CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

CSDS 600: Deep Generative Models

Inversion issue in GANs

Yu Yin (yu.yin@case.edu)
Case Western Reserve University

https://yin-yu.github.io/

Generative Adversarial Network (GAN)

- Improved GANs
  - Problems in GANs
  - Wasserstein GAN (WGAN)
  - WGAN -GP
- Selected GANs
  - Conditional GAN
  - CycleGAN, DualGAN, DiscoGAN
  - High-Resolution Image Generation: Progressive GAN, StyleGAN

CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Interpretable Latent Space and Inversion Issue in GANs

- Walking on the Latent Space
  - Supervised approach
  - Unsupervised approach
- Inversion of real images
  - Optimization-based method
  - Encoder-based method
  - Pivot-tuning

Progress for Image Generation

2014

GAN

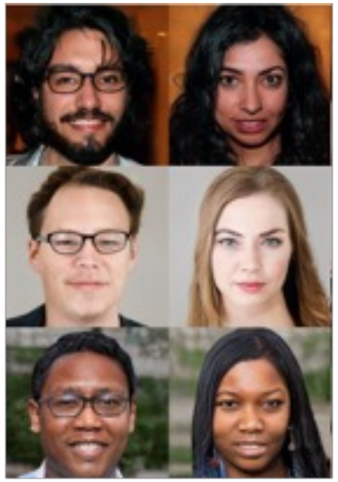
2015

DCGAN

2017

PG-GAN

2018



StyleGAN

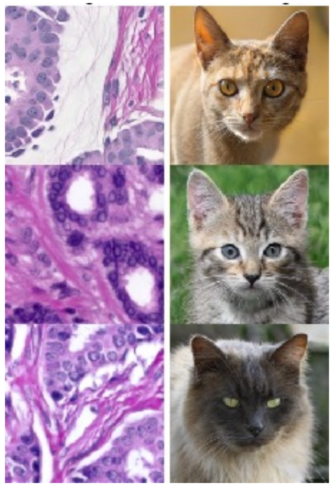
2018

BigGAN

2019

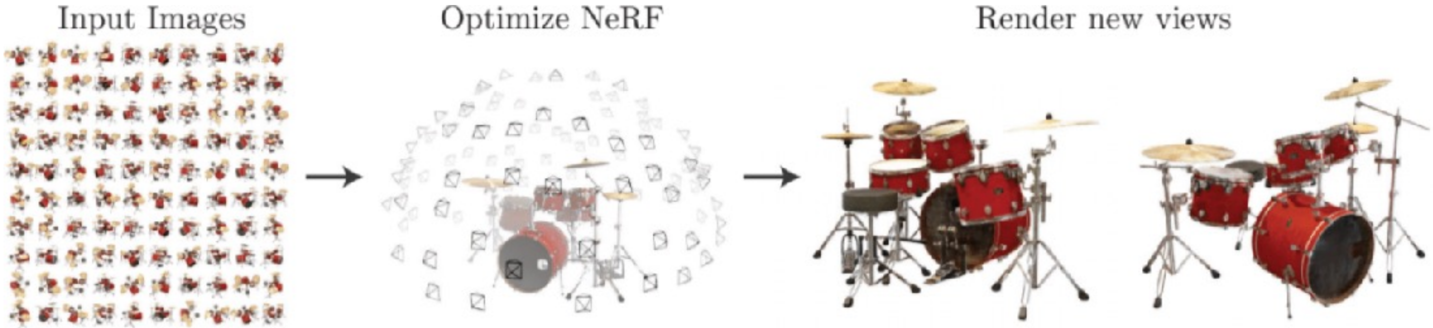
StyleGAN2

2020



StyleGAN-Ada

2020: NeRF (Neural Radiance Fields)

