Image-to-Image Translation用GAN手法のサーベイ

2018/08/29

産業技術総合研究所 人工知能研究センター
金崎 朝子

この資料について

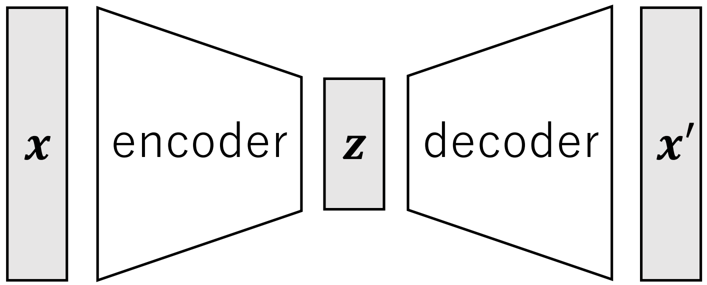
- 画像変換系のGANについて調べる機会があったので、せっかくなのでまとめました。
- なおECCV2018以降はノーチェック。
- アーキテクチャとコスト関数にのみフォーカス。
- 手法の比較が目的のため、詳しい説明は省略。
- 初心者向け

1. 基礎的な手法の紹介

Autoencoder (AE)

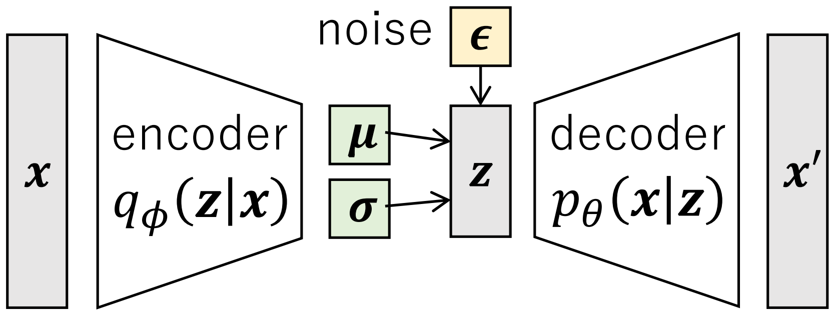
$$\mathcal{L}_{\text{AE}} = \mathbb{E}_{\mathbf{x}}[\|\mathbf{x} - \mathbf{x}'\|^2].$$

$\mathcal{L}$：目的関数



Variational Autoencoder (VAE)

- D. P. Kingma, M.Welling, “**Auto-encoding variational bayes**,” in: Proceedings of International Conference on Learning Representations (ICLR), 2014.
- D. J. Rezende, S. Mohamed, D. Wierstra, “**Stochastic backpropagation and approximate inference in deep generative models**,” in: Proceedings of International Conference on Machine Learning (ICML), 2014.



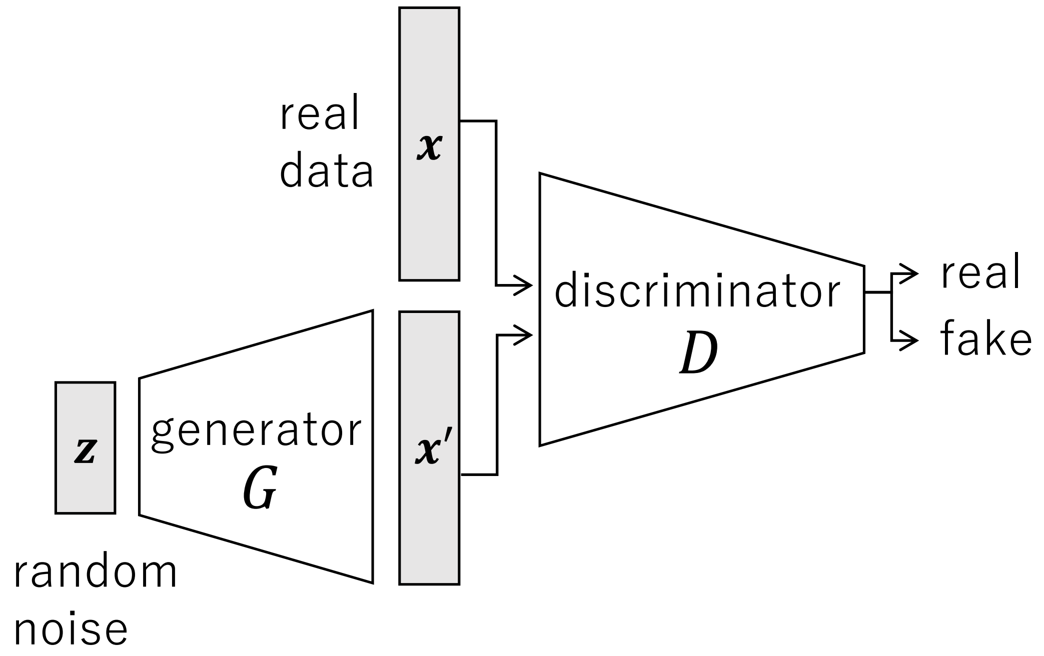
$$\mathcal{L}_{\mathrm{VAE}} = D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})\right) - \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})].$$

KLダイバージェンス

$$\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2\mathbf{I})$$

Generative Adversarial Nets (GAN)

- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "**Generative adversarial nets**," in: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2014.

