

Generative Adversarial Networks の基礎・発展・応用

NTTコミュニケーション科学基礎研究所
金子 卓弘

2019年7月28日@MIRU2019チュートリアル

解説記事

208

日本音響学会誌 74 卷 4 号 (2018), pp. 208–218

解説

Generative Adversarial Networks の基礎と応用*

金子卓弘 (日本電信電話株式会社)**

PDF無料ダウンロード可

https://www.jstage.jst.go.jp/article/jasj/74/4/74_208/_pdf

金子のWebページ 「Takuhiro Kaneko NTT」で検索



Takuhiro Kaneko
Research Scientist
NTT Communication Science Laboratories, NTT Corporation
kaneko.takuhiro at lab.ntt.co.jp
[Google Scholar]

デモ・コードあり

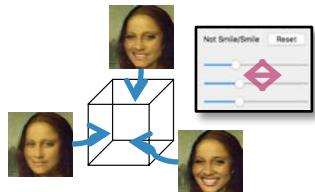
<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/index.html>

自己紹介



2014. 4～ NTTコミュニケーション科学基礎研究所
2017. 4～ 東京大学（社会人博士、現在在学中）

MIRU

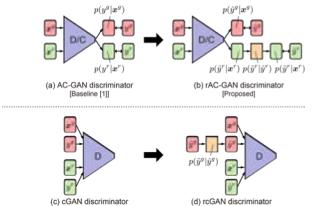


CFGAN
[CVPR 2017]

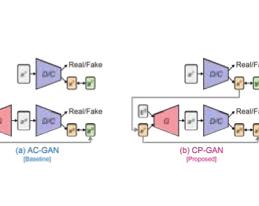


DTLC-GAN
[CVPR 2018]

操作可能な画像生成



rGAN⁺
[CVPR 2019]

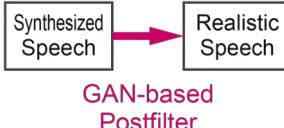


CP-GAN⁺
[BMVC 2019]

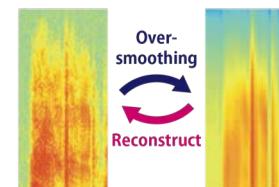
不完全データに頑健な画像生成

[†]社会人博士の研究成果

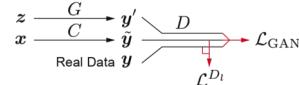
KIKU



GAN-PF
[ICASSP 2017]



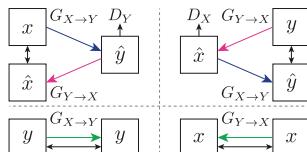
GAN-PF for STFT
[Interspeech 2017]



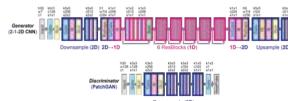
GAN-VC
[Interspeech 2017]

高品質な音声合成・音声変換

CycleGAN-VC



CycleGAN-VC
[EUSIPCO 2018]



CycleGAN-VC2
[ICASSP 2019]

StarGAN-VC2

StarGAN-VC2
[Interspeech 2019]

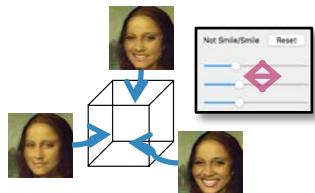
ノンパラレル・マルチドメイン音声変換

自己紹介

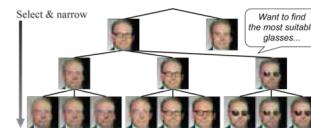


2014. 4～ NTTコミュニケーション科学基礎研究所
2017. 4～ 東京大学（社会人博士、現在在学中）

MIRU

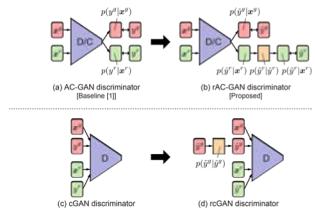


CF-GAN
[CVPR 2017]



DTLC-GAN
[CVPR 2018]

操作可能な画像生成



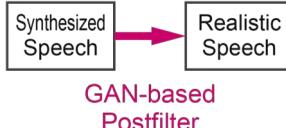
rGAN[†]
[CVPR 2019]

共通する技術：GAN

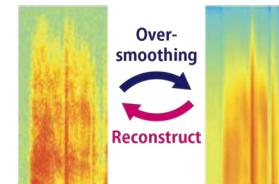
CP-GAN[†]
[BMVC 2019]

不完全データに頑健な画像生成

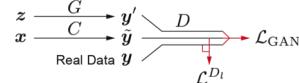
KIKU



GAN-PF
[ICASSP 2017]

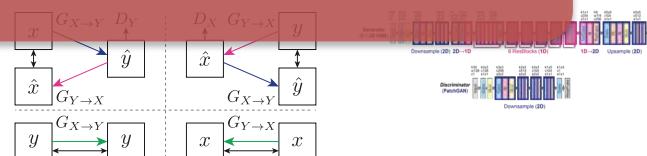


GAN-PF for STFT
[Interspeech 2017]

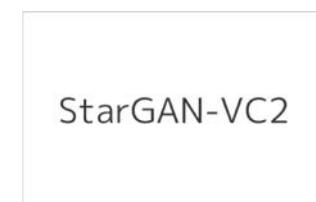


GAN-VC
[Interspeech 2017]

高品質な音声合成・音声変換



CycleGAN-VC
[EUSIPCO 2018]



StarGAN-VC2
[Interspeech 2019]

ノンパラレル・マルチドメイン音声変換

[†]社会人博士の研究成果

GANによる画像生成とは？



Q1. 計算機が生成したのはどれ？

- 生成 = 計算機が適当に与えた「乱数」から作り出したもの



A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.
T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

GANによる画像生成とは？

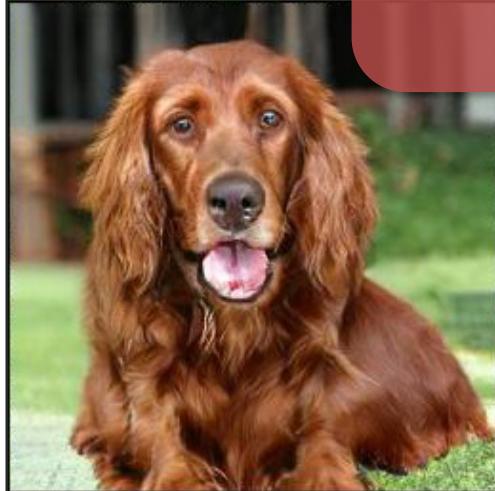


Q1. 計算機が生成したのはどれ？ → **BigGAN/StyleGAN**

- 生成 = 計算機が適当に与えた「乱数」から作り出したもの

正解は全部

高精細な画像を
ランダムに**生成可能**



BigGAN

StyleGAN

StyleGAN

A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.
T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

GANによる画像生成とは？



Q2. ○○に変換するには？



絵画 by Monet

写真



?



ウマ

シマウマ



?

J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

GANによる画像生成とは？

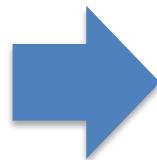


Q2. ○○に変換するには？ → **CycleGAN**



絵画 by Monet

写真



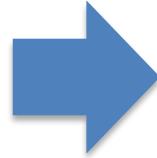
自動変換
が可能
シマウマ



ウマ



写真



シマウマ

J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

GANによる画像生成とは？



Q3. 好みの写真に修正するには？

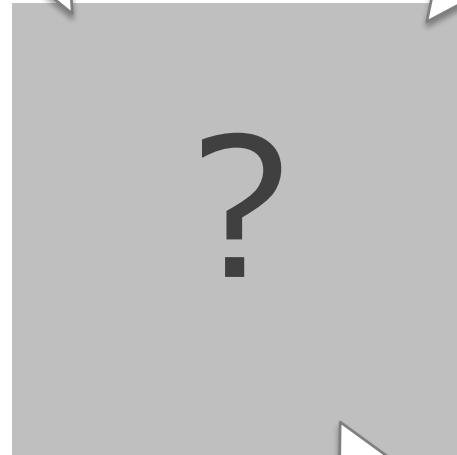


編集



長髪？

男性化？



？

笑顔？

GANによる画像生成とは？



Q3. 好みの写真に修正するには？ → **CFGAN**



デモ動画：<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/gac/index.html>

T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.

GANによる画像生成とは？



画像生成



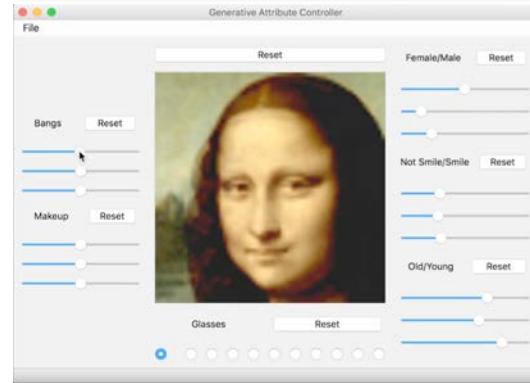
BigGAN [Brock+ICLR2019]
StyleGAN [Karras+CVPR2019]

画像変換



CycleGAN [Zhu+ICCV2017]

画像編集



CFGAN [Kaneko+CVPR2017]

GANによる画像生成とは？



画像生成



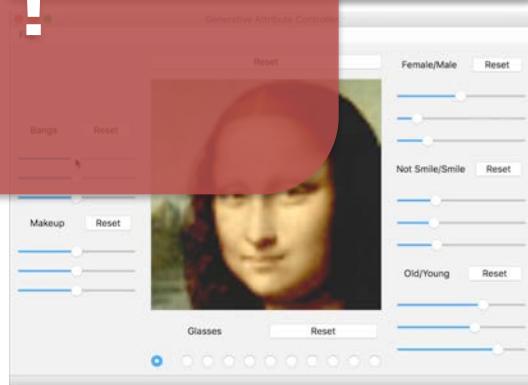
BigGAN [Brock+ICLR2019]
StyleGAN [Karras+CVPR2019]

様々な応用が
可能な技術！



CycleGAN [Zhu+ICCV2017]

画像編集



CFGAN [Kaneko+CVPR2017]

目次



**GAN
の基礎**

**GAN
の発展**

**GAN
の応用**

目次



**GAN
の基礎**

**GAN
の発展**

**GAN
の応用**

GANの位置付け

- 生成モデルとは？
- 深層生成モデル
- AR, Flow, VAE, GANの比較

GANの理論

- 目的関数
- 更新アルゴリズム
- 理論的解釈

GANの位置付け

- 生成モデルとは？
- 深層生成モデル
- AR, Flow, VAE, GANの比較

GANの理論

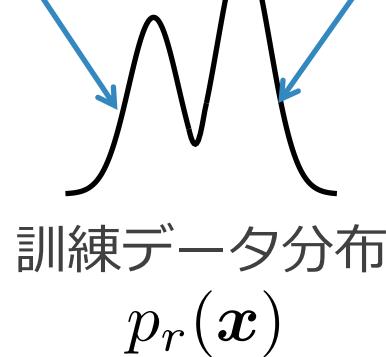
- 目的関数
- 更新アルゴリズム
- 理論的解釈

生成モデルとは？

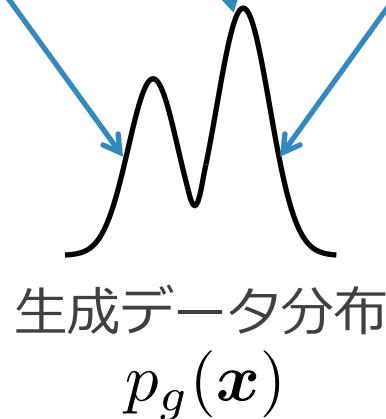
訓練データ



生成データ



一致させたい…



Pictures: A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.

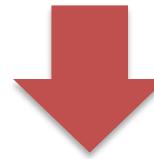
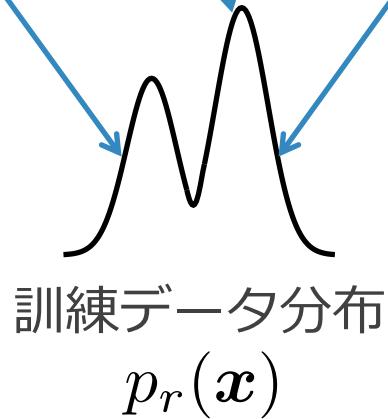
生成モデルの学習の難しさ



生成データ



高次元データ（画像・動画など）
では複雑な形状
→人手で設計するのは困難



深層学習の表現力の高さを活用できないか?
→**深層生成モデル**

Pictures: A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.

深層生成モデル



代表的な深層生成モデル (**AR**, **Flow**, **VAE**, **GAN**)

$p_g(\mathbf{x})$ の表現方法

陽に表現

陰に表現

近似なし

近似あり

AR

Autoregressive Model

連鎖律 (Chain Rule)

$$p_g(\mathbf{x}) = \prod_{i=1}^n p_g(x_i | x_1, \dots, x_{i-1})$$

Flow

Flow-based Model

可逆生成

$$z = f(\mathbf{x}) \quad \mathbf{x} = f^{-1}(z)$$

VAE

Variational Autoencoder

変分近似

$$\begin{aligned} \log p_g(\mathbf{x}) &\geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}) \\ &= -KL(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})) \\ &\quad + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})} [\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})] \end{aligned}$$

GAN

Generative Adversarial Networks

敵対的学習

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

I. Goodfellow., "NIPS 2016 Tutorial: Generative Aversarial Networks," NIPS 2016.

深層生成モデル



代表的な深層生成モデル (**AR, Flow, VAE, GAN**)

$p_g(x)$ の表現方法

陽に表現

近似なし

AR

Autoregressive Model

連鎖律 (Chain Rule)

$$p_g(x) = \prod_{i=1}^n p_g(x_i|x_1, \dots, x_{i-1})$$

Flow

Flow-based Model

可逆生成

$$z = f(x) \quad x = f^{-1}(z)$$

陰に表現

近似あり

VAE

Variational Autoencoder

変分近似

$$\begin{aligned}\log p_g(x) &\geq \mathcal{L}(\theta, \phi; x) \\ &= -KL(q_\phi(z|x)\|p(z)) \\ &\quad + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]\end{aligned}$$

GAN

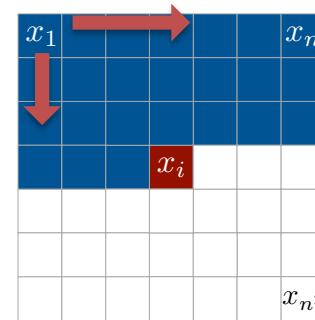
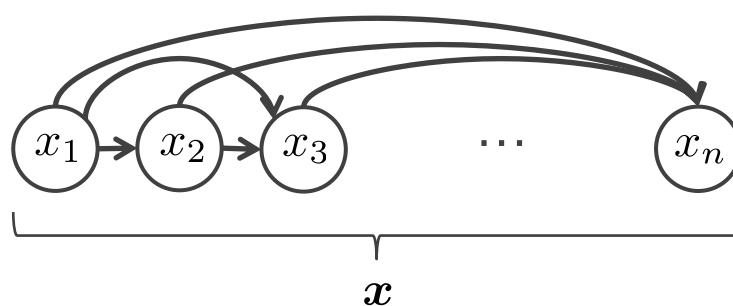
Generative Adversarial Networks

敵対的学習

$$\begin{aligned}\min_G \max_D \mathbb{E}_{x \sim p_r(x)} [\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]\end{aligned}$$

Autoregressive Model [van den Oord+2016]

- 連鎖律 (Chain Rule) を用いて n 次元のベクトル x を 1次元の確率分布 x_i に分解



全体では
複雑な分布

$$p_g(x) = \prod_{i=1}^n p_g(x_i | x_1, \dots, x_{i-1})$$

個々の分布に分解
し扱いやすくする

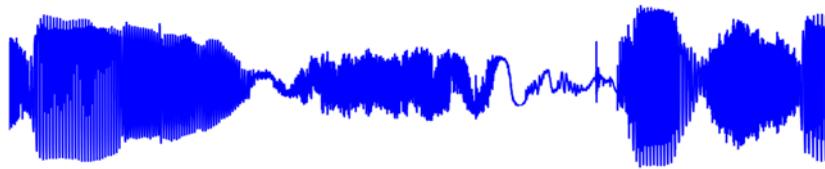
- 各因子の関係をニューラルネットワークを用いてモデル化
 - RNN, CNNなど

A. van den Oord et al., "Pixel Recurrent Neural Networks," ICML 2016.

ARの特徴（長所）



モデル	AR 連鎖律
サンプル クオリティ	😊 高
サンプリング コスト	😢 高
データ生成の コントロール	😢 困難
モデルの 定量的評価	😊 対数尤度
ネットワーク 構造制限	😢 強



例：WaveNet
[van den Oord+2016]

再帰的なデータ生成

潜在的な表現なし

$p_g(x)$ を近似なしに最適化

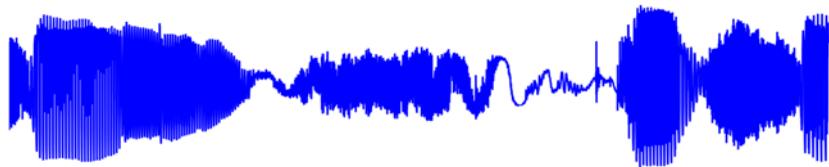
再帰的な表現が必要

A. van den Oord et al., "WaveNet: A Generative Model for Raw Audio," arXiv 2016.

ARの特徴（短所）



モデル	AR 連鎖律
サンプル クオリティ	😊 高
サンプリング コスト	😢 高
データ生成の コントロール	😢 困難
モデルの 定量的評価	😊 対数尤度
ネットワーク 構造制限	😢 強



例：WaveNet
[van den Oord+2016]

再帰的なデータ生成

潜在的な表現なし

$p_g(x)$ を近似なしに最適化

再帰的な表現が必要

A. van den Oord et al., "WaveNet: A Generative Model for Raw Audio," arXiv 2016.

深層生成モデル



代表的な深層生成モデル (AR, Flow, VAE, GAN)

$p_g(x)$ の表現方法

陽に表現

近似なし

AR

Autoregressive Model

連鎖律 (Chain Rule)

$$p_g(x) = \prod_{i=1}^n p_g(x_i|x_1, \dots, x_{i-1})$$

Flow

Flow-based Model

可逆生成

$$z = f(x) \quad x = f^{-1}(z)$$

陰に表現

近似あり

VAE

Variational Autoencoder

変分近似

$$\begin{aligned}\log p_g(x) &\geq \mathcal{L}(\theta, \phi; x) \\ &= -KL(q_\phi(z|x)\|p(z)) \\ &\quad + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]\end{aligned}$$

GAN

Generative Adversarial Networks

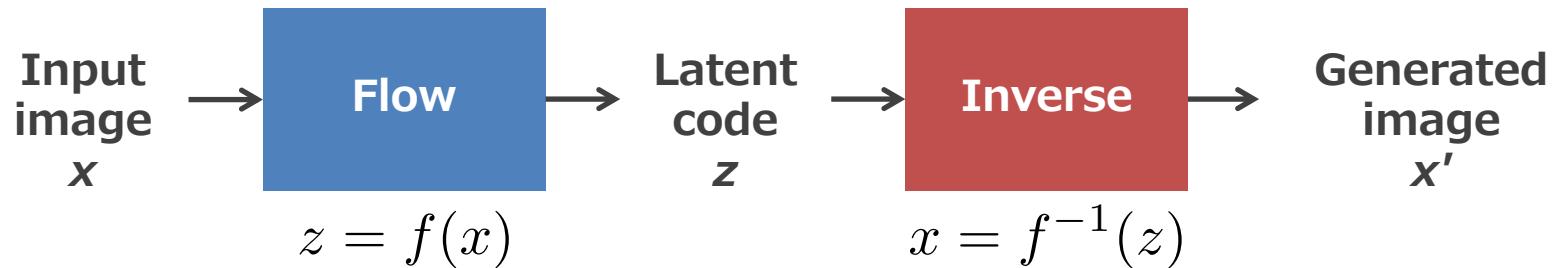
敵対的学習

$$\begin{aligned}\min_G \max_D \mathbb{E}_{x \sim p_r(x)} [\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]\end{aligned}$$

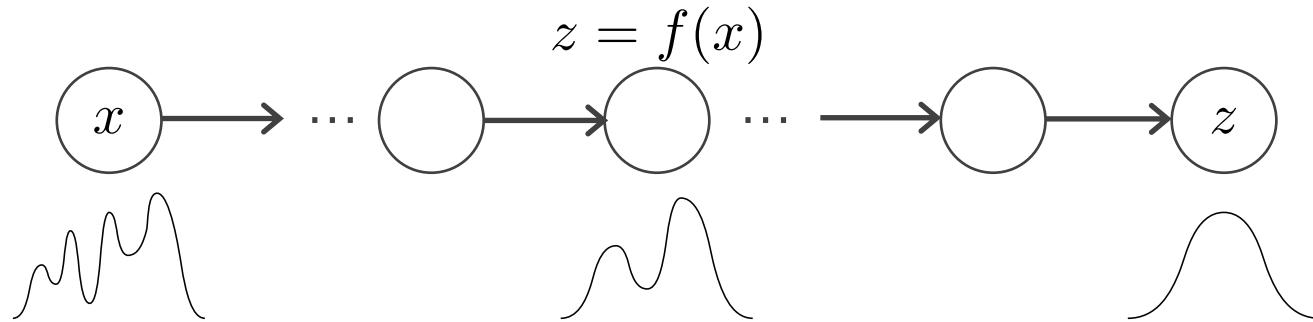
I. Goodfellow., "NIPS 2016 Tutorial: Generative Aversarial Networks," NIPS 2016.

Flow-based model [Dinh+2015, Dinh+2017]

- 逆変換可能なネットワーク (f) を利用
→ 推論 ($z = f(x)$) と生成 ($x = f^{-1}(z)$) を一つのモデルで実現



- 逆変換可能な層を多層にわたって用いることにより徐々に近づける



L. Dinh et al., "NICE: Non-linear Independent Components Estimation," ICLR 2015.

L. Dinh et al., "Density estimation using Real NVP," ICLR 2017.

Flowの特徴 (長所)



モデル	可逆生成
サンプル クオリティ	😊 高
サンプリング コスト	😊 低
データ生成の コントロール	😊 容易
モデルの 定量的評価	😊 対数尤度
ネットワーク 構造制限	😢 強



例 : Glow
[Kingma+2018]

1回のフォワードプロセスで生成可能

潜在的な表現あり

$p_g(x)$ を近似なしに最適化

全層が逆変換可能であることが必要

D. P. Kingma & P. Dhariwal, "Glow: Generative Flow with Invertible 1×1 Convolutions," NeurIPS 2018.

Flowの特徴 (短所)



モデル	可逆生成
サンプル クオリティ	😊 高
サンプリング コスト	😊 低
データ生成の コントロール	😊 容易
モデルの 定量的評価	😊 対数尤度
ネットワーク 構造制限	😢 強



例 : Glow
[Kingma+2018]

1回のフォワードプロセスで生成可能

潜在的な表現あり

$p_g(x)$ を近似なしに最適化

全層が逆変換可能であることが必要

D. P. Kingma & P. Dhariwal, "Glow: Generative Flow with Invertible 1×1 Convolutions," NeurIPS 2018.

深層生成モデル



代表的な深層生成モデル (AR, Flow, VAE, GAN)

$p_g(\mathbf{x})$ の表現方法

陽に表現

近似なし

AR

Autoregressive Model

連鎖律 (Chain Rule)

$$p_g(\mathbf{x}) = \prod_{i=1}^n p_g(x_i | x_1, \dots, x_{i-1})$$

Flow

Flow-based Model

可逆生成

$$z = f(\mathbf{x}) \quad \mathbf{x} = f^{-1}(z)$$

陰に表現

近似あり

VAE

Variational Autoencoder

変分近似

$$\begin{aligned}\log p_g(\mathbf{x}) &\geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}) \\ &= -KL(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})) \\ &\quad + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})} [\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})]\end{aligned}$$

GAN

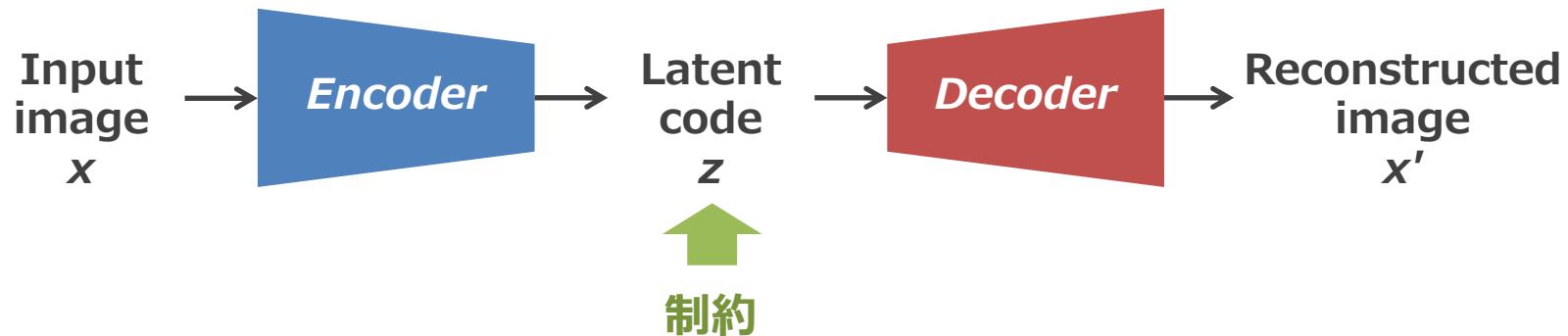
Generative Adversarial Networks

敵対的学習

$$\begin{aligned}\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] \\ + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]\end{aligned}$$

Variational Autoencoder [Kingma+2014, Rezende+2014]

- 推論 (*Encoder*) と生成 (*Decoder*) を二つのネットワークで近似



- 変分近似をして、下方限界を最大化

$$\log p_g(x) \geq \mathcal{L}(\theta, \phi; x)$$

直接扱う
のは困難

変分近似をして扱いやすくする

$$= -KL(q_\phi(z|x) \| p(z)) + \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)]$$

潜在変数の制約 Encoder Decoder

※サンプリングには
Reparameterization Trickを利用

D. P. Kingma & M. Welling, "Auto-encoding Variational Bayes," ICLR 2014.

D. J. Rezende et al., "Stochastic Backpropagation and Approximate Inference in Deep Generative Models," ICML 2014.

VAEの特徴 (長所)



モデル	VAE	変分推論
サンプル クオリティ		低
サンプリング コスト		低
データ生成の コントロール		容易
モデルの 定量的評価		下方限界
ネットワーク 構造制限		弱

VAE

データの分布 $p(x)$ を陽に近似 (Gaussianなど)
→ 統計的な平均化によりぼやける傾向

GAN

1回のフォワードプロセスで生成可能

潜在的な表現あり

最適化対象を陽に表現

様々な拡張が可能

A. B. L. Larsen et al., "Autoencoding beyond Pixels Using a Learned Similarity Metric," ICML 2016.

VAEの特徴 (短所)



モデル	VAE 変分推論	
サンプル クオリティ	 低	データの分布 $p(x)$ を陽に近似 (Gaussianなど) → 統計的な平均化によりぼやける傾向
サンプリング コスト	 低	1回のフォワードプロセスで生成可能
データ生成の コントロール	 容易	潜在的な表現あり
モデルの 定量的評価	 下方限界	最適化対象を陽に表現
ネットワーク 構造制限	 弱	様々な拡張が可能

A. B. L. Larsen et al., "Autoencoding beyond Pixels Using a Learned Similarity Metric," ICML 2016.

深層生成モデル



代表的な深層生成モデル (AR, Flow, VAE, GAN)

$p_g(x)$ の表現方法

陽に表現

陰に表現

近似なし

近似あり

AR

Autoregressive Model

連鎖律 (Chain Rule)

$$p_g(x) = \prod_{i=1}^n p_g(x_i|x_1, \dots, x_{i-1})$$

Flow

Flow-based Model

可逆生成

$$z = f(x) \quad x = f^{-1}(z)$$

VAE

Variational Autoencoder

変分近似

$$\begin{aligned}\log p_g(x) &\geq \mathcal{L}(\theta, \phi; x) \\ &= -KL(q_\phi(z|x)\|p(z)) \\ &\quad + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]\end{aligned}$$

GAN

Generative Adversarial Networks

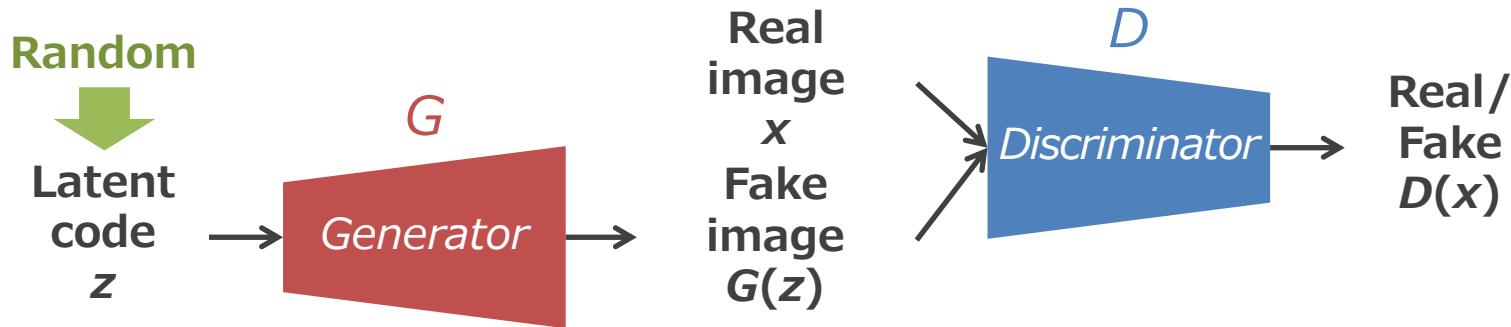
敵対的学習

$$\min_G \max_D \mathbb{E}_{x \sim p_r(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

