

# Commutative and Nonlinear Image Manipulation for Deep Generative Model (深層生成モデルのための 可換かつ非線形な画像編集)

---

青嶋 雄大

# 背景

深層生成モデル (e.g., GAN, VAE)

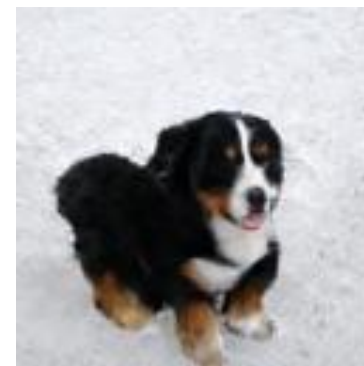
○高精度に画像を生成可能.

✗生成画像の意味的な編集は困難.

➤深層生成モデルのための画像編集に取り組む.



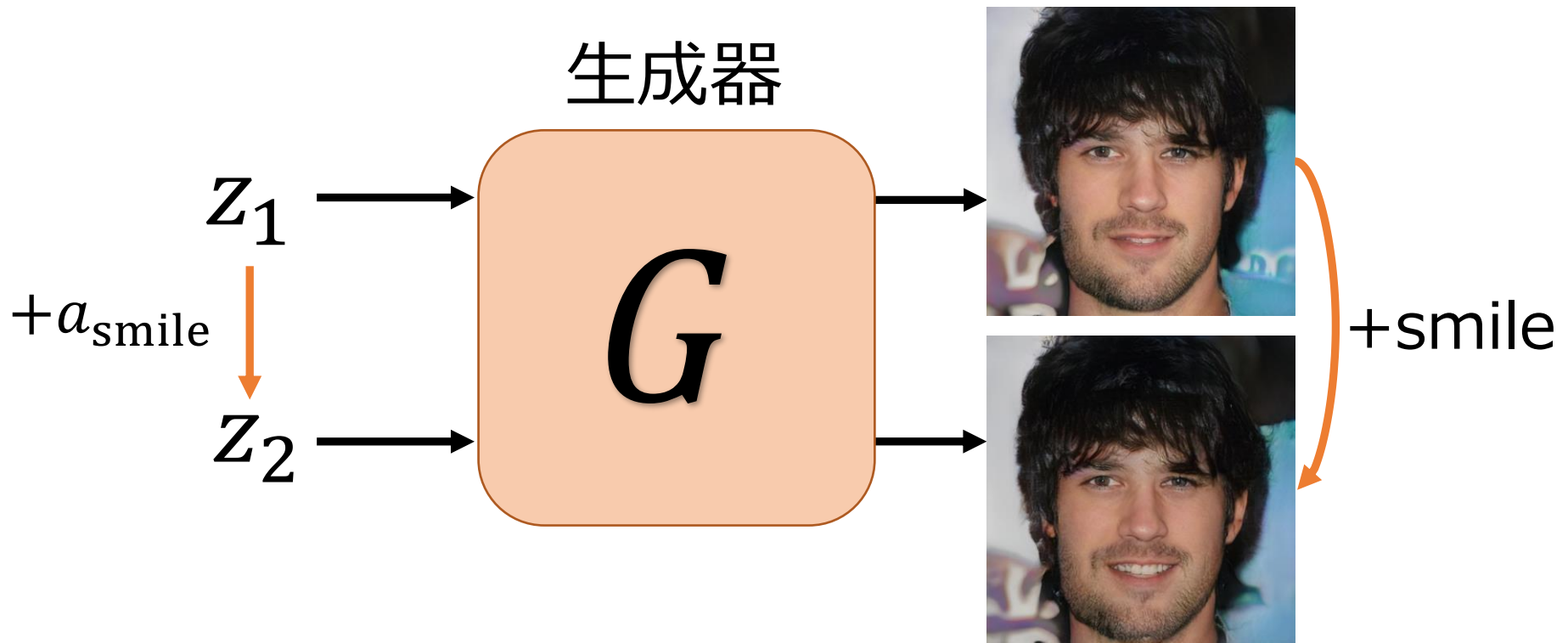
顔の向き



犬の大きさ

# 背景

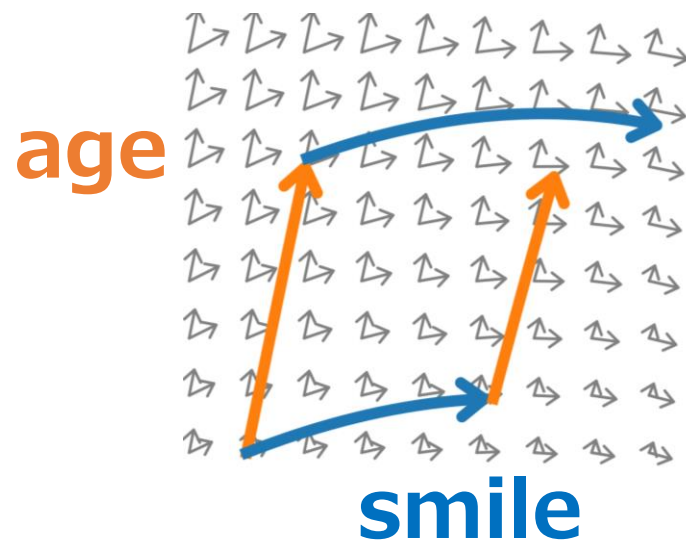
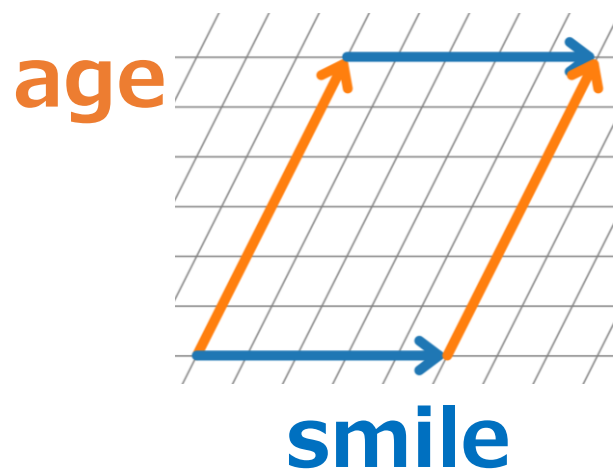
深層生成モデルは潜在変数 $z$ から画像を生成  
➤ 潜在変数を操作することで画像編集が可能.



# 背景

## 潜在変数の操作による画像編集

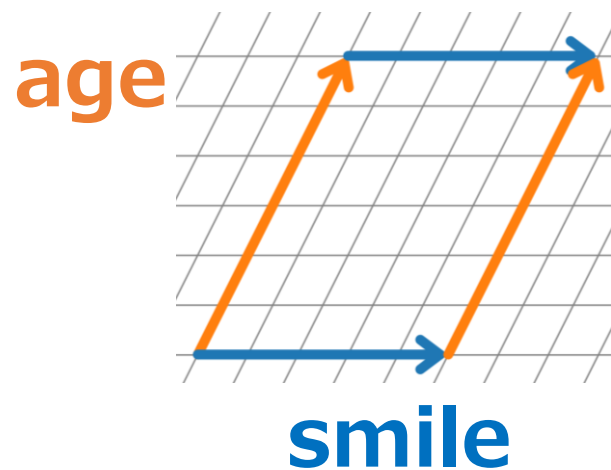
- 潜在空間における意味的な経路を自動で抽出.
- 抽出された経路に沿って潜在変数を操作.
- 操作した潜在変数を生成器に入力して画像を生成.



