Deep Curvilinear Editing: Commutative and Nonlinear Image Manipulation for Pretrained Deep Generative Model

Takehiro Aoshima, Takashi Matsubara (Osaka University)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

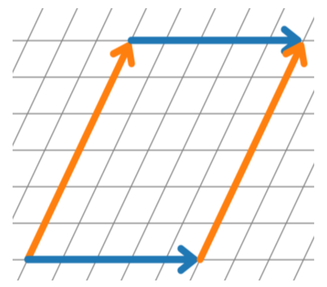
Introduction

Background

- Deep generative models are known for their ability to produce high-quality images.
- They do not provide an inherent way to edit images semantically.
- Several studies have proposed to find linear or nonlinear semantic paths in the latent space of pretrained models.

Related work

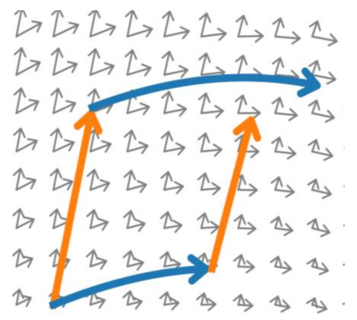
Linear methods (e.g., [1])



✓ provide commutative edits.

✗ fail to discover nonlinear semantic paths.

Nonlinear methods (e.g., [2])

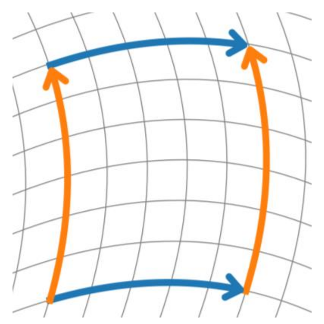


✓ discover nonlinear semantic paths.

✗ do not provide commutative edits.

Our solution

The proposed method



✓ discovers nonlinear semantic paths.

✓ provides commutative edits.

References

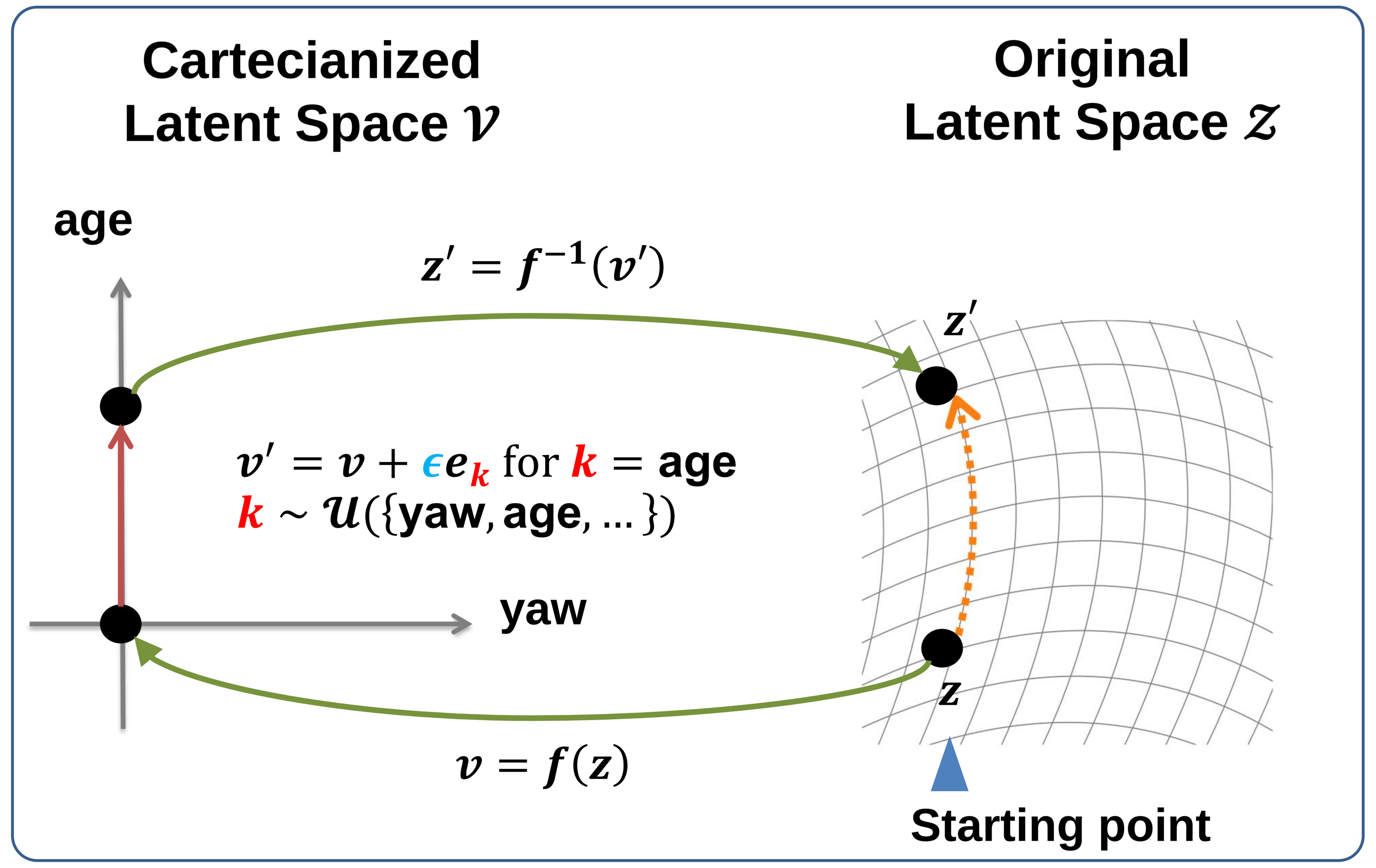
[1] Voynov and Babenko. "Unsupervised Discovery of Interpretable Directions in the GAN Latent Space." In: ICML 2020.