I. Goodfellow., "NIPS 2016 Tutorial: Generative Aversarial Networks," NIPS 2016.

Generative Adversarial Networks [Goodfellow+2014]

- 生成器 (Generator) と識別器 (Discriminator) を敵対的に学習



- Min-Max最適化 (Generator VS Discriminator)

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

※ 詳細は次ページから

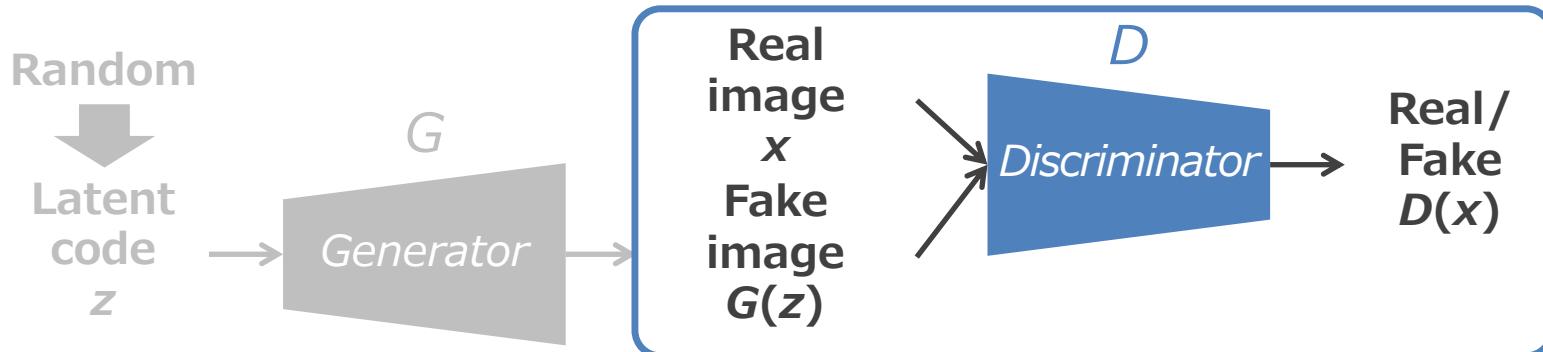
I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GAN (Discriminatorの最適化)



Generative Adversarial Networks [Goodfellow+2014]

- 生成器 (Generator) と識別器 (Discriminator) を敵対的に学習



Real/Fakeの識別

- Min-Max最適化 (Generator VS Discriminator)

Real/Fakeの二値識別

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

Generatorに騙されないように最適化

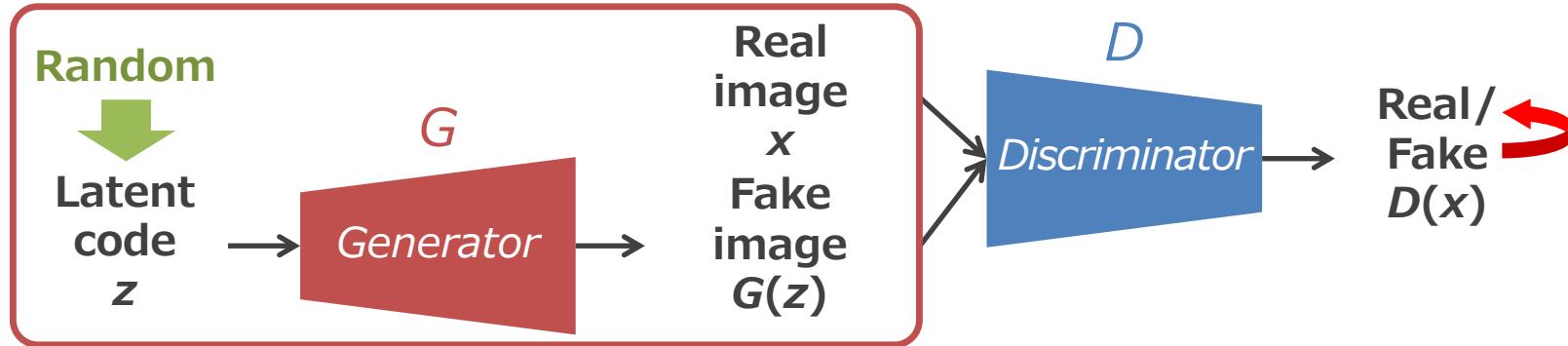
I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GAN (Generatorの最適化)



Generative Adversarial Networks [Goodfellow+2014]

- 生成器 (Generator) と識別器 (Discriminator) を敵対的に学習



乱数から画像生成

- Min-Max最適化 (Generator VS Discriminator)

Fakeでなくする

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

Discriminatorを騙せるように最適化

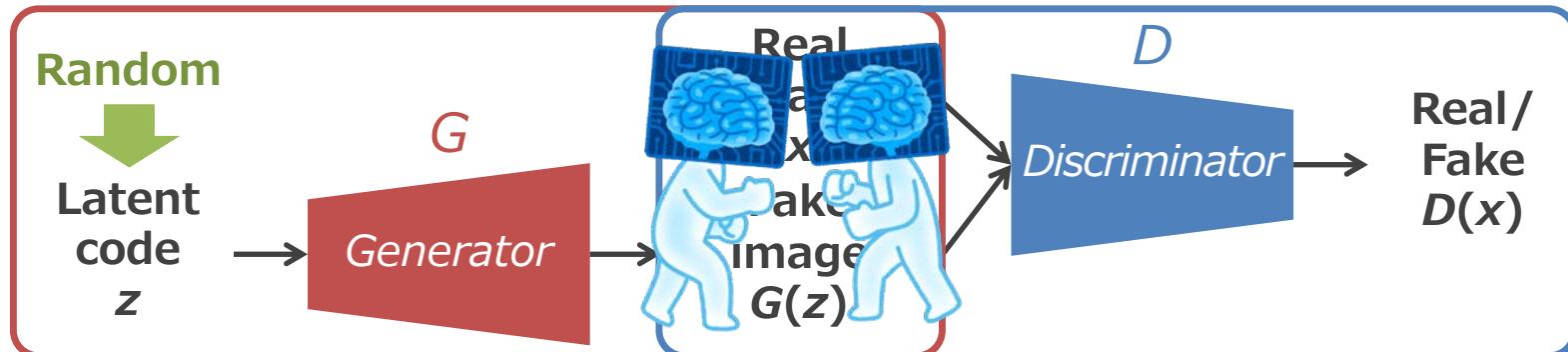
I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GAN (最適化のまとめ)



Generative Adversarial Networks [Goodfellow+2014]

- 生成器 (Generator) と識別器 (Discriminator) を敵対的に学習



乱数から画像生成

Real/Fakeの識別

- Min-Max最適化 (Generator VS Discriminator)

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

GeneratorとDiscriminatorをMin-Maxで競争させながら最適化

I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GANの特徴 (長所)



モデル	GAN 敵対的学習
サンプル クオリティ	😊 高
サンプリング コスト	😊 低
データ生成の コントロール	😊 容易
モデルの 定量的評価	😢 困難
ネットワーク 構造制限	😊 弱



例 : BigGAN, StyleGAN
[Brock+2019] [Karras+2019]

1回のフォワードプロセスで生成可能

潜在的な表現あり

Min-Maxのため明確な基準なし

様々な拡張が可能

A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.

T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

GANの特徴 (短所)



モデル	GAN 敵対的学習
サンプル クオリティ	😊 高
サンプリング コスト	😊 低
データ生成の コントロール	😊 容易
モデルの 定量的評価	😢 困難
ネットワーク 構造制限	😊 弱



例 : BigGAN, StyleGAN
[Brock+2019] [Karras+2019]

1回のフォワードプロセスで生成可能

潜在的な表現あり

Min-Maxのため明確な基準なし

様々な拡張が可能

A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.

T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

AR, Flow, VAE, GANの特徴まとめ



	AR	Flow	VAE	GAN
モデル	連鎖律	可逆生成	変分推論	敵対的学習
サンプル クオリティ	高	高	低	高
サンプリング コスト	高	低	低	低
データ生成の コントロール	困難	容易	容易	容易
モデルの 定量的評価	対数尤度	対数尤度	下方限界	困難
ネットワーク 構造制限	強	強	弱	弱

※どのモデルも近年研究が盛んであり、各短所を克服しようとする試みあり

GANの位置付け

- 生成モデルとは？
- 深層生成モデル
- AR, Flow, VAE, GANの比較

GANの理論

- 目的関数
- 更新アルゴリズム
- 理論的解釈

GANの目的関数



Min-Max最適化

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

Realの確率

Realでない確率

実画像のサンプリング

乱数のサンプリング

生成画像

実画像に関する項

生成画像に関する項

I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GANの更新アルゴリズム



交互更新アルゴリズム

- Step 1. G を固定して D を勾配法で更新

$$\max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

- Step 2. D を固定して G を勾配法で更新

理論 $\min_G \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$

下限が制限されて
いないため学習困難

実際 $\max_G \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log D(G(\mathbf{z}))]$

上限は制限 (0)

Non-Saturating Loss

I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

GANの理論的解釈



Generatorを固定したときの最適なDiscriminator

$$D_G^*(\mathbf{x}) = \frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})}$$

Real/Fakeの
密度比推定

最適なDiscriminator下の目的関数

$$\begin{aligned} & \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D_G^*(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D_G^*(G(\mathbf{z})))] \\ &= -\log(4) + KL \left(p_r(\mathbf{x}) \middle\| \frac{p_r(\mathbf{x}) + p_g(\mathbf{x})}{2} \right) + KL \left(p_g(\mathbf{x}) \middle\| \frac{p_r(\mathbf{x}) + p_g(\mathbf{x})}{2} \right) \\ &= -\log(4) + 2 \cdot JSD(p_r(\mathbf{x}) \| p_g(\mathbf{x})) \end{aligned}$$

Jensen-Shannon Divergenceで
 $p_r(\mathbf{x})$ と $p_g(\mathbf{x})$ を近づける

※ 収束性も保証

- ただし、この理論では以下を想定
 - GeneratorとDiscriminatorが十分な表現力を持つ
 - データサイズが十分大きい

実際には困難

I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

目次



**GAN
の基礎**

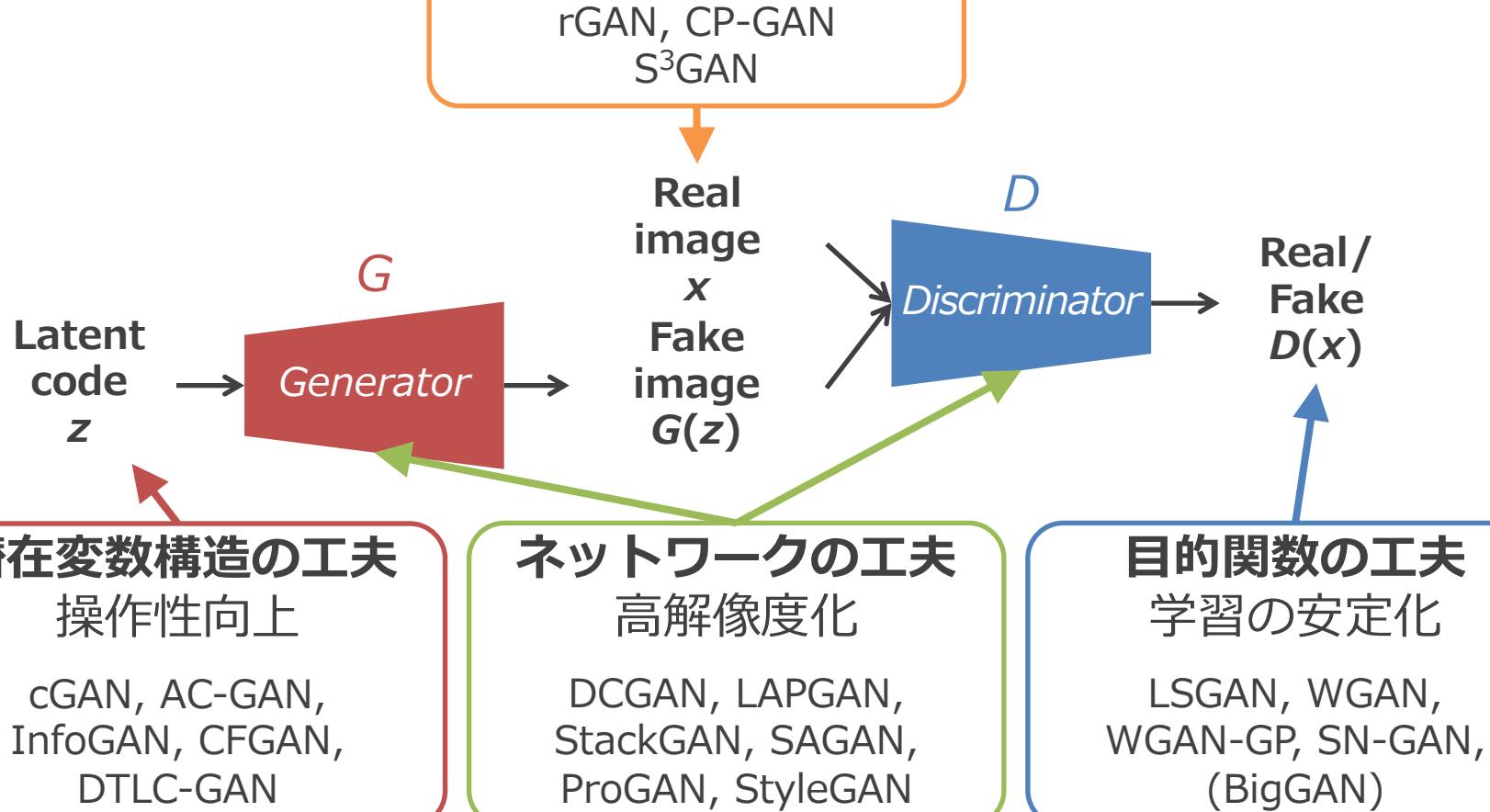
**GAN
の発展**

**GAN
の応用**

GANの発展



4つの発展

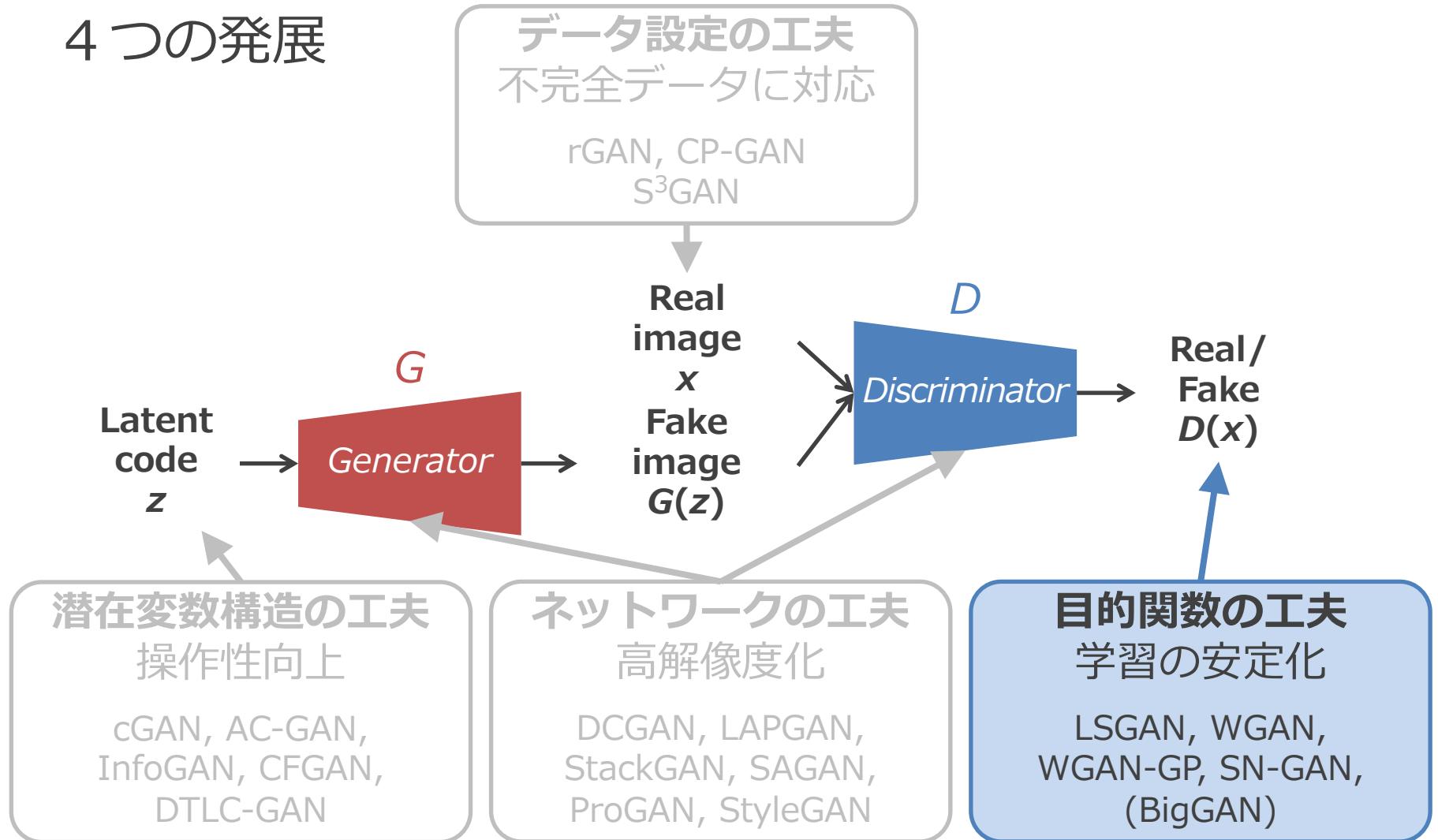


※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

GANの発展



4つの発展



※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

NaïveなGANの課題1



Min-Max最適化のため学習が不安定

- 理想的な条件下ではGANは収束 [Goodfellow+2014]
 - GeneratorとDiscriminatorが十分な表現力を持つ
 - データサイズが十分大きい
 - しかし、実際には以下の条件
 - パラメタライズされたGeneratorとDiscriminator
 - データサイズは限定的
- 学習は不安定化
- Oscillation (振動): 長く学習しても良くなるとは限らない
 - Mode Collapse: 次スライド

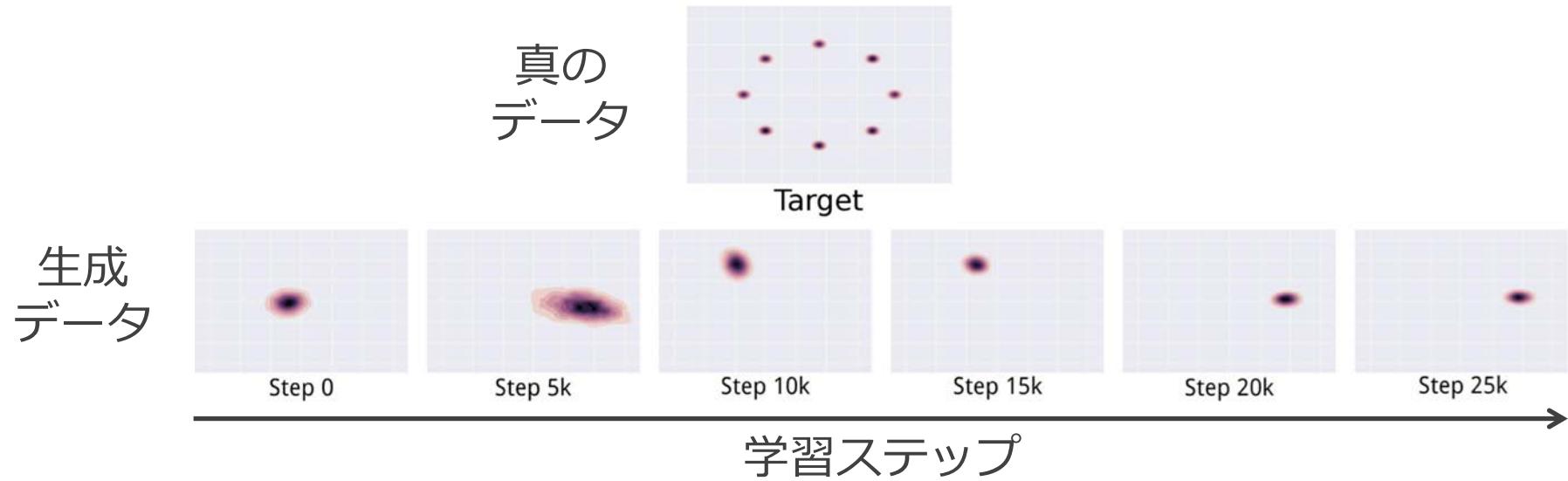
I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014.

学習失敗時によく見られる現象



Mode Collapse [Metz+2017]

- 全サンプルが特定箇所に集中



- Discriminator: Real/Fakeの識別可
- Generator: Realデータに近づける

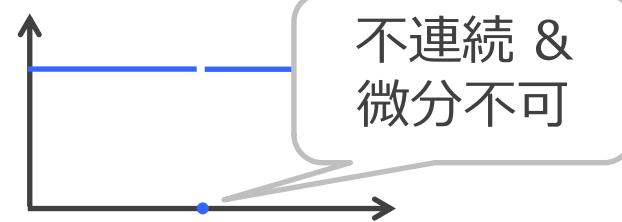
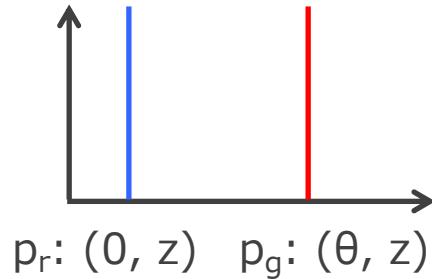
NaïveなGANでは
良し悪しの判断不可

RealとFakeの分布間距離がうまく測れていない？

L. Metz et al., "Unrolled Generative Adversarial Networks," ICLR 2017.

NaïveなGAN

- Jensen-Shannon Divergence (JSD) を利用
→ 二つの分布に重なりがない時に不連続 & 微分不可 [Arjovsky+2017]



$$JSD(p_r, p_g) = \begin{cases} \log 2 & \text{if } \theta \neq 0 \\ 0 & \text{if } \theta = 0 \end{cases}$$

改善案

- 距離基準を変えてみてはどうか?
 - LSGAN: Chi-square Divergenceを利用
 - WGAN: Wasserstein (Earth Mover) Distanceを利用

M. Arjovsky et al., "Wasserstein Generative Adversarial Networks," ICML 2017.

Least Squares GAN [Mao+2017]

- Sigmoid Cross Entropyの代わりにLeast Squares Lossを利用

Discriminator

$$\text{GAN} \quad \min_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})}[-\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[-\log(1 - D(G(\mathbf{z})))]$$



$$\text{LSGAN} \quad \min_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})}[(D(\mathbf{x}) - 1)^2] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[D(G(\mathbf{z}))^2]$$

Realは1に近づける Fakeは0に近づける

Generator

$$\text{GAN} \quad \min_G \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[-\log D(G(\mathbf{z}))]$$



$$\text{LSGAN} \quad \min_G \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[(D(G(\mathbf{z})) - 1)^2]$$

Fakeを1 (Real) に近づける

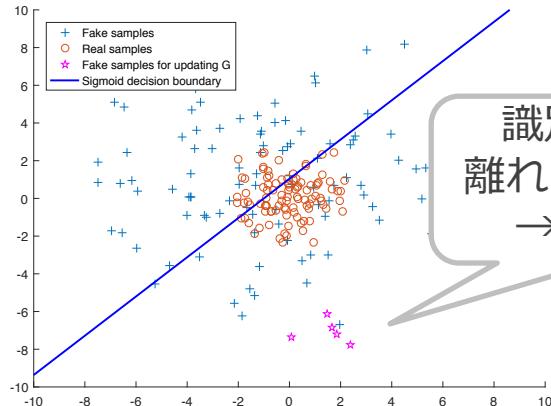
※ LSGANは $p_r(\mathbf{x}) + p_g(\mathbf{x})$ と $2p_g(\mathbf{x})$ のChi-square Distanceを最小化

X. Mao et al., "Least Squares Generative Adversarial Networks," ICCV 2017.

LSGAN (cont.)



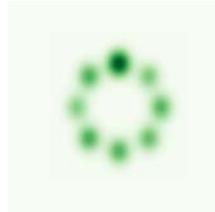
直感的解釈



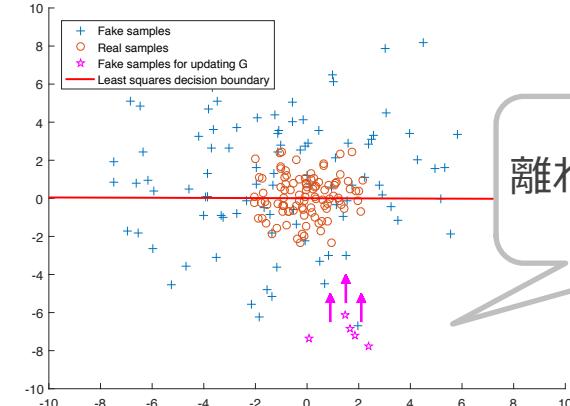
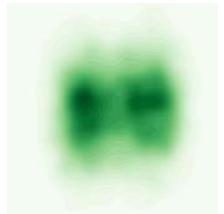
GAN: Sigmoid Cross Entropy

実験結果

真の
データ

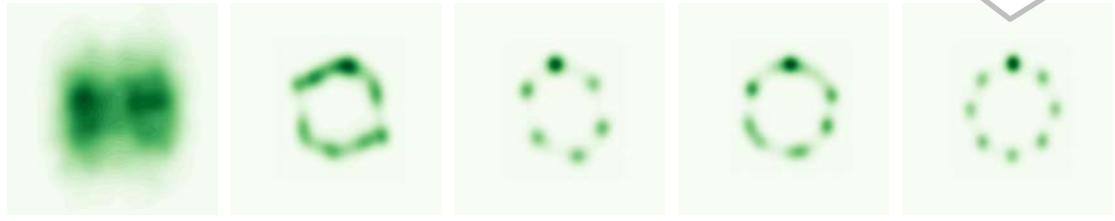


生成
データ



LSGAN: Least Squares Loss

Mode Collapseを回避



学習ステップ

X. Mao et al., "Least Squares Generative Adversarial Networks," ICCV 2017.

Wasserstein GAN [Arjovsky+2017]

- Classifier関数の代わりにCritic関数を利用

Discriminator (Critic)

$$\text{GAN} \quad \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$



$$\text{WGAN} \quad \max_D \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} [D(\mathbf{x})] - \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [D(G(\mathbf{z}))]$$

RealとFakeの距離をなるべく正確に測る

Generator

$$\text{GAN} \quad \max_G \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log D(G(\mathbf{z}))]$$



$$\text{WGAN} \quad \max_G \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [D(G(\mathbf{z}))]$$

距離を最小化

DはLipschitz制約を満たすことが必要
→ Weight Clipping

※ WGANは $p_g(\mathbf{x})$ と $p_r(\mathbf{x})$ のEarth Mover Distanceを最小化

M. Arjovsky et al., "Wasserstein Generative Adversarial Networks," ICML 2017.

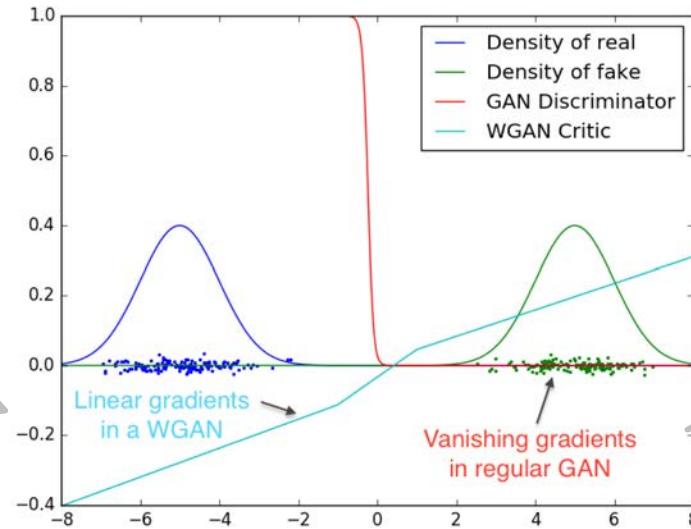
WGAN (cont.)