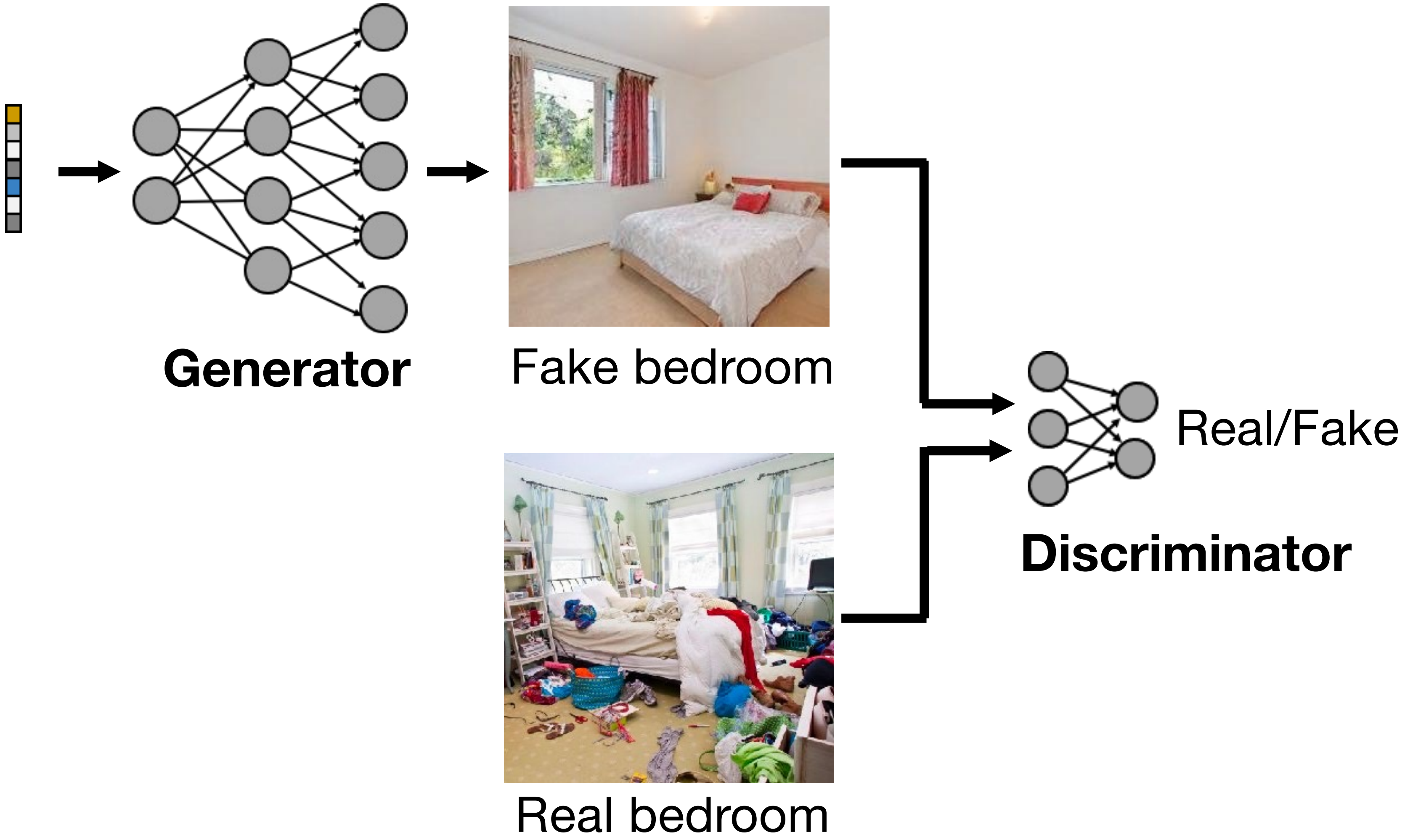


2021: OpenAI DALLE (VQ-VAE)