# 関連研究

## 線形な経路を用いる手法

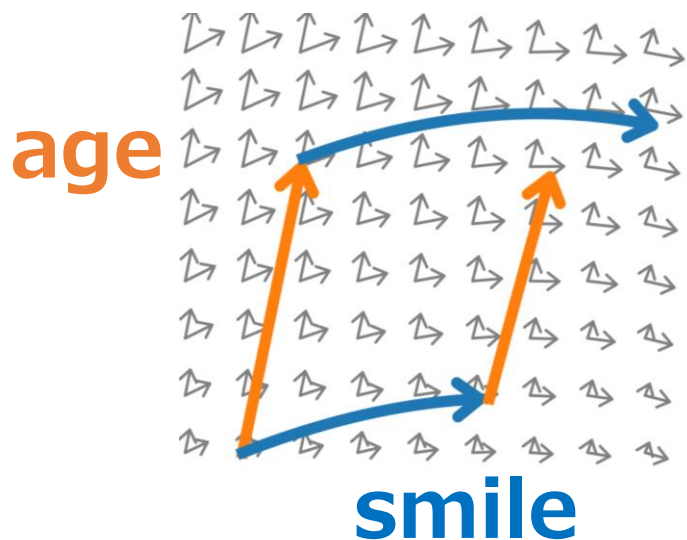
- 潜在空間上の斜交座標系を学習.
- 斜交座標系の軸に沿って潜在変数进行操作.
- 複数属性の可換な編集が可能.  
(順番によらない)
- ✗ 意味的な経路の抽出に失敗することがある.



# 関連研究

非線形な経路を用いる手法

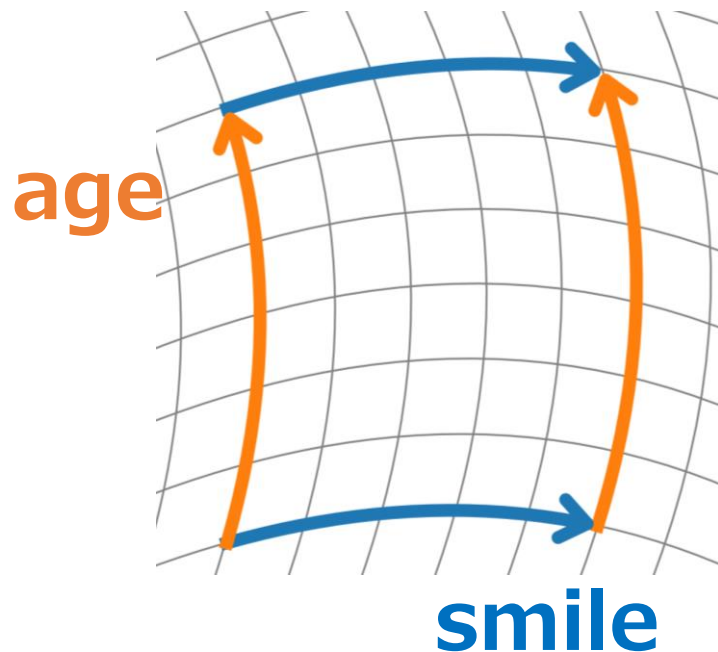
- ベクトル場を学習.
- ベクトル場に沿って潜在変数を操作.
- より柔軟な経路の抽出が可能.
- ✗複数属性の可換な編集ができない.



# 提案手法

意味的な曲線座標系を抽出する手法を提案

- **曲線座標系は可換なベクトル場.**
- 複数属性の可換な編集が可能.
- 柔軟な経路の抽出が可能.



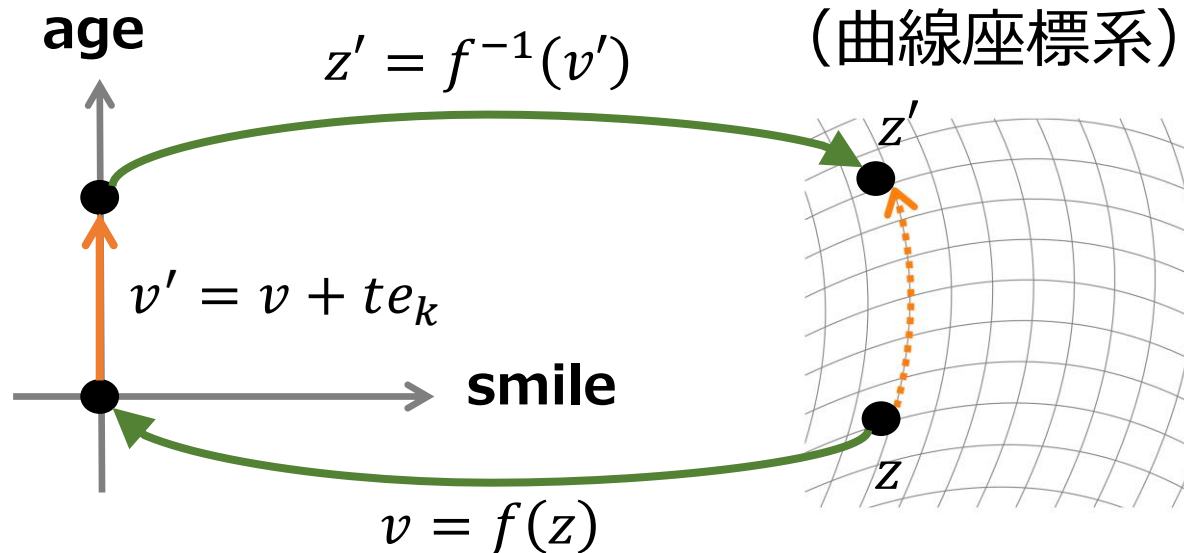
# 提案手法

曲線座標系の軸に沿った操作のために,

- 潜在空間  $z$  と同じ次元数のユークリッド空間  $\mathcal{V}$  を用意.
- 滑らかで可逆な写像  $f: z \rightarrow \mathcal{V}$  により潜在空間  $z$  上に曲線座標系を定義.

ユークリッド空間  $\mathcal{V}$   
(デカルト座標系)

潜在空間  $z$   
(曲線座標系)



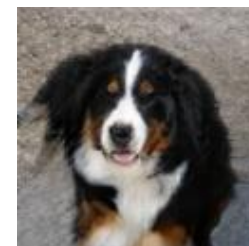
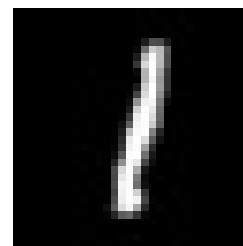


# 実験

事前学習済みGANのための教師なし学習の  
フレームワーク [1]を用いて写像 $f$ を学習.

