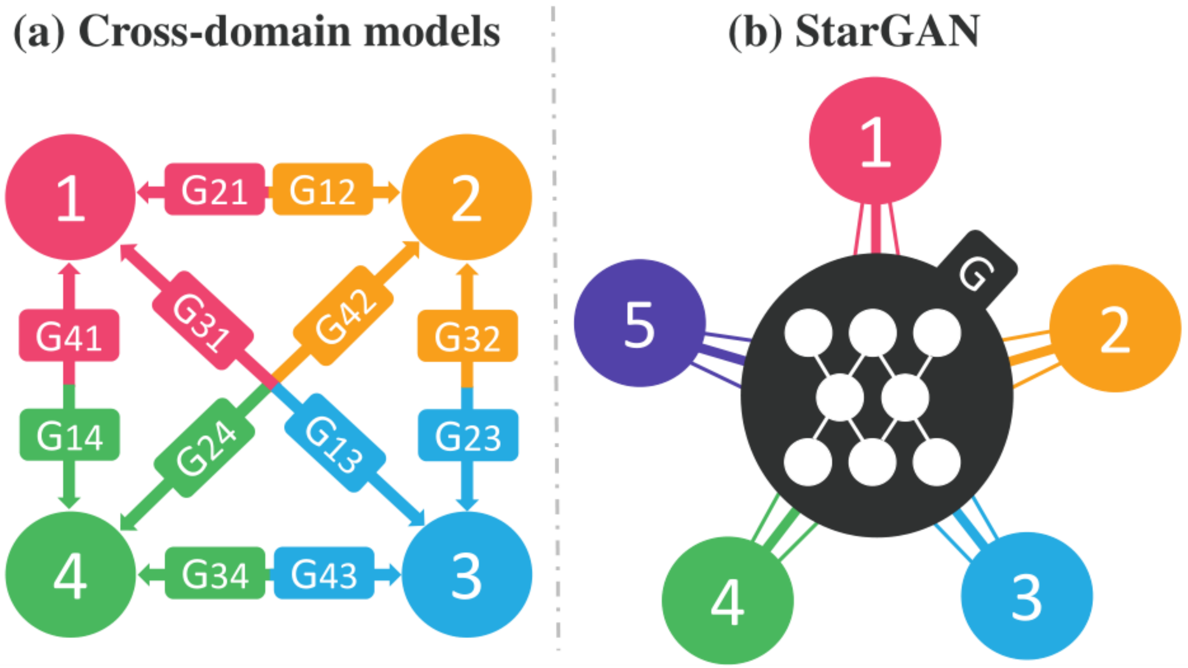
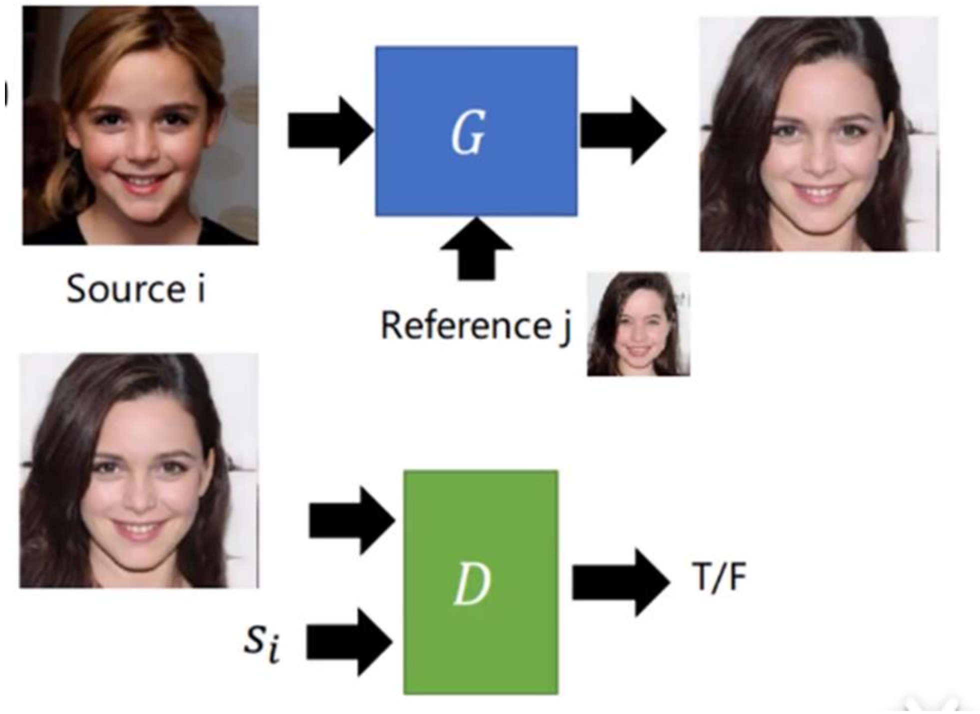