[2] Tzelepis *et al*. "WarpedGANSpace: Finding Non-Linear RBF Paths in GAN Latent Space." In: ICCV 2021.

Method

Deep Curvilinear Editing (DeCurvEd)

- An N-dimensional latent space $\mathcal{Z}$.
- An N-dimensional Euclidean space $\mathcal{V}$ (Cartecianized latent space).
- A bijective function $f: \mathcal{Z} \to \mathcal{V}$ to define a curvilinear coordinate on $\mathcal{Z}$ by transforming a Cartesian coordinate on $\mathcal{V}$.
- Edit a latent code z as
  1. Get a mapped latent code $v = f(z)$.
  2. Edit the mapped latent code $v' = v + \epsilon e_k$.
  3. Get an edited latent code $z' = f^{-1}(v')$.
- This edit is nonlinear and commutative because curvilinear coordinates are equivalent to commuting vector fields.
- We name the proposed method Deep Curvilinear Editing (DeCurvEd).
- DeCurvEd is available for any generative models.

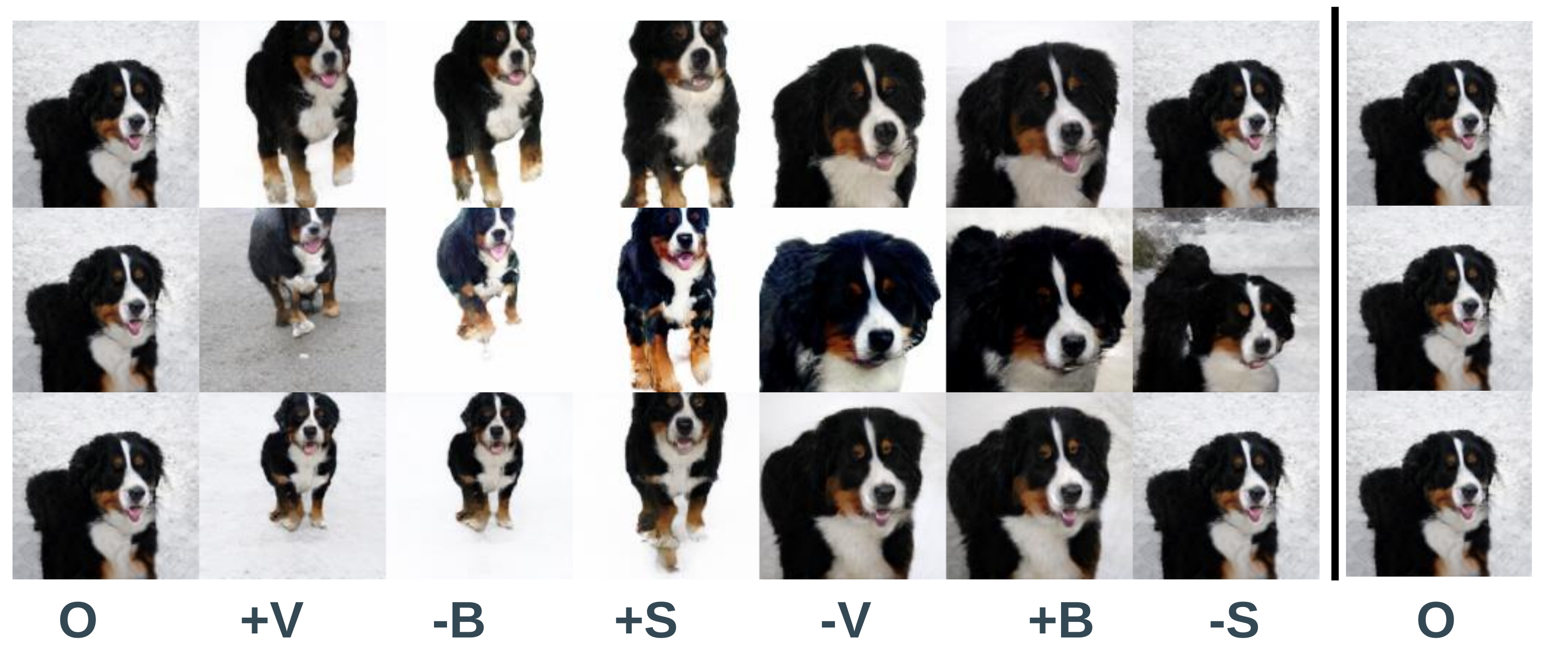


Experimental Results

Commutativity

- Linear method is commutative, but the quality is inferior.
- Nonlinear method offers a better quality, but it is non-commutative.