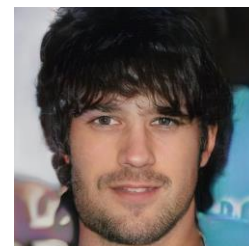
## 比較手法

- LinearGANSpace (線形) [1]
- WarpedGANSpace (非線形) [2]



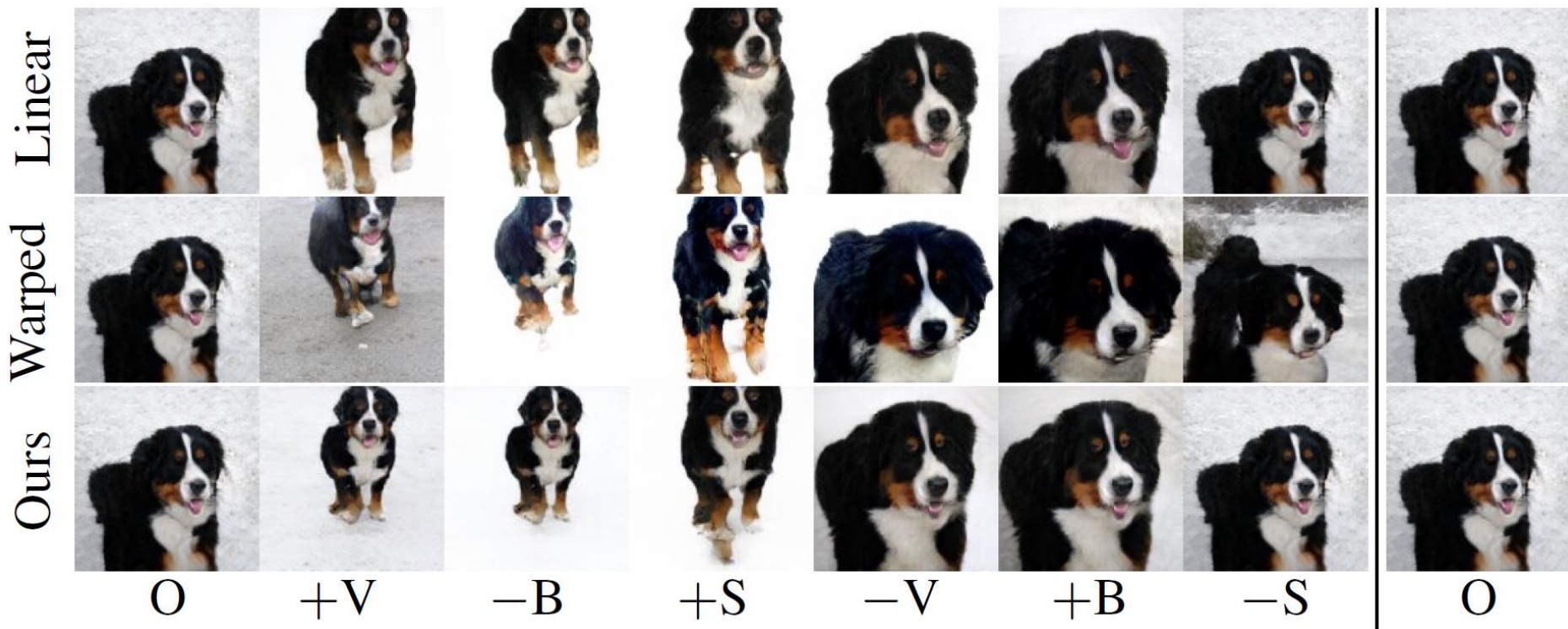
## 事前学習済みGAN

- MNISTで学習したSNGAN.
- AnimeFacesで学習したSNGAN.
- ILSVRCで学習したBigGAN.
- CelebA-HQで学習したProgGAN.
- CelebA-HQで学習したStyleGAN2.



# 結果 (可換性の検証)

- 複数の属性に順番に編集した結果.



O: 元画像, V: 縦移動, B: 背景の濃さ, S: 大きさ

# 結果 (可換性の検証)

- 複数の属性に順番に編集した結果.

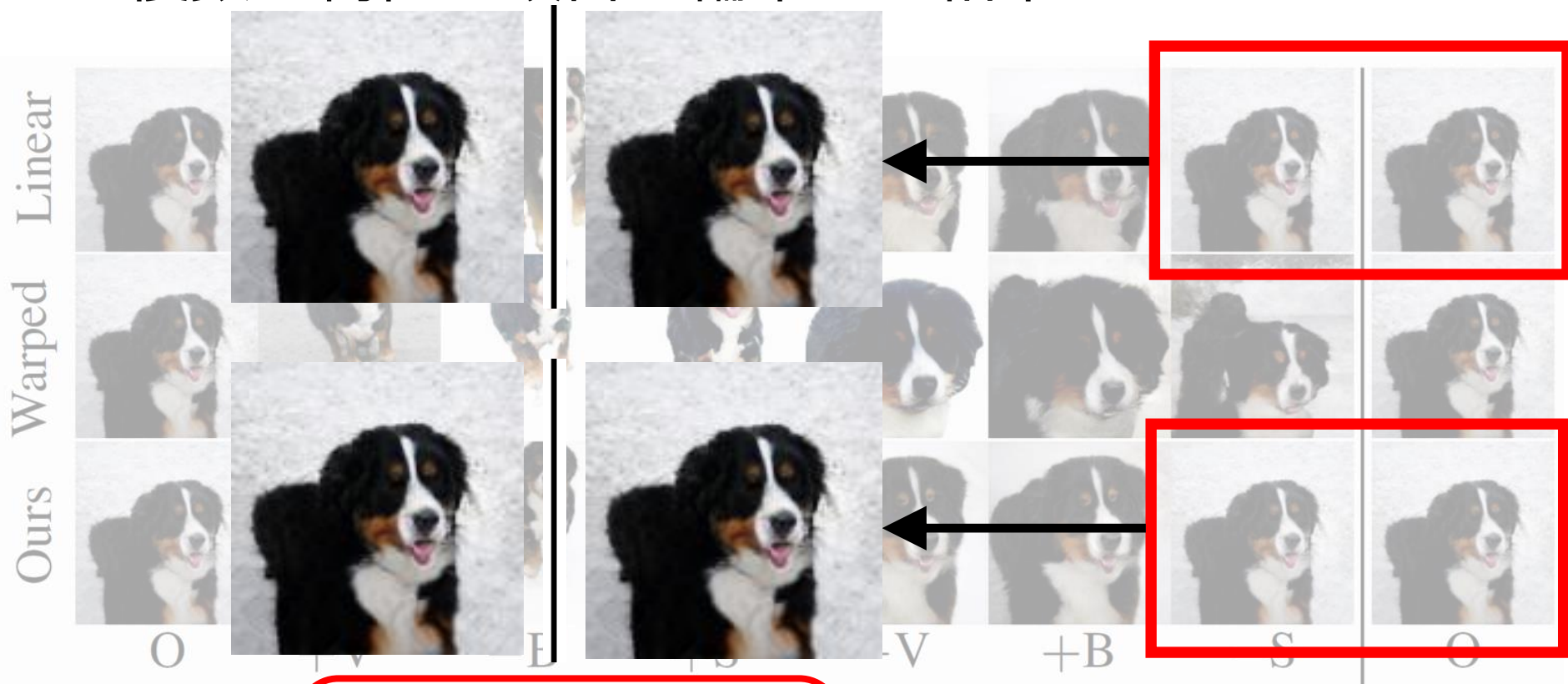


元の背景や大きさに  
戻っていない

背景の濃さ, S: 大きさ

# 結果 (可換性の検証)

- 複数の属性に順番に編集した結果.



