1 STARGAN

Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation [1]

1.1 Abstract

Image to Multi Domain Image ができる GAN。経験則的に顔の部分と表情の変換には効果的なモデル。

1.2 Intro

1.2.1 Dataset

- CelebA[2]：10,177 人のセレブ、202,599 サイズ、40 種類の表情のデータセット
- RaFD[3]：67 人の 8 種類の表情のデータセット

1.2.2 Compare

既存の multi domain モデルは、k 個のドメインに対して $k(k-1)$ 個の generator を学習させる必要がある。が、StarGAN は一個だけでいい (1.1)。

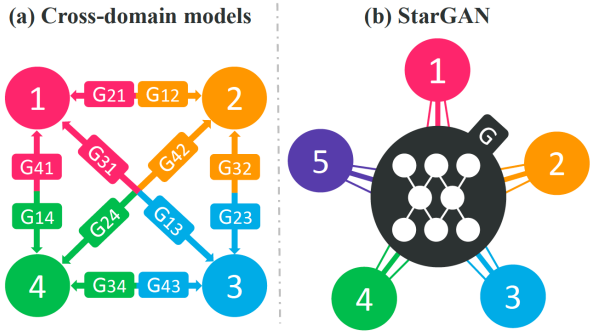

図 1.1: Compare

- GAN
- Conditional GANs
- Image to Image Translation. pix2pix, cGANs, UNIT, CoGAN, CycleGAN, DiscoGAN

1.3 Method

GOAL: to train A Single G that learns mappings among multiple domains.

HOW: to achive this, we train G to traslate an input image x into an output image y conditioned on the target domain label c, $G(x, c) \to y$

We randomly generate the target domain label c so that G learns to flexibly translate the input image.

We introduce an auxiliary(補助的な) classifier that allows a single discriminator to control multiple domains.

That is, our discriminator D produces probability distributions over both sources and domain labels, $D : x \to \{D_{src}(x), D_{cls}(x)\}$. D はinput image x を $\{real/fake, domains\}$ にマップする。

G tries to minimize the objetive, while D tries to maximize it.

Domain Classification Loss. For a given input image x and a target domain label c, our goal is to translate x into an ouput image y, which is properly classified to the taget domain c.

objective: a domain classification loss of real images used to optimize D, and a domain classficiation loss of fake images used to optimze G.

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

L_{cls}^{r} を学習データを使って minimize することで、D は与えられた画像に適切なラベルをつけるようになる。

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

G は L_{cls}^{f} を minimize することで、G によって生成された画像が目的ドメインとして判別されるようになう。

Reconstruction Loss. By minimizing the *adversarial* and *classification* losses(2 つの loss), G は realistic で classified な画像を生成できる。However, minimizing the losses does not guarantee that ドメインに変換しつつ、入力画像の質を維持する。これを緩和するため、we apply a cycle consistency loss to the generator.

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

これは、target domain c を設定して生成される $G(x, c)$ を、元の domain c' に生成し直すということ。Reconstuction するということだね。ちなみに、L1 ノルム使ってる。

Full Objective.

$$L_D = -L_{adv} + \lambda_{cls} L^r_{cls}$$

$$L_G = L_{adv} + \lambda_{cls} L^f_{cls} + \lambda_{rec} L_{rec}$$

We use $\lambda_{cls} = 1,\ \lambda_{rec} = 10$

参考文献

[1] Choi, Yunjey and Choi, Minje and Kim, Munyoung and Ha, Jung-Woo and Kim, Sunghun and Choo, Jaegul. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. In *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June, 2018.

[2] Ziwei Liu and Ping Luo and Xiaogang Wang and Xiaoou Tang. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*. December, 2015.

[3] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24(8)*, 1377—1388. DOI: 10.1080/02699930903485076