Figure 4. Facial attribute transfer results on the CelebA dataset. The first column shows the input image, next four columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. H: Hair color, G: Gender, A: Aged.

[1] Choi Y , Choi M , Kim M , et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[J].

Stargan





如果只能训练一对一的模型，会导致两个问题

- 训练低效，每次训练耗时很大
- 训练效果有限，因为一个领域转换单独训练的话就不能利用其它领域的数据来增大泛化能力

Stargan

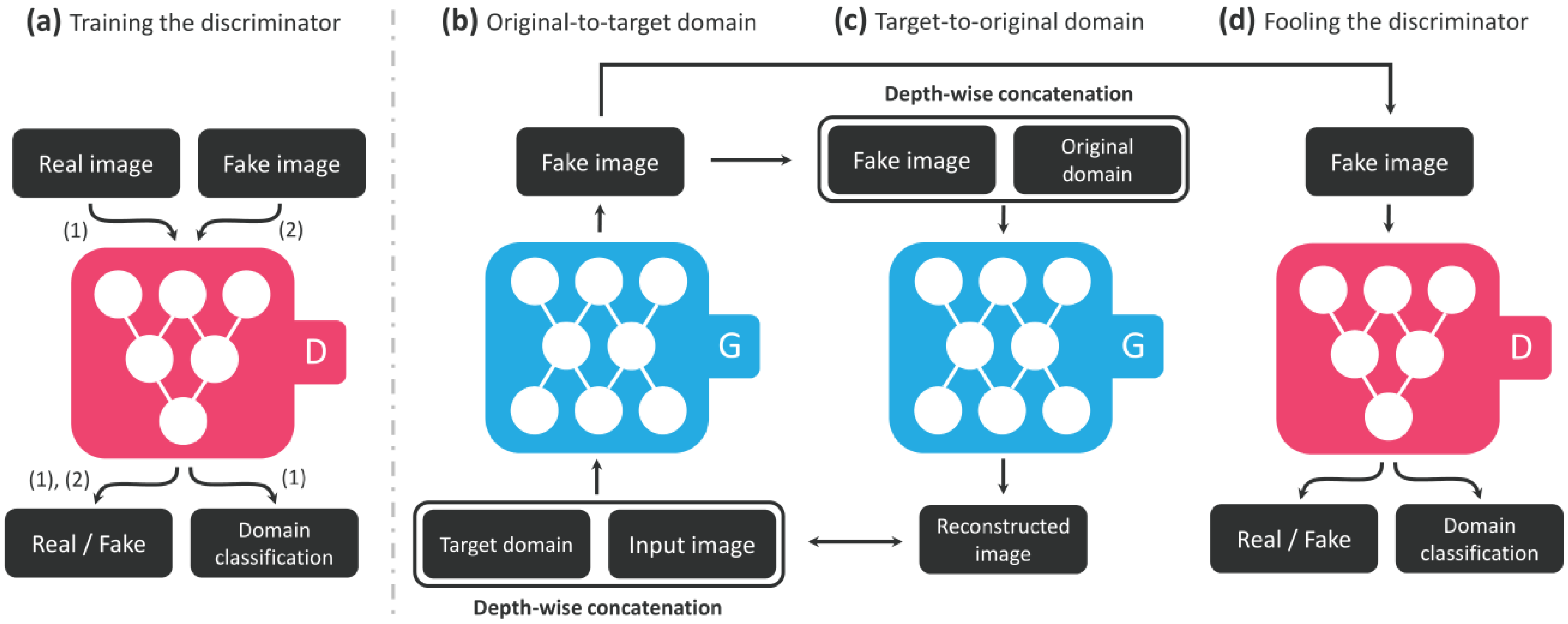


Figure 3. Overview of StarGAN, consisting of two modules, a discriminator D and a generator G. **(a)** D learns to distinguish between real and fake images and classify the real images