O: テクスチャ, V: 背景の濃さ, S: 大きさ

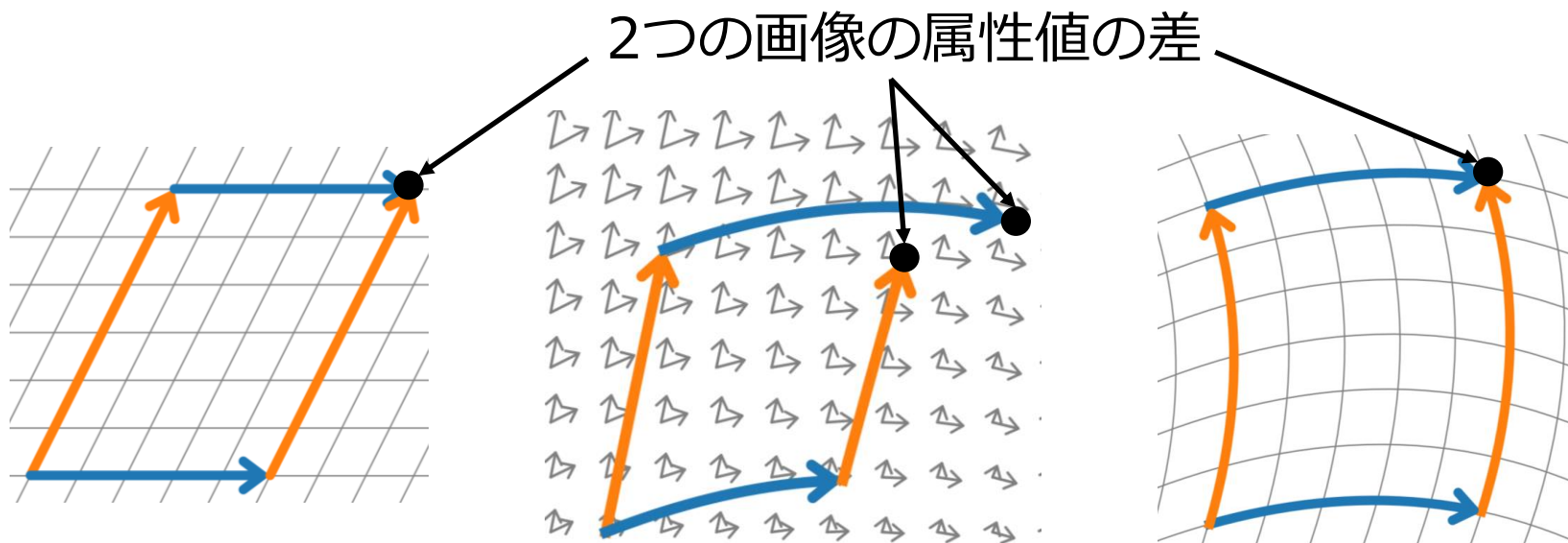
可換な編集が  
できている

# 結果 (可換性の検証)

元画像と元に戻した画像の属性値の差 (%).

CelebA-HQ+StyleGAN2	A+G	R+P	B+Y
LinearGANSpace	<b>0.01 / 0.05</b>	<b>0.02 / 0.07</b>	<b>0.02 / 0.15</b>
WarpedGANSpace	11.40 / 6.62	3.15 / 3.46	1.28 / 2.22
提案手法	<u>0.07 / 0.35</u>	<u>0.05 / 0.62</u>	<u>0.08 / 0.55</u>

A: “age”, G: “gender”, R: “race” B: “bangs”, P: “pitch”, Y: “yaw”.

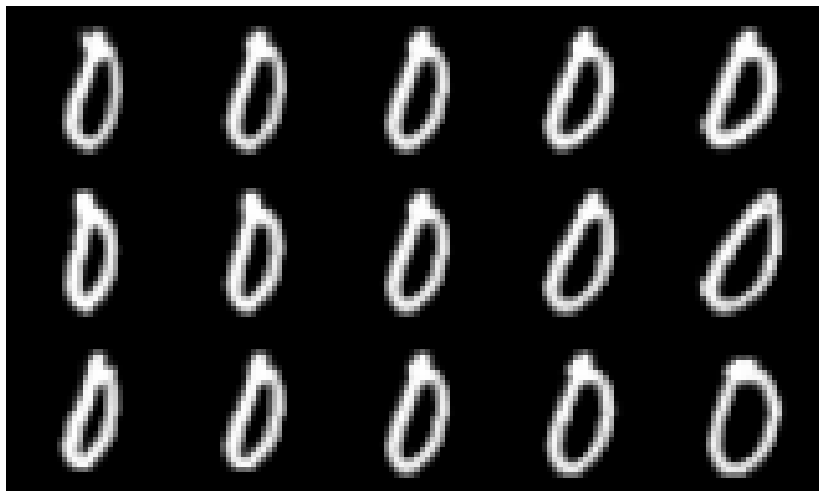


# 結果 (画像編集の質の検証)

色を十分に編集  
できていない

回転している

提案手法  
Warped Linear



MNIST, width.

提案手法  
Warped Linear



AnimeFaces, hair color.

顔が細く  
なっている

提案手法  
Warped Linear

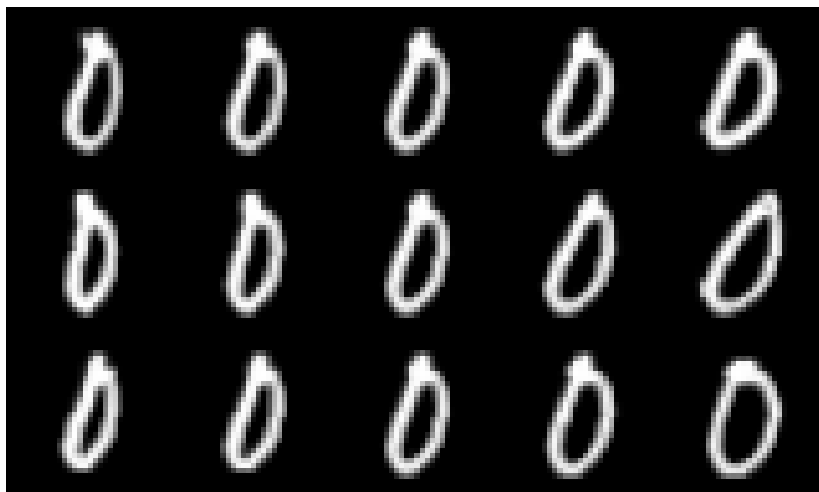


ProgGAN, smile.

Yaw方向に  
回転している

# 結果 (画像編集の質の検証)

提案手法  
Warped  
Linear



MNIST, width.

横幅のみ  
変えている

提案手法  
Warped  
Linear



AnimeFaces, hair color.

髪の色のみ  
変えている

提案手法  
Warped  
Linear



ProgGAN, smile.

笑顔のみ  
変えている

# 結果 (画像編集の質の検証)

対象の属性以外の変化量 (%).