Arm chair in shape of avocado

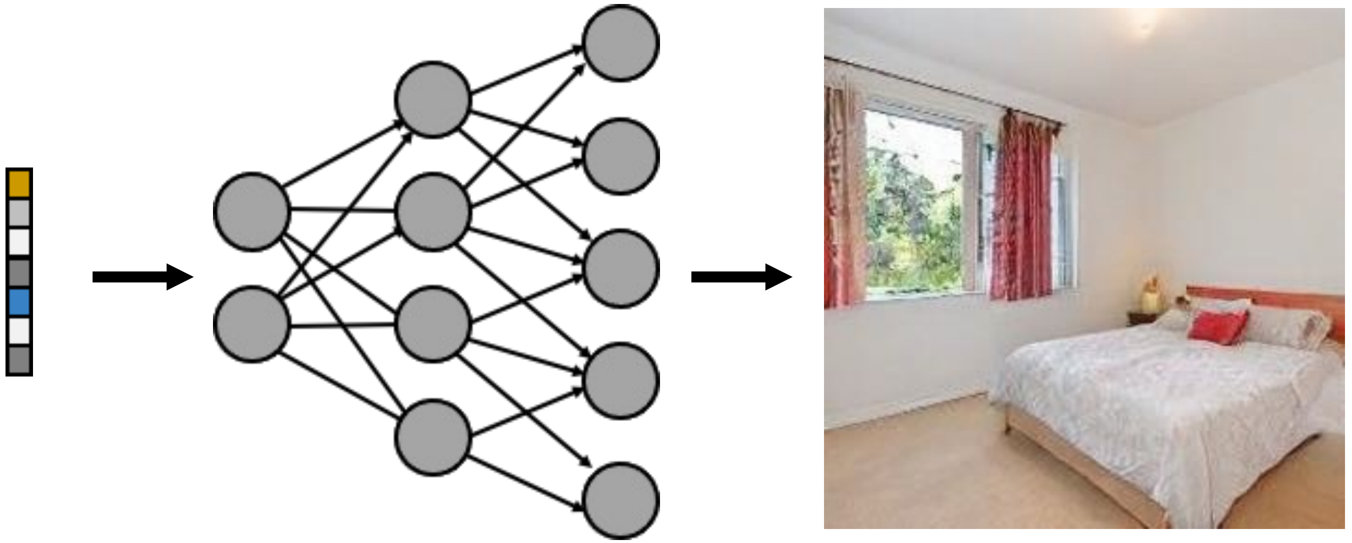
GAN

Adversarial Training

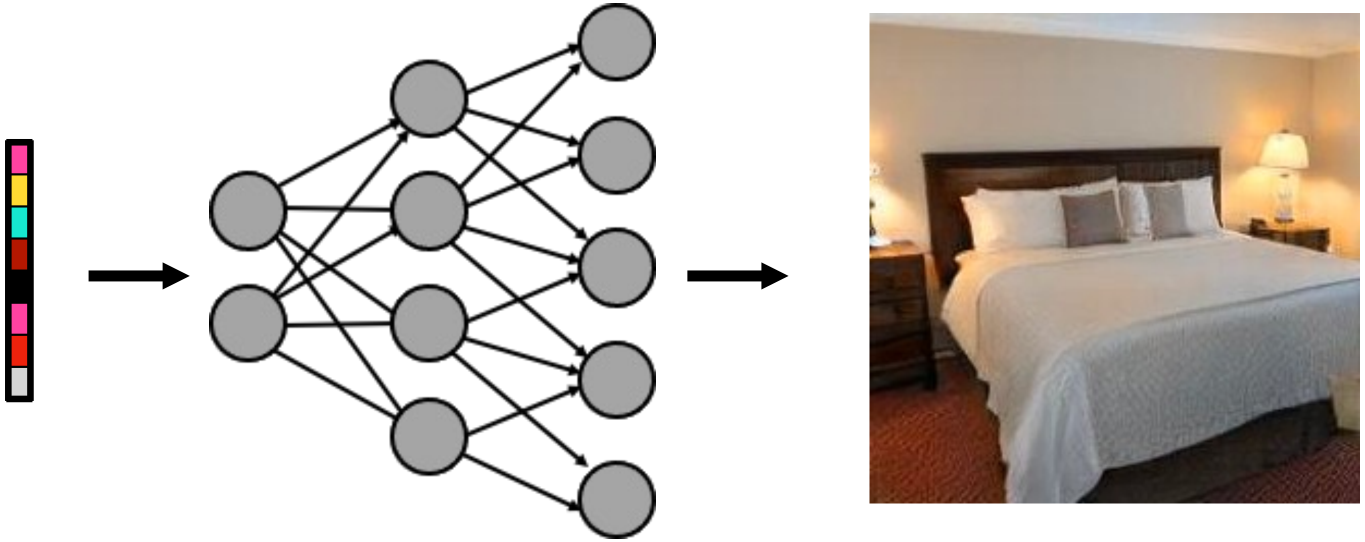


Goodfellow et al. NeurIPS'14

Neural Image Generation

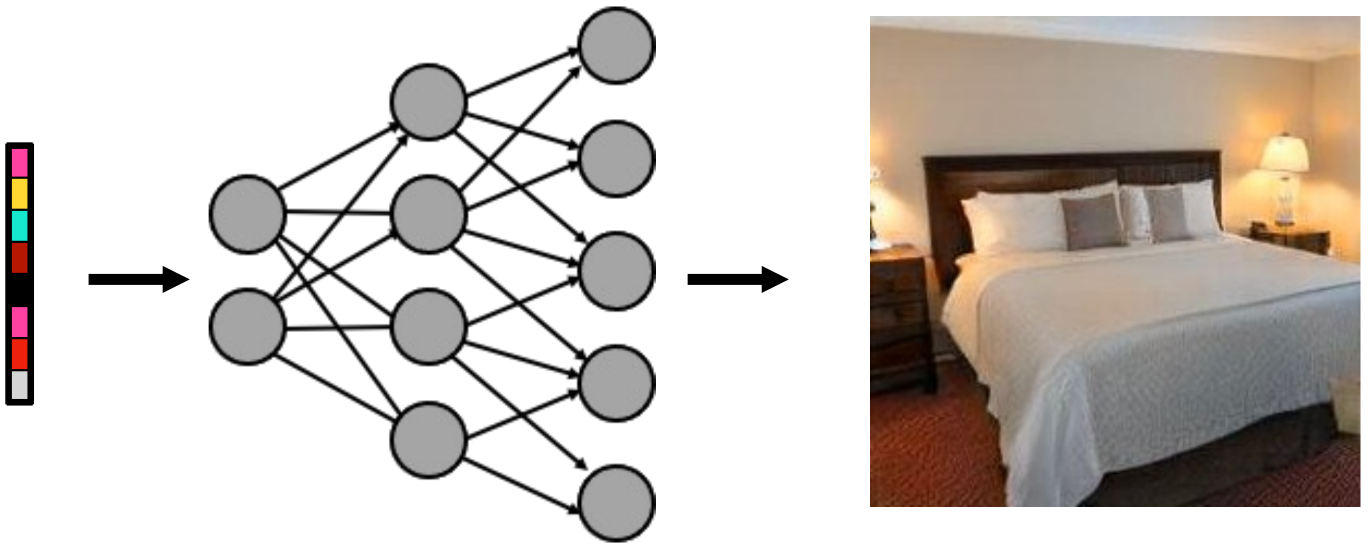


Neural Image Generation

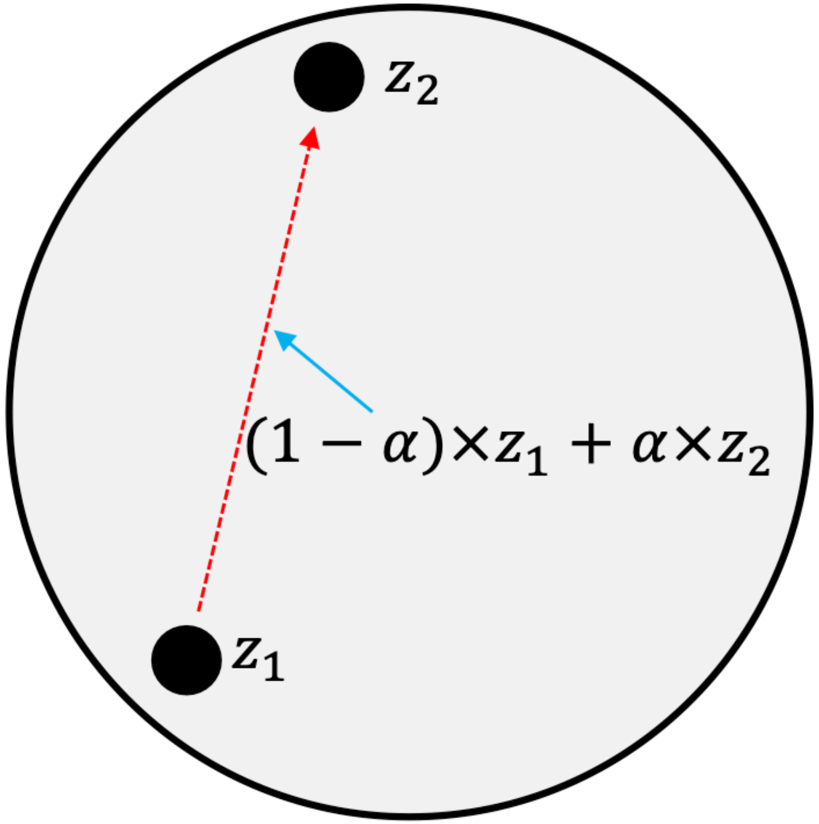


How to Steer Neural Image Generation?

- Interpret the generative representations with human understandable concepts
- Put human in the loop of AI content creation



Walking on the Latent Space

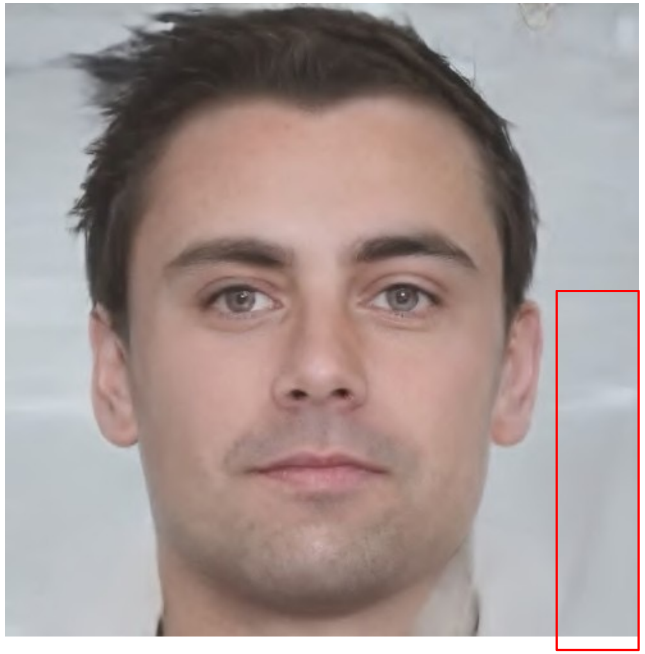


Linear walk on latent space

- Start point z_1
- End point z_2
- Step size $\alpha \in [0, 1]$
- Synthesised image $\hat{x} = G(z)$