直感的解釈

WGAN:
線形の勾配

Wasserstein距離は
緩い制約のもと
連続・微分可能と
いう良い性質を持つ



GAN:
勾配消失

実験結果

生成
データ



Mode Collapseを回避

M. Arjovsky et al., "Wasserstein Generative Adversarial Networks," ICML 2017.



WGANの課題

- D にLipschitz制約を課すためWeight Clipping (きつい制約)

$$1\text{-Lipschitz } \|f(\mathbf{x}) - f(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y}$$

→ モデルを過度に簡易化

Weight Clippingにより f の範囲を制約

WGAN Gradient Penalty [Gulrajani+2017]

- 背景 : f が連続で微分可能 $\rightarrow f$ が1-Lipschitzになる必要十分条件は

$$\|\nabla f\| \leq 1 \text{ everywhere}$$

- 提案 : Gradient Penalty (GP) (緩い制約)

$$\lambda \mathbb{E}_{\hat{\mathbf{x}} \sim p_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})} [(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2]$$

RealとFakeの間でランダムにサンプリング \rightarrow “everywhere”をカバー

- 财献 : 様々なモデル (FC, Conv, ResNet) でも学習可能に

I. Gulrajani et al., "Improved Training of Wasserstein GANs," NIPS 2017.

Spectral Normalization GAN [Miyato+2018]

- 背景：ニューラルネットワークは重み W と活性化関数 a の組合せ各層がLipschitz $\rightarrow f$ もLipschitz

$$f(\mathbf{x}) = W^{L+1}(a_L(W^L(a_{L-1}(W^{L-1}(\dots a_1(W^1\mathbf{x})\dots)))))$$

活性化関数は元からLipschitz

W をLipschitzにするには？

- 提案：Spectral Normalization (SN)

$$W_{\text{SN}} = W/\sigma(W)$$

$\sigma(W)$ ：最大特異値

- 貢献：複雑なデータ (ImageNet) でも学習可能に



T. Miyato et al., "Spectral Normalization for Generative Adversarial Networks," ICLR 2018.

Large Scale GAN [Brock+2019]

- 計算機パワーを活かして、
バッチサイズ（一度の学習で見るデータ量）
とチャンネルサイズ（モデルの表現能力）
を上げて学習
→ 学習が安定化し、より高精細な画像の生成が可能に



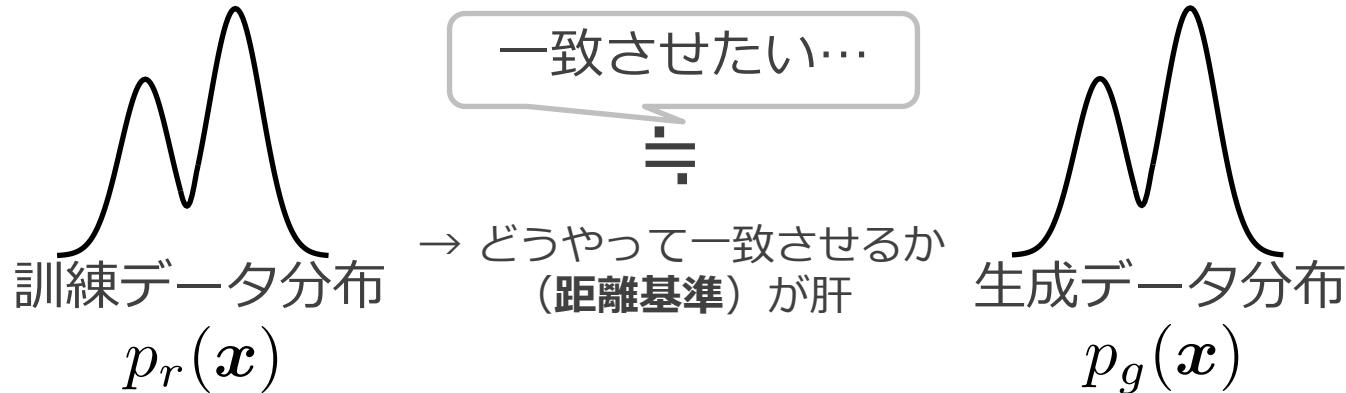
A. Brock et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," ICLR 2019.

目的関数のまとめ

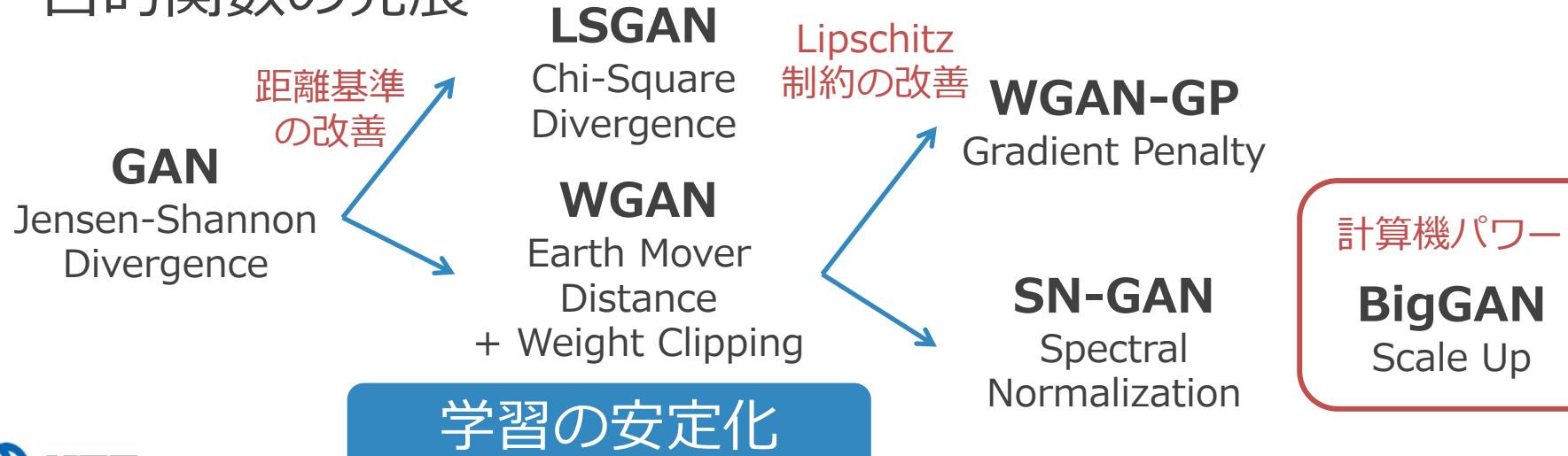


NaïveなGANの課題

- Min-Max最適化のため学習が不安定



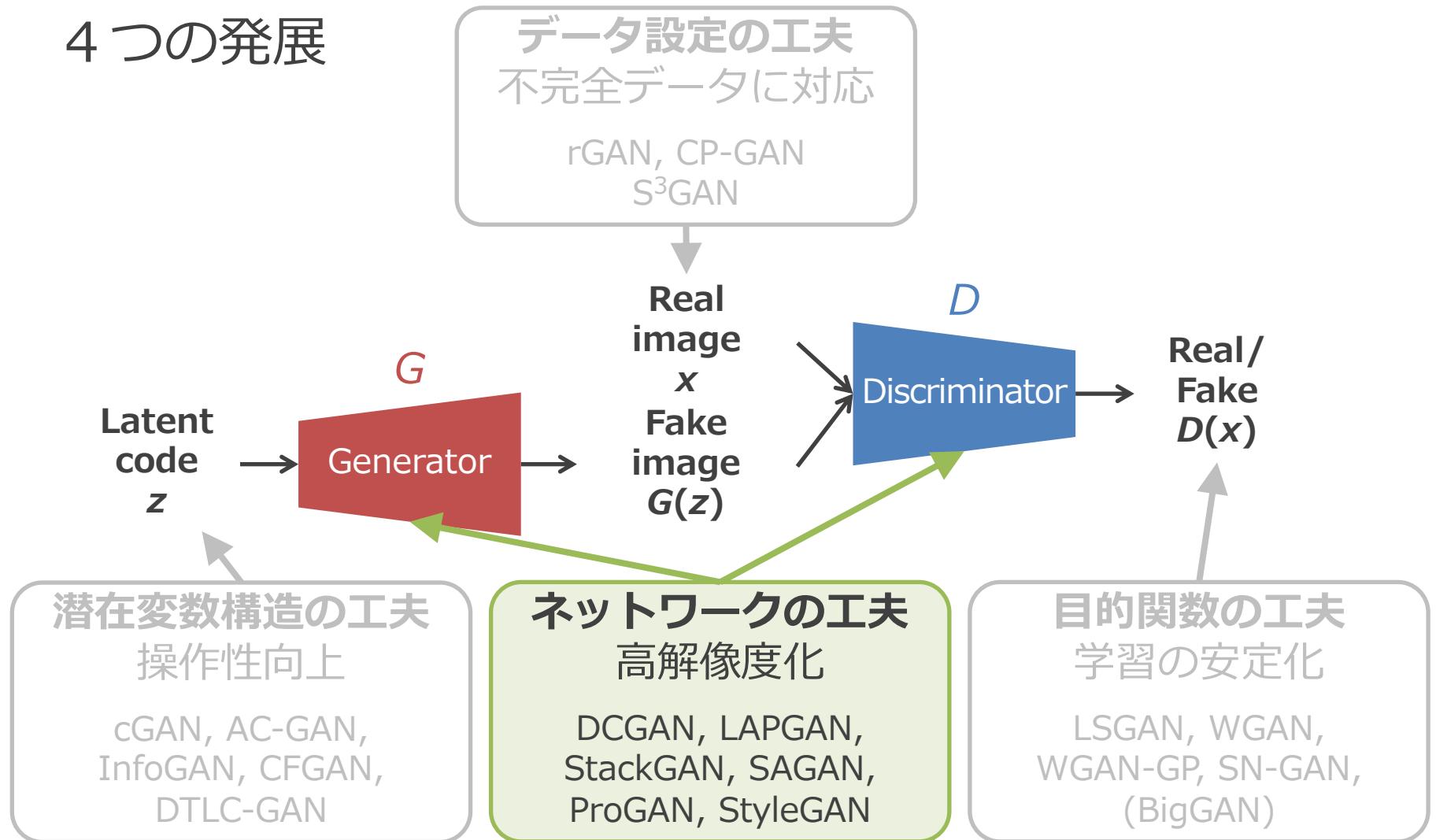
目的関数の発展



GANの発展



4つの発展



※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

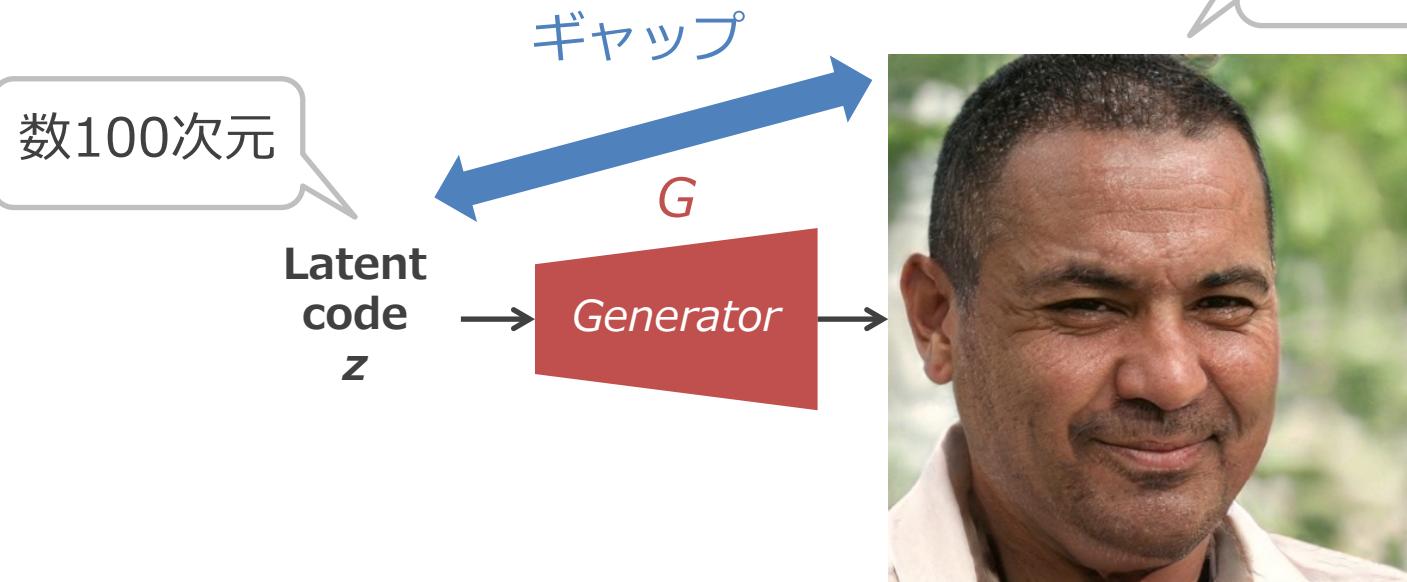
NaïveなGANの課題2



入力と出力のギャップ

- 非常に複雑な関数を学習することが必要
 - 特に、解像度が高くなればなるほどギャップ大

1024x1024の場合
約100万次元

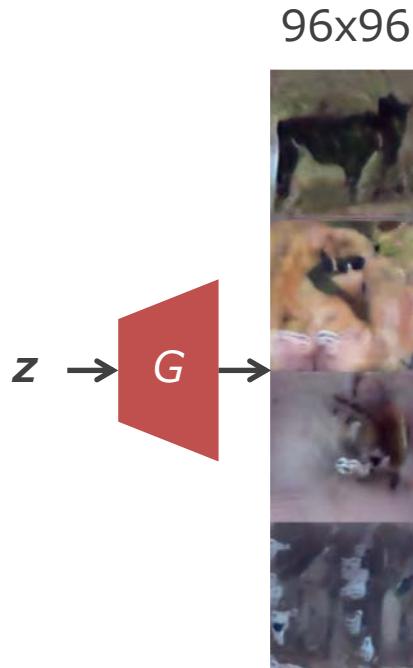


生成しやすいネットワーク構造にすることが重要

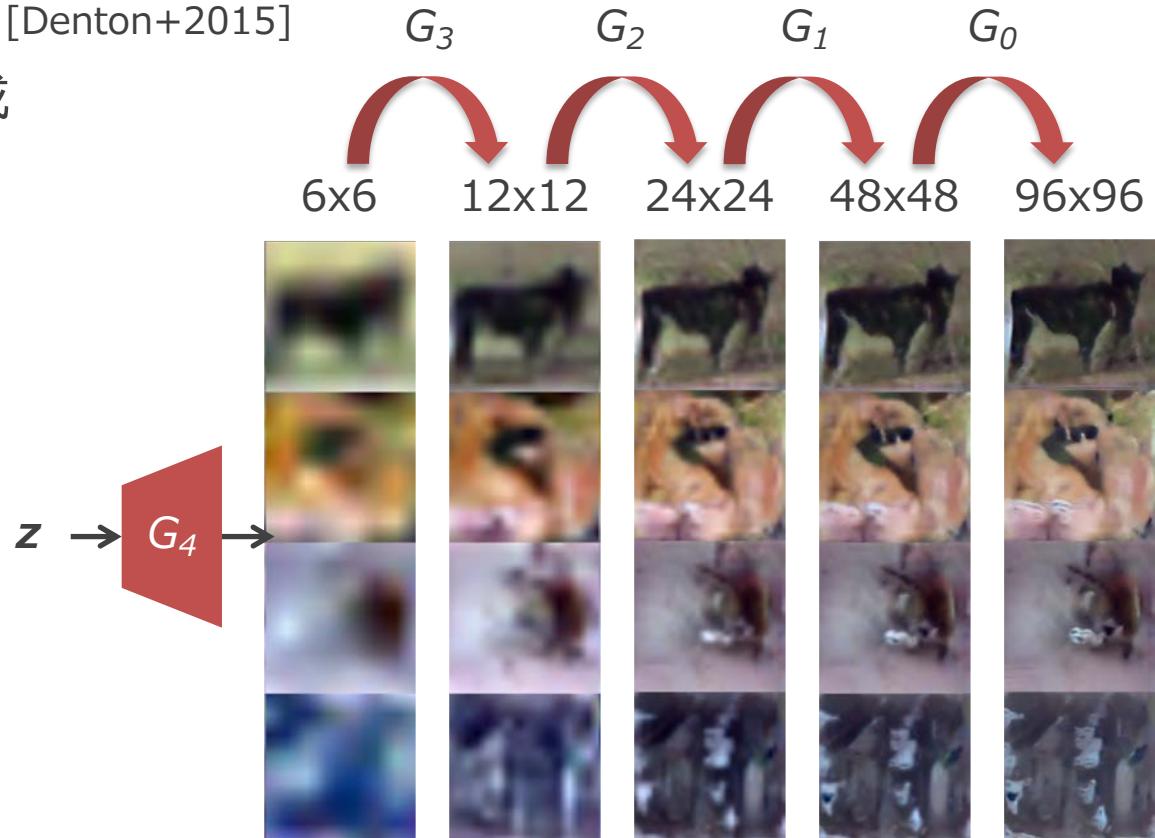
Pictures: T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

Laplacian GAN [Denton+2015]

- 段階的に画像生成



NaïveなGAN
 G : 複雑な関数が必要



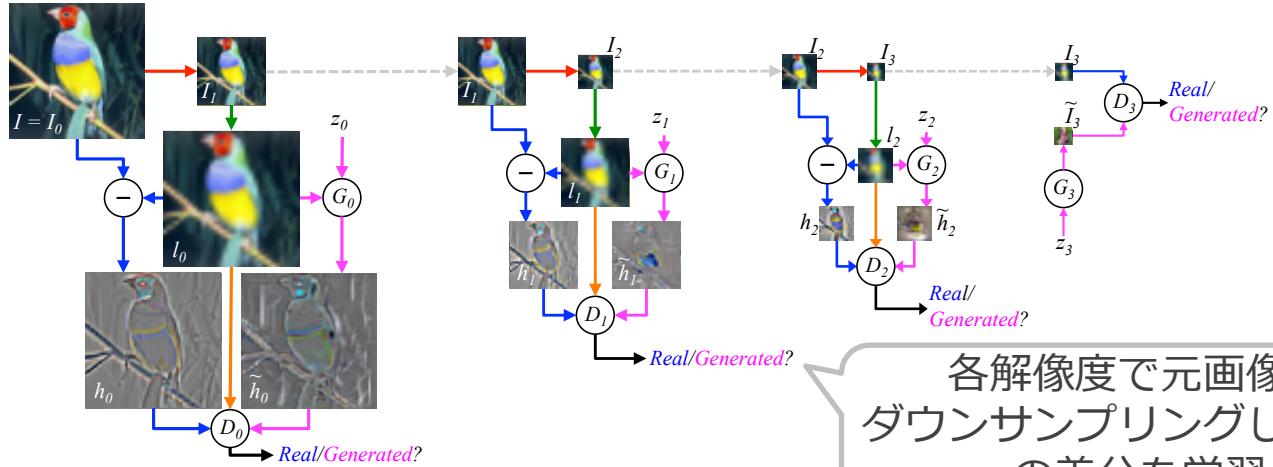
LAPGAN (問題を分割)
 G_0, \dots, G_4 : 比較的簡易な関数でよい

E. Denton et al., "Deep Generative Image Modeling using a Laplacian Pyramid of Adversarial Networks," NIPS 2015.

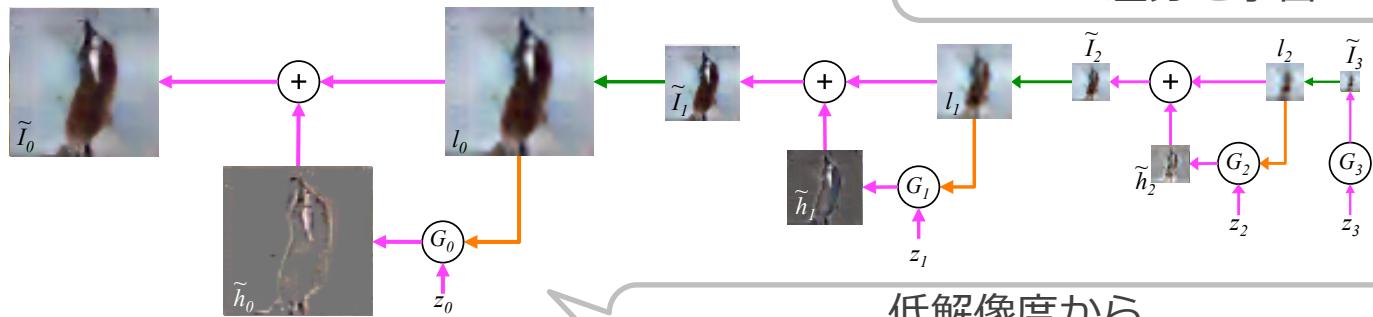
LAPGAN (cont.)



学習 プロセス



生成 プロセス



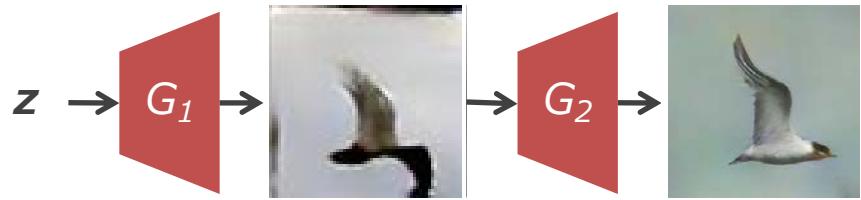
低解像度から
「アップサンプリングして差分を追加」
というプロセスを繰り返す

※ Virtual Tuning Test (生成画像が本物かどうかの主観評価) で4割の画像で騙すことに成功

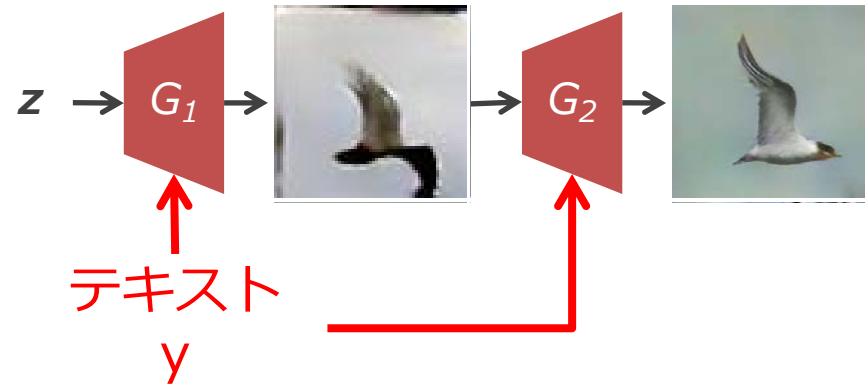
E. Denton et al., "Deep Generative Image Modeling using a Laplacian Pyramid of Adversarial Networks," NIPS 2015.

Stacked GAN [Zhang+2017]

- 補助情報（テキスト）を活用しながら段階的に画像生成



補助情報なし
 $p(x)$ を学習
複雑な分布



補助情報あり (StackGAN)
 $p(x|y)$ を学習
学習対象を限定可能
→比較的簡易な分布

StackGAN (cont.)



画像生成例

StackGAN
Stage-I
64x64

This bird is white with some black on its head and wings, and has a long orange beak



This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face



This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments



StackGAN
Stage-II
256x256



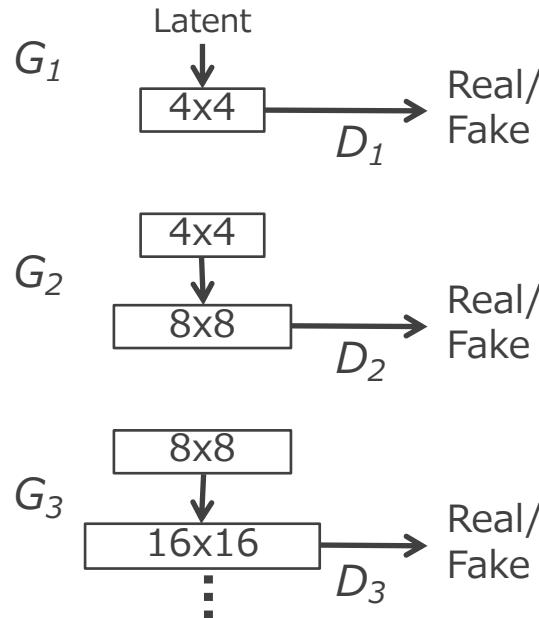
Vanilla GAN
256x256



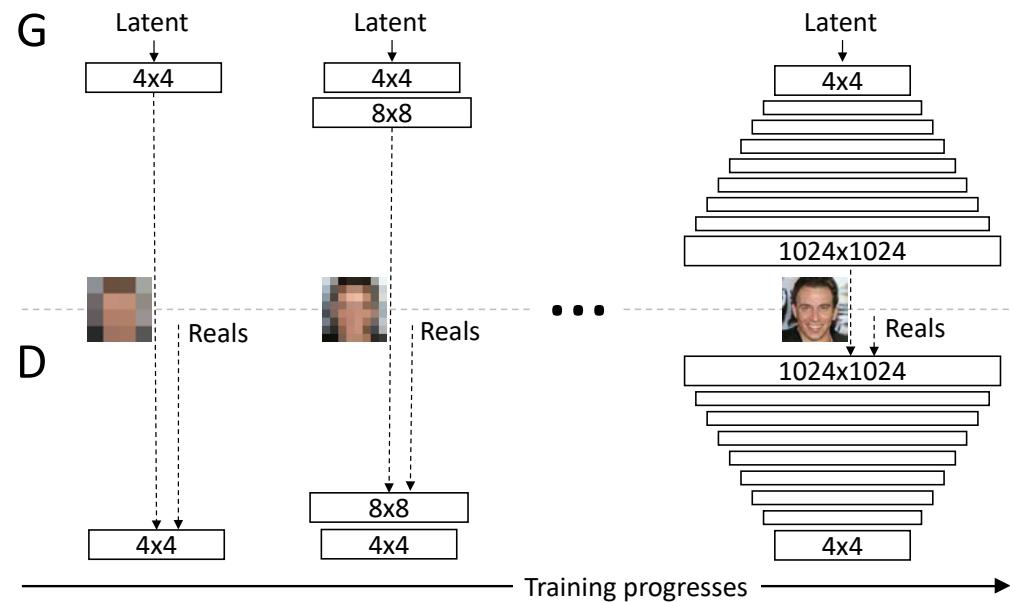
H. Zhang et al., "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks," ICCV 2017.

Progressive Growing of GANs [Karras+2018]

- 単一のネットワーク内で、層を追加しながら高解像度化



LAPGAN, StackGAN
各解像度ごとにGとDが必要
→ パラメータ、学習時間が大



ProGAN
単一のGとD
→ パラメータ、学習時間が小

T. Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation," ICLR 2018.

ProGAN (cont.)



画像生成例



1024x1024の高解像度画像の生成が可能

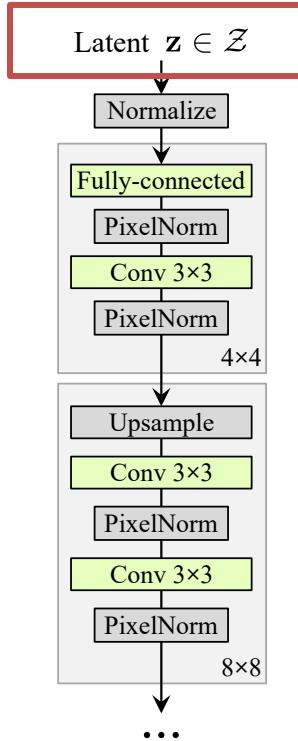
T. Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation," ICLR 2018.

StyleGAN

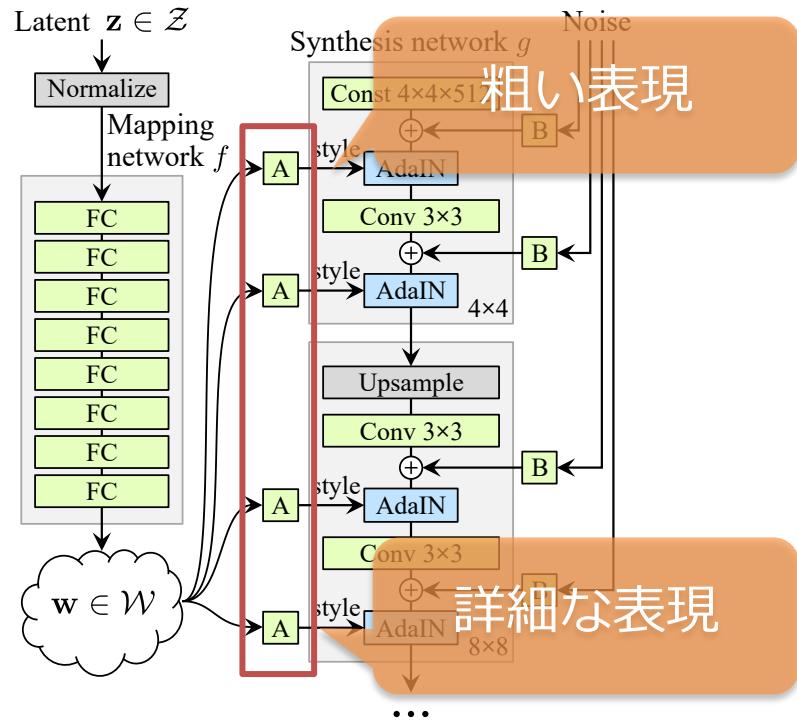


Style-Based Generator for GANs [Karras+2019]

- Generatorの入力を多段階化



最初の入力で全ての表現が必要
→ 学習困難、制御に限界



各層で表現を分担
→ 学習緩和、多段階で制御可

T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

StyleGAN (cont.)



画像生成例



さらに高精細な画像の生成が可能

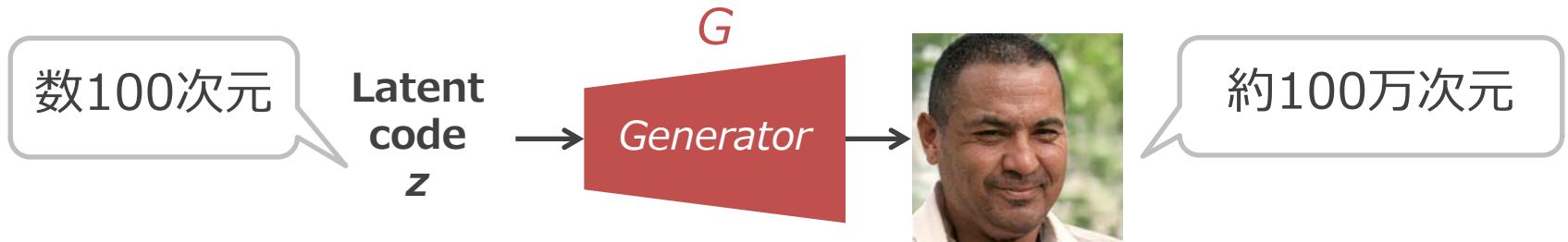
T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR 2019.