to its corresponding domain. **(b)** G takes in as input both the image and target domain label and generates an fake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

Star-Gan

- 我们的目标是训练一个生成器G，它学习多个域之间的映射。为此，我们训练G将输入图像x和目标域标签c，即G（x，c）→y，得到输出图像y。
- 对抗损失（Adversarial Loss）

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

Star-Gan

- 域分类损失（Domain Classification Loss）

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

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

- 重构损失（Reconstruction Loss）

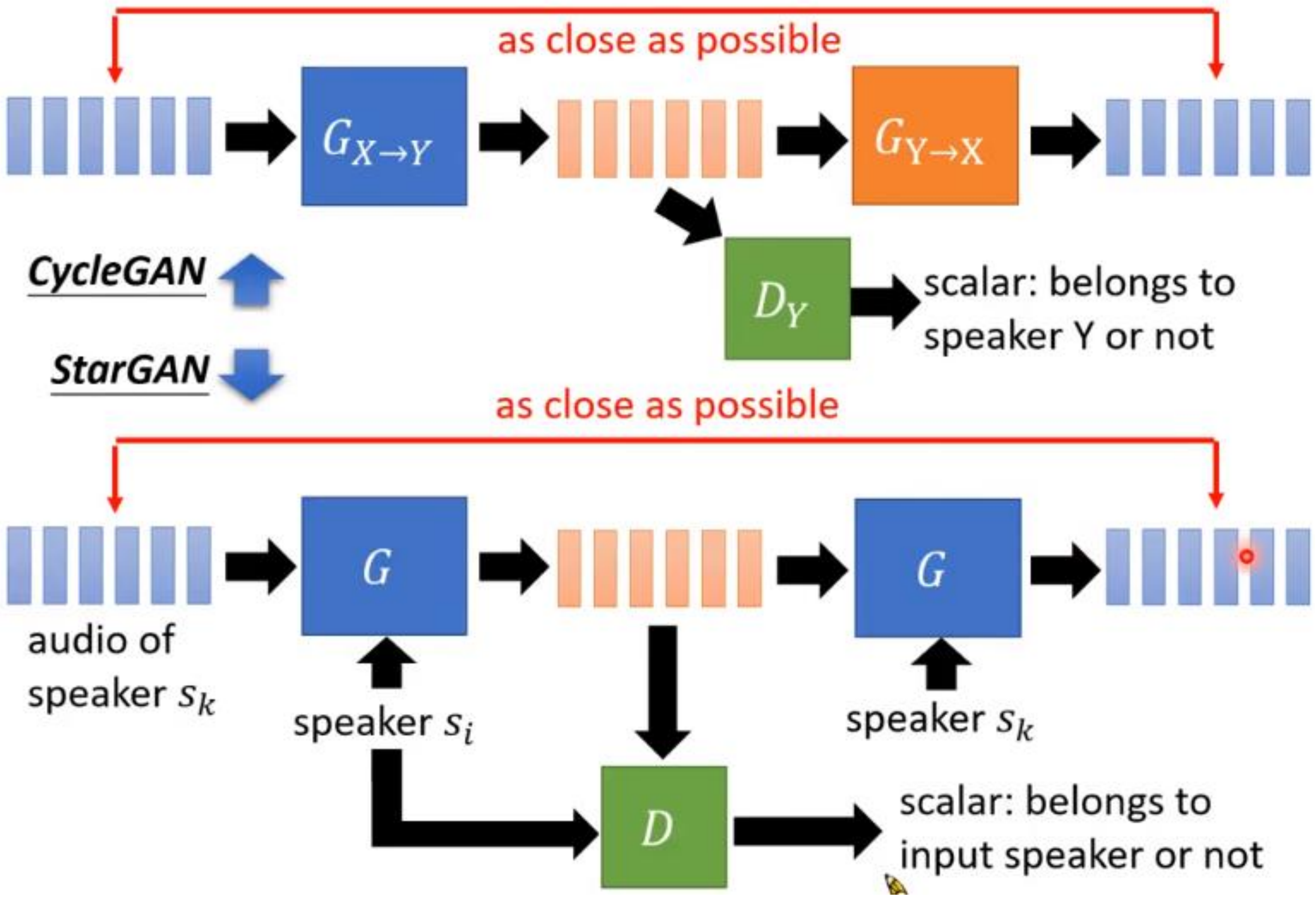
$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x,c),c')||_1]$$