Walking on the Latent Space

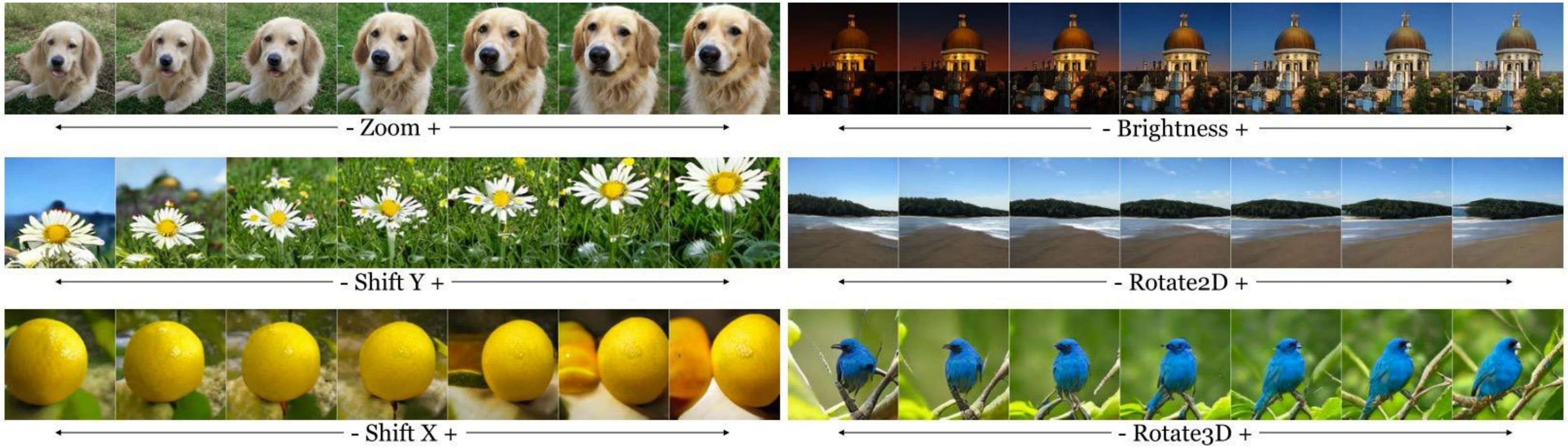
- Random Linear Walk on the Latent Space of StyleGAN (a big GAN 2019)




artifact in the space

Walking on the Latent Space

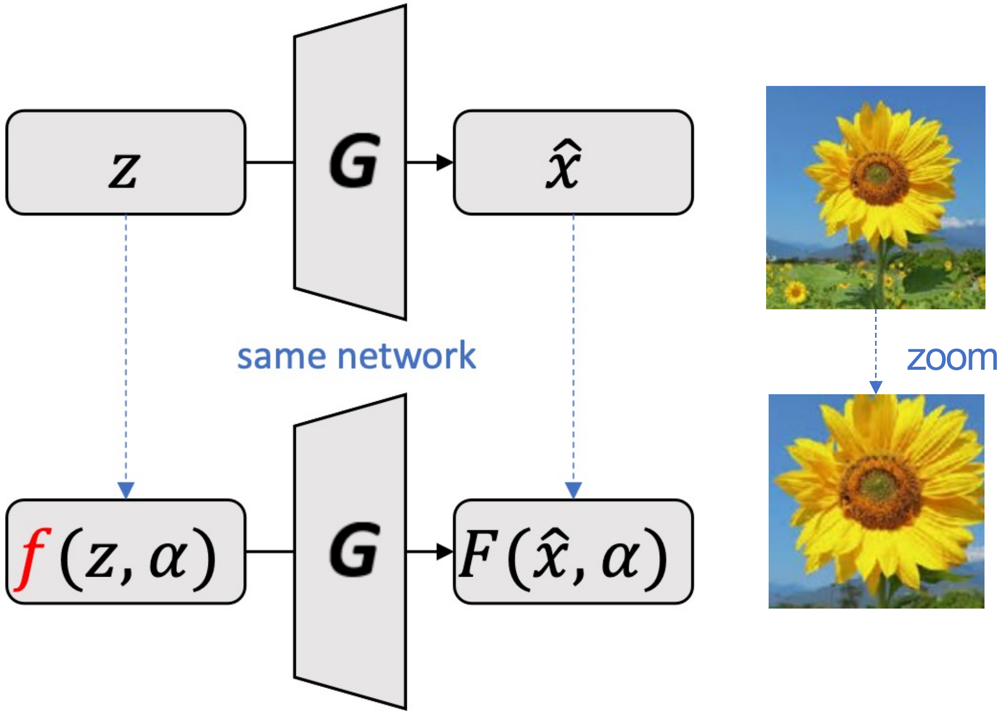
- Beyond Random Walk: How to Control the Walking on the Latent Space?
  - Given the prior knowledge: the transformation functions (zoom, shift ..) on image space
  - Find the corresponding function on the latent space



On the "steerability" of generative adversarial networks. Jahanian, Ali. Chai, Lucy. Isola, Phillip. ICLR 2020