ネットワーク構造のまとめ



NaïveなGANの課題

- 入力と出力のギャップ（特に、高解像度化したいときに課題）



ネットワーク構造の発展

LAPGAN

段階的に生成 →

96x96



StackGAN

補助情報も利用 →

256x256



ProGAN

層を追加しながら生成 →

1024x1024



StyleGAN

入力を多段階化 →

1024x1024

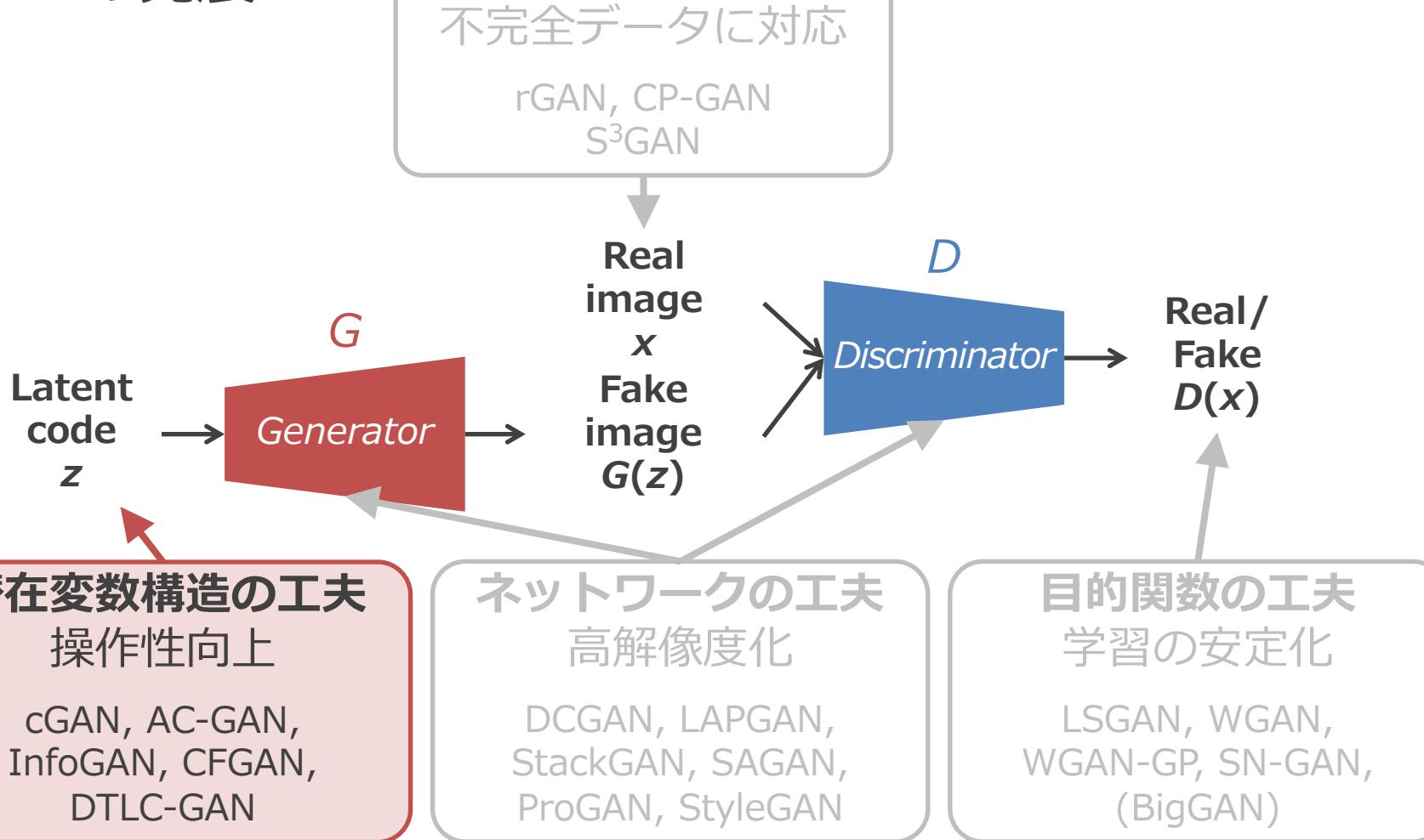


高解像度化

GANの発展



4つの発展



※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

NaïveなGANの課題3



画像生成のコントロールが困難



入力は乱数 → 表現は混在
(例：個人性、属性は混在)

画像生成例

?

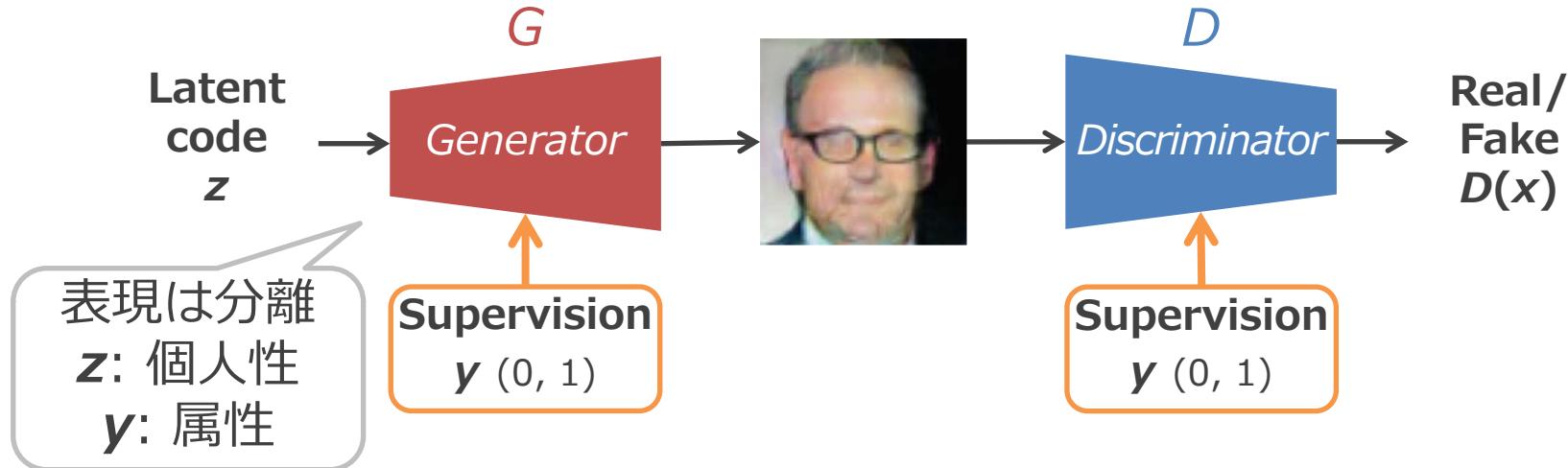


どんな画像が生成されるかはランダム

Pictures: T. Kaneko et al., "Generative Adversarial Image Synthesis with Decision Tree Latent Controller," CVPR 2018.
T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.

Conditional GAN [Mirza+2014]

- 属性の有無 (例 : 笑顔かどうか) を表す教師情報 y を導入
→ ON/OFFの制御

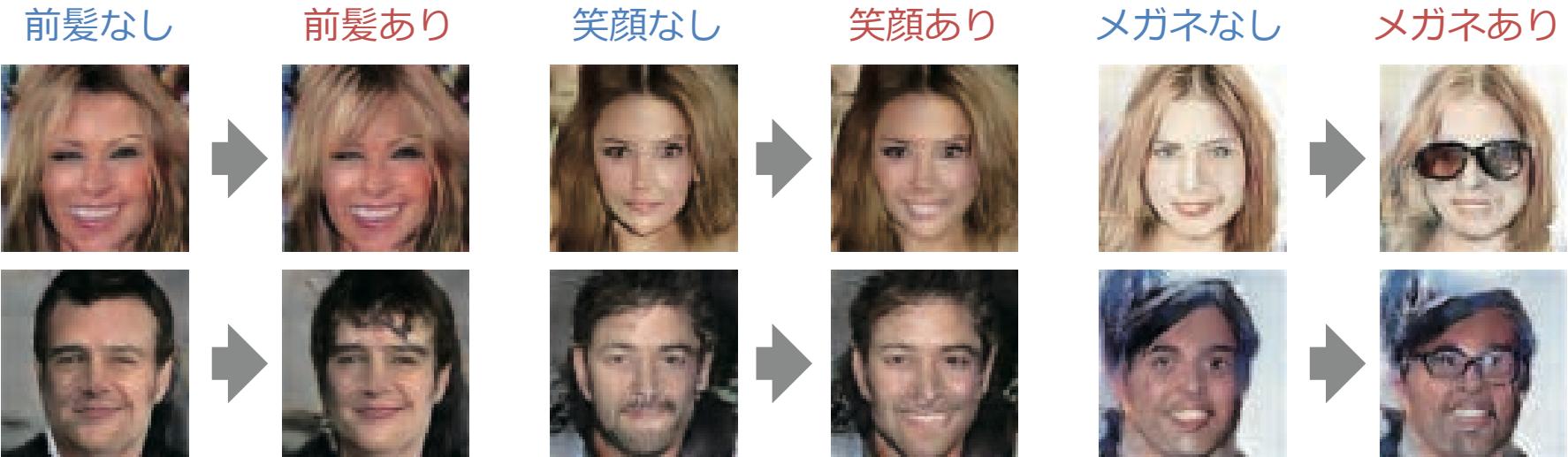


M. Mirza & S. Osindero, "Conditional Generative Adversarial Nets," arXiv 2014.

cGAN (cont.)



画像生成例

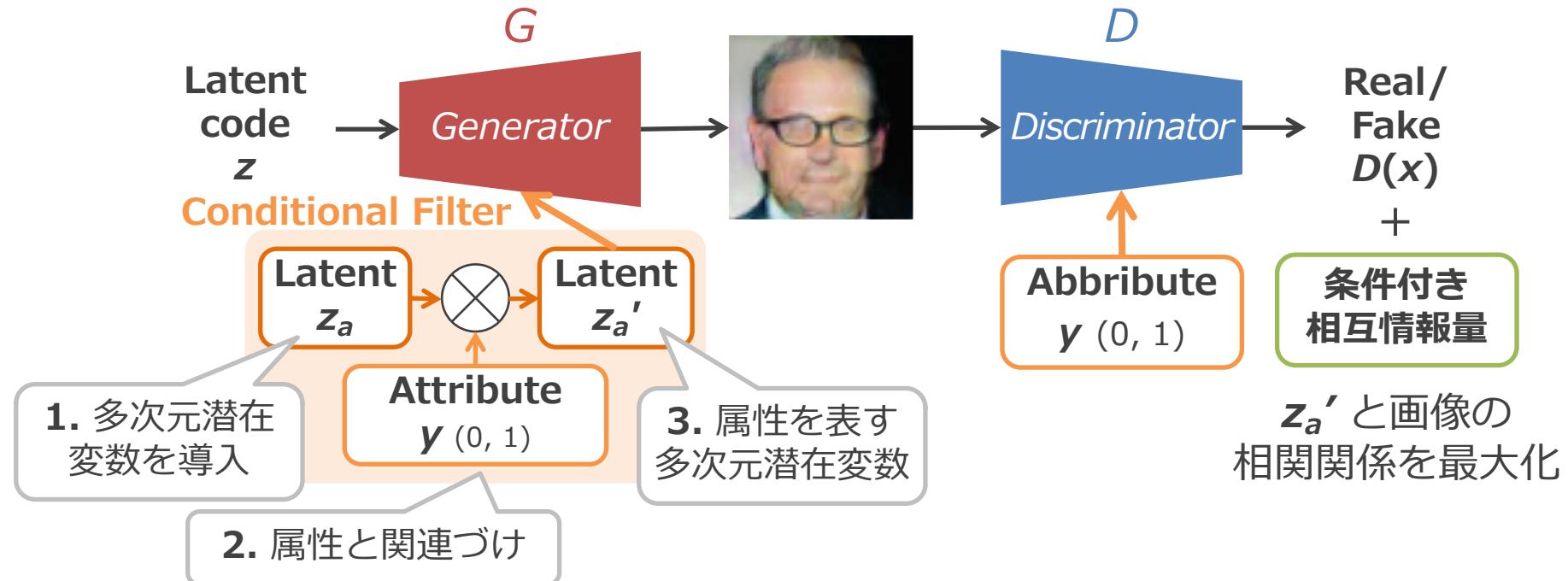


教師情報に基づく制御 (属性のON/OFF) に限定

T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.

Conditional Filtered GAN [Kaneko+2017]

1. 新たに多次元潜在変数 z_a を導入
2. 教師情報 y でフィルタリングすることで属性と関連づけ
3. 属性を表す対次元潜在変数 z_a' を構成 → 多次元的に制御

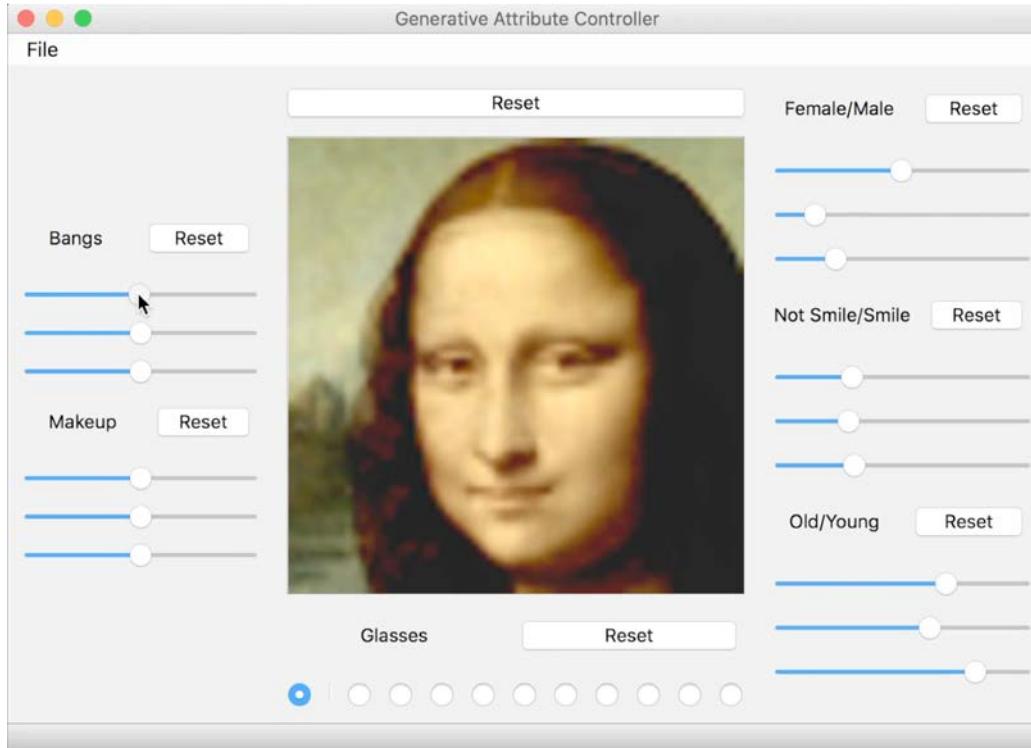


T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.

CFGAN (cont.)



画像生成例



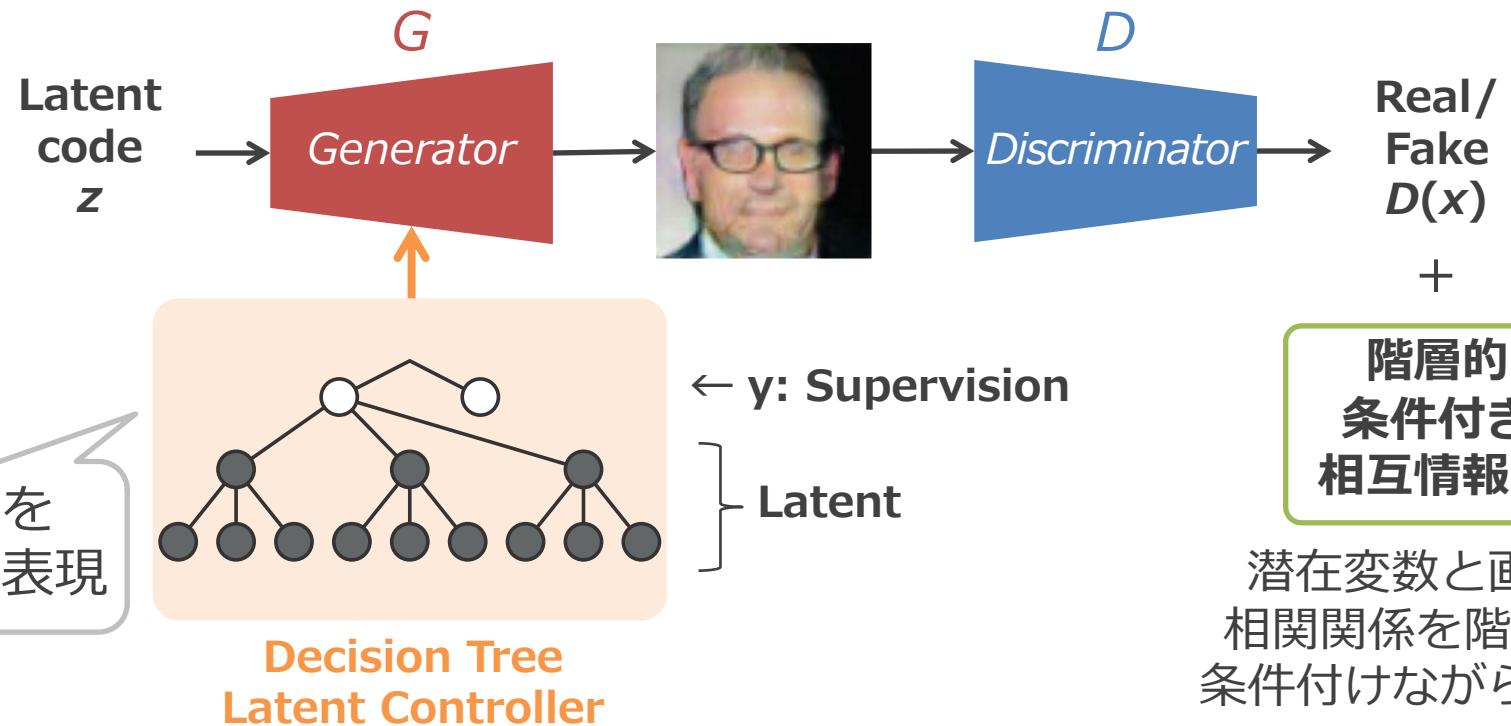
デモ動画：<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/gac/index.html>

多次元的に制御可能

T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.

Decision Tree Latent Controller GAN [Kaneko+2018]

- 階層的なサンプリング構造を導入 → 階層的に制御



T. Kaneko et al., "Generative Adversarial Image Synthesis with Decision Tree Latent Controller," CVPR 2018.

DTLC-GAN (cont.)



画像生成例

メガネの
有無



ラフな
カテゴリ



詳細な
カテゴリ



※ 教師情報は最上位 (メガネの有無) のみ
詳細なカテゴリはデータから発見

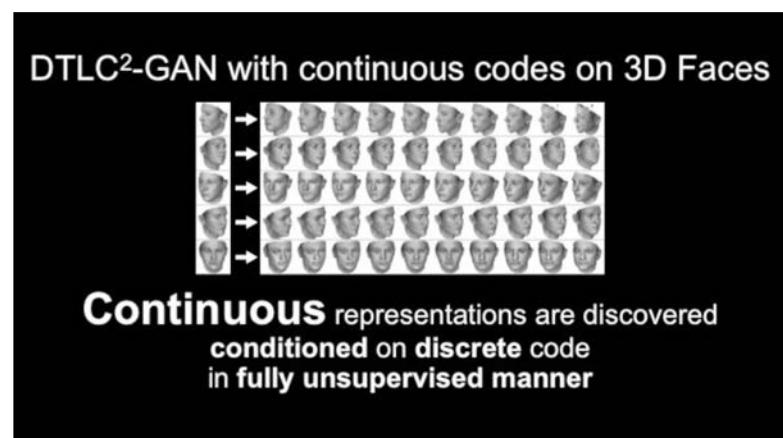
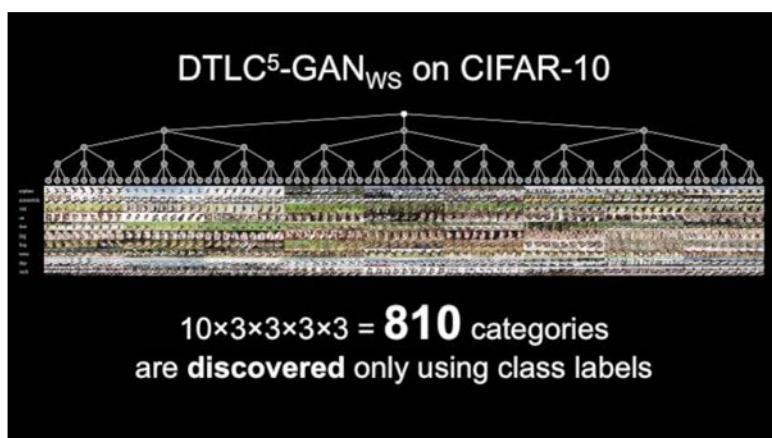
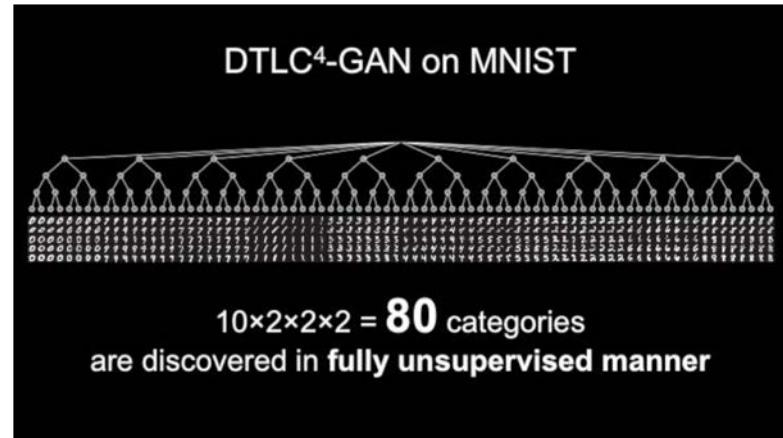
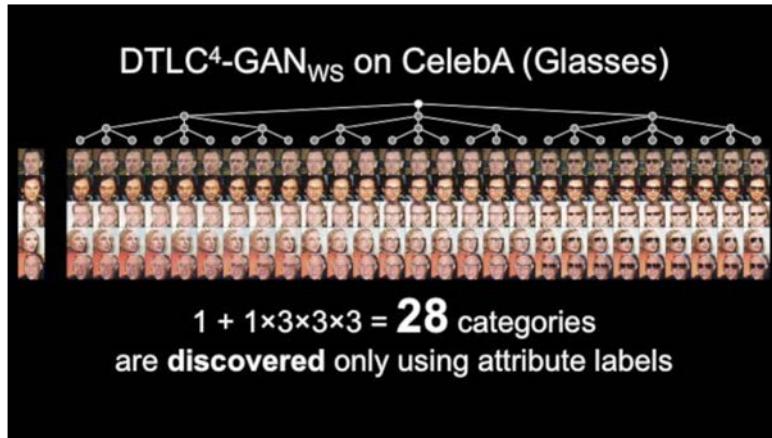
階層的に制御可能

T. Kaneko et al., "Generative Adversarial Image Synthesis with Decision Tree Latent Controller," CVPR 2018.

DTLC-GAN (cont.)



様々なデータに適用可能



デモ動画：<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/dtlc-gan/index.html>

T. Kaneko et al., "Generative Adversarial Image Synthesis with Decision Tree Latent Controller," CVPR 2018.

潜在変数構造のまとめ



NaïveなGANの課題

- 画像生成のコントロールが困難

潜在変数構造の発展

操作性向上

層数	0	1	2, 3, ...
教師なし	GAN [Goodfellow+2014] 分離なし	InfoGAN [Chen+2016] 一層の潜在変数	DTLC-GAN [Kaneko+2018]
(弱) 教師あり	cGAN [Mirza+2014] AC-GAN [Odena+2017] 教師情報に制約	CFGAN [Kaneko+2017] 	多階層の潜在変数

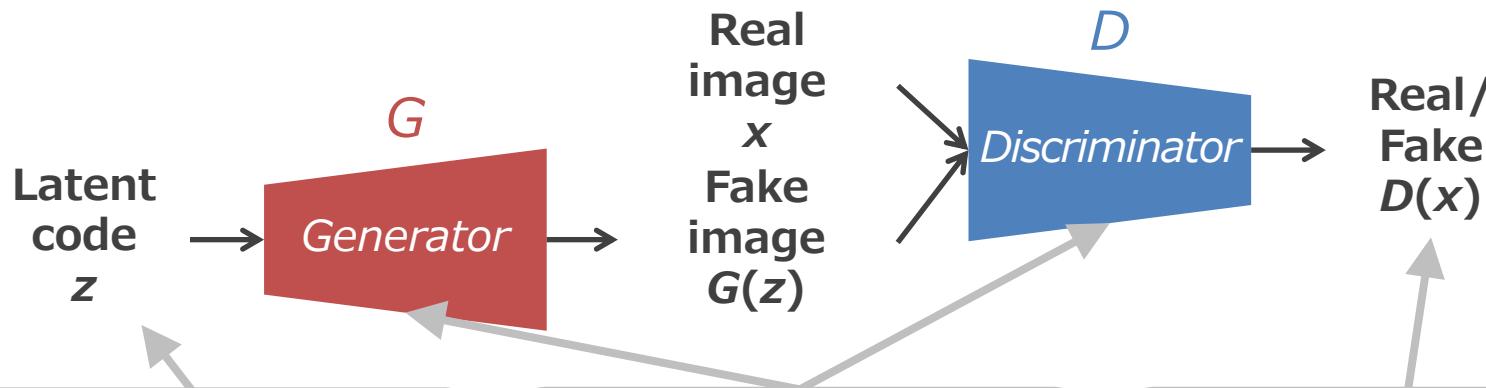
GANの発展



4つの発展

データ設定の工夫
不完全データに対応

rGAN, CP-GAN
S³GAN



潜在変数構造の工夫
操作性向上

cGAN, AC-GAN,
InfoGAN, CFGAN,
DTLC-GAN

ネットワークの工夫
高解像度化

DCGAN, LAPGAN,
StackGAN, SAGAN,
ProGAN, StyleGAN

目的関数の工夫
学習の安定化

LSGAN, WGAN,
WGAN-GP, SN-GAN,
(BigGAN)

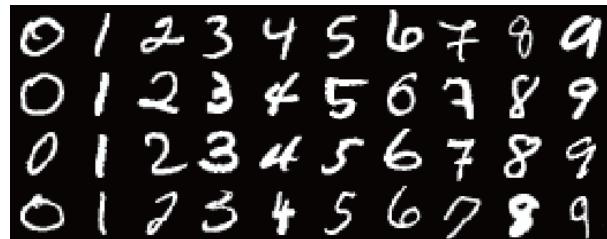
※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

NaïveなGANの課題4①



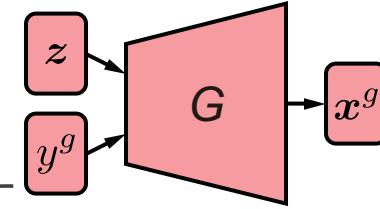
学習データは完全であることを仮定

例 1：ラベルがCleanであることを仮定



完全 (Clean)

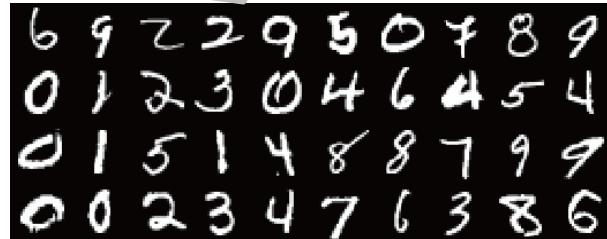
データ通りに
モデル化



Generator

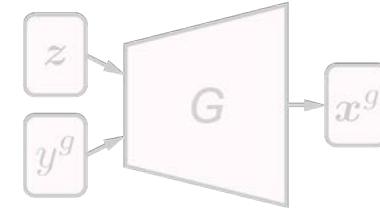
アノテーションミスなど

データの半数のラベルがフリップ



不完全 (Noisy)

どう学習？

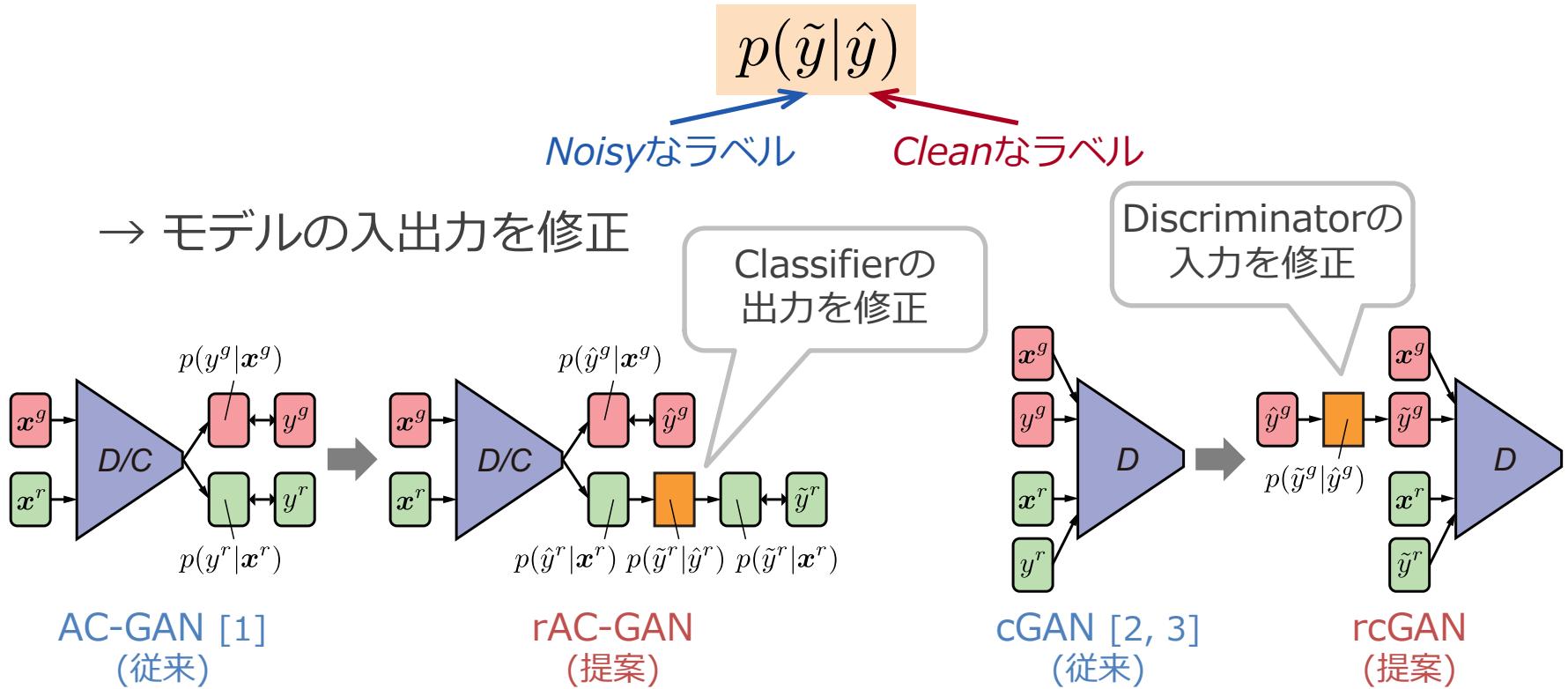


Generator

T. Kaneko et al., "Label-Noise Robust Generative Adversarial Networks," CVPR 2019.

Label-Noise Robust GAN [Kaneko+2019]

- Noise Transition Modelを従来のAC-GAN [1] とcGAN [2, 3] に導入



T. Kaneko et al., "Label-Noise Robust Generative Adversarial Networks," CVPR 2019.

[1] A. Odena et al., "Conditional Image Synthesis With Auxiliary Classifier GANs," ICML 2017.