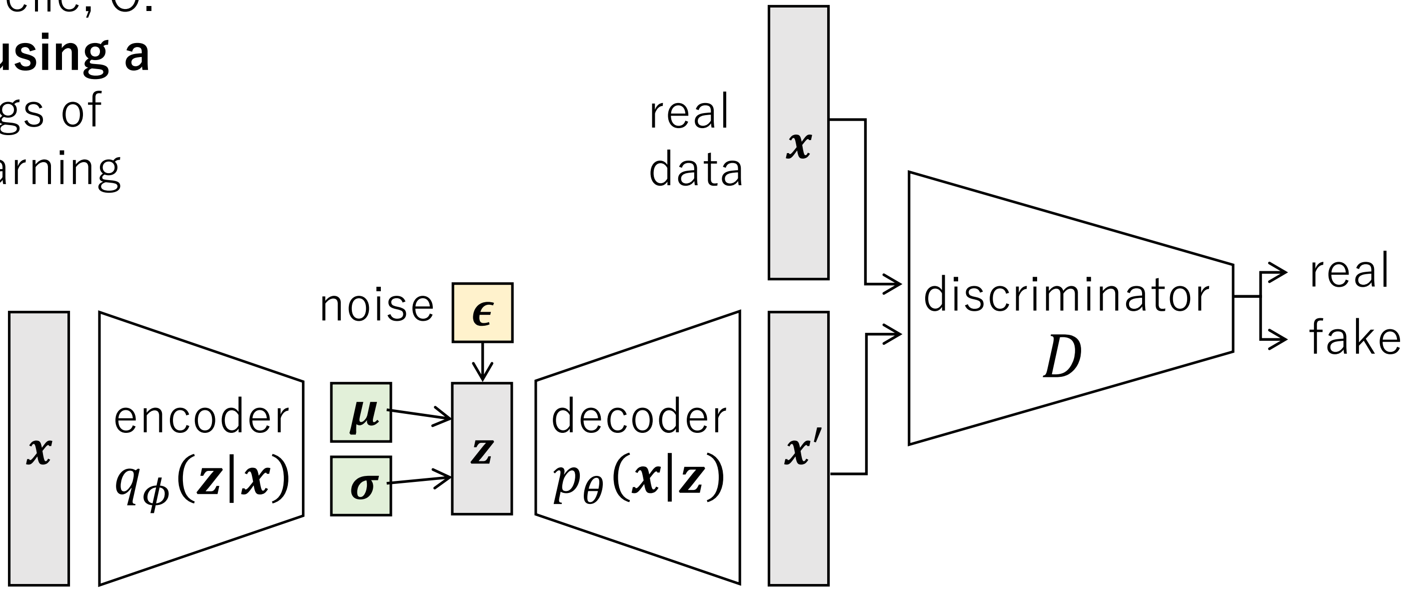


$$\min_G \max_D \mathcal{L}_{\mathrm{GAN}}$$ ←GANのmin, maxはこのノリなので以下省略

$$\mathcal{L}_{\mathrm{GAN}} = \mathbb{E}_{\mathbf{x}\sim p_{\mathrm{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))].$$

VAE-GAN

- A. B. L. Larsen, S. K. Sønderby, H. Larochelle, O. Winther, “**Autoencoding beyond pixels using a learned similarity metric**,” in: Proceedings of International Conference on Machine Learning (ICML), 2016.



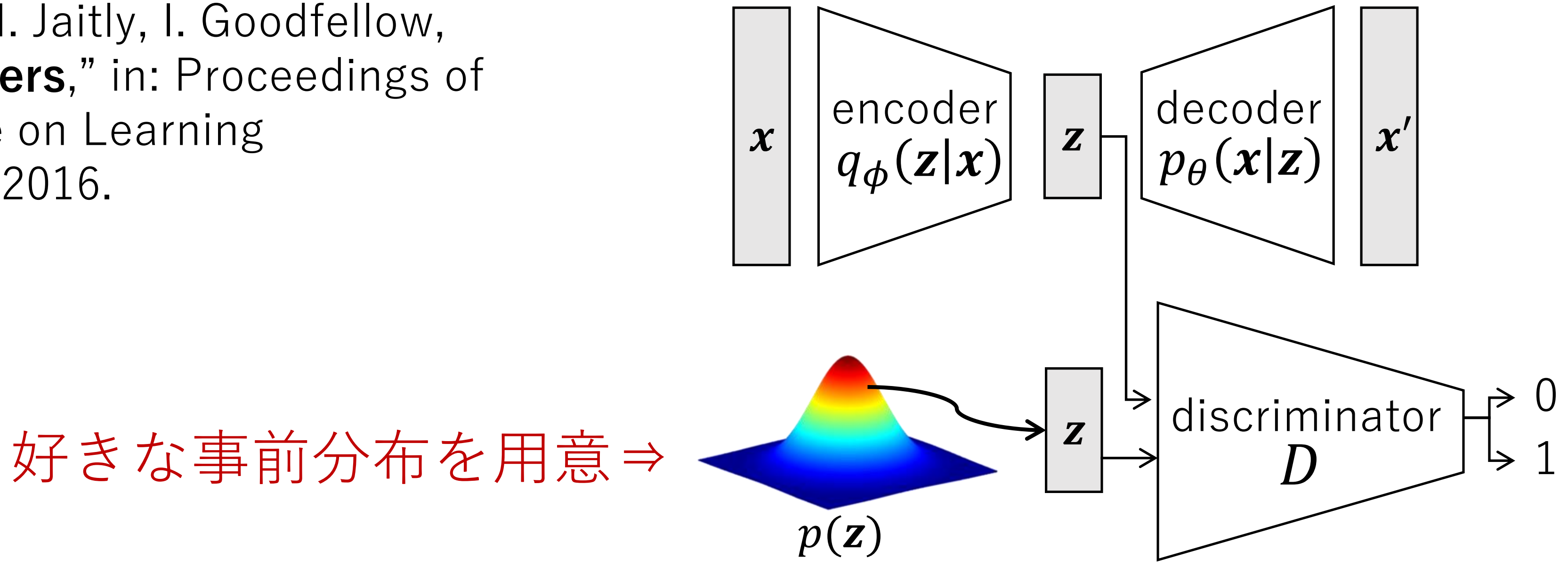
$$\mathcal{L}_{\text{VAE}-\text{GAN}} = \mathcal{L}_{\text{VAE}^*} + \mathcal{L}_{\text{GAN}}$$