- 完整目标（Full Objective）在所有实验中，我们使用λcls= 1和λrec= 10

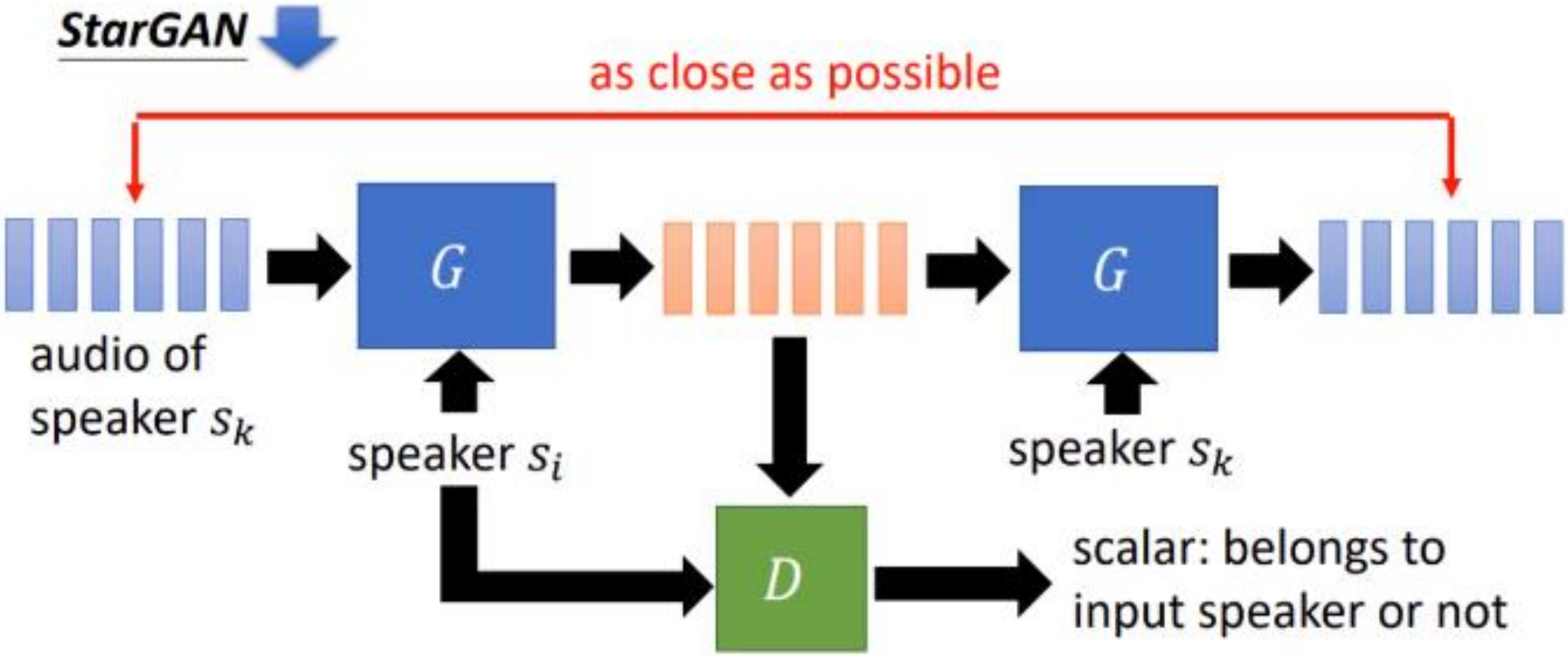
$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}_{cls}^{r},$$