Walking on the Latent Space

- Transformation on Image Space == Transformation on Latent Space



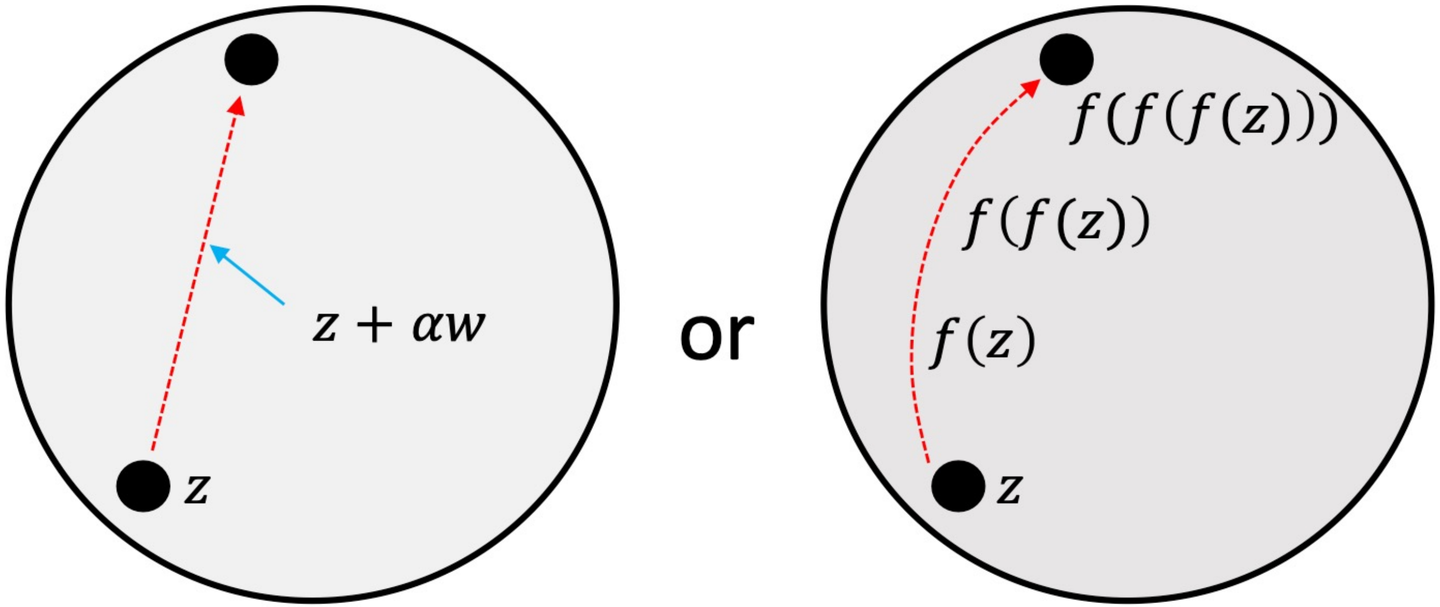
- Given
  - image transformation function F
    - shifting, zooming, brightness …
    - α controls the degree
  - pre-trained generator G

- Find latent transformation f

On the “steerability” of generative adversarial networks. Jahanian, Ali. Chai, Lucy. Isola, Phillip. ICLR 2020

Walking on the Latent Space

- Latent Transformation Function



Linear walk on latent space

Non-linear walk on latent space

Function f is a neural networks

On the “steerability” of generative adversarial networks. Jahanian, Ali. Chai, Lucy. Isola, Phillip. ICLR 2020

Interpretation Approaches

- Supervised approach:

  use labels or trained classifiers to probe the representation of the generator

- Unsupervised approach:

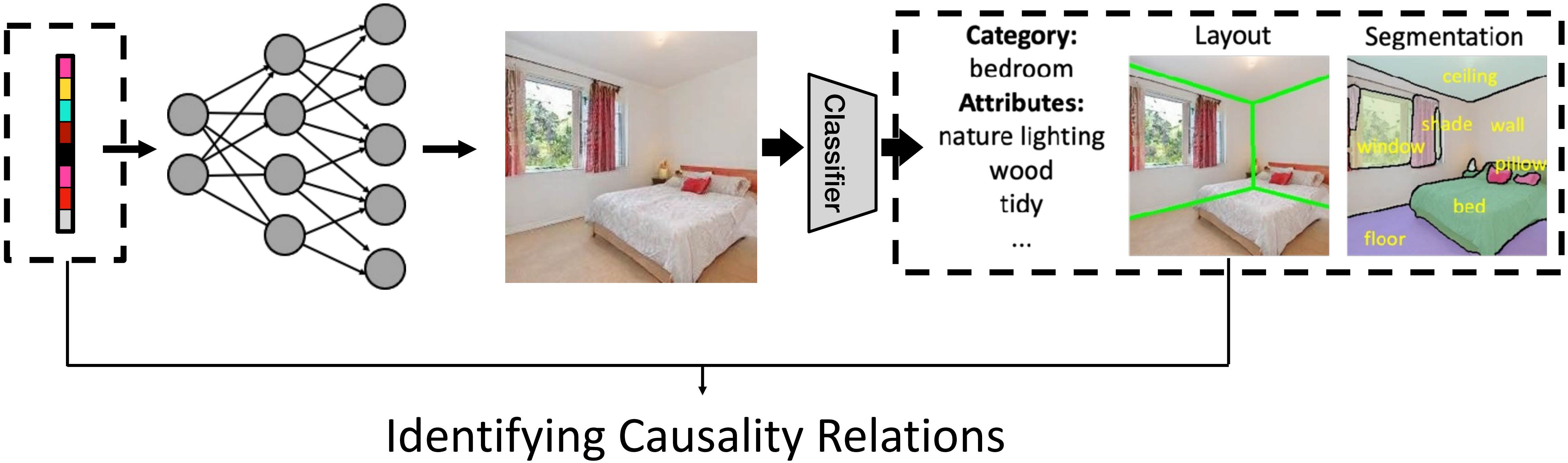
  identify the controllable dimensions of generator without labels/classifiers

Interpretation Approaches

- Supervised approach:
  use labels or trained classifiers to probe the representation of the generator
- Unsupervised approach:
  identify the controllable dimensions of generator without labels/classifiers