$$\mathcal{L}_{\text{VAE}^*} = D_{\text{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})\right) - \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(D(\mathbf{x})|\mathbf{z})],$$

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))].$$

Adversarial Autoencoder (AAE)

- A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, "**Adversarial autoencoders**," in: Proceedings of International Conference on Learning Representations (ICLR), 2016.



$$\mathcal{L}_{\text{AAE}} = \mathcal{L}_{\text{AE}} + \mathcal{L}_{\text{GAN}}$$

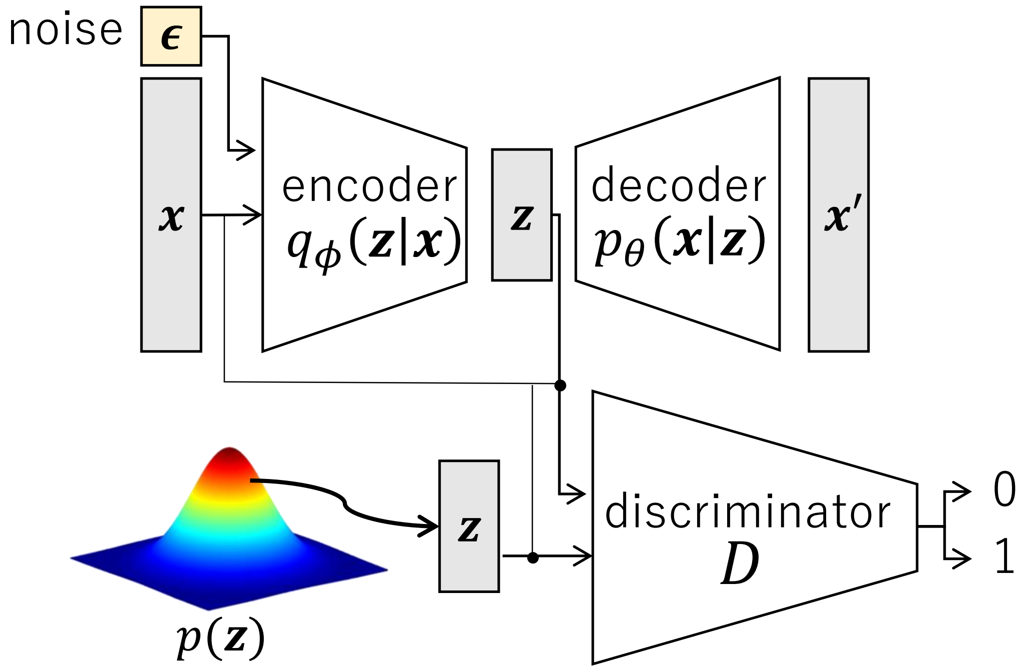
$$\mathcal{L}_{\text{AE}} = \mathbb{E}_{\mathbf{x}}[\|\mathbf{x} - \mathbf{x}'\|^2].$$

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log D(\mathbf{z})] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}[\log(1 - D(\mathbf{z}))].$$

Adversarial Variational Bayes (AVB)

- L. Mescheder, S. Nowozin, A. Geiger, "**Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks**," in: Proceedings of International Conference on Machine Learning (ICML), 2017.