- Ours method is commutative and offers the best quality.

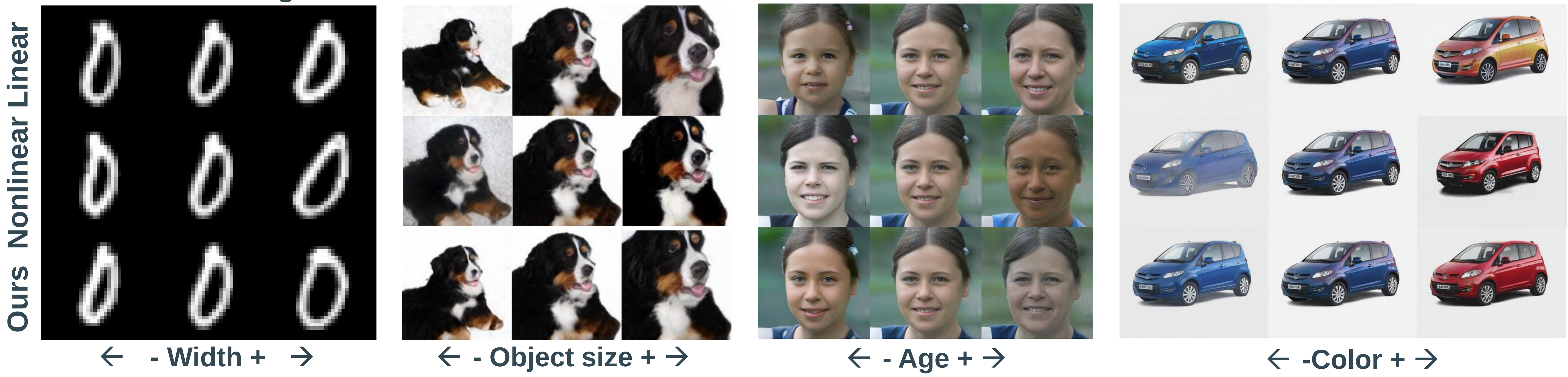
O: original, D: dark colored-hair, L: hair length.
SNGAN + AnimeFaces.



O: original, V: vertical position, S: object size, B: background.
BigGAN + ILSVRC.

Visualization results

- Linear and nonlinear methods exhibit undesirable side effects (e.g., age affects face color).
- Ours edits images without severe side effects.



MNIST + SNGAN | ILSVRC + BigGAN | CelebA-HQ + StyleGAN2 | LSUN Car + StyleGAN2

Identity error

- We calculate identity error to evaluate the disentanglement.
- Ours has the lowest error for two out of six attributes, the second lowest errors for the remaining.

	A	G	R	B	P	Y	Avg.
Linear	<u>26.1</u>	**5.5**	**19.1**	47.4	26.4	24.7	29.9
Nonlinear	27.6	56.2	33.6	<u>6.3</u>	**14.6**	**8.4**	<u>29.3</u>
Ours	**21.1**	<u>15.4</u>	<u>25.3</u>	**6.0**	<u>18.9</u>	<u>9.6</u>	**19.2**

A: age, G: gender, R: race, B: bangs, P: pitch: Y: yaw, Avg.: average. CelebA-HQ + StyleGAN2