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

Cyclegan & Stargan



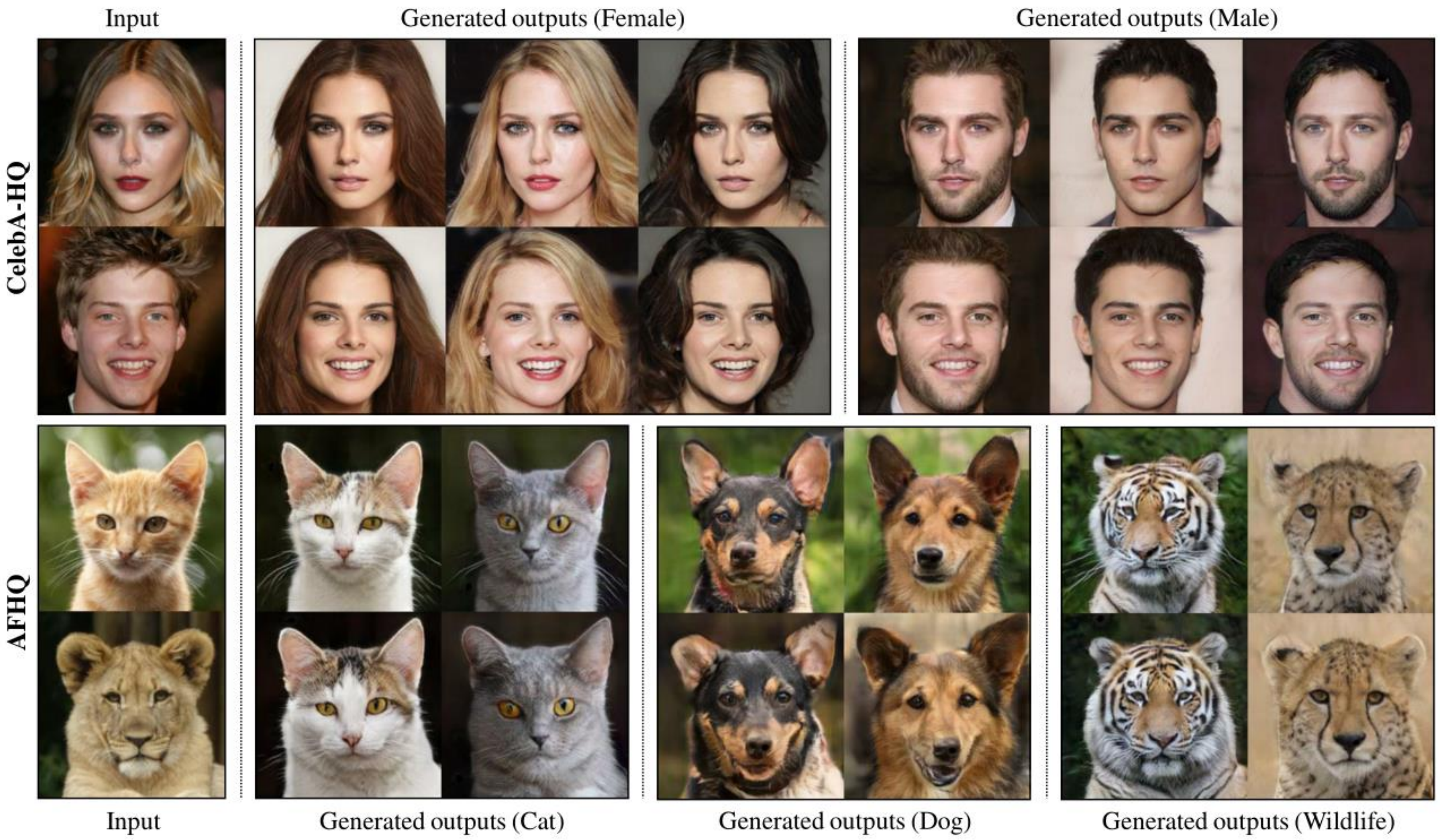
Stargan (Voice conversion)