[2] M. Mirza & S. Osindero, "Conditional Generative Adversarial Nets," arXiv 2014.

[3] T. Miyato & M. Koyama, "cGANs with Projection Discriminator," ICLR 2018.

rGAN (cont.)



画像生成例

6 9 2 2 9 5 0 7 8 9
0 1 2 3 0 4 6 4 5 4
0 1 5 1 4 8 8 7 9 9
0 0 2 3 4 7 6 3 8 6
2 1 6 3 4 0 6 7 8 9



学習データ
(Noisy)

Train

0 6 6 6 9 5 6 9 6 9
4 1 4 4 4 4 4 7 4 9
0 1 2 3 4 5 6 7 8 9
0 6 2 3 6 5 6 6 8 6
0 1 2 3 4 5 6 4 8 4



Noisyなラベル
に条件づけ

cGAN
(従来)

0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9



Cleanなラベル
に条件づけ

rcGAN
(提案)

コード : <https://github.com/takuhirok/rGAN/>

ラベルノイズに頑健な画像生成が可能

T. Kaneko et al., "Label-Noise Robust Generative Adversarial Networks," CVPR 2019.

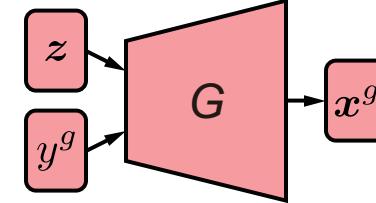
NaïveなGANの課題4②



学習データは完全であることを仮定

- 例2：Class-Separateであることを仮定

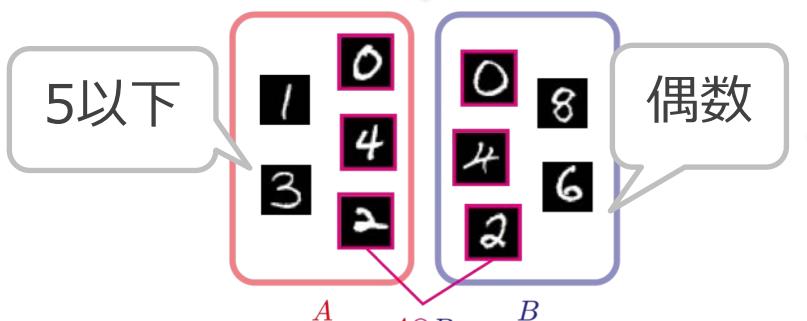
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9



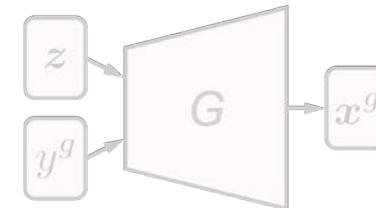
離散的な
完全 (Class-Separate) ラベルに基づき Generator
モデル化

ラベルの基準が曖昧

どう学習？



不完全 (Class-Overlapping)



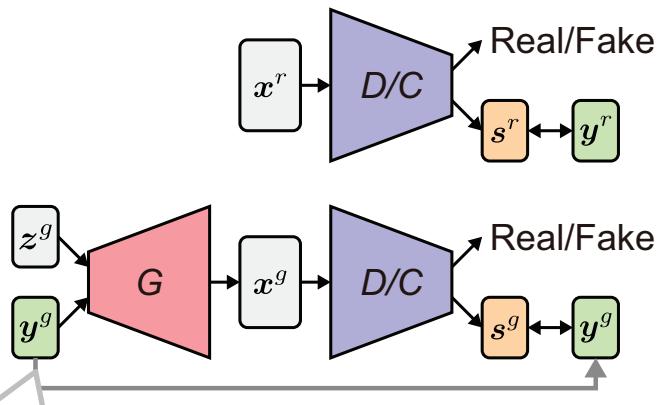
Generator

T. Kaneko et al., "Class-Distinct and Class-Mutual Image Generation with GANs," BMVC 2019.

Classifier's Posterior GAN [Kaneko+2019]

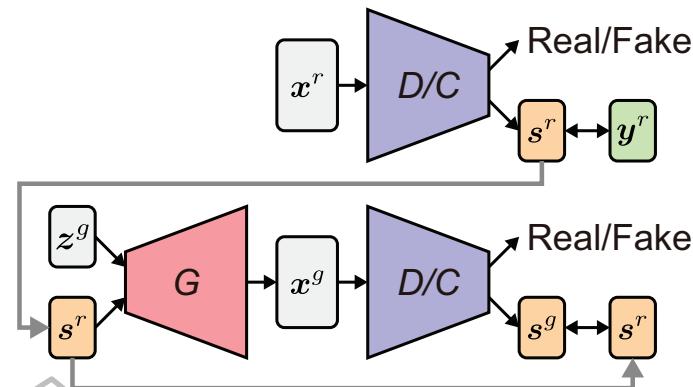
- Classifier's Posteriorを用いてAC-GAN [Odena+2017] の
入力と目的関数を再設計

AC-GAN (従来)



離散的な
ラベルに条件づけ

CP-GAN (提案)



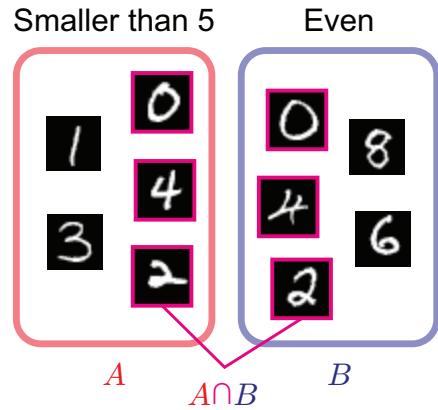
クラス間の関係を表す
Classifier's Posteriorに条件づけ

T. Kaneko et al., "Class-Distinct and Class-Mutual Image Generation with GANs," BMVC 2019.
A. Odena et al., "Conditional Image Synthesis With Auxiliary Classifier GANs," ICML 2017.

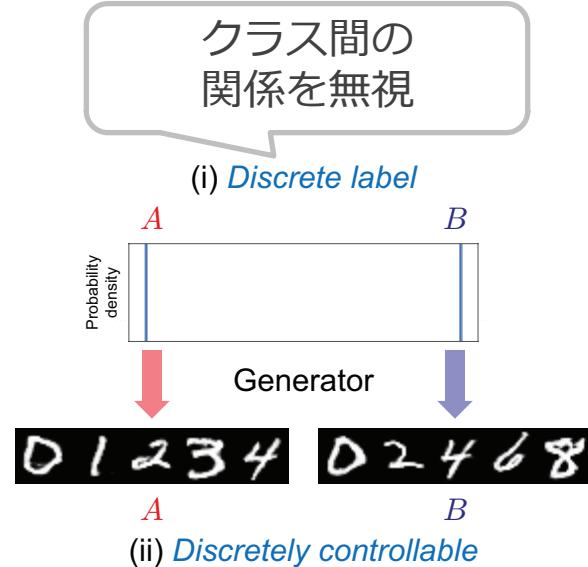
CP-GAN (cont.)



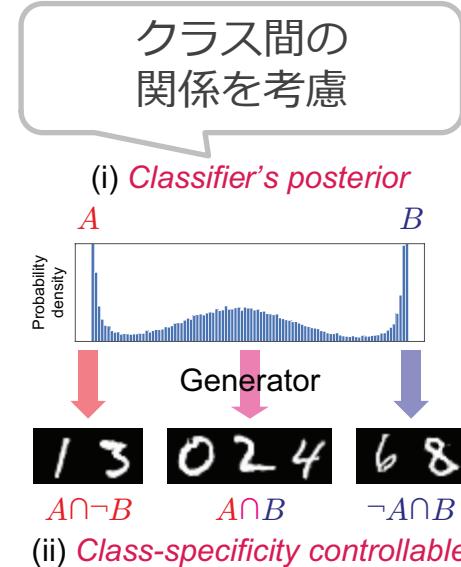
画像生成例



学習データ
(Class-Overlapping)



AC-GAN
(従来)



CP-GAN
(提案)

コード : <https://github.com/takuhirok/CP-GAN/>

クラス間の関係を考慮した画像生成が可能

T. Kaneko et al., "Class-Distinct and Class-Mutual Image Generation with GANs," BMVC 2019.

NaïveなGANの課題4③



学習データは完全であることを仮定

- 例3：全データにラベル付与を仮定



Butterfly



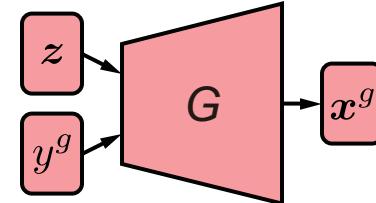
Dog



Burger

ラベルを
用いて
モデル化

データリソースが不十分



Generator

どう学習？



Butterfly

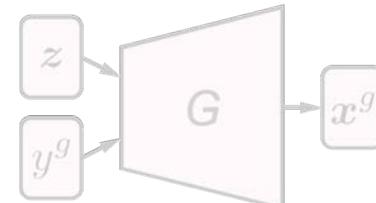


?



?

不完全 (少数ラベル)



Generator

M. Lucic et al., "High-Fidelity Image Generation With Fewer Labels," ICML 2019.

Semi-Supervised GAN with Self-Supervision [Lucic+2019]

- ・ 準教師あり学習と自己教師あり学習をGANに導入

生成画像例

BigGAN
(従来)
全ラベルを使用



S³GAN
(提案)
10%のラベルを使用



少数ラベルだけでも全ラベルありに匹敵する画像生成が可能

M. Lucic et al., "High-Fidelity Image Generation With Fewer Labels," ICML 2019.

データ設定のまとめ



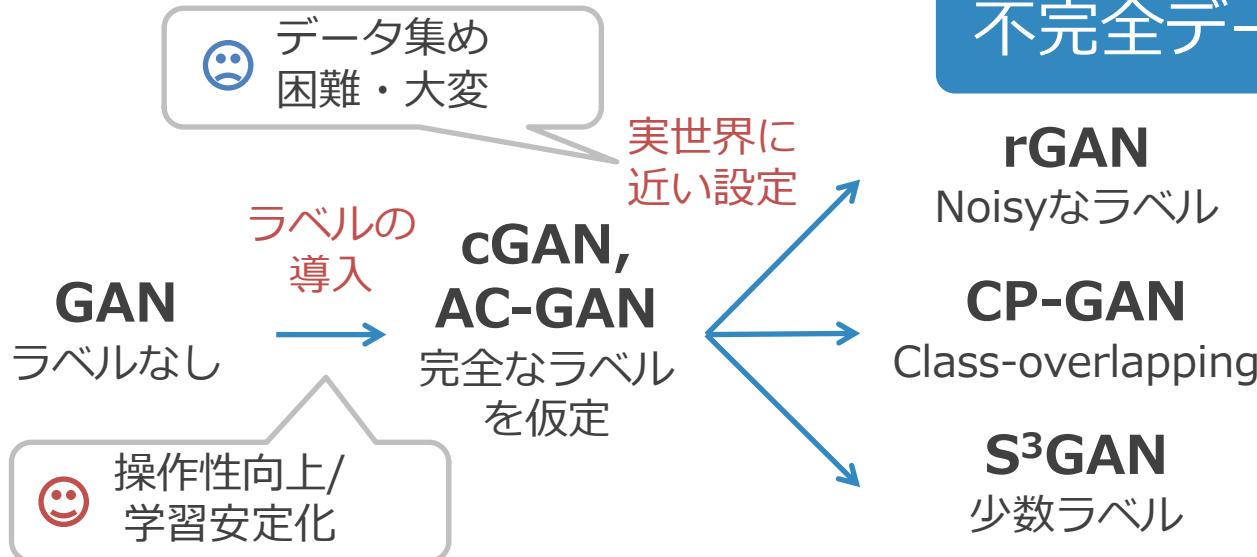
NaïveなGANの課題

- 学習データは完全であることを仮定

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

- ①Clean
- ②Class-Separate
- ③全データにラベルを仮定

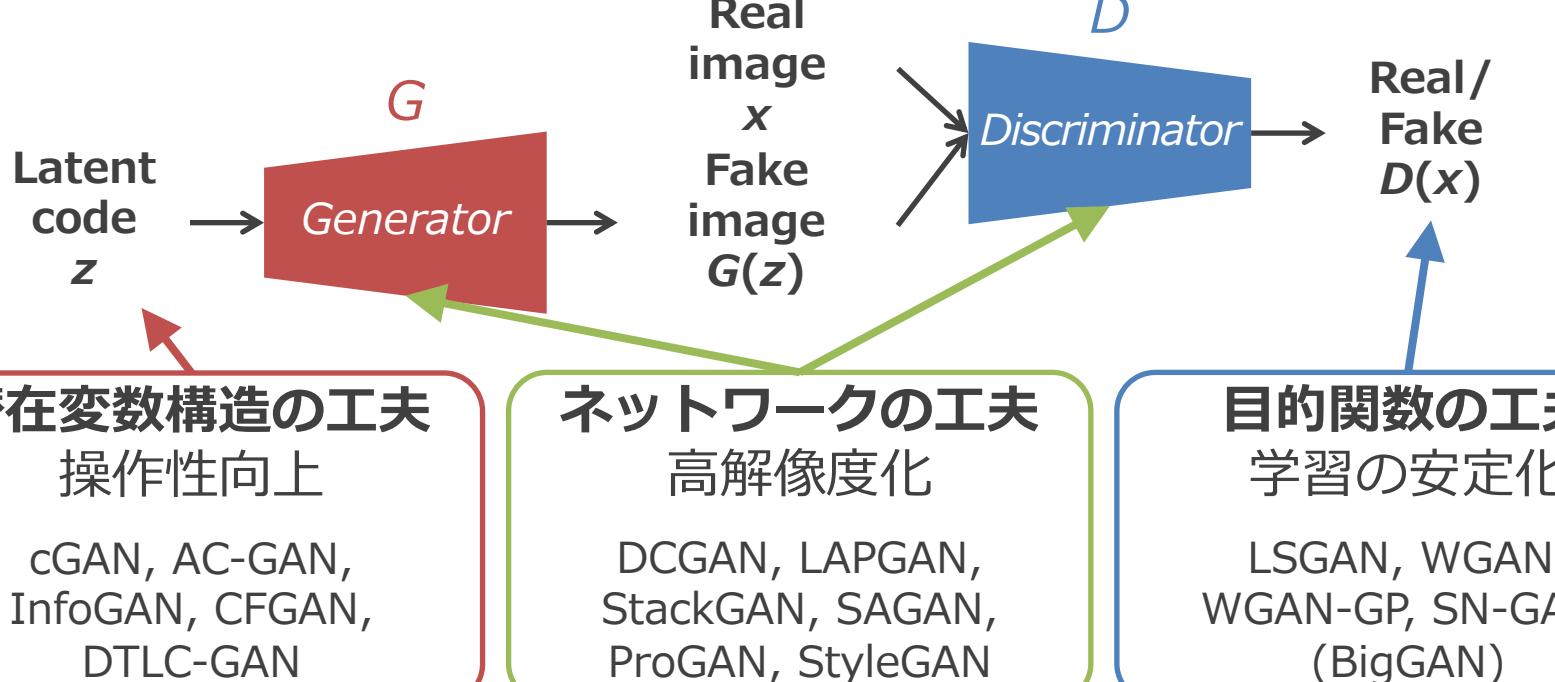
データ設定の発展



GANの発展 (まとめ)



4つの発展



※GANの研究は近年非常に盛んであり他にも様々な面白い試みあり

目次



**GAN
の基礎**

**GAN
の発展**

**GAN
の応用**

GANの応用①: 画像の場合

- 高品質なデータ変換
- ペアデータなしデータ変換

GANの応用②: 他モーダルでは?

- 高品質なデータ変換
- ペアデータなしデータ変換



GANの応用①: 画像の場合

- 高品質なデータ変換
- ペアデータなしデータ変換

GANの応用②: 他モーダルでは?

- 高品質なデータ変換
- ペアデータなしデータ変換

GANの応用①：画像の場合



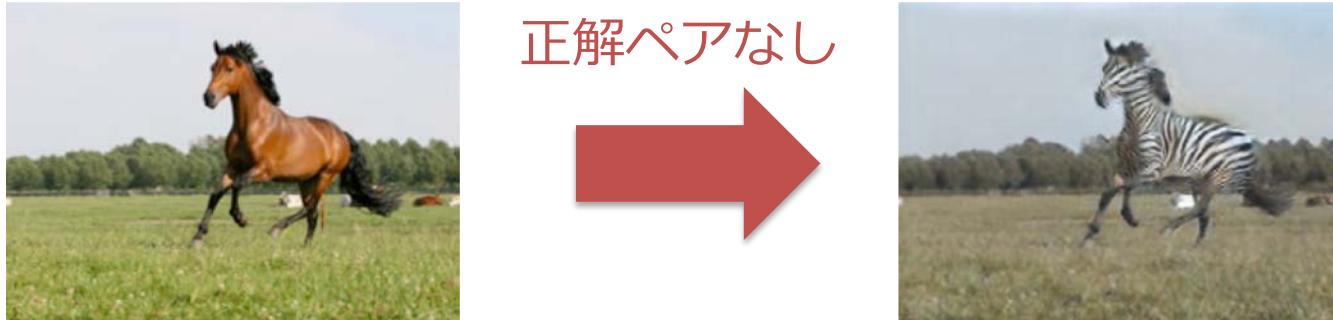
2つの代表例

1. 高品質なデータ変換



C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

2. ペアデータなしデータ変換



J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

GANの応用①：画像の場合



2つの代表例

1. 高品質なデータ変換



C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

2. ペアデータなしデータ変換



J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

データ変換とは？



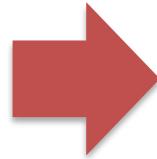
データ変換



絵画 by Monet

変換器を学習により獲得したい

写真

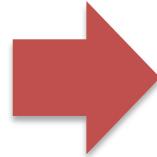


写真



ウマ

シマウマ



シマウマ

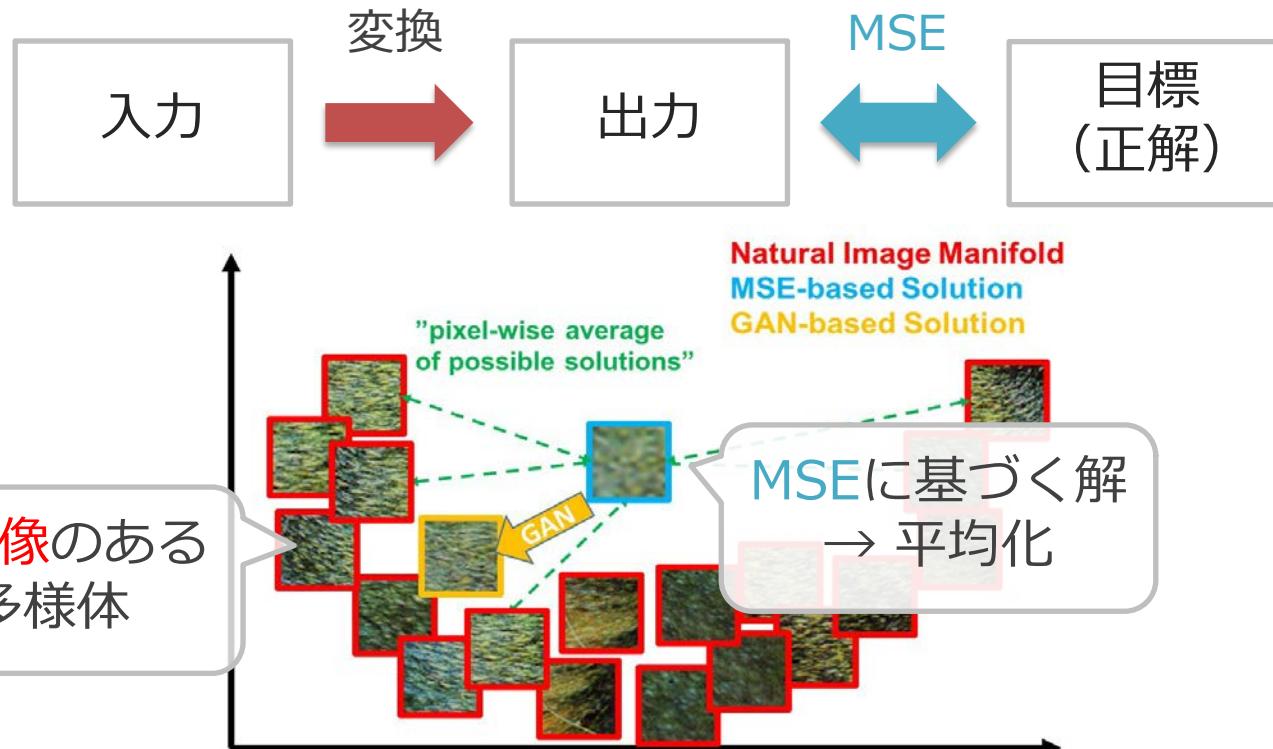
J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

従来の統計的データ変換手法



既定の分布に基づく目的関数

- ガウシアン分布、最小二乗誤差など
→ 複数の解候補の中で平均化



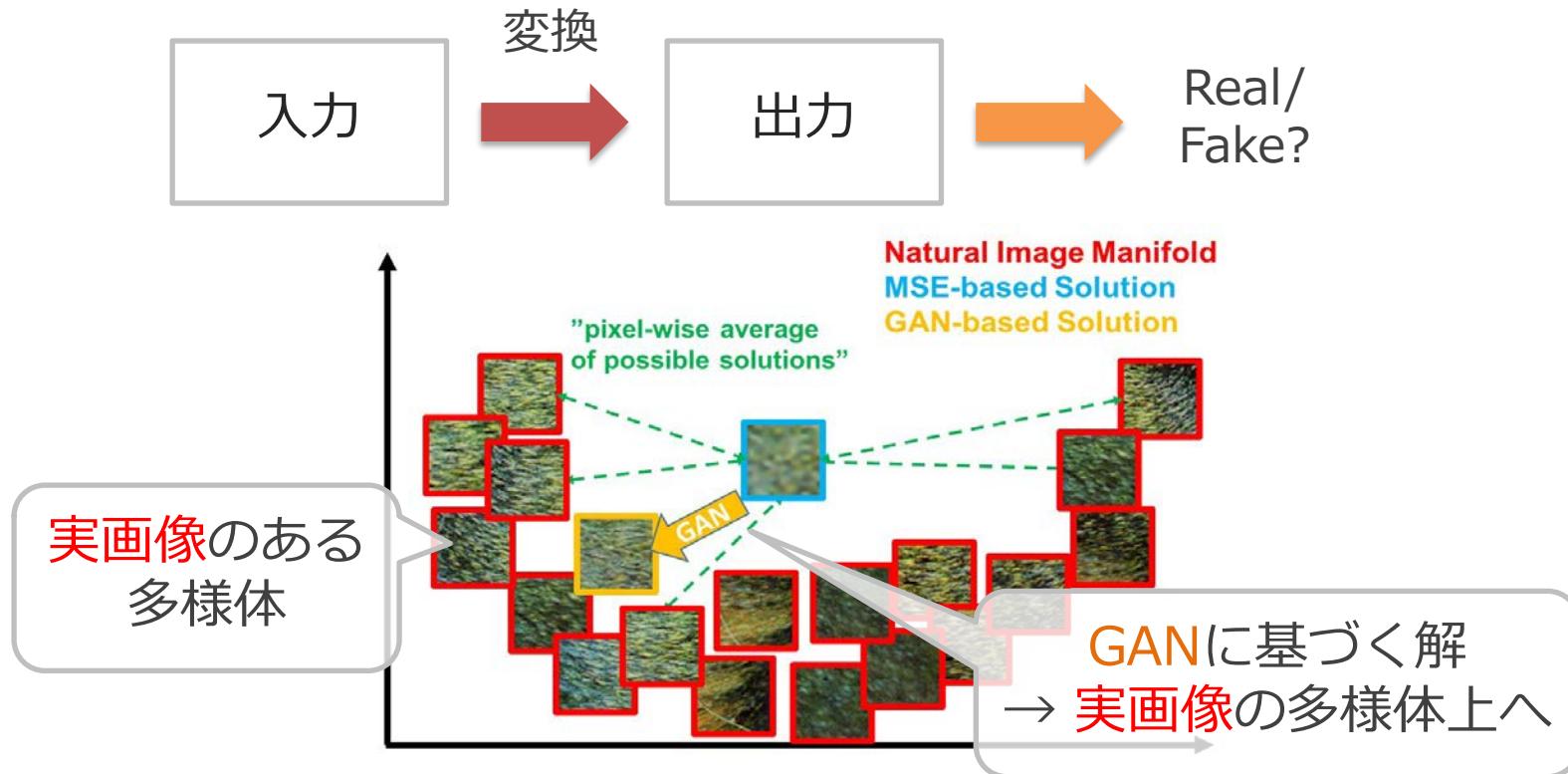
C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

GANベースのデータ変換手法



GANに基づく目的関数

- 分布を陽に仮定する必要なし
→ Real/Fakeという基準で解を発見 → 平均化の影響を緩和可



C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

例1: SRGAN



Super Resolution GAN [Ledig+2017]

- 目的: 超解像



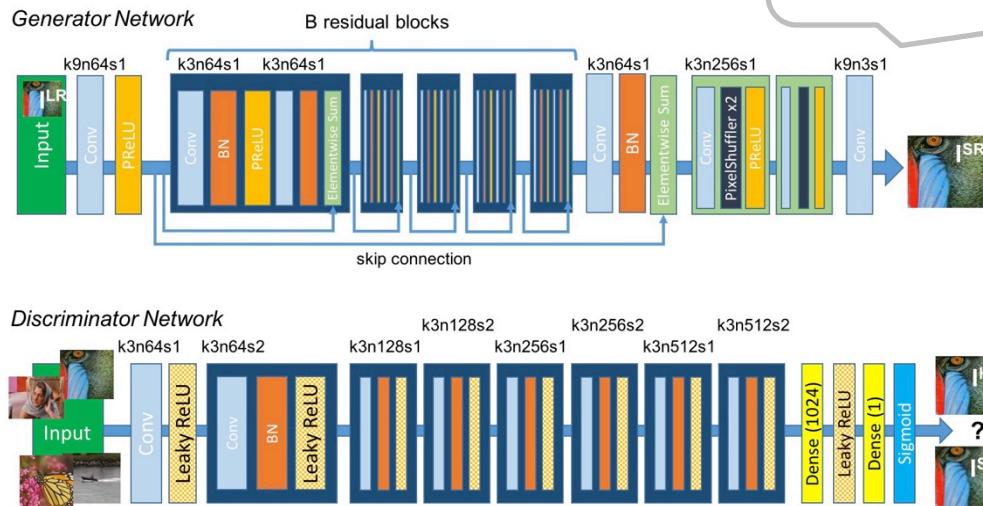
C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

例1: SRGAN (cont.)

目的関数

- Content Loss
 - MSE Loss: 画像(ピクセル)空間で距離を計測
 - VGG Loss (Perceptual Loss): VGGの特徴量空間で距離を計測
- Adversarial Loss
 - Real/Fakeの識別

Residual構造と
Skip Connectionで入力を保持



C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

例1: SRGAN (cont.)