※AAEはAVBの特殊なケース

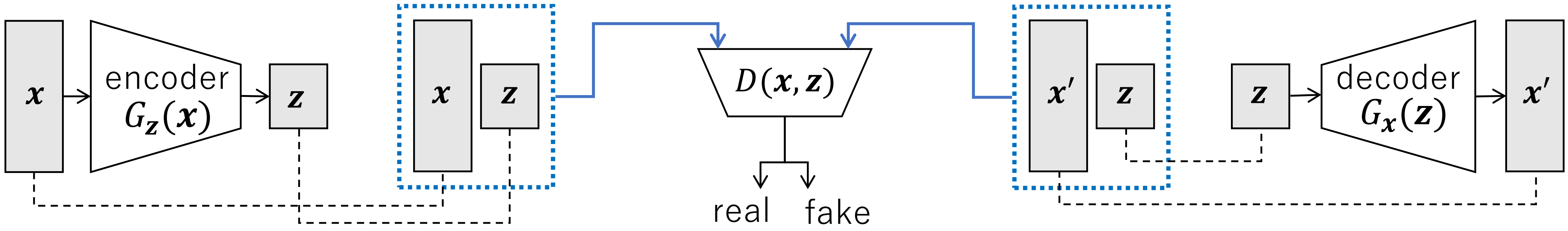


$$\sigma(t) \equiv (1 + e^{-t})^{-1}$$

$$\max_D \; \mathbb{E}_{p_{\mathrm{data}}(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \log \sigma(D(\mathbf{x}, \mathbf{z})) + \mathbb{E}_{p_{\mathrm{data}}(\mathbf{x})} \mathbb{E}_{p(\mathbf{z})} \log(1 - \sigma(D(\mathbf{x}, \mathbf{z}))),$$

$$\max_\theta \max_\phi \mathbb{E}_{p_{\mathrm{data}}(\mathbf{x})} \mathbb{E}_\epsilon \left(-D(\mathbf{x}, \mathbf{z}_\phi(\mathbf{x}, \epsilon)) + \log p_\theta(\mathbf{x}|\mathbf{z}_\phi(\mathbf{x}, \epsilon))\right).$$

ALI, BiGAN　※同じモデル



ALI

- V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, A. Courville, "**Adversarially learned inference**," in: Proceedings of International Conference on Learning Representations (ICLR), 2017.

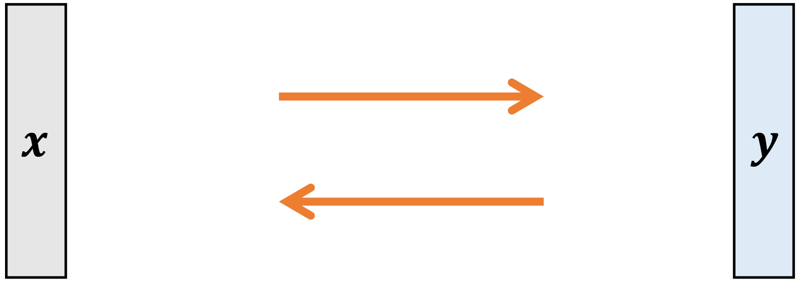
BiGAN

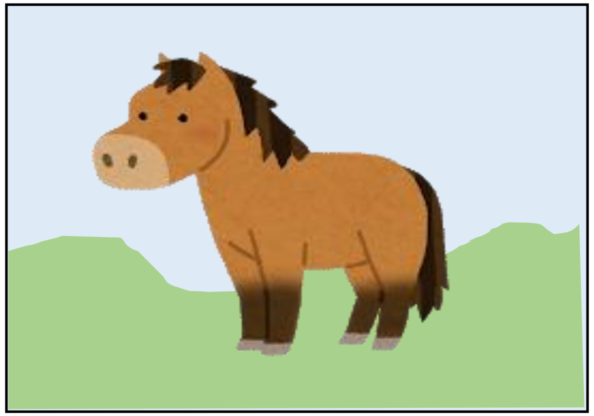
- J. Donahue, P. Krähenbühl, T. Darrell, "**Adversarial feature learning**," in: Proceedings of International Conference on Learning Representations (ICLR), 2017.

$$\mathcal{L}_{\mathrm{BiGAN}} = \mathbb{E}_{q(\mathbf{x})}[\log D(\mathbf{x}, G_{\mathbf{z}}(\mathbf{x}))] + \mathbb{E}_{p(\mathbf{z})}[\log(1 - D(G_{\mathbf{x}}(\mathbf{z}), \mathbf{z}))].$$