Supervised approach

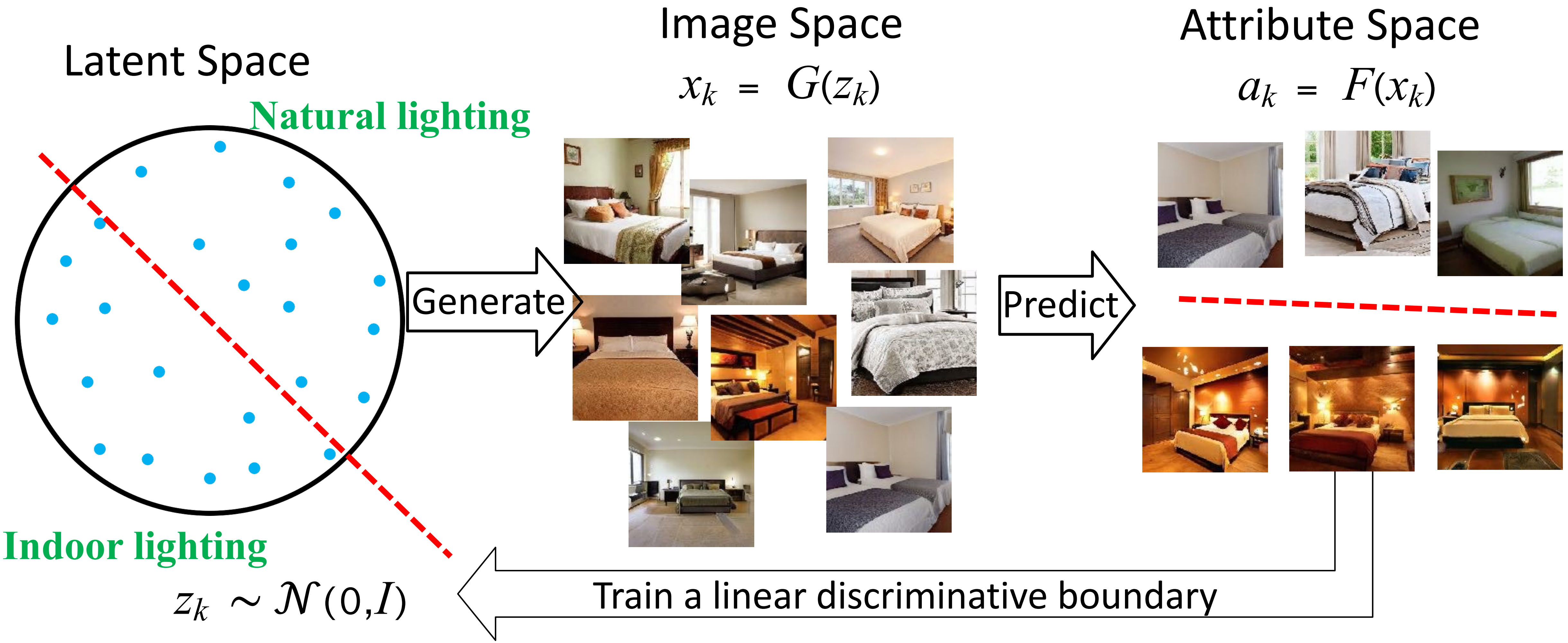
- Probing latent space with linear classifier



Yang, Shen, Zhou. Semantic Hierarchy Emerges in Deep Generative Representations for Scene Synthesis. IJCV, 2020