変換例



Bicubic補間



SRGAN



元画像

高品質な超解像が可能

C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

例2：pix2pix



Image-to-Image Translation with CGAN [Isola+2017]

- 目的: ペアデータありの汎用的な画像変換



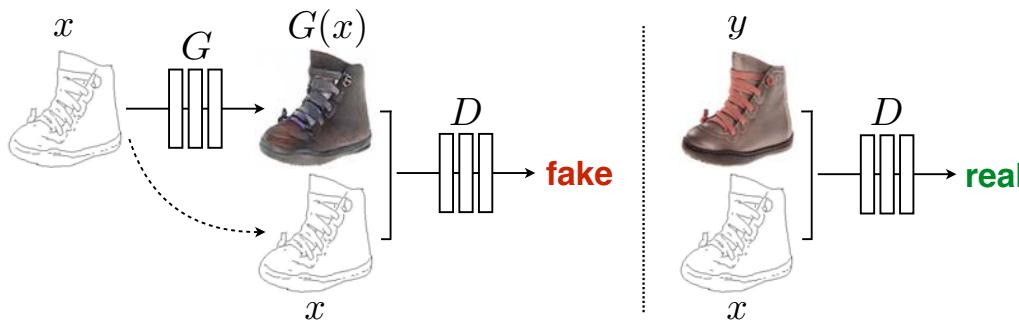
P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Networks," CVPR 2017.

J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

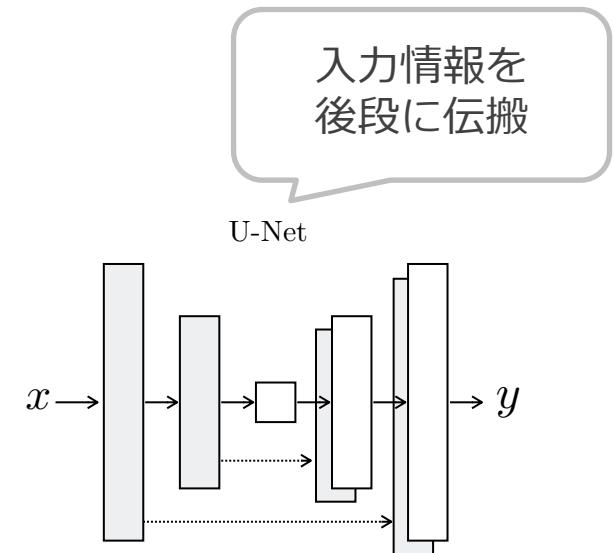
例2: pix2pix (cont.)

目的関数

- Content Loss
 - L1 Loss: 画像(ピクセル)空間で距離を計測
- Adversarial Loss
 - Real/Fakeの識別



入力画像に条件づけされたConditional GAN



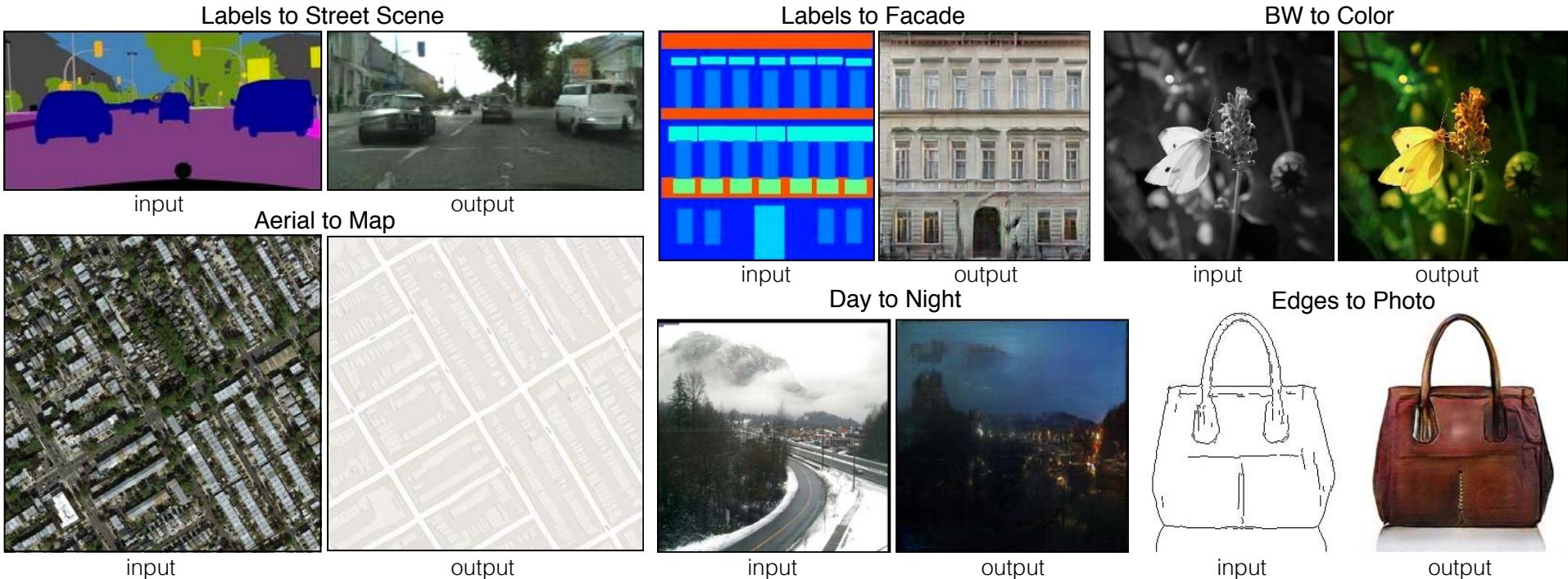
Generatorの構造

P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Networks," CVPR 2017.

例2: pix2pix (cont.)



変換例



様々なデータで高品質な画像変換が可能

P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Networks," CVPR 2017.

例3: pix2pixHD



High-Resolution pix2pix [Wang+2018]

- 目的: ①高解像度化 ②一对多変換

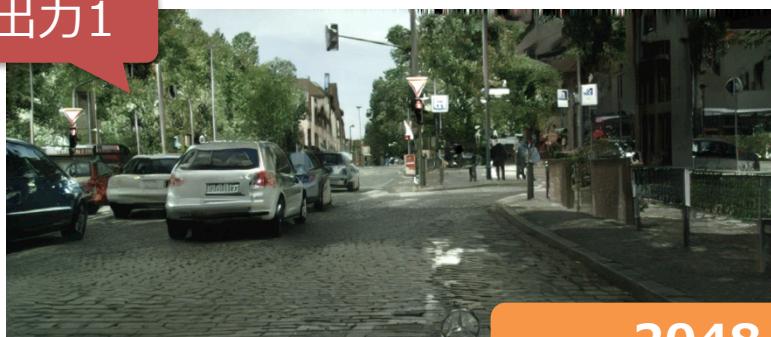
入力



出力2



出力1



出力3



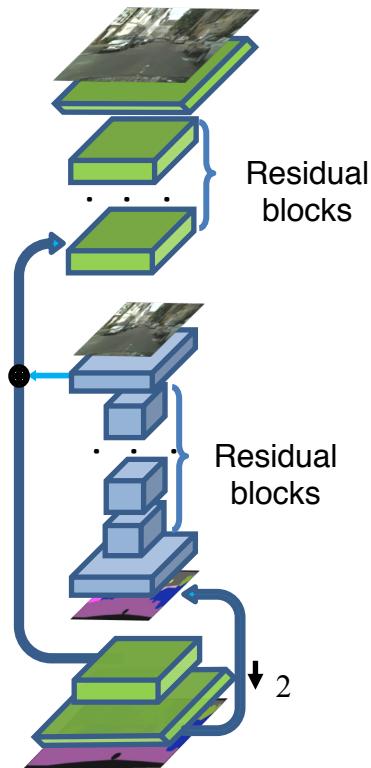
2048 x 1024

T.C. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018.

例3: pix2pixHD (cont.)

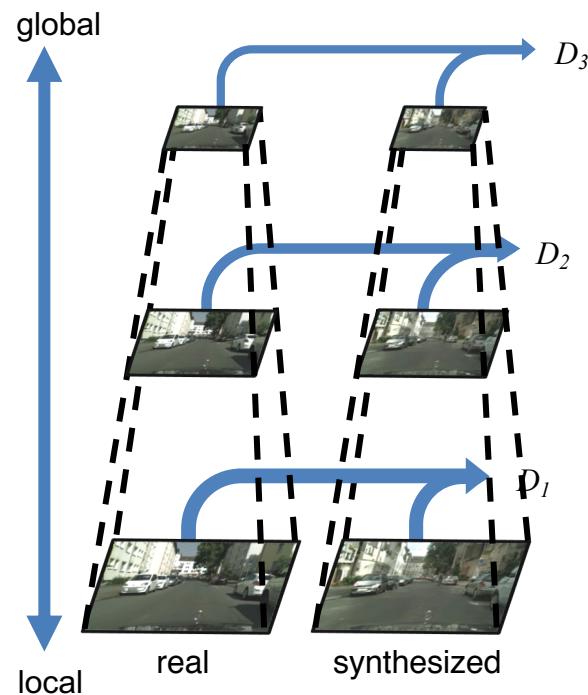
手法 (高解像度化)

Coarse-to-fine Generator



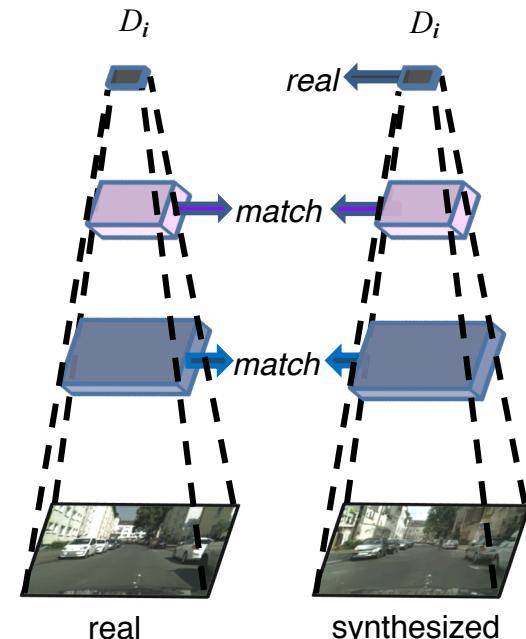
段階的に生成

Multi-scale Discriminators



多段階でReal/Fakeの識別

Robust Objective



多段階でRealデータに近接

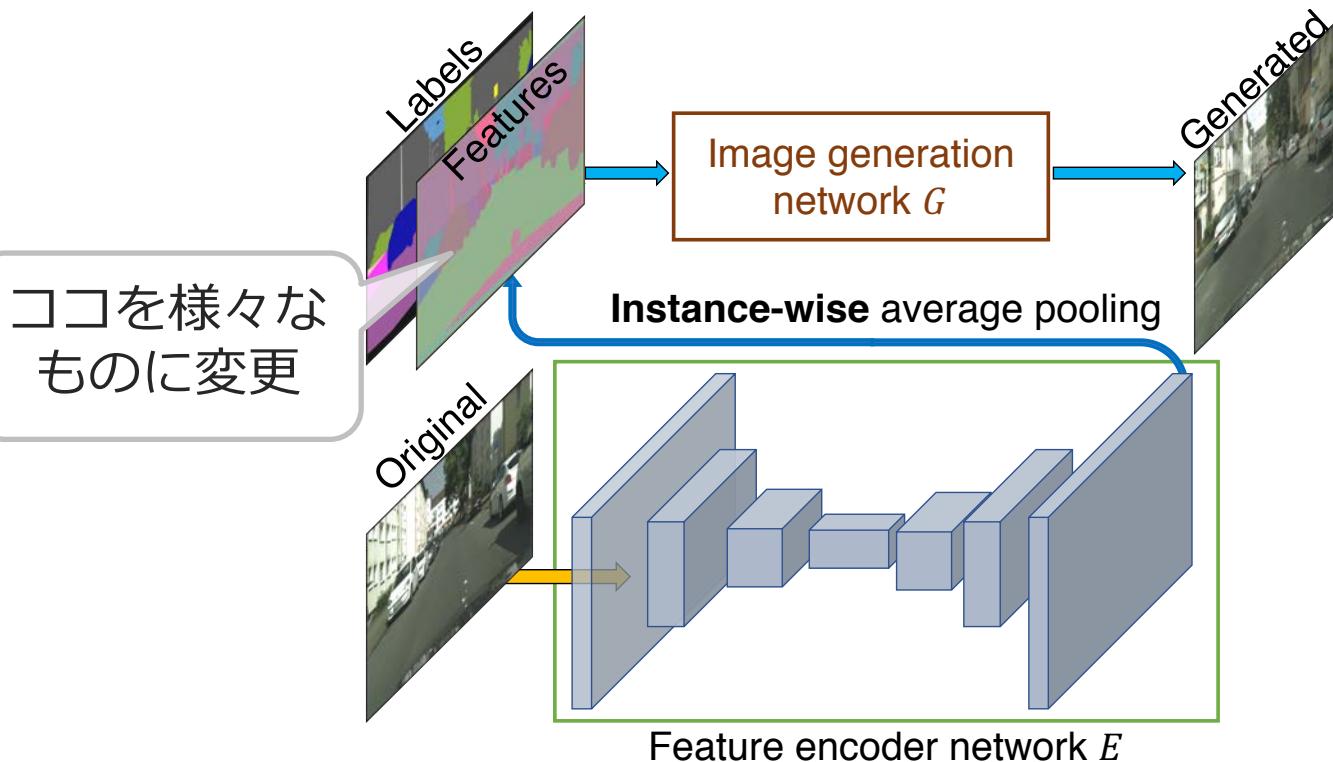
T.C. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018.

例3: pix2pixHD (cont.)



手法（一対多変換）

- 特徴量に条件づけして画像を変換
- 特徴量を変えることで様々な画像の生成が可

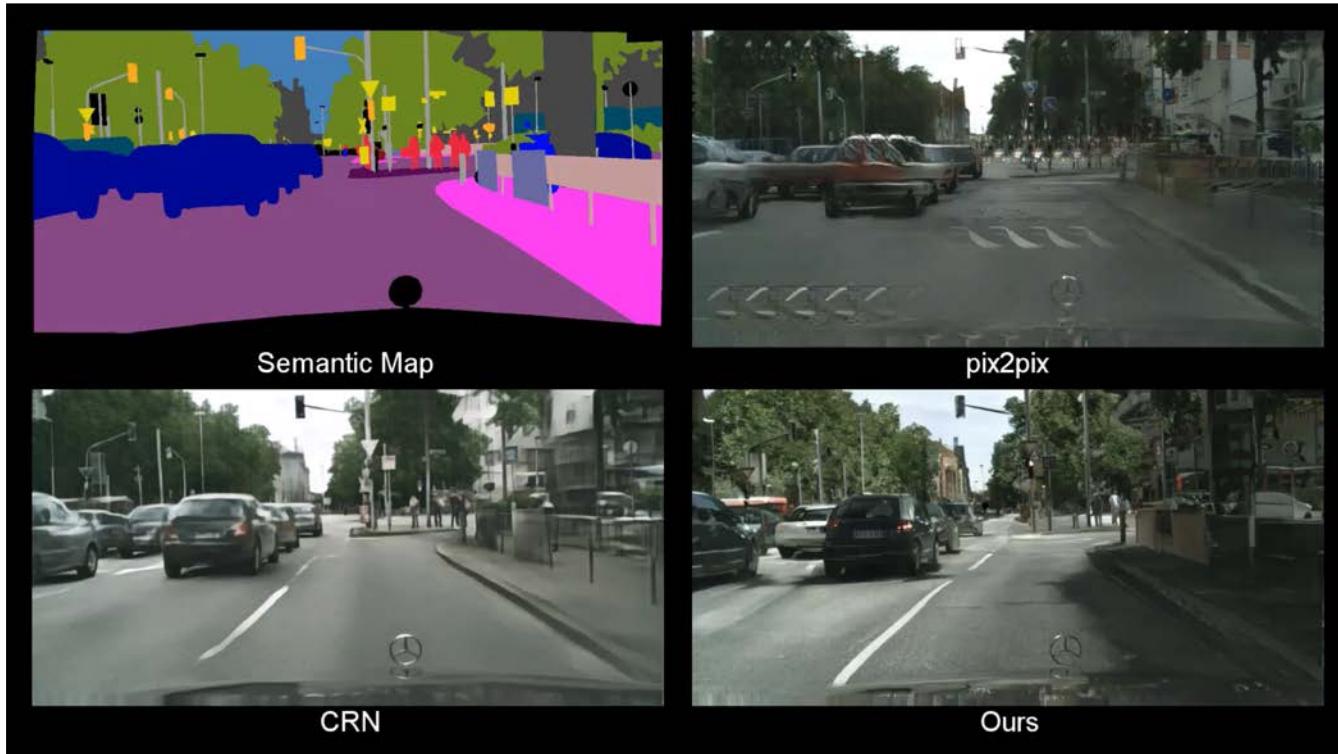


T.C. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018.

例3: pix2pixHD (cont.)



変換例（高解像度画像の生成）



デモ動画：<https://tcwang0509.github.io/pix2pixHD/>

さらに高品質な画像変換が可能

T.C. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018.

例3: pix2pixHD (cont.)



変換例（一对多変換）



デモ動画：<https://tcwang0509.github.io/pix2pixHD/>

一对多変換が可能

T.C. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," CVPR 2018.

例4: vid2vid



Video-to-Video Synthesis [Wang+2018]

- 目的: 高品質な動画から動画への変換



デモ動画 : <https://tcwang0509.github.io/vid2vid/>

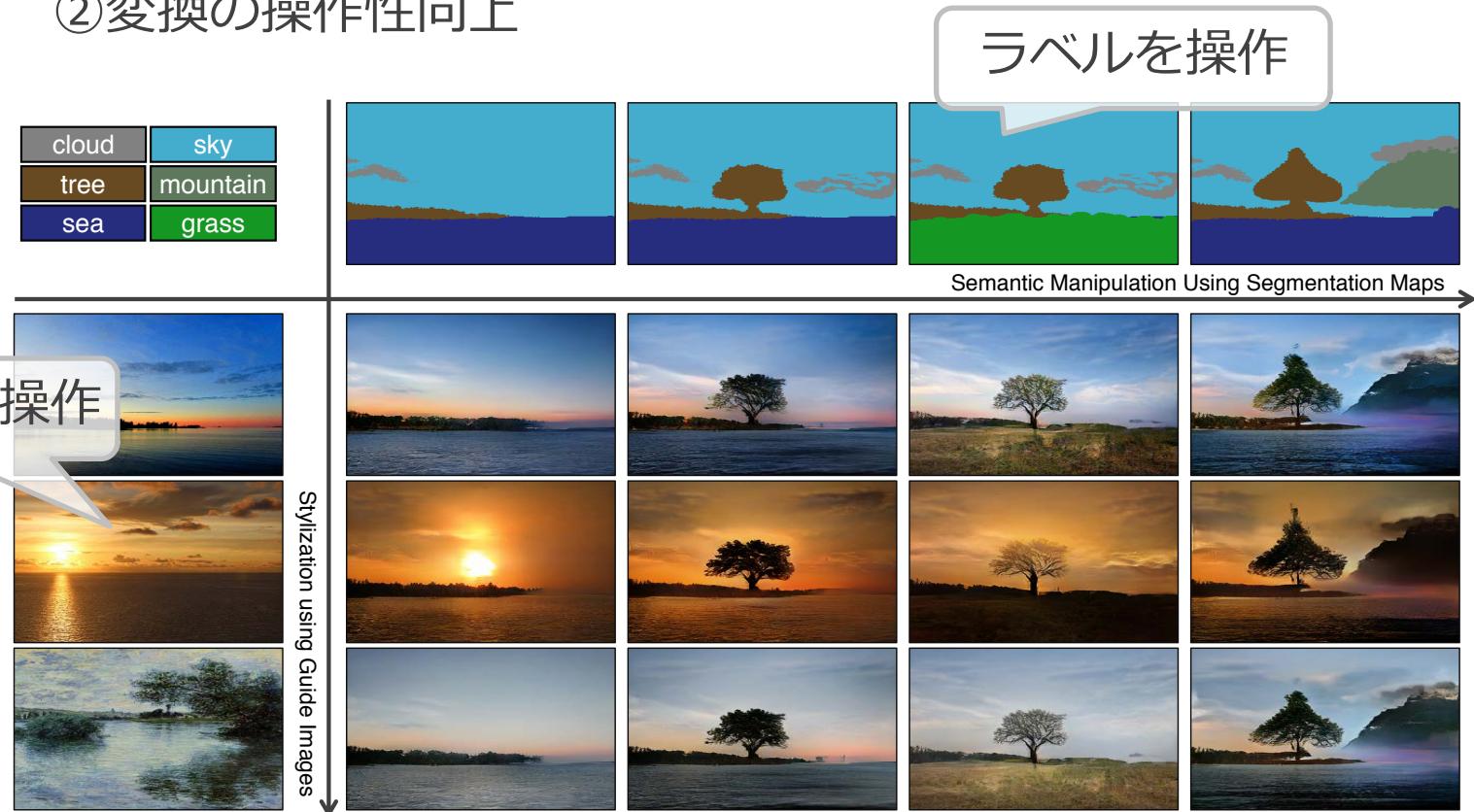
時系列にわたる一貫性も実現

T.C. Wang et al., "Video-to-Video Synthesis," NIPS 2018.

例5: SPADE

SPatially-Adaptive (DE)normalization [Park+2019]

- 目的: ①さらに高品質な「Semantic Segmentation → 画像」変換
②変換の操作性向上



T. Park et al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," CVPR 2019.

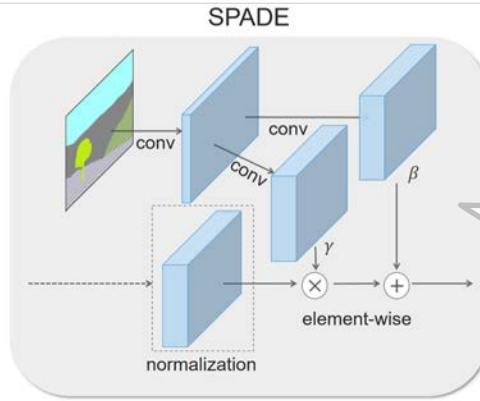
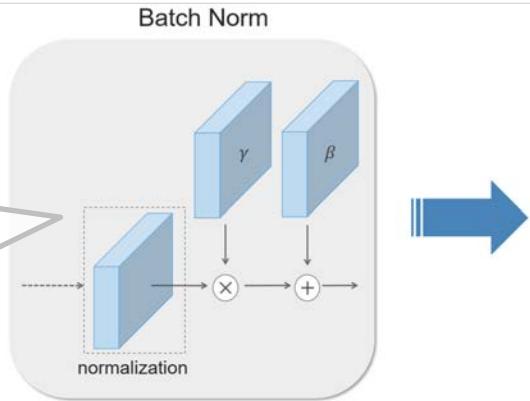
例5: SPADE (cont.)



手法

- SPADE

従来
空間情報を無視

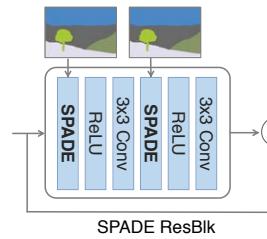


提案
空間情報に適応的

→ Semantic Segmentationの空間情報を保持した変換を実現

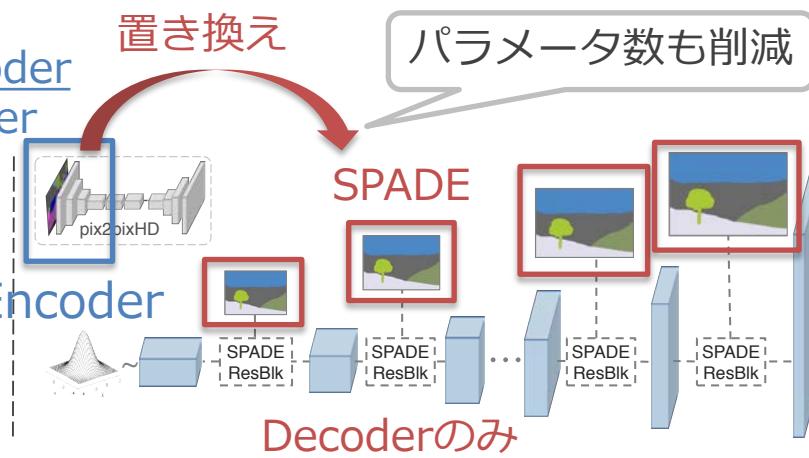
- SPADE Generator

従来: Encoder
& Decoder



置き換え

パラメータ数も削減

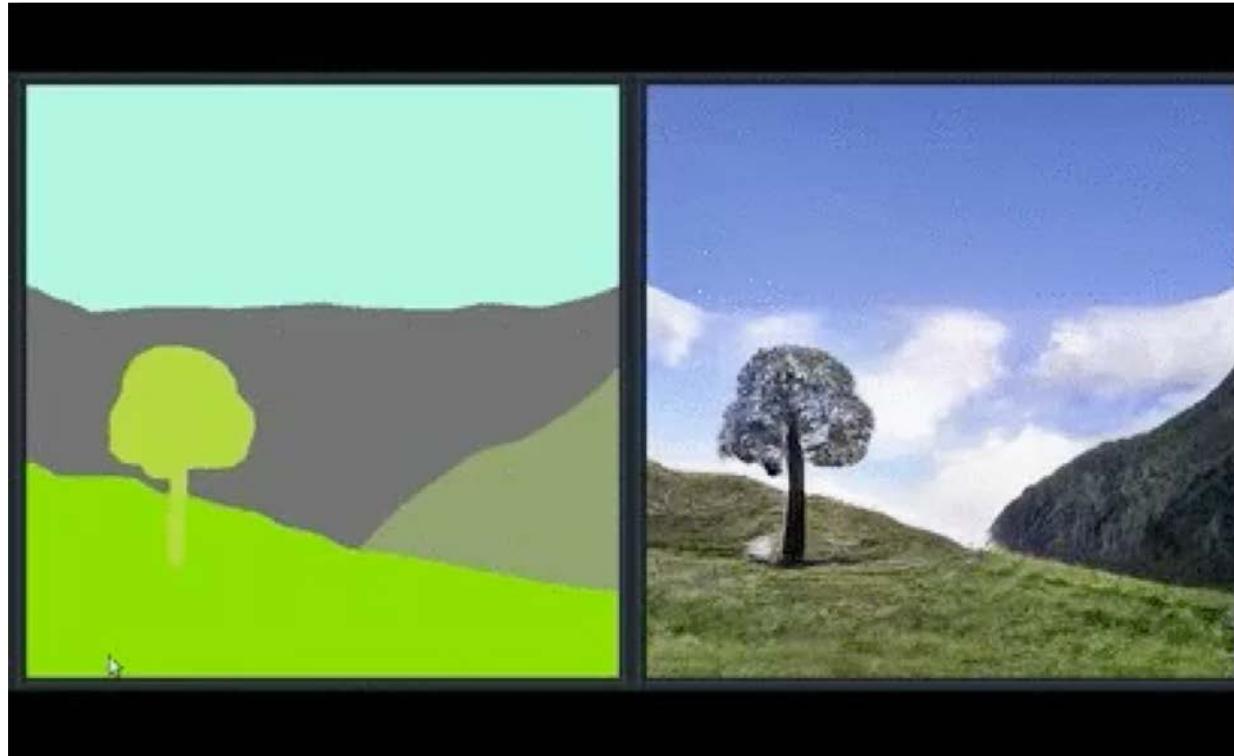


T. Park et al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," CVPR 2019.

例5: SPADE (cont.)



変換例



デモ動画：<https://nvlabs.github.io/SPADE/>

インタラクティブにラベル・スタイルの変換可能

T. Park et al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," CVPR 2019.

GANの応用①：画像の場合



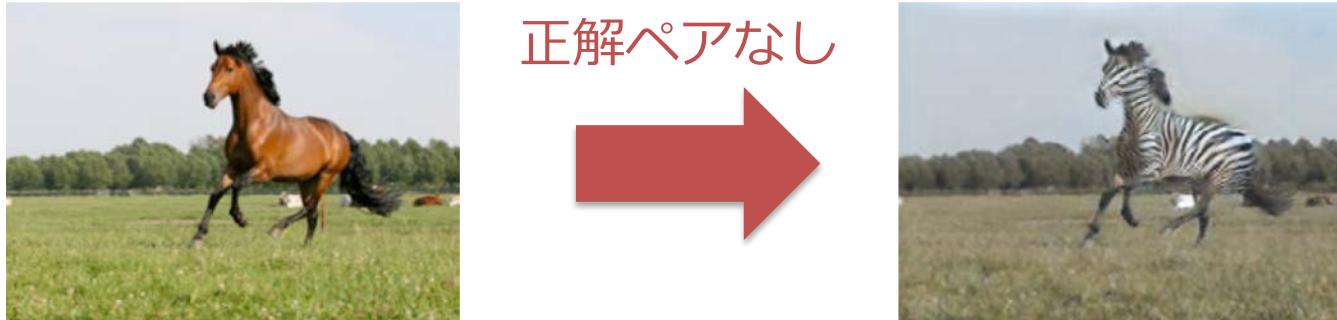
2つの代表例

1. 高品質なデータ変換



C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," CVPR 2017.