- 计算出mel倒频谱系数，训练G之后，我们可以将输入话语的声学特征序列x转换为y = G(x, c)
- 生成器（Generator）：将声学特征序列x视为具有1个通道的Q×N大小的图像，并使用2D CNN构造
- 判别器（Real/Fake Discriminator）：使用门控CNN设计D，该门将声学特征序列y和属性标签c作为输入，并产生一系列概率，这些概率测量y的每个片段成为属性c的真实语音特征的可能性。最终输出D（y，c）由所有这些概率的乘积给出（使用PatchGAN的想法）
- 领域分类器（Domain Classifier）：使用门控CNN的域分类器C，该分类器采用声学特征序列y并生成一系列类概率分布，该序列测量了y的每个片段属于属性c的可能性。

[2] Kameoka H，Kaneko T，Tanaka K , et al. StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks[J]. IEEE, 2018.

Stargan v2



[3] Choi Y , Uh Y , Yoo J , et al. StarGAN v2: Diverse Image Synthesis for Multiple Domains[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

Stargan v2

- Overview

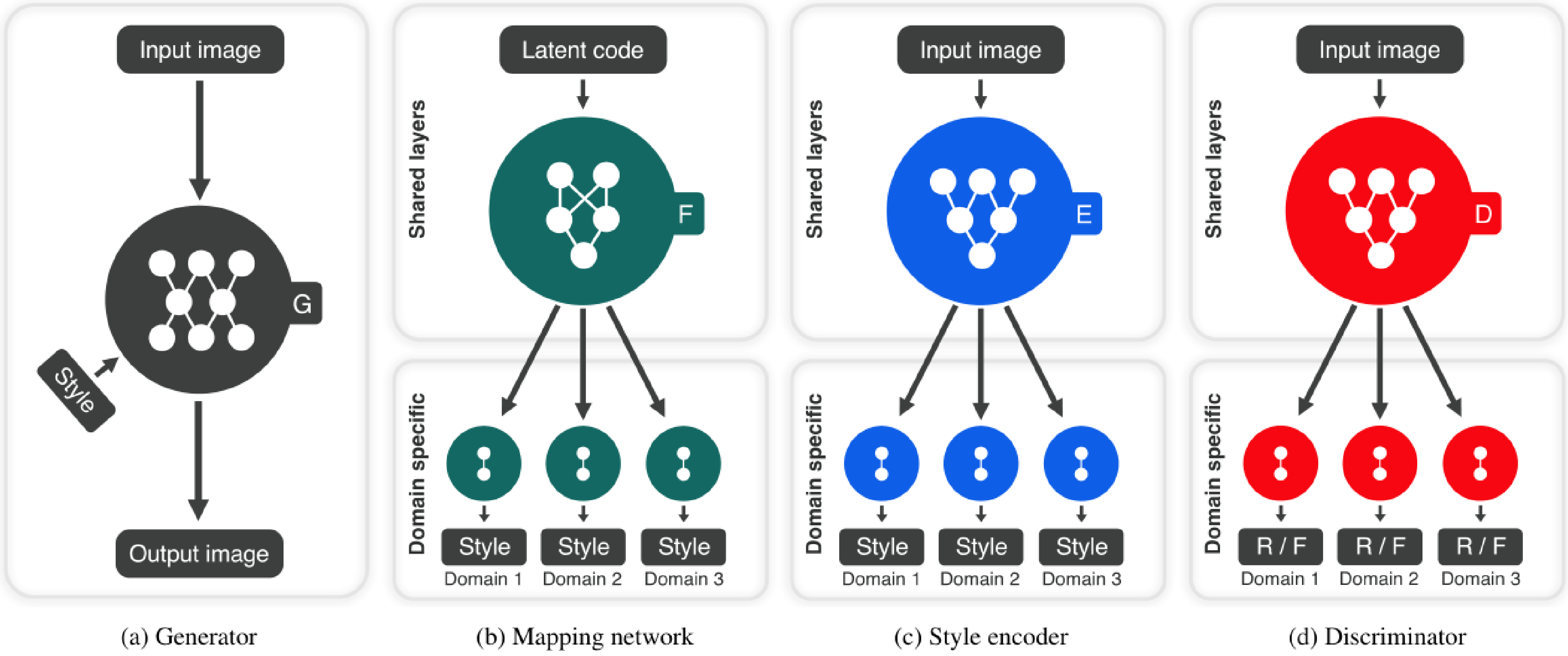


Figure 2. Overview of StarGAN v2, consisting of four modules. **(a)** The generator translates an input image into an output image reflecting the domain-specific style code. **(b)** The mapping network transforms a latent code into style codes for multiple domains, one of which is randomly selected during training. **(c)** The style encoder extracts the style code of an image, allowing the generator to perform reference-guided image synthesis. **(d)** The discriminator distinguishes between real and fake images from multiple domains. Note that all modules except the generator contain multiple output branches, one of which is selected when training the corresponding domain.

Stargan v2

给定一个图象 $\mathbf{x} \in \mathcal{X}$ 它的原始领域 $y \in \mathcal{Y}$.

- Adversarial objective

随机地采样隐空间 $\mathbf{z} \in \mathcal{Z}$ 和目标领域 $\widetilde{y} \in \mathcal{Y}$ ，生成style code $\widetilde{\mathbf{s}} = F_{\widetilde{y}}(\mathbf{z})$