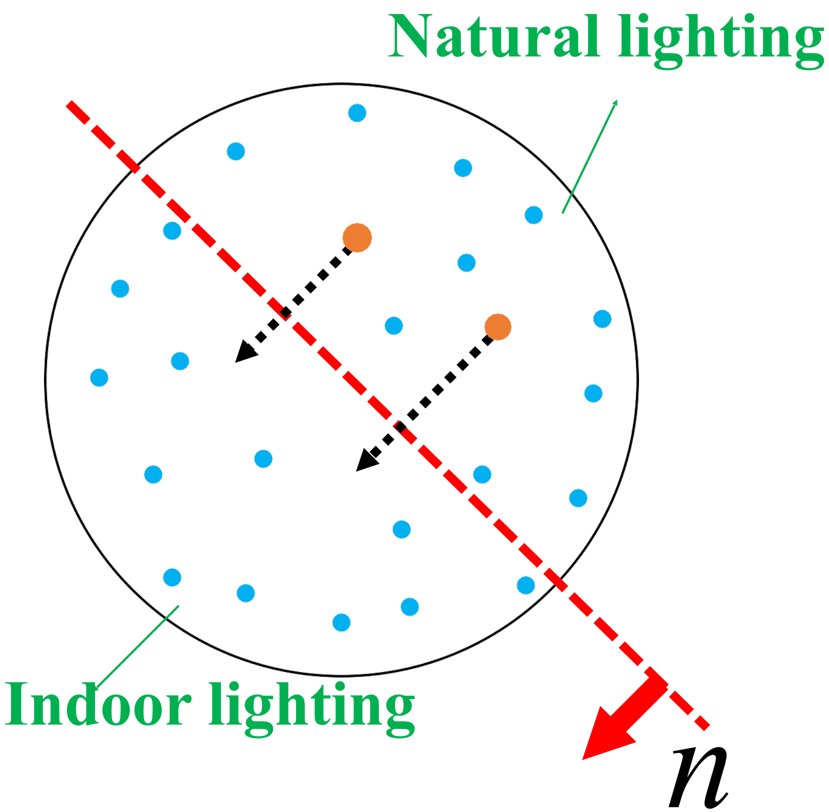
Supervised approach

- Identifying Causality Relations



Supervised approach

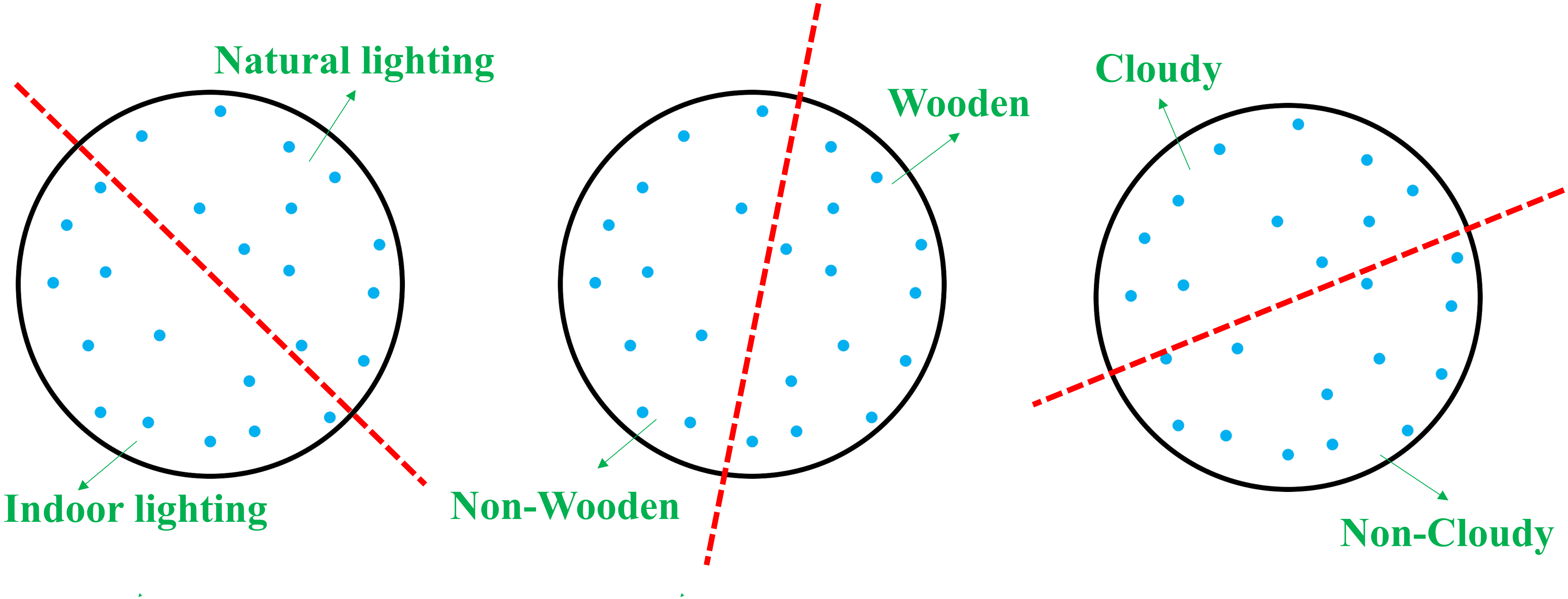
- Linear Manipulation on Latent Code



$$G(z_k) \dashrightarrow G(z_k + \lambda n)$$

Supervised approach

- Various Attribute Boundaries in Latent Space



Supervised approach

- Steering Generative Model: Changing Indoor lighting

Supervised approach

- Steering Generative Model: Adding clouds

Supervised approach

- **InterFaceGAN**: Probing latent space of face GAN with linear classifier

Age

Gender

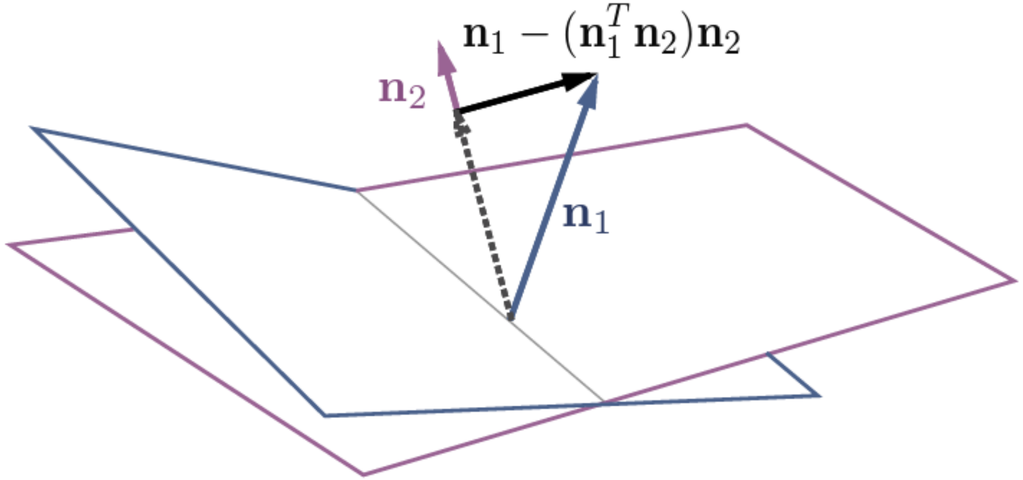
Pose

Artifact

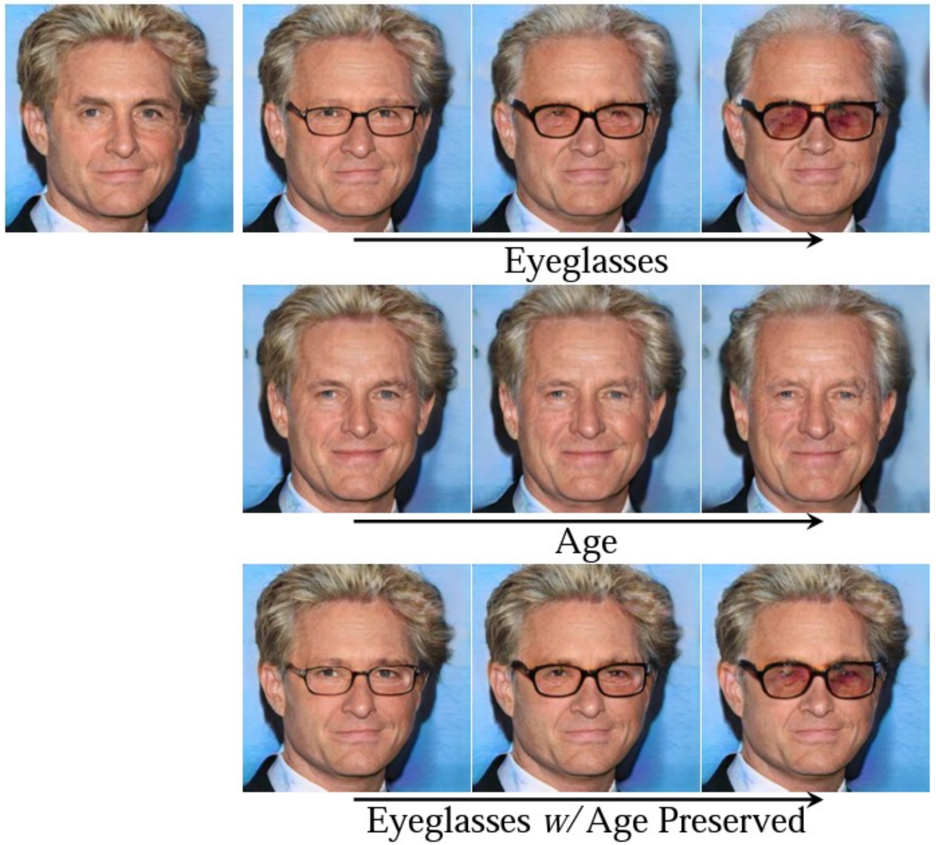
Shen, Yang, Tang, Zhou. InterFaceGAN: interpreting the disentangled face representation learned by GANs. CVPR'20 and IEEE TPAMI