2. ペアデータなしデータ変換



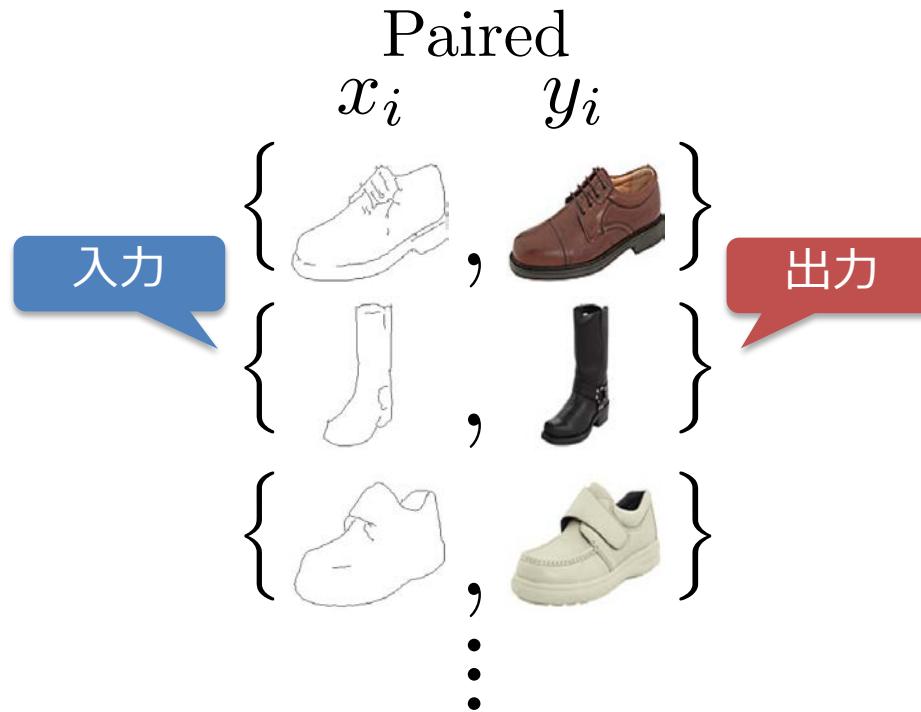
J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

従来のデータ変換



ペアデータありのデータ変換

- 😊 変換関数を学習しやすい
 - 😢 データを集めるのが大変（ペアデータが存在しない場合もあり）



J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

実現したいデータ変換

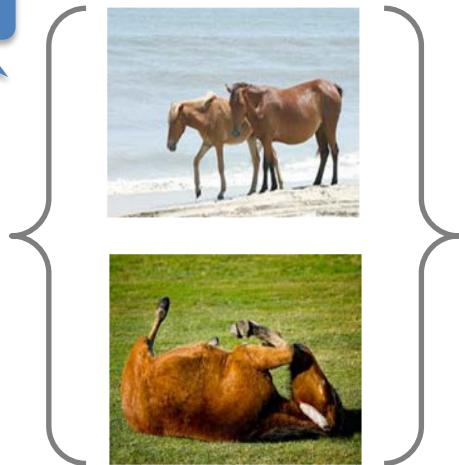


ペアデータなしのデータ変換

😊 データを収集しやすい

😢 どうやって学習する？ → GANを活用

入力



ペアデータ
なし

Horse

出力



Zebra

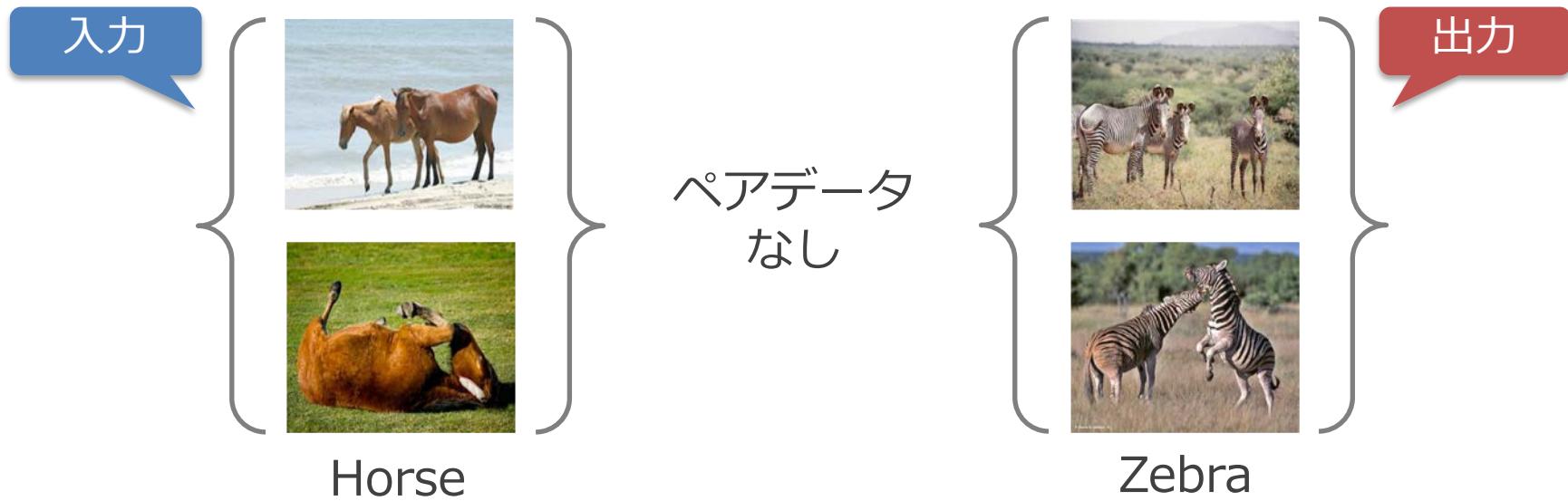
J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

例1: CycleGAN



CycleGAN [Zhu+2017] (別名DiscoGAN [Kim+2017], DualGAN [Yi+2017])

- 目的: ペアデータなしの画像変換



J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

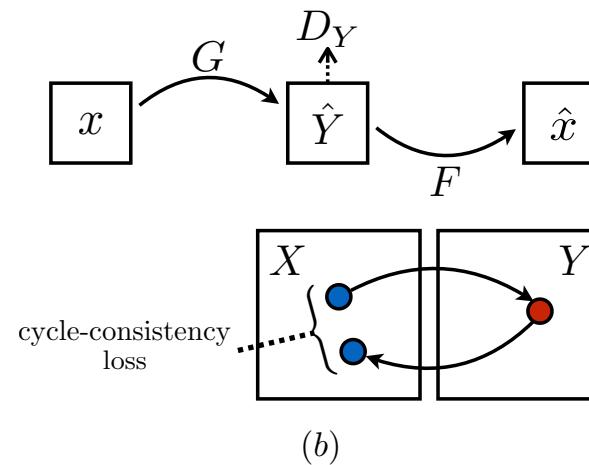
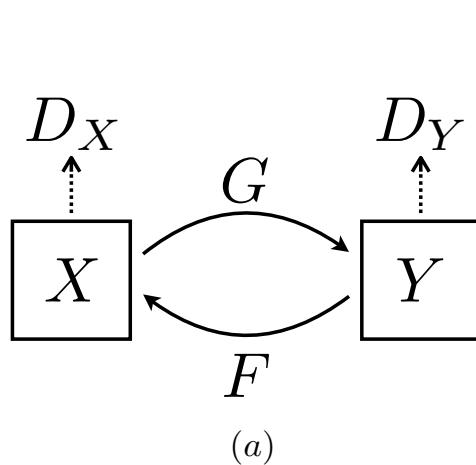
T. Kim et al., "Learning to Discover Cross-domain Relations with Generative Adversarial Networks," ICML 2017.

Z. Yi et al., "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," ICCV 2017.

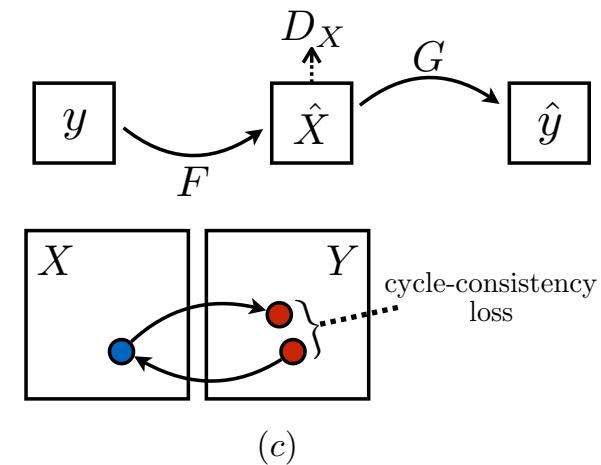
例1: CycleGAN (cont.)

目的関数

- Adversarial Loss
 - Real/Fakeの識別
- Cycle-consistency Loss
 - 順→逆変換、逆→順変換すると元に戻るという制約
 - 擬似的にペアデータを発見



Adversarial Loss



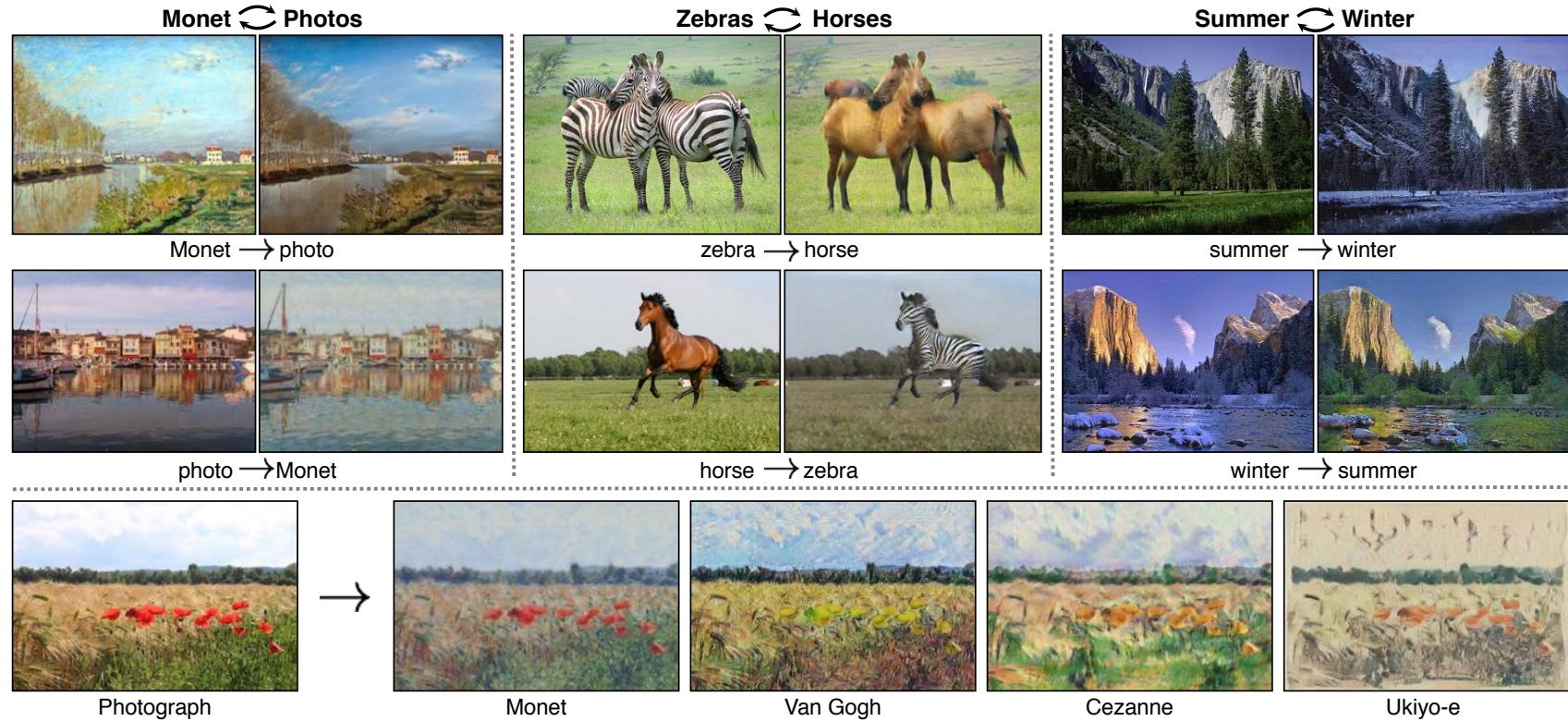
Cycle-consistency Loss

J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

例1: CycleGAN (cont.)



変換例



様々なデータでペアデータなし画像変換が可能

J.Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV 2017.

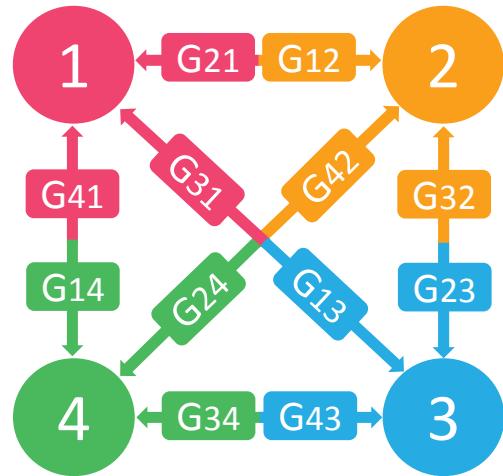
例2: StarGAN



StarGAN [Choi+2018]

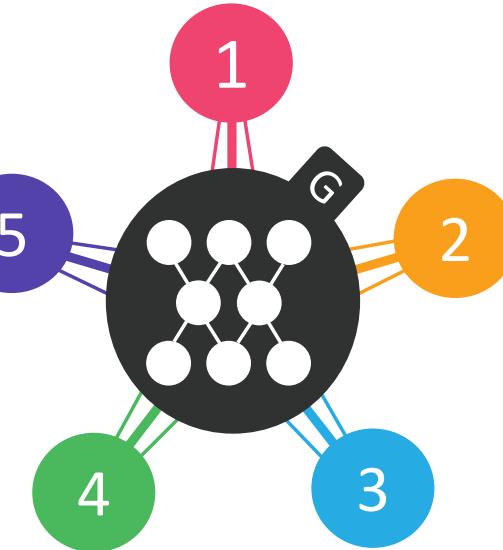
- 目的: マルチドメインでペアデータなしの画像変換

(a) Cross-domain models



従来手法
ドメインの組合せ分
変換モデルが必要

(b) StarGAN



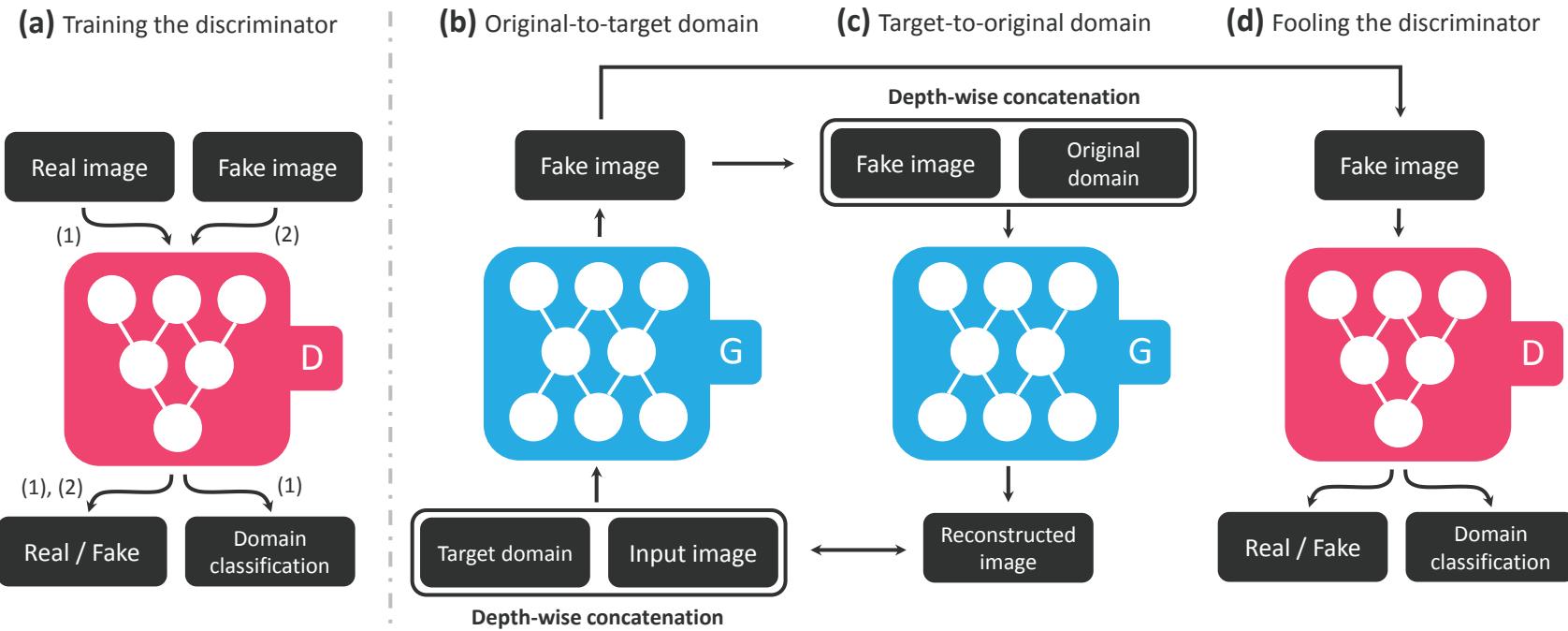
StarGAN
単一のモデル
で実現

Y. Choi et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," CVPR 2018.

例2: StarGAN (cont.)

手法

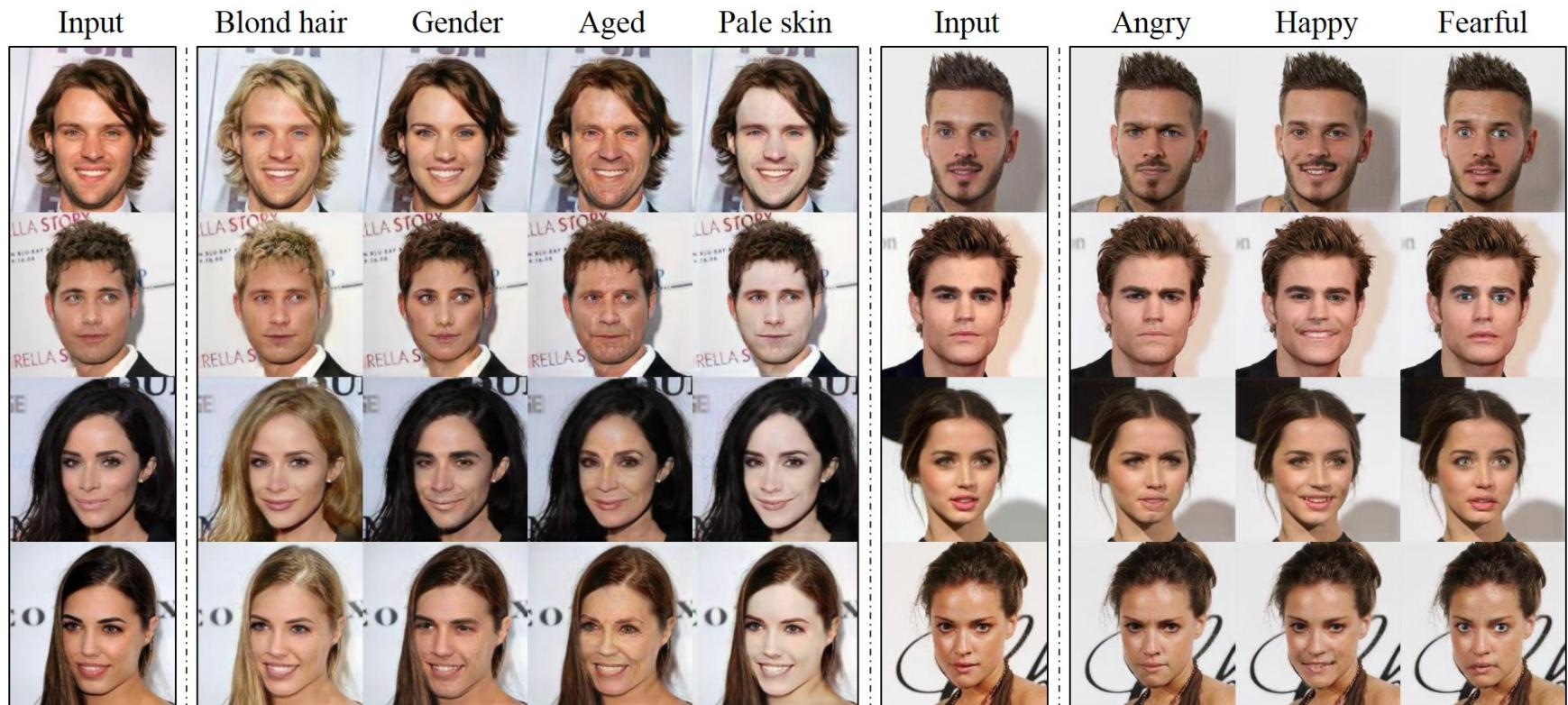
- CycleGANを条件付き設定に拡張
 - Generator: Domain Codeで条件付け
 - Discriminator: Domain Classifierと一緒に学習



Y. Choi et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," CVPR 2018.

例2: StarGAN (cont.)

変換例



単一のモデルでマルチドメイン画像変換が可能

Y. Choi et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," CVPR 2018.



GANの応用①: 画像の場合

- 高品質なデータ変換
- ペアデータなしデータ変換

GANの応用②: 他モーダルでは?

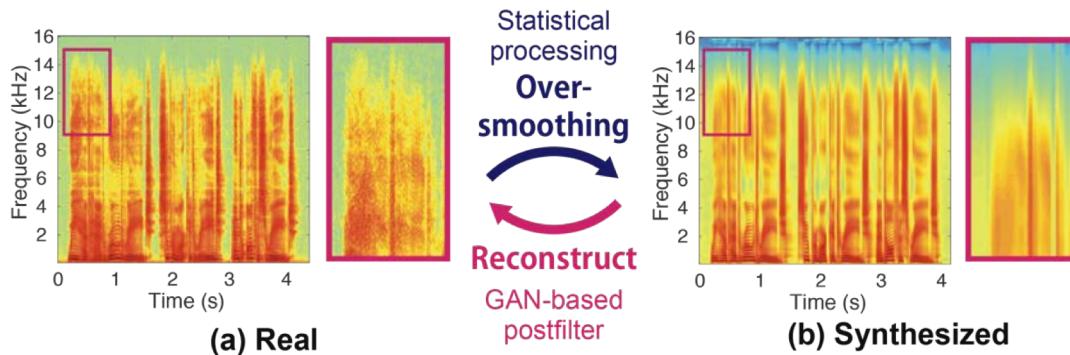
- 高品質なデータ変換
- ペアデータなしデータ変換

GANの応用②: 他モーダルでは?



2つの代表例

1. 高品質なデータ変換



T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

2. ペアデータなしデータ変換



話者A



ペアデータなし
で変換器を学習



話者B

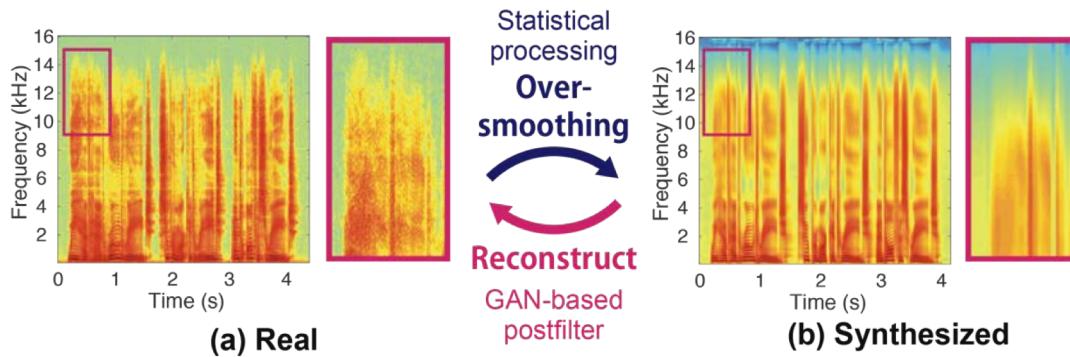
T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.

GANの応用②: 他モーダルでは?



2つの代表例

1. 高品質なデータ変換



T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

2. ペアデータなしデータ変換



話者A



ペアデータなし
で変換器を学習



話者B

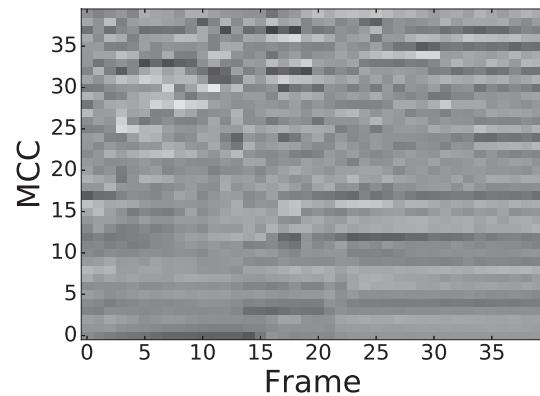
T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.

例1: GAN-PF



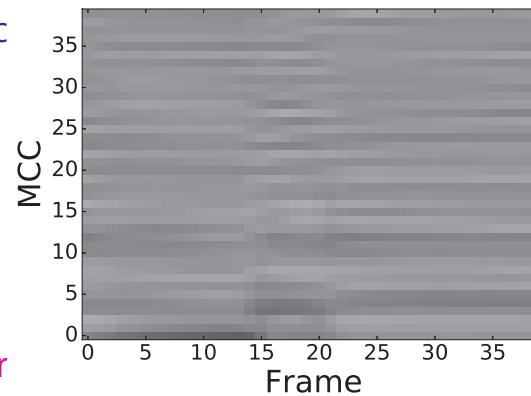
GAN-based Postfilter [Kaneko+2017]

- 目的: 統計的テキスト合成音声/音声変換における過剰な平滑化の緩和
 - 従来手法では、規定の分布(ガウシアン分布、最小二乗誤差)で近似
→ データは過剰に平滑化される傾向



(a) Natural Speech

Statistical Parametric
Speech Synthesis
Over-smoothing
Reconstruct
GAN-based Postfilter



(b) Synthesized Speech

"Spectral Textures" in Mel-cepstral domains

※ テキスト合成音声: テキストから音声を合成
音声変換: 音声Aを音声Bに変換

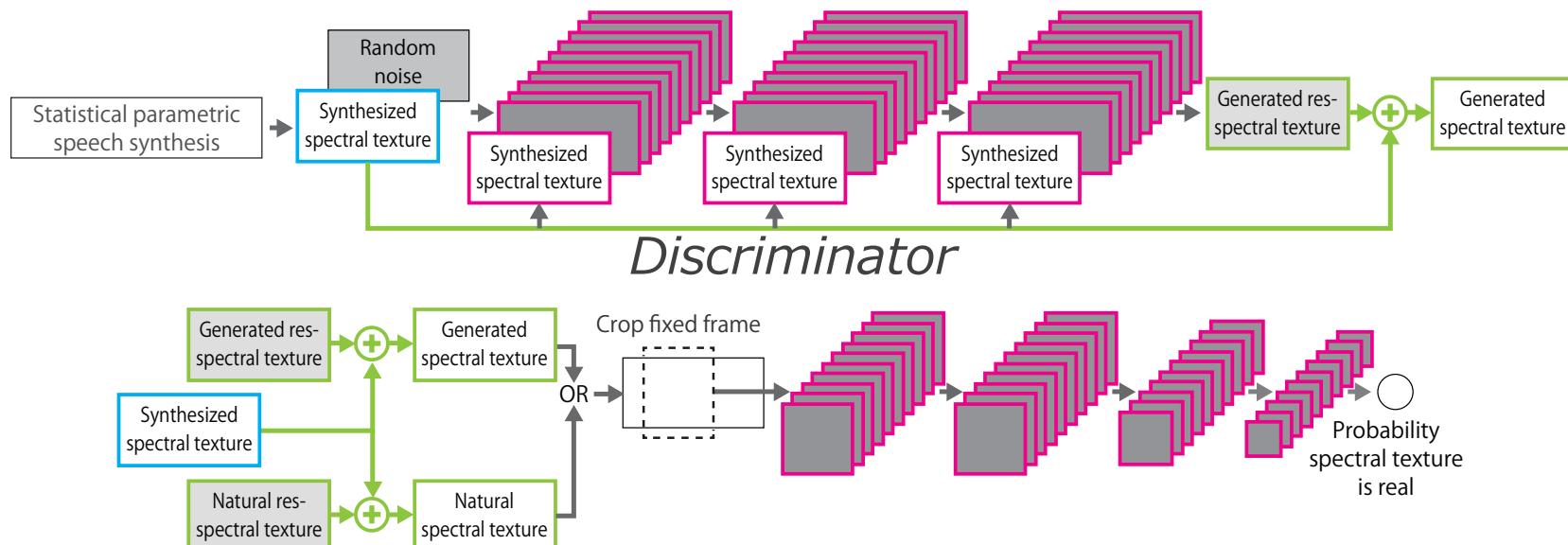
T. Kaneko et al., "Generative Adversarial Network-based Postfilter for Statistic Parametric Speech Synthesis," ICASSP 2017.

例1: GAN-PF (cont.)