CelebA-HQ+StyleGAN2	Target $k$	Side Effect Errors $l$ [%]					
		A	G	R	B	P	Y
LinearGANSpace	A	<u>100</u>	59	37	63	41	61
	G	28	<u>100</u>	16	78	20	17
	R	61	52	<u>100</u>	71	24	19
	B	<b>175</b>	<b>172</b>	78	<u>100</u>	70	64
	P	71	<b>90</b>	43	76	<u>100</u>	57
	Y	58	55	43	<b>94</b>	36	<u>100</u>
WarpedGANSpace	A	<u>100</u>	51	63	<b>111</b>	59	23
	G	75	<u>100</u>	<b>94</b>	<b>124</b>	<b>236</b>	57
	R	63	64	<u>100</u>	<b>131</b>	73	25
	B	23	27	22	<u>100</u>	15	21
	P	41	44	30	80	<u>100</u>	41
	Y	30	30	22	<b>97</b>	23	<u>100</u>
提案手法	A	<u>100</u>	80	45	<b>137</b>	60	37
	G	62	<u>100</u>	50	84	61	40
	R	65	56	<u>100</u>	60	37	23
	B	40	38	15	<u>100</u>	14	19
	P	60	52	36	76	<u>100</u>	44
	Y	41	62	21	79	21	<u>100</u>

A : “age”, G : “gender”, R : “race” B : “bangs”, P : “pitch” Y : “yaw”.



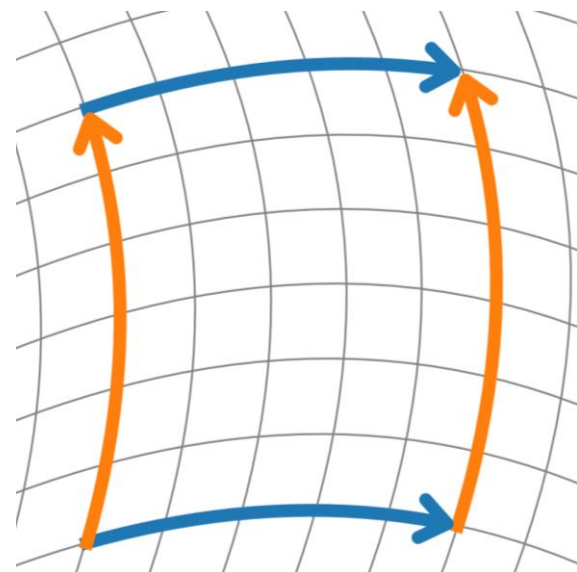
# 結論

深層生成モデルのための画像編集の手法として、  
潜在空間における曲線座標系を学習することを提案

- 複数属性の可換な編集が可能.
- より質の高い画像編集が可能.

今後の課題

- 他の深層生成モデルに適用.



# Appendix

---

# 関連研究

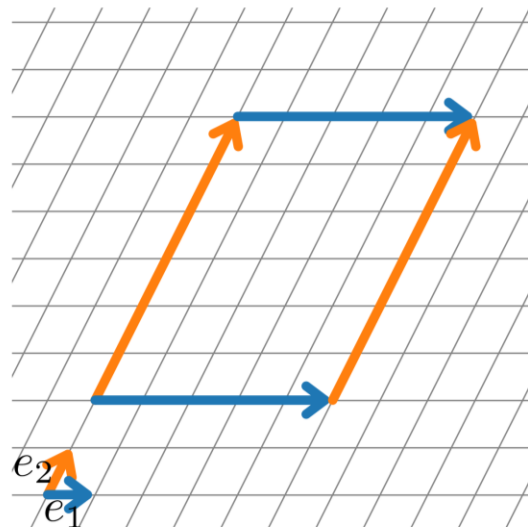
線形な経路を用いる手法 (e.g., [1])

- 以下のように潜在変数を操作

$$z' = z + tAe_k$$

$z$ : 潜在変数,  $t$ : 実数,  $A$ : 行列,  
 $e_k$ : 単位ベクトル,  $k$ : 属性

- 行列 $A$ を学習する



# 関連研究

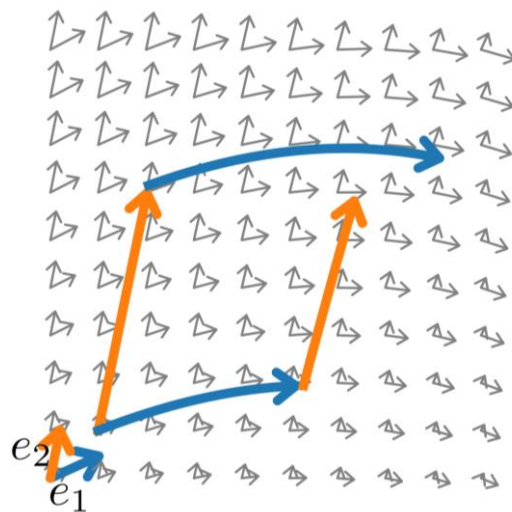
非線形な経路を用いる手法 (e.g., [2])

- 以下のように潜在変数 $z$ を操作.

$$z' = z + \int_0^t X_k(z(\tau)) d\tau$$

$X_k$ : ベクトル場

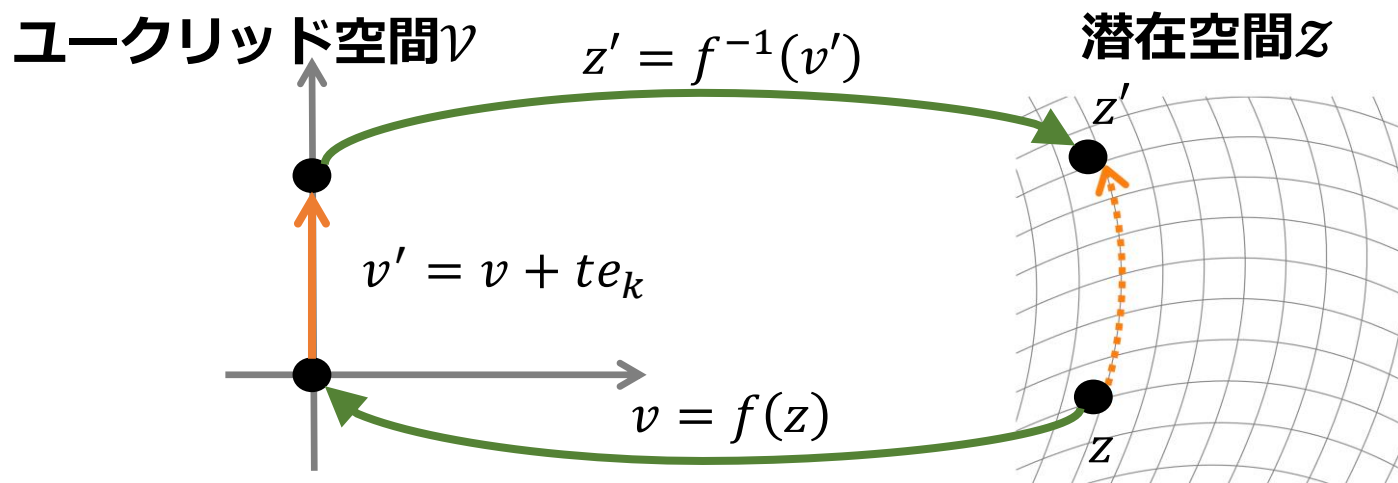
- ベクトル場 $X_k$ を学習する.
- 数値積分法を用いることで積分が可能.