Supervised approach

- **InterFaceGAN**: Probing latent space of face GAN with linear classifier



Conditional manipulation



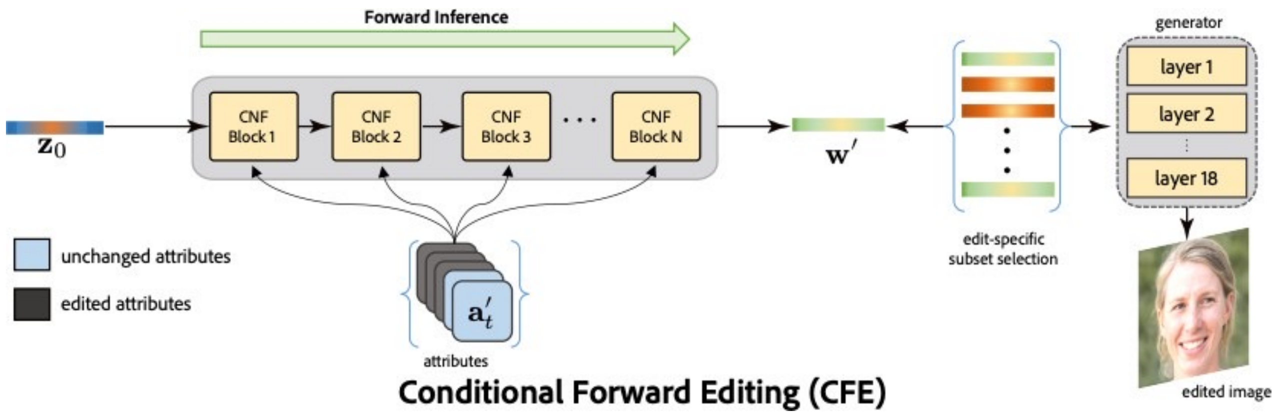
Supervised approach

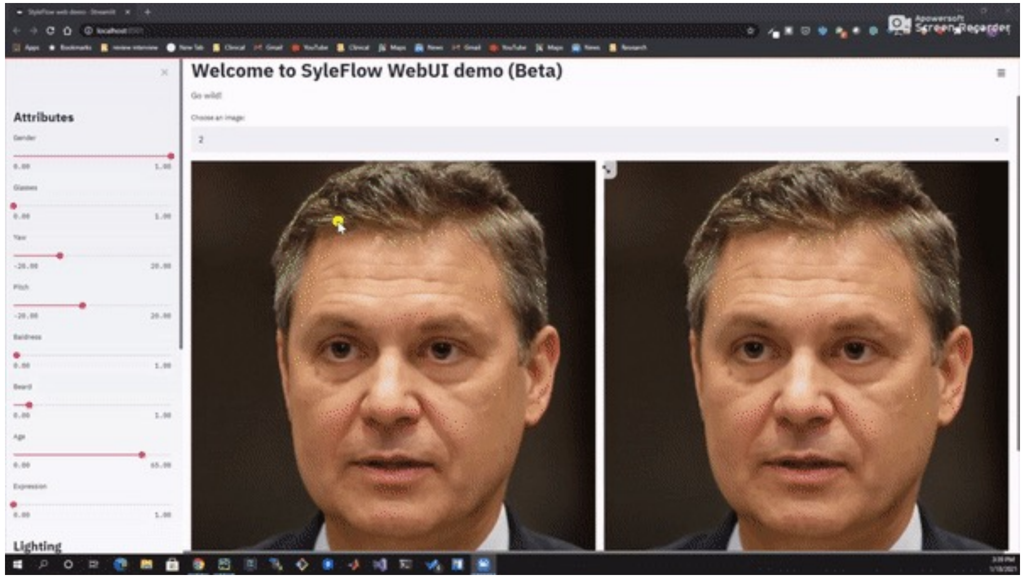
- **StyleFlow:** StyleGAN + flow-based conditional model
  - Previous work assumes the linear manipulation model:

$$I' = G(w + \lambda n_a), w = F(z), z \sim \mathcal{N}(0,1)$$

  - StyleFlow: Replace the MLP with an invertible flow model conditioned on attributes

$$w = \Phi(z, a), z \sim \mathcal{N}(0,1)$$





Abdal et al. StyleFlow: Attribute-Conditioned Exploration of StyleGAN-Generated Images Using Conditional Continuous Normalizing Flows. SIGGRAPH'21

Challenges for Supervised Approach

- How to expand the annotated dictionary size?
- How to further disentangle the relevant attributes?
- How to align latent space with image region attributes?

Interpretation Approaches

- Supervised approach:

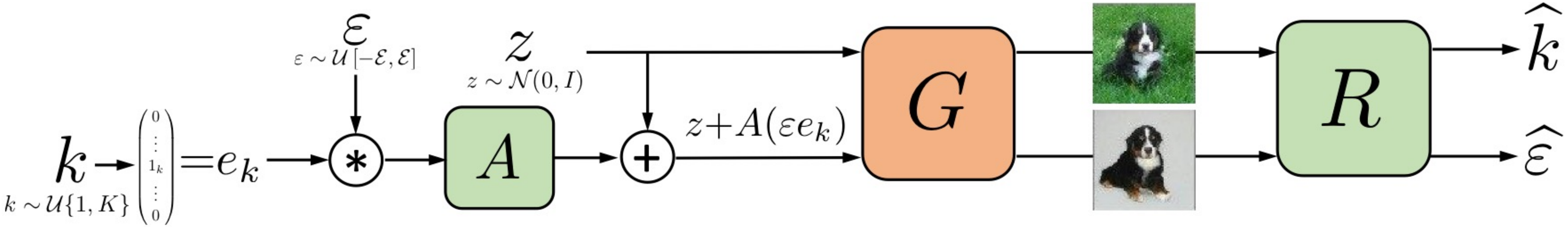
  use labels or trained classifiers to probe the representation of the generator

- Unsupervised approach:

  identify the controllable dimensions of generator without labels/classifiers

Unsupervised approach

- Unsupervised Discovery of Interpretable Directions