モデル

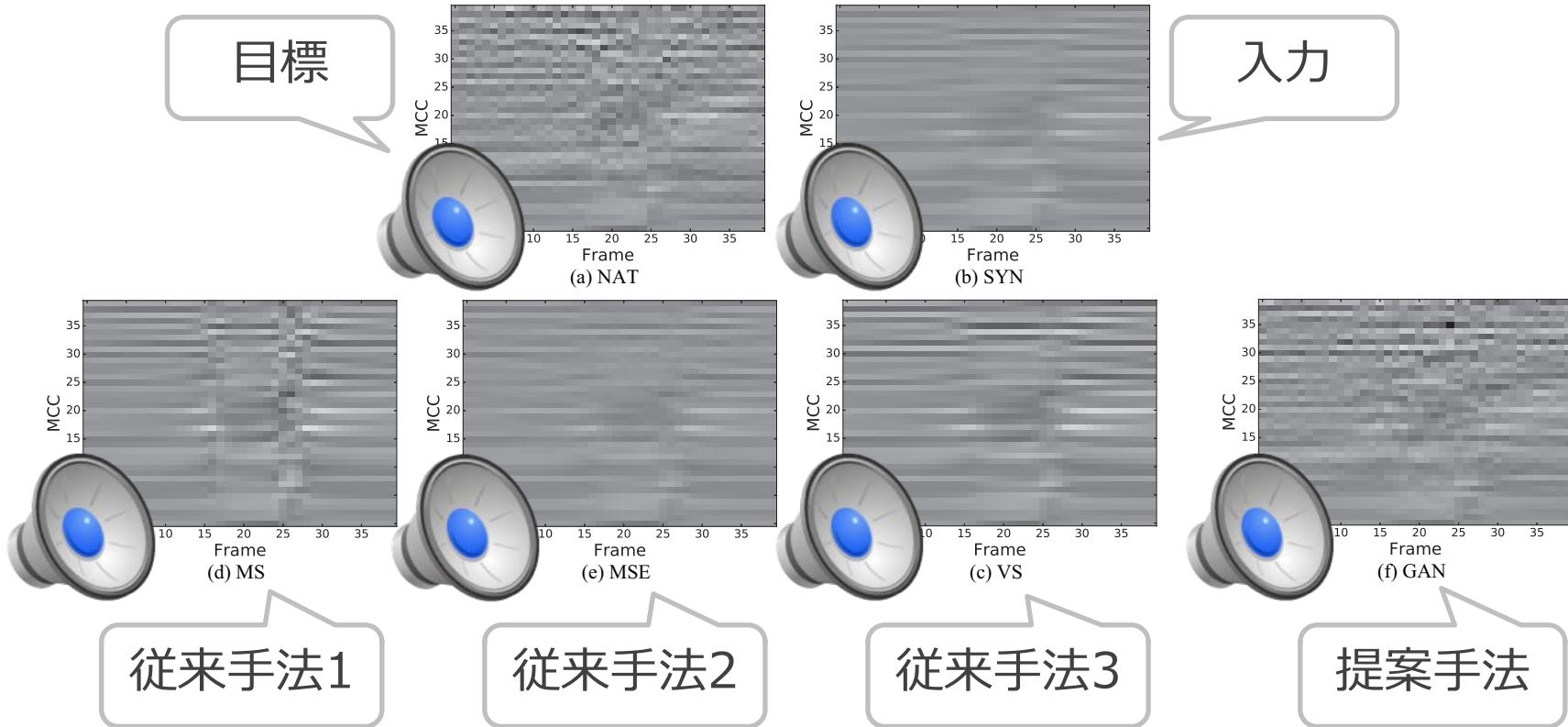
- **Conditional:** 入力に応じて変換
 - **Residual:** 入力情報を保持
 - **Convolutional:** 時間・周波数構造 (Spectral Texture) をモデル化
 - 特にFully Convolutional Networks [Long+2015] を用いて任意長に対応
- Generator*



T. Kaneko et al., "Generative Adversarial Network-based Postfilter for Statistic Parametric Speech Synthesis," ICASSP 2017.
J. Long et al., "Fully Convolutional Networks for Semantic Segmentation," CVPR 2015.

例1: GAN-PF (cont.)

変換例 (メルケプストラムを変換)



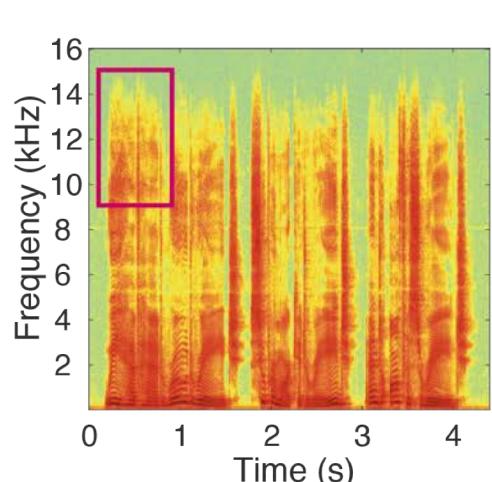
音声特徴量でも自然性を改善するのにGANが有効

T. Kaneko et al., "Generative Adversarial Network-based Postfilter for Statistic Parametric Speech Synthesis," ICASSP 2017.

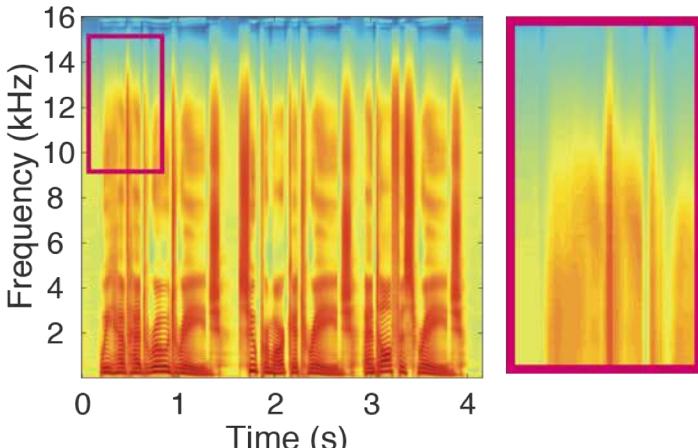
例2: GAN-PF for STFT

GAN Postfilter for STFT Spectrograms [Kaneko+2017]

- 目的: スペクトログラムの統計的モデル化による過剰な平滑化の緩和
 - 挑戦: 先の例では 10^1 オーダー → ここでは 10^3 オーダー



(a) Real



(b) Synthesized

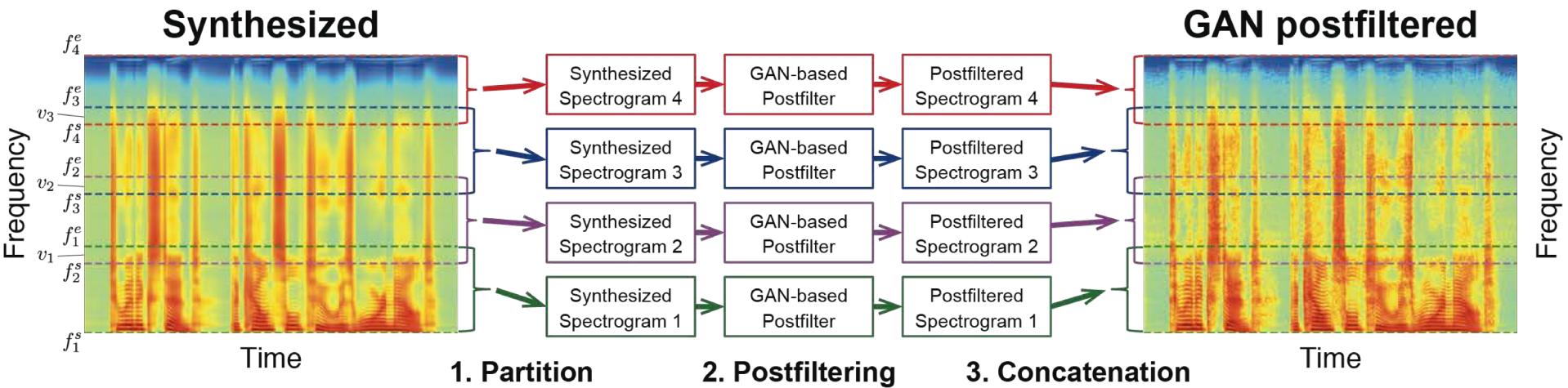
T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

例2: GAN-PF for STFT (cont.)



手法

1. 帯域分割 (高次元データを扱うため)
2. GAN Postfilter [Kaneko+2017]
3. 窓関数をかけて結合



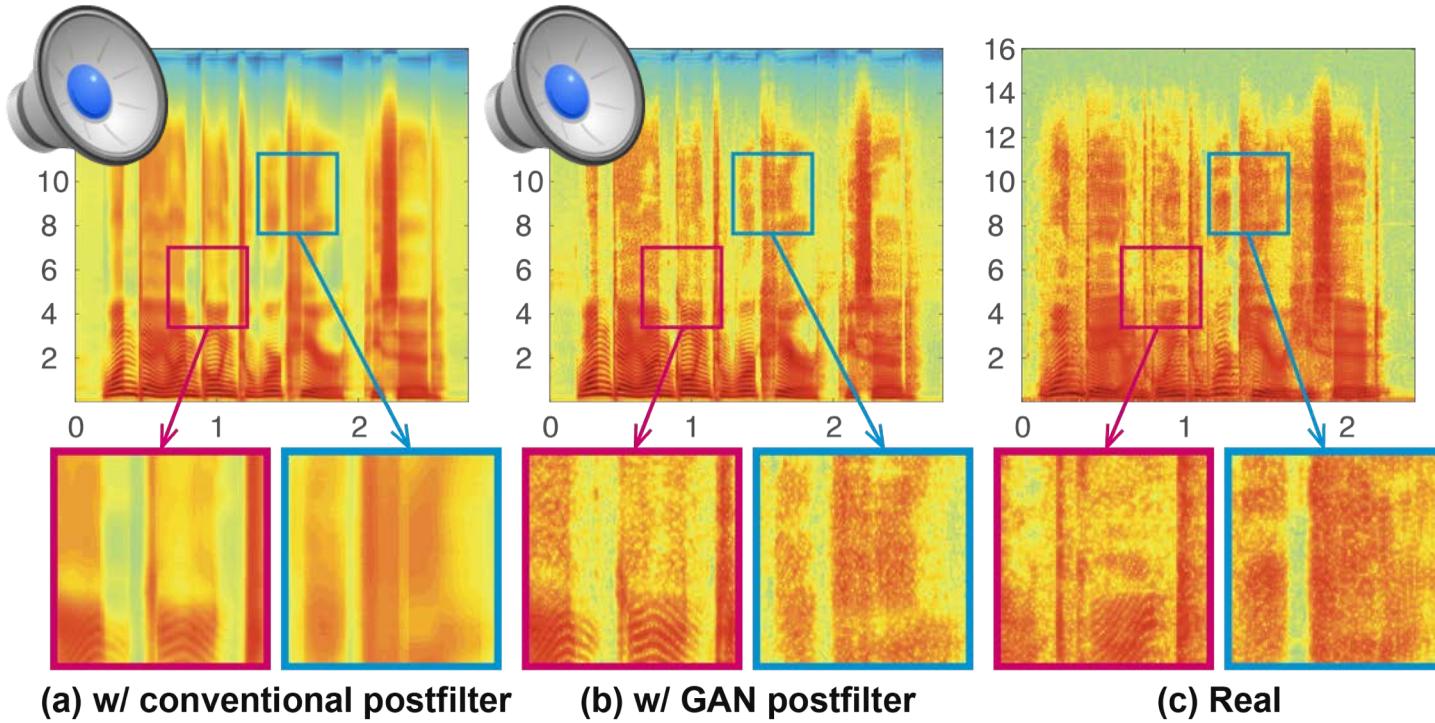
T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

T. Kaneko et al., "Generative Adversarial Network-based Postfilter for Statistic Parametric Speech Synthesis," ICASSP 2017.

例2: GAN-PF for STFT (cont.)



変換例 (STFTスペクトログラムを変換)



サンプル音声 : http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/ganp_stft/index.html

スペクトログラム (高次元) でも自然性の改善にGANが有効

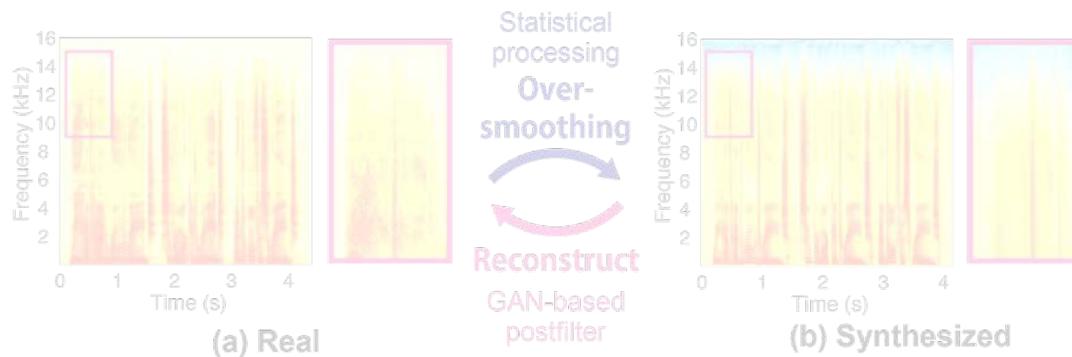
T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

GANの応用②: 他モーダルでは?



2つの代表例

1. 高品質なデータ変換



T. Kaneko et al., "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Interspeech 2017.

2. ペアデータなしデータ変換



話者A



ペアデータなし
で変換器を学習



話者B

T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.

例1: CycleGAN-VC v1/v2



CycleGAN-VC [Kaneko+2017] / CycleGAN-VC2 [Kaneko+2019]

- 目的: ペアデータなしの音声変換
 - 話者Aと話者Bで異なる発話内容



T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.
T. Kaneko et al., "CycleGAN-VC2: Improved CycleGAN-based Non-parallel Voice Conversion," ICASSP 2019.

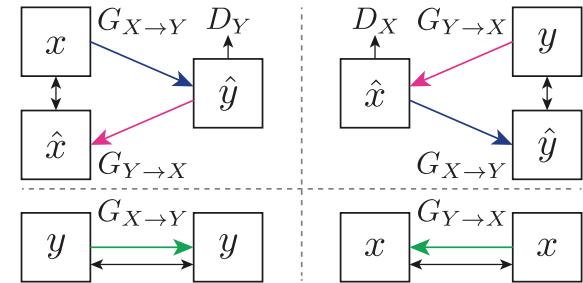
例1: CycleGAN-VC v1/v2 (cont.)



目的関数 (CycleGAN-VC)

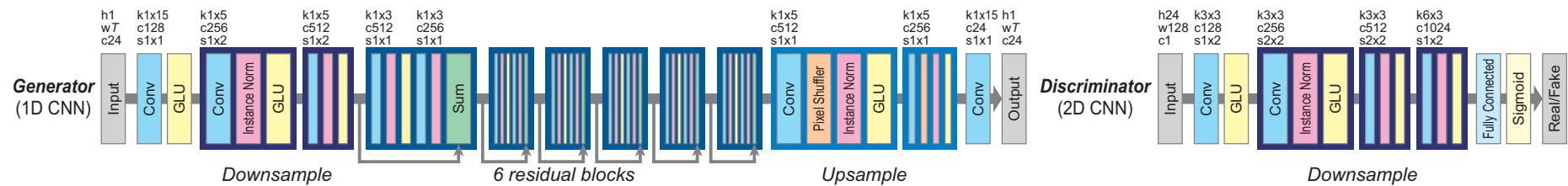
- CycleGANのLoss
 - Adversarial Loss + Cycle-consistency Loss
- Identity-Mapping Loss [Taigman+2017]
 - 言語情報の保持

CycleGAN-VC



モデル (CycleGAN-VC)

- Gated CNN [Dauphin+2017] : 時系列構造のモデル化



T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.

Y. Taigman et al. "Unsupervised Cross-domain Image Generation," ICLR 2017.

Y. N. Dauphin et al., "Language Modeling with Gated Convolutional Networks," ICML 2017.

例1: CycleGAN-VC v1/v2 (cont.)



変換例 (音声特徴量を変換) (CycleGAN-VC2)



※ 学習時はペアデータなし
データ量は各話者81文 (5分程度)

サンプル音声 : <http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/cyclegan-vc2/index.html>

音声データでもCycleGANでペアデータなし変換可能

T. Kaneko & K. Kameoka, "Parallel-Data-Free Voice Conversion Using Cycle-Consistent Adversarial Networks," arXiv 2017.
T. Kaneko et al., "CycleGAN-VC2: Improved CycleGAN-based Non-parallel Voice Conversion," ICASSP 2019.

例2: WaveCycleGAN v1/v2

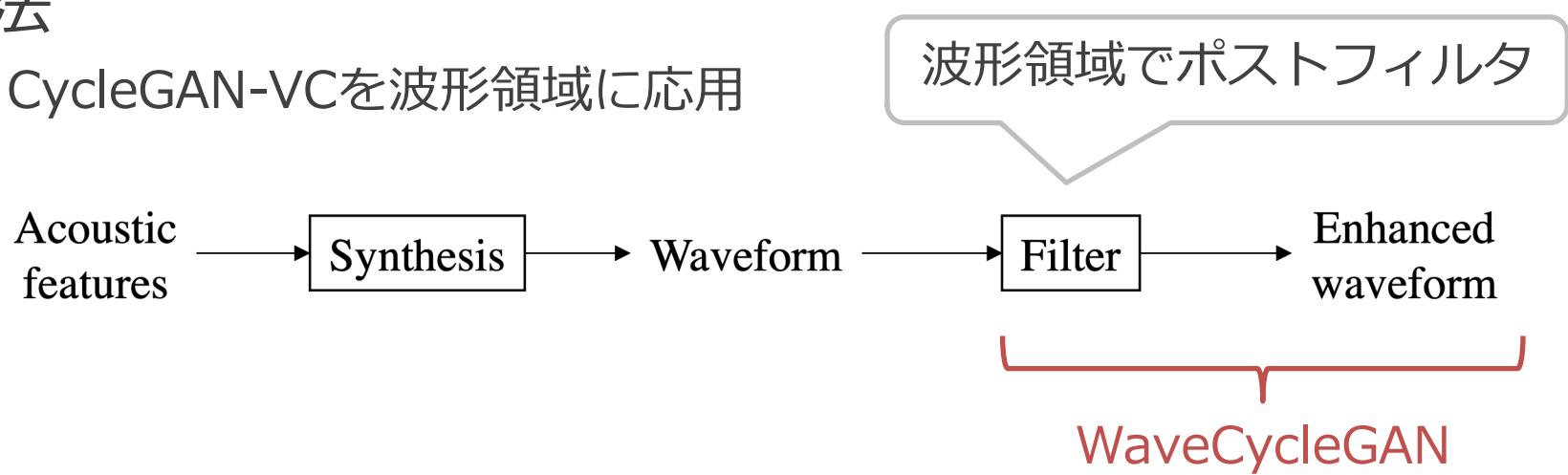


WaveCycleGAN [Tanaka+2018] / WaveCycleGAN2 [Tanaka+2019]

- 目的: 統計的テキスト合成音声/音声変換における過剰な平滑化の緩和
 - 波形領域で変換することで、ボコーダーによる影響を緩和
 - ペアデータなしで解くことで、アライメント処理不要

手法

- CycleGAN-VCを波形領域に応用



K. Tanaka et al. (T. Kaneko共著), "Speech-to-Natural Speech Waveform Conversion Using Cycle-Consistent Adversarial Networks," SLT 2018.

K. Tanaka et al. (T. Kaneko共著), "WaveCycleGAN2: Time-domain Neural Post-filter for Speech Waveform Generation," arXiv 2019.

例2: WaveCycleGAN v1/v2 (cont.)



変換例 (波形を変換) (WaveCycleGAN2)



サンプル音声 : <http://www.kecl.ntt.co.jp/people/tanaka.ko/projects/wavecyclegan2/index.html>

波形領域でもCycleGANでペアデータなし変換可能

K. Tanaka et al. (T. Kaneko共著), "Speech-to-Natural Speech Waveform Conversion Using Cycle-Consistent Adversarial Networks," SLT 2018.

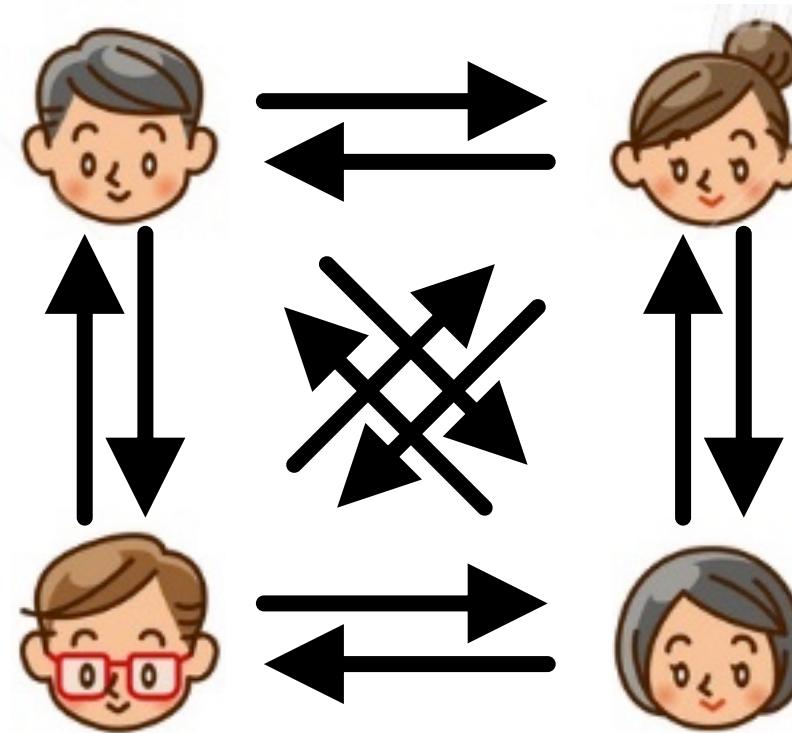
K. Tanaka et al. (T. Kaneko共著), "WaveCycleGAN2: Time-domain Neural Post-filter for Speech Waveform Generation," arXiv 2019.

例3: StarGAN-VC v1/v2



StarGAN-VC [Kameoka+2018] / StarGAN-VC2 [Kaneko+2019]

- 目的: マルチソース/ターゲットでパラレルデータなしの音声変換



H. Kameoka et al. (T. Kaneko共著), "StarGAN: Non-parallel Many-to-Many Voice Conversion with Star Generative Adversarial Networks," SLT 2018.

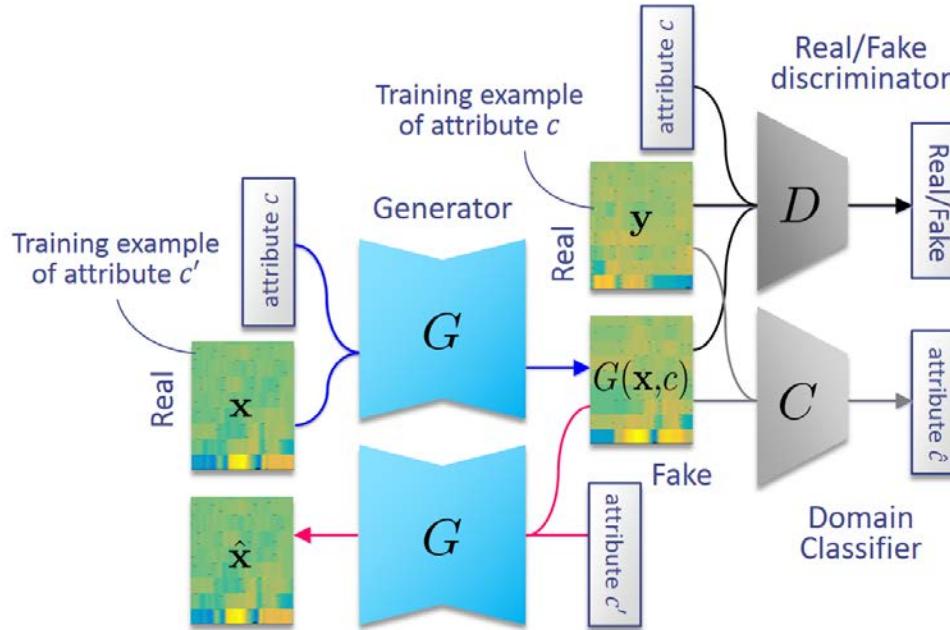
T. Kaneko et al. "Rethinking Conditional Methods for StarGAN-Based Voice Conversion," Interspeech 2019.

例3: StarGAN-VC v1/v2 (cont.)



手法 (StarGAN-VC1)

- StarGANを応用
 - Generator: 話者コードで条件付け
 - Discriminator: 話者識別器と一緒に学習



H. Kameoka et al. (T. Kaneko共著), "StarGAN: Non-parallel Many-to-Many Voice Conversion with Star Generative Adversarial Networks," SLT 2018.

T. Kaneko et al. "Rethinking Conditional Methods for StarGAN-Based Voice Conversion," Interspeech 2019.

例3: StarGAN-VC v1/v2 (cont.)



変換例 (メルケプストラムを変換) (StarGAN-VC2)



※ 単一のモデルで全ての変換を実現

サンプル音声 : <http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/stargan-vc2/index.html>

音声データでもStarGANでマルチドメイン変換可能

H. Kameoka et al. (T. Kaneko共著), "StarGAN: Non-parallel Many-to-Many Voice Conversion with Star Generative Adversarial Networks," SLT 2018.

T. Kaneko et al. "Rethinking Conditional Methods for StarGAN-Based Voice Conversion," Interspeech 2019.

まとめ

基礎：GANとは？

- 騙す G VS 騙されない D ：敵対的に学習



発展：様々な拡張モデルが提案

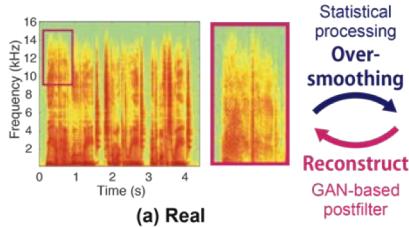
潜在変数構造の工夫
操作性向上

ネットワークの工夫
高解像度化

目的関数の工夫
学習の安定化

データ設定の工夫
不完全データに対応

応用：GANによって精度向上・新しいタスクの実現



今後も面白いことを
どんどん実現できる
可能性あり！

高品質なデータ変換 [Kaneko+2017] ペアデータなしデータ変換 [Kaneko+2017]

解説記事

208

日本音響学会誌 74 卷 4 号 (2018), pp. 208–218

解説

Generative Adversarial Networks の基礎と応用*

金子卓弘 (日本電信電話株式会社)**

PDF無料ダウンロード可

https://www.jstage.jst.go.jp/article/jasj/74/4/74_208/_pdf

金子のWebページ 「Takuhiro Kaneko NTT」で検索

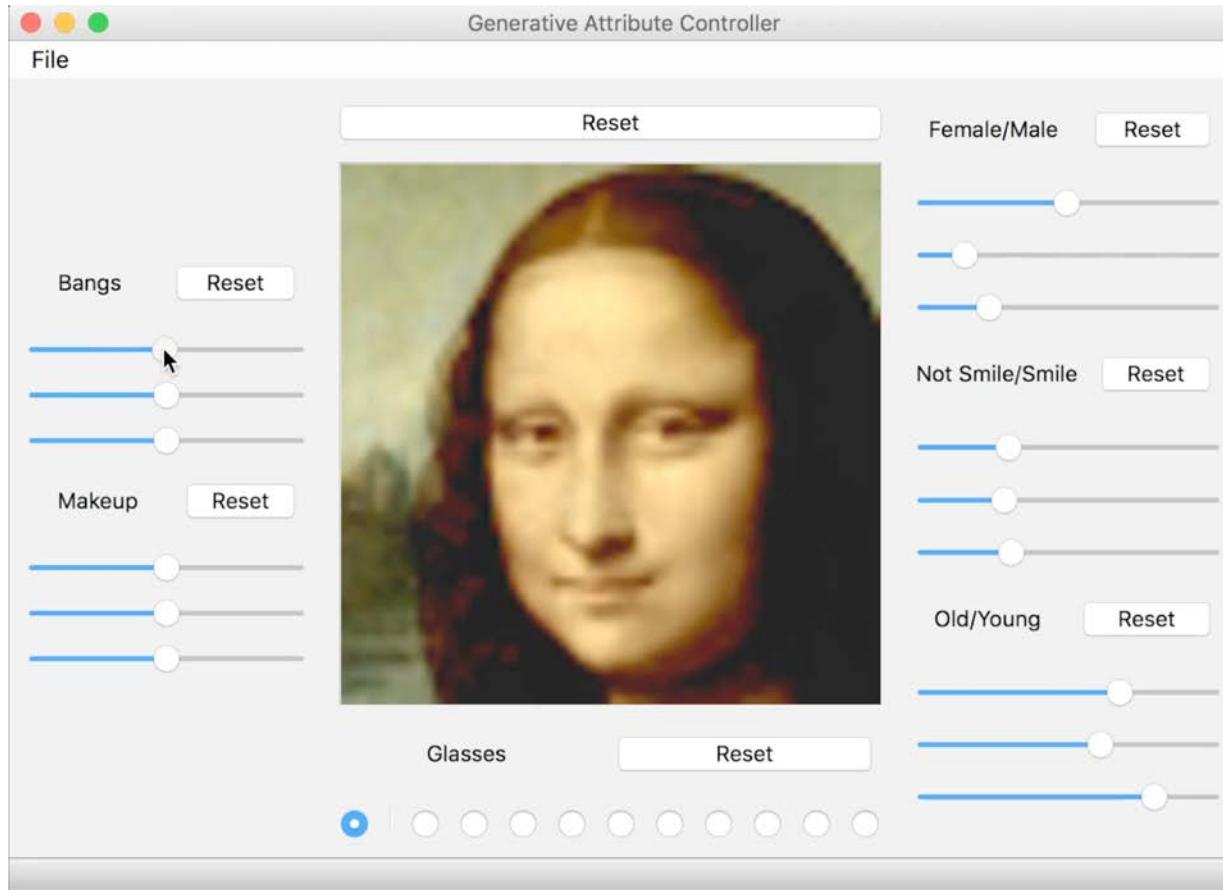


Takuhiro Kaneko
Research Scientist
NTT Communication Science Laboratories, NTT Corporation
kaneko.takuhiro at lab.ntt.co.jp
[Google Scholar]

デモ・コードあり

<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/index.html>

Q & A



CFGAN Demo

デモ動画：<http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/gac/index.html>

T. Kaneko et al., "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," CVPR 2017.