# 提案手法

提案手法による編集が可換であることの証明

$$\begin{aligned}\psi_k(v) &= v + e_k, & \phi_k(z) &= f * \psi_k * f^{-1}(z) \text{ とすると,} \\ \phi_k * \phi_l &= f * \psi_k * f^{-1} * f * \psi_l * f^{-1} \\ &= f * \psi_k * \psi_l * f^{-1} \\ &= f * \psi_l * \psi_k * f^{-1} \\ &= f * \psi_l * f^{-1} * f * \psi_k * f^{-1} \\ &= \phi_l * \phi_k\end{aligned}$$

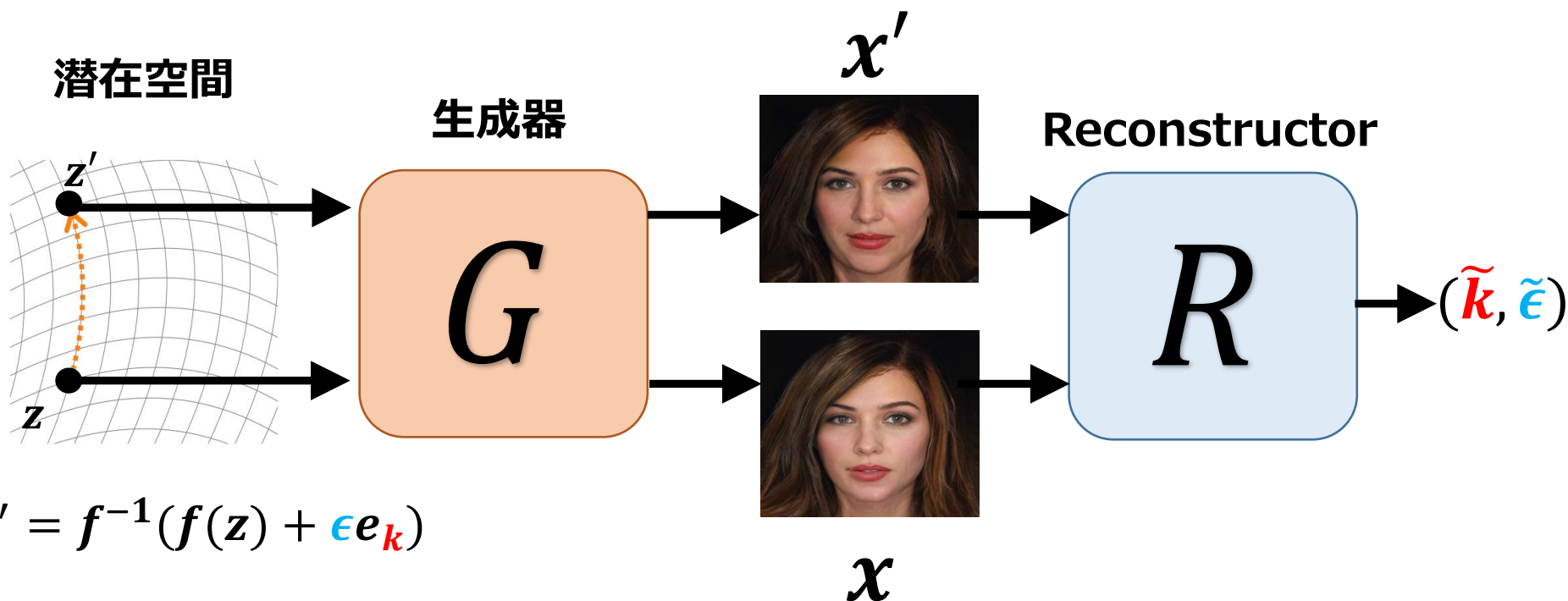


# 提案手法

実験に使用した学習フレームワーク

学習する関数

- 滑らかで可逆な写像  $f$  (Normalizing flow)
- Reconstructor  $R$  ( $(\tilde{k}, \tilde{\epsilon})$ を予測)



# 提案手法

## 損失関数

$$\min_{f,R} \mathbb{E} \left[ \underline{\mathcal{L}_{\text{cls}}(k, \tilde{k})} + \lambda \mathcal{L}_{\text{reg}}(\epsilon, \tilde{\epsilon}) + \alpha \left( \log \det \left| \frac{\partial f}{\partial \mathbf{z}} \right| \right)^2 \right]$$

- $\mathcal{L}_{\text{cls}}(k, \tilde{k})$  はクロスエントロピー
  - 分類問題を簡単にするために**分離された経路**を抽出
- これらの一部が**意味的な経路**

# 提案手法

## 損失関数

$$\min_{f, R} \mathbb{E} \left[ \mathcal{L}_{\text{cls}}(k, \tilde{k}) + \lambda \mathcal{L}_{\text{reg}}(\epsilon, \tilde{\epsilon}) + \alpha \left( \log \det \left| \frac{\partial f}{\partial \mathbf{z}} \right| \right)^2 \right]$$

- $\mathcal{L}_{\text{reg}}(\epsilon, \tilde{\epsilon})$  は絶対値誤差
- 抽出される経路が連続的な変化をもたらすようになる



← Shift along direction →

図引用: [1] Voynov+, ICML2020.

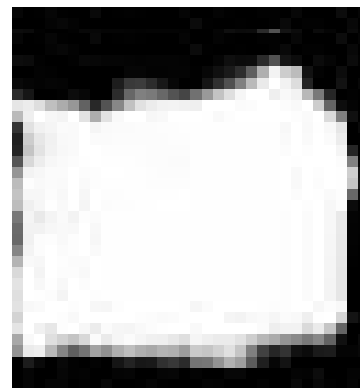
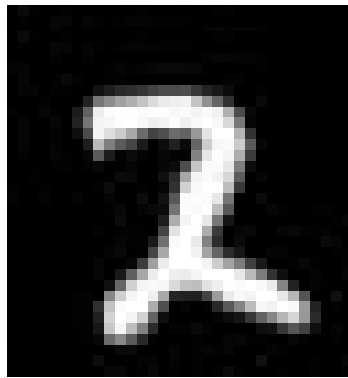
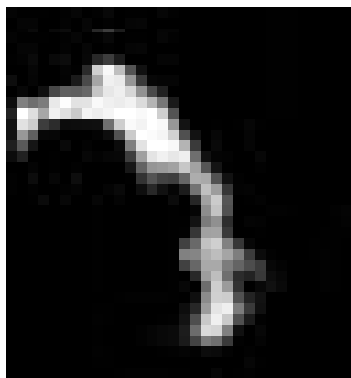