$$\begin{aligned}\mathcal{L}_{adv} = &\mathbb{E}_{\mathbf{x},y}\left[\log D_y(\mathbf{x})\right] + \\ &\mathbb{E}_{\mathbf{x},\widetilde{y},\mathbf{z}}[\log\left(1 - D_{\widetilde{y}}(G(\mathbf{x},\widetilde{\mathbf{s}}))\right)],\end{aligned}$$

- Style reconstruction

$$\mathcal{L}_{sty} = \mathbb{E}_{\mathbf{x},\widetilde{y},\mathbf{z}}\left[||\widetilde{\mathbf{s}} - E_{\widetilde{y}}(G(\mathbf{x},\widetilde{\mathbf{s}}))||_1\right].$$

- Style diversification

$$\mathcal{L}_{ds} = \mathbb{E}_{\mathbf{x},\widetilde{y},\mathbf{z}_1,\mathbf{z}_2}\left[\|G(\mathbf{x},\widetilde{\mathbf{s}}_1) - G(\mathbf{x},\widetilde{\mathbf{s}}_2)\|_1\right],$$

Stargan v2

- Preserving source characteristics

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{x},y,\widetilde{y},\mathbf{z}}\left[||\mathbf{x} - G(G(\mathbf{x},\widetilde{\mathbf{s}}),\hat{\mathbf{s}})||_1\right],$$

- Full objective

$$\begin{aligned}\min_{G,F,E}\max_{D} \quad & \mathcal{L}_{adv} + \lambda_{sty}\,\mathcal{L}_{sty} \\ & -\lambda_{ds}\,\mathcal{L}_{ds} + \lambda_{cyc}\,\mathcal{L}_{cyc},\end{aligned}$$

Stargan v2


Jonathan Fly @jonathanfly · 19小时
The good news is StarGAN v2 code is out.
The better news is if you put a human into the animal model you do in fact get out a feline version of the human, and it's even wearing a suit.
github.com/clovaai/starga... h/t @ak92501


Jonathan Fly @jonathanfly · 19小时
Dogs seem to work better in human to animal translation. Maybe they have a wider range of facial expressions? They certainly look more comfortable wearing a suit.

Stargan (Voice conversion)

- 英语实验（VTCK数据集）
- 原说话人声音　　目标说话人声音　　转换后声音
- 藏文实验