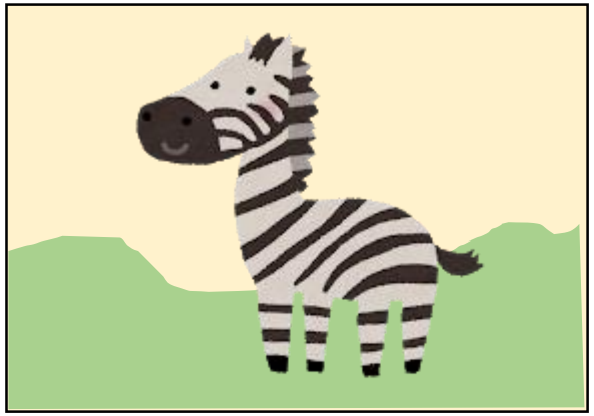
2. Image-to-Image Translation

画像から画像への変換





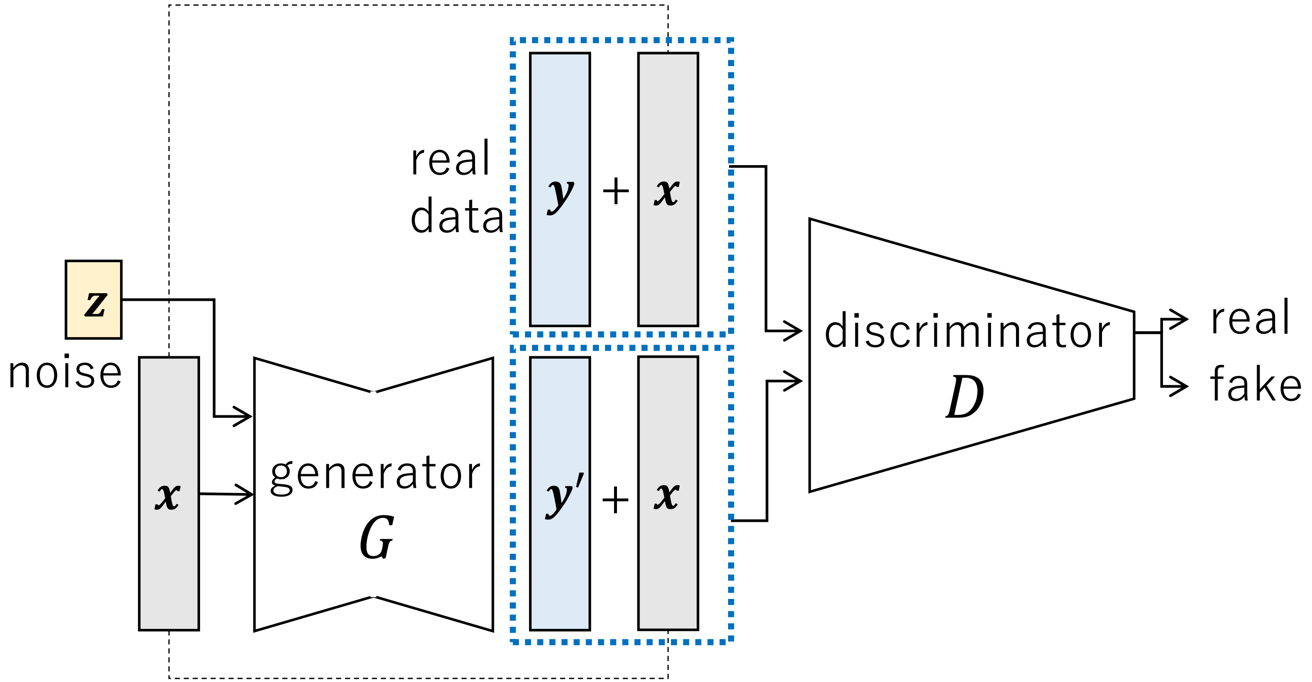
ソースドメイン



ターゲットドメイン

pix2pix

- P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, "**Image-to-image translation with conditional adversarial networks**," in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.



conditional GAN

$$\mathcal{L}_{\text{pix2pix}}(G, D) = \mathcal{L}_{\text{cGAN}}(G, D) + \lambda \mathcal{L}_{L_1}(G),$$

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{\mathbf{x},\mathbf{y}}[\log D(\mathbf{x}, \mathbf{y})] + \mathbb{E}_{\mathbf{x},\mathbf{z}}[\log(1 - D(\mathbf{x}, G(\mathbf{x}, \mathbf{z})))],$$

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{z}}[\|\mathbf{y} - G(\mathbf{x}, \mathbf{z})\|_1].$$

pix2pixHD

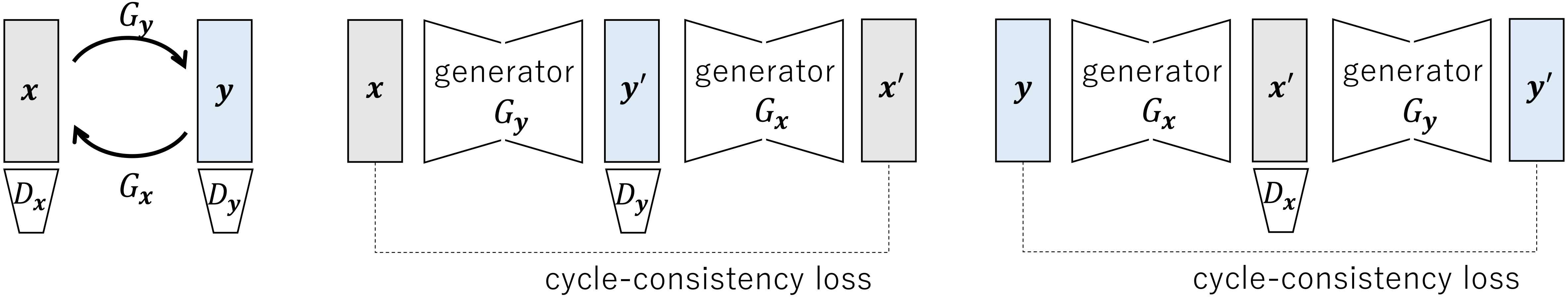
- T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, “**High-resolution image synthesis and semantic manipulation with conditional gans**,” in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

マルチスケールのGとDを使うことで高解像度に

$$\min_{G}\left(\left(\max_{D_1,D_2,D_3}\sum_{k=1,2,3}\mathcal{L}_{\mathrm{cGAN}}(G,D_k)\right)+\lambda\sum_{k=1,2,3}\mathcal{L}_{\mathrm{FM}}(G,D_k)\right)$$

FM: Feature Matching

CycleGAN, DiscoGAN, DualGAN　※同じモデル



CycleGAN
J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, "**Unpaired image-to-image translation using cycle-consistent adversarial networks**," in: Proceedings of International Conference on Computer Vision (ICCV), 2017.