# 提案手法

## 損失関数

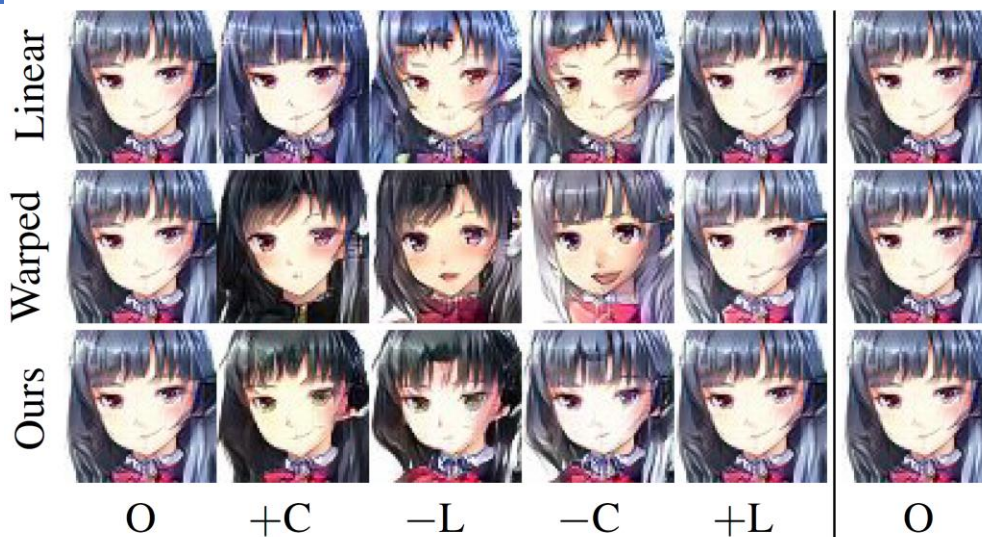
$$\min_{f, R} \mathbb{E} \left[ \mathcal{L}_{\text{cls}}(k, \tilde{k}) + \lambda \mathcal{L}_{\text{reg}}(\epsilon, \tilde{\epsilon}) + \alpha \left( \log \det \left| \frac{\partial f}{\partial \mathbf{z}} \right| \right)^2 \right]$$

- $\log \det \left| \frac{\partial f}{\partial \mathbf{z}} \right|$  はヤコビアン対数の対数
- 小さな操作で画像が大きく変化するのを防ぐ

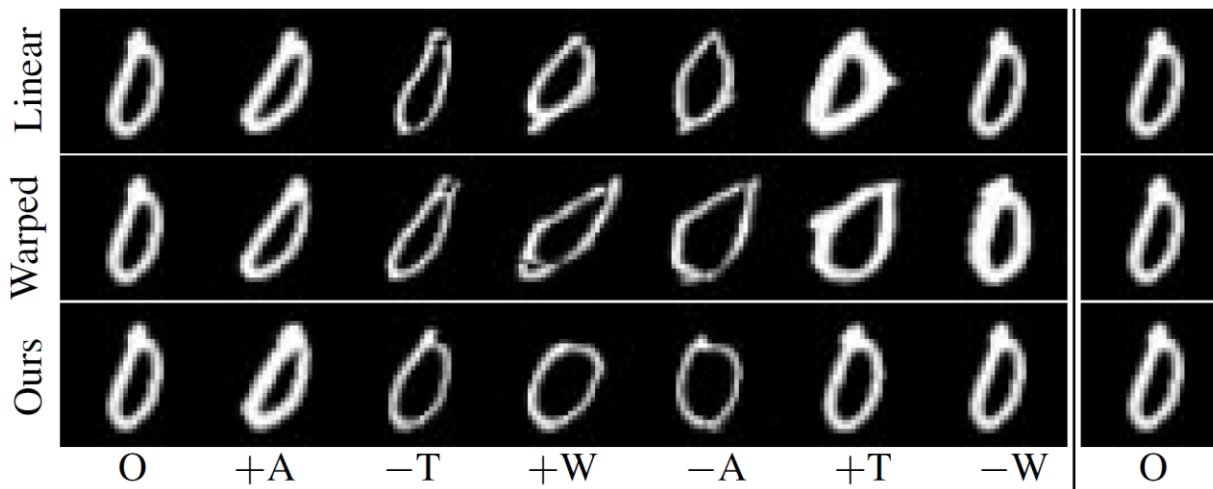


← - Shift along direction + →

# 結果 (可換性の検証)



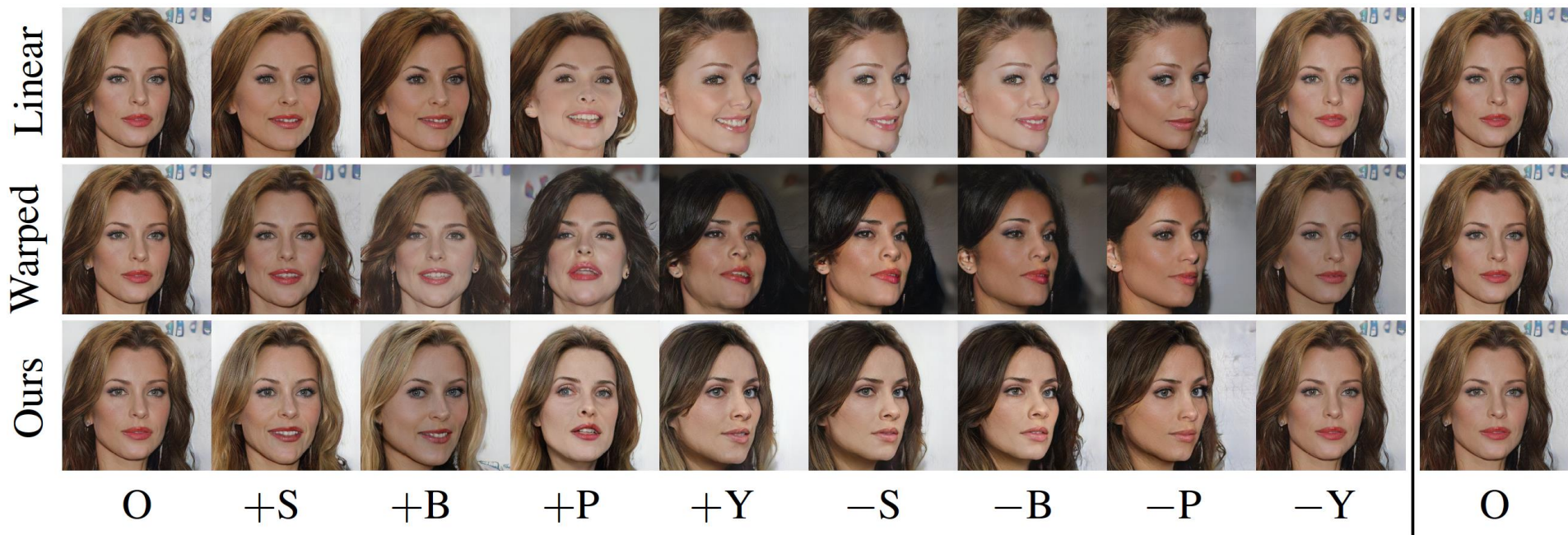
O: original, C: hair color, L: hair length



O: original, A: angle, T: thickness, W: width

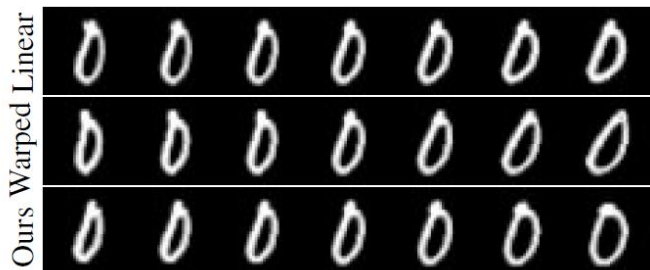
# 結果 (可換性の検証)

- 複数の属性に順番に編集した結果.