DiscoGAN
T. Kim, M. Cha, H. Kim, J. K. Lee, J. Kim, "**Learning to discover cross-domain relations with generative adversarial networks**," in: Proceedings of International Conference on Machine Learning (ICML), 2017.

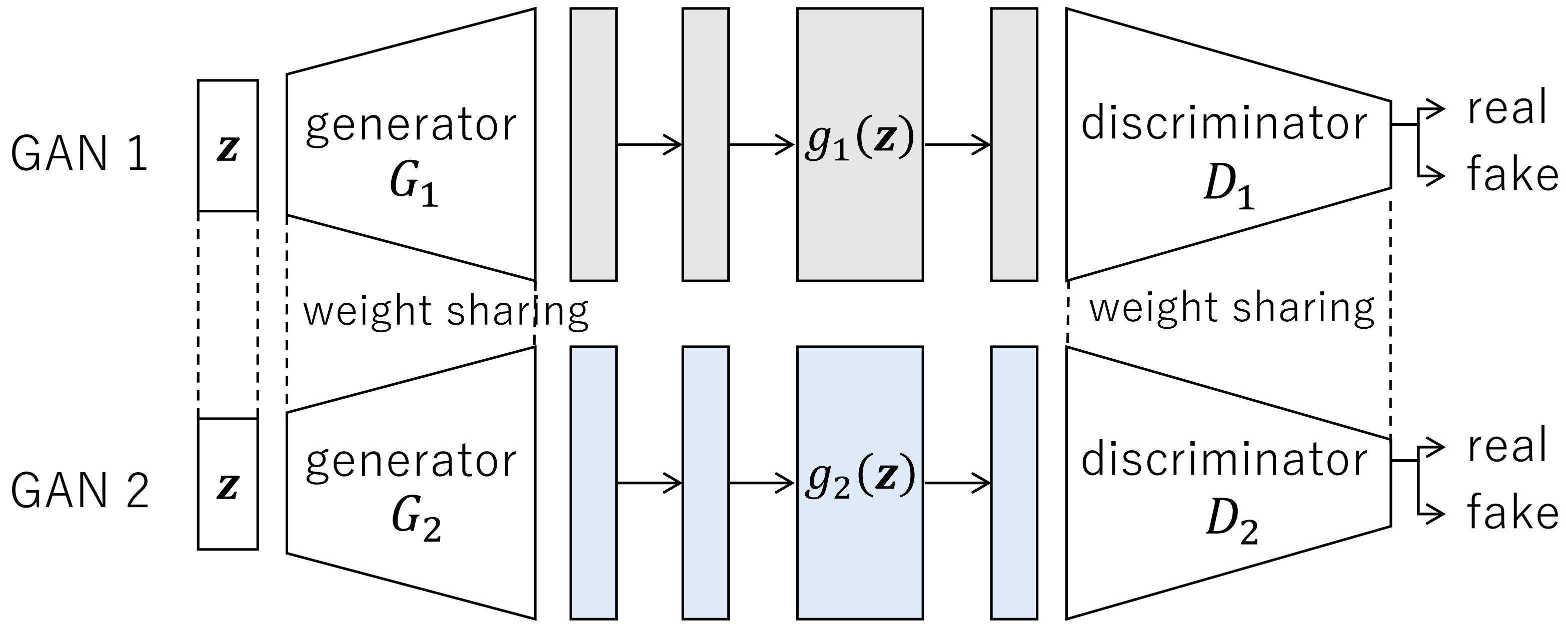
DualGAN
Z. Yi, H. Zhang, P. Tan, M. Gong, "**Dualgan: Unsupervised dual learning for image-to-image translation**," in: Proceedings of International Conference on Computer Vision (ICCV), 2017.

$$\mathcal{L}_{\text{cycleGAN}}(G_{\mathbf{x}}, G_{\mathbf{y}}, D_{\mathbf{x}}, D_{\mathbf{y}}) = \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}(\mathbf{y})}[\log D_{\mathbf{y}}(\mathbf{y})] + \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log(1 - D_{\mathbf{y}}(G_{\mathbf{y}}(\mathbf{x})))] + \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log D_{\mathbf{x}}(\mathbf{x})] + \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}(\mathbf{y})}[\log(1 - D_{\mathbf{x}}(G_{\mathbf{x}}(\mathbf{y})))] + \lambda \underline{\mathcal{L}_{\text{cyc}}(G_{\mathbf{x}}, G_{\mathbf{y}})}.$$

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

CoGAN

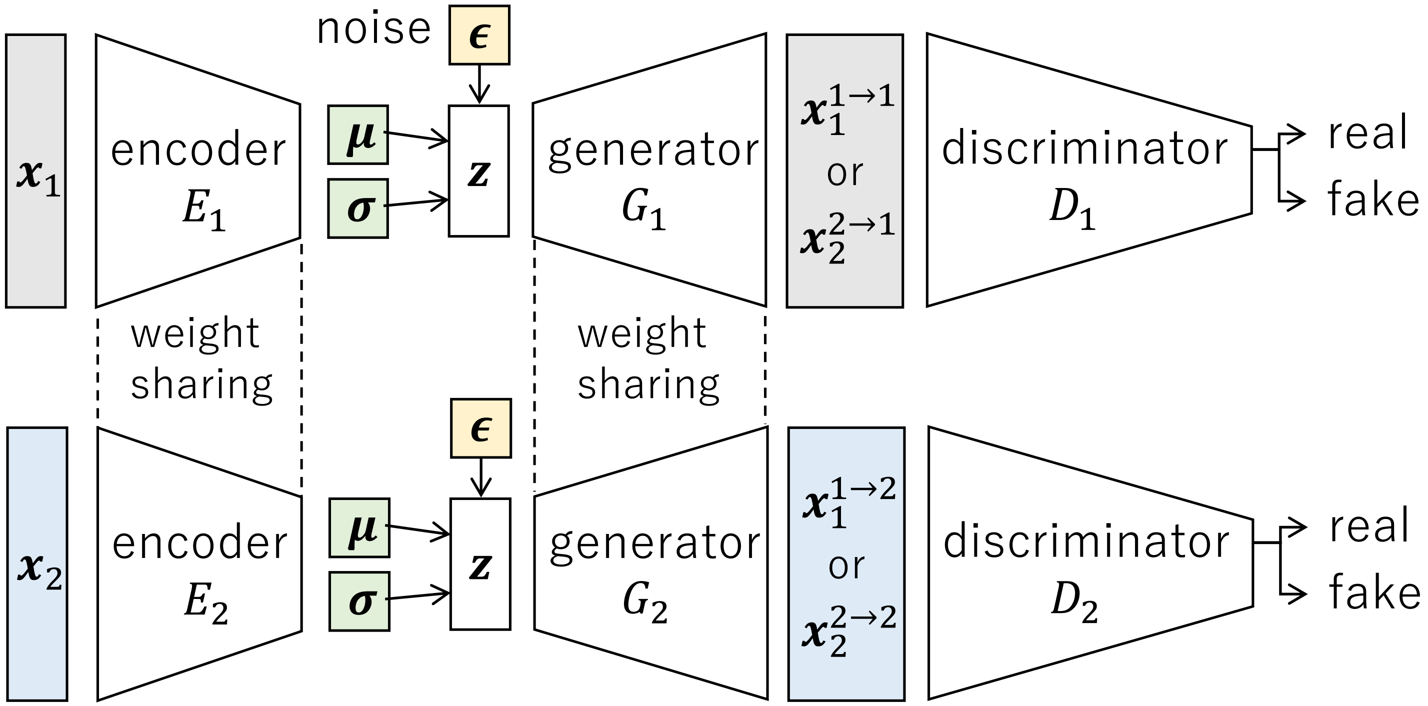
- M.-Y. Liu, O. Tuzel, “**Coupled generative adversarial networks**,” in: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2016.



$$\mathcal{L}_{\mathrm{CoGAN}}(G_1, G_2, D_1, D_2) = \mathbb{E}_{\mathbf{x}_2 \sim p_{\mathrm{data}}(\mathbf{x}_2)}[\log D_2(\mathbf{x}_2)] + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}[\log(1 - D_2(G_2(\mathbf{z})))] + \mathbb{E}_{\mathbf{x}_1 \sim p_{\mathrm{data}}(\mathbf{x}_1)}[\log D_1(\mathbf{x}_1)] + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}[\log(1 - D_1(G_1(\mathbf{z})))].$$

UNIT

- M.-Y. Liu, T. Breuel, J. Kautz, "**Unsupervised image-to-image translation networks**," in: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2017