O: original, S: Smiling, B: Bangs, P: Pitch, Y: Yaw

# 結果 (side effect)



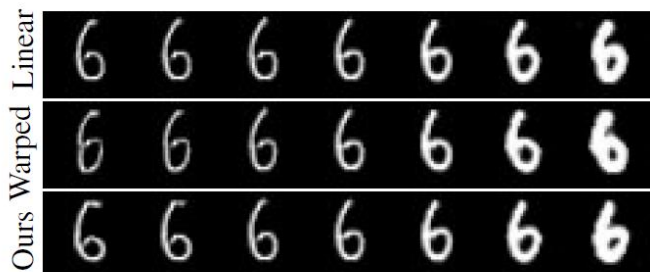
(a) MNIST, "width".



(c) AnimeFaces, "hair color".



(e) ILSVRC, "object size".



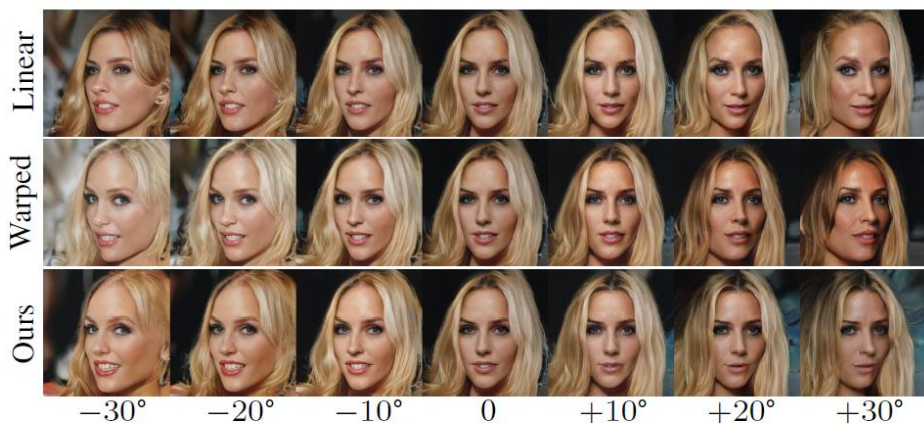
(b) MNIST, "thickness".



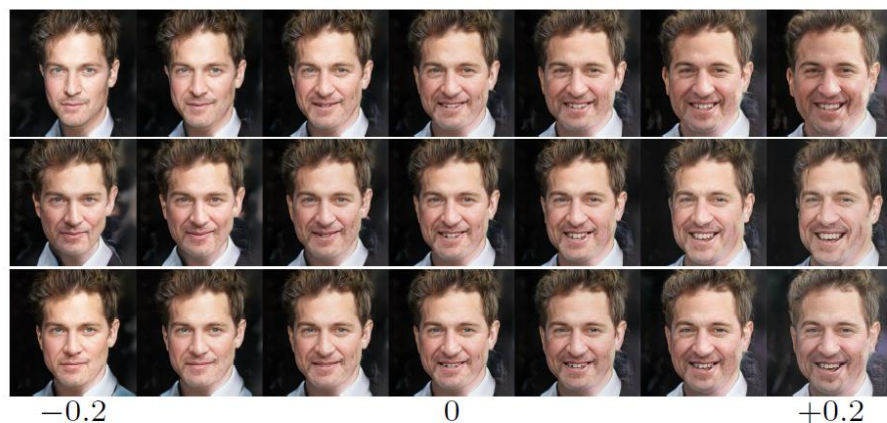
(d) AnimeFaces, "hair length".



(f) ILSVRC, "vertical position".

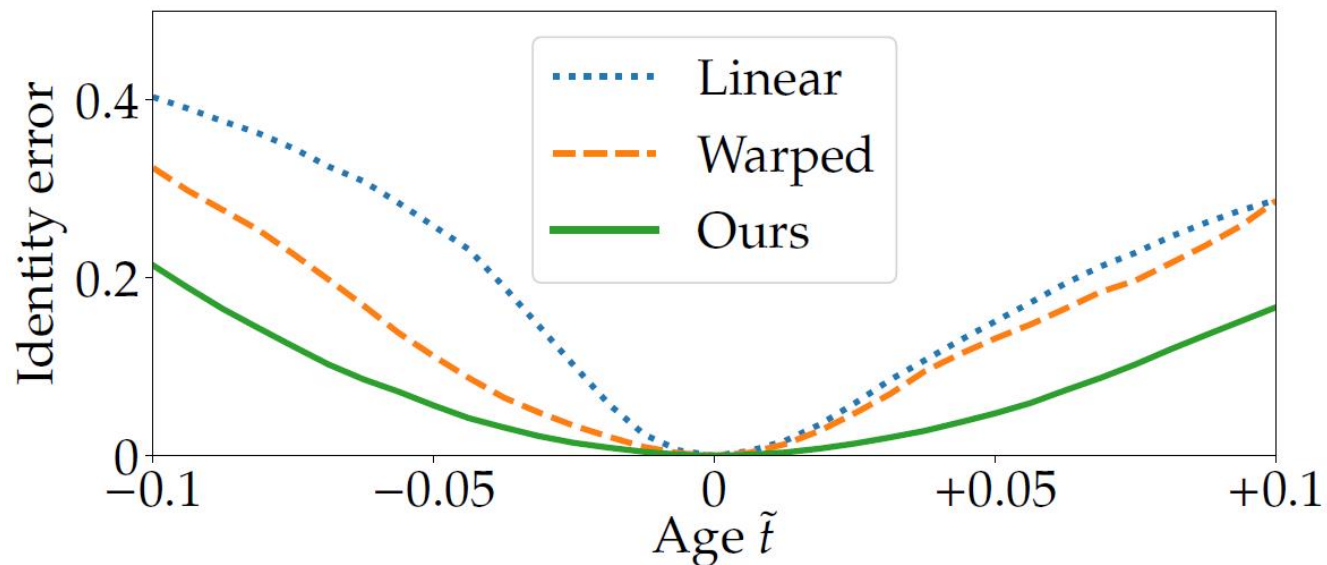


(g) ProgGAN, "yaw".



(h) ProgGAN, "smile".

# 結果 (同一性)



# 結果 (同一性)

同一性の誤差(%).

	<b>A</b>	<b>G</b>	<b>R</b>	<b>B</b>	<b>P</b>	<b>Y</b>	<b>Avg.</b>
LinearGANSpace	<u>26.1</u>	<b>5.5</b>	<b>19.1</b>	47.4	26.4	24.7	29.9
WarpedGANSpace	27.6	56.2	33.6	<u>6.3</u>	<b>14.6</b>	<b>8.4</b>	<u>29.3</u>
CurvilinearGANSpace (ours)	<b>21.1</b>	<u>15.4</u>	<u>25.3</u>	<b>6.0</b>	<u>18.9</u>	<u>9.6</u>	<b>19.2</b>

A: “age”, G: “gender”, R: “race” B: “bangs”, P: “pitch”, Y: “yaw”,  
Avg.: average.