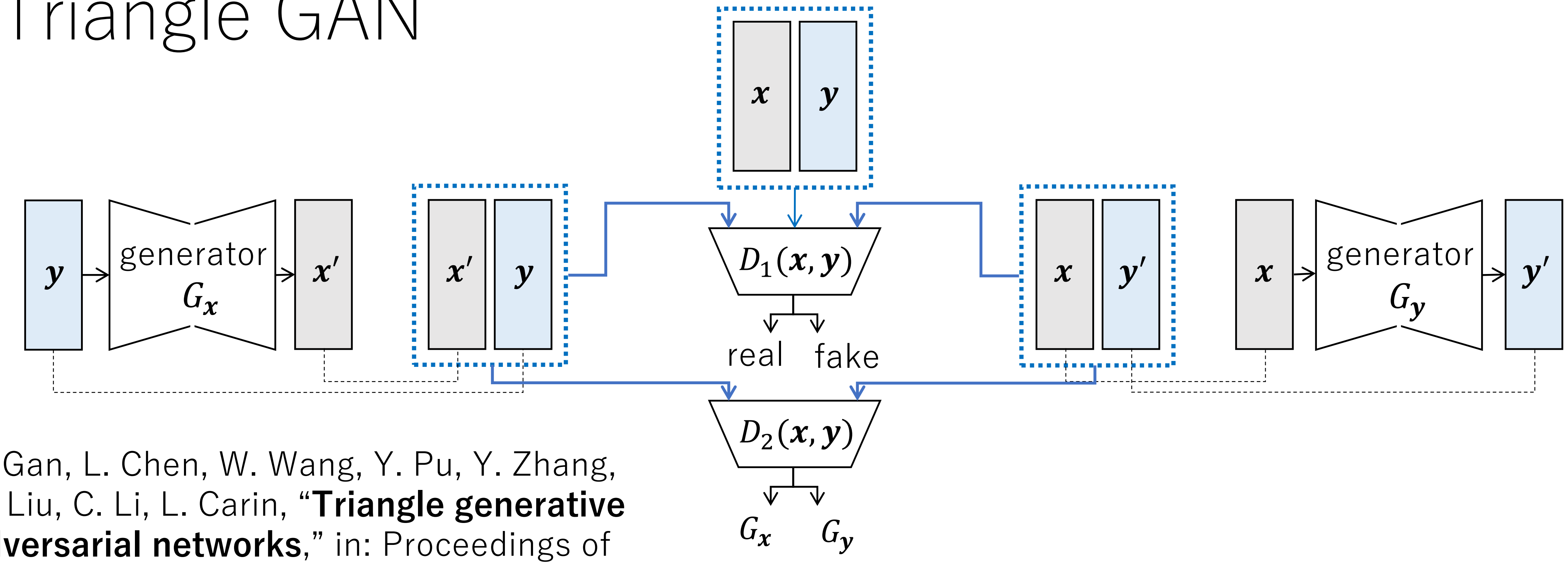
VAEとCoGANの組み合わせ

$$\mathcal{L}_{\text{UNIT}} = \lambda_1 \mathcal{L}_{\text{VAE}}(E_1, G_1) + \lambda_2 \mathcal{L}_{\text{GAN}}(E_2, G_1, D_1) + \lambda_3 \underline{\mathcal{L}_{\text{CC}}(E_1, G_1, E_2, G_2)} +$$
$$\lambda_1 \mathcal{L}_{\text{VAE}}(E_2, G_2) + \lambda_2 \mathcal{L}_{\text{GAN}}(E_1, G_2, D_2) + \lambda_3 \underline{\mathcal{L}_{\text{CC}}(E_2, G_2, E_1, G_1)}.$$

$$\underline{\mathcal{L}_{\text{CC}}(E_a, G_a, E_b, G_b)} = D_{\text{KL}}\left(q_a(\mathbf{z}_a|\mathbf{x}_a)||p_\theta(\mathbf{z})\right) + D_{\text{KL}}\left(q_b(\mathbf{z}_b|\mathbf{x}_a^{a\to b})||p_\theta(\mathbf{z})\right) -$$
$$\mathbb{E}_{\mathbf{z}_b \sim q_b(\mathbf{z}_b|\mathbf{x}_a^{a\to b})}[\log p_{G_a}(\mathbf{x}_a|\mathbf{z}_b)].$$

cycle consistency

Triangle GAN



- Z. Gan, L. Chen, W. Wang, Y. Pu, Y. Zhang, H. Liu, C. Li, L. Carin, “**Triangle generative adversarial networks**,” in: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2017.

cGANとBiGAN（= ALI）の組み合わせ

$$\mathcal{L}_{\Delta-\mathrm{GAN}} = \mathcal{L}_{\mathrm{cGAN}} + \mathcal{L}_{\mathrm{BiGAN}},$$

$$\mathcal{L}_{\mathrm{cGAN}} = \mathbb{E}_{p(\mathbf{x},\mathbf{y})}[\log D_1(\mathbf{x},\mathbf{y})] + \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x}',\mathbf{y})}[\log(1 - D_1(\mathbf{x}',\mathbf{y}))] + \mathbb{E}_{p_{\mathbf{y}}(\mathbf{x},\mathbf{y}')}[\log(1 - D_1(\mathbf{x},\mathbf{y}'))],$$

$$\mathcal{L}_{\mathrm{BiGAN}} = \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x}',\mathbf{y})}[\log D_2(\mathbf{x}',\mathbf{y})] + \mathbb{E}_{p_{\mathbf{y}}(\mathbf{x},\mathbf{y}')}[\log(1 - D_2(\mathbf{x},\mathbf{y}'))].$$

今回紹介しなかったもの

- Domain Transfer Network (DTN) [Taigman+, ICLR2017]
https://arxiv.org/abs/1611.02200

- BicycleGAN [Zhu+, NIPS2017]
https://junyanz.github.io/BicycleGAN/

- Unsupervised Image-to-Image Translation with Generative Adversarial Networks [Dong+, NIPS2017]
https://arxiv.org/pdf/1701.02676.pdf

- Triple GAN [Li+, NIPS2017]: Triangle GANと間違われやすいが画像間変換ではない
https://arxiv.org/abs/1703.02291

- StarGAN [Choi+, CVPR2018]: マルチドメインを同時に扱える
https://github.com/yunjey/StarGAN

おわりに

- 同時期に全く同じ手法が複数出ていて地獄
- 名前が似てても全く別物のものが多いのでちゃんと読んだ方がいい
- 間違い等があればご連絡くださいm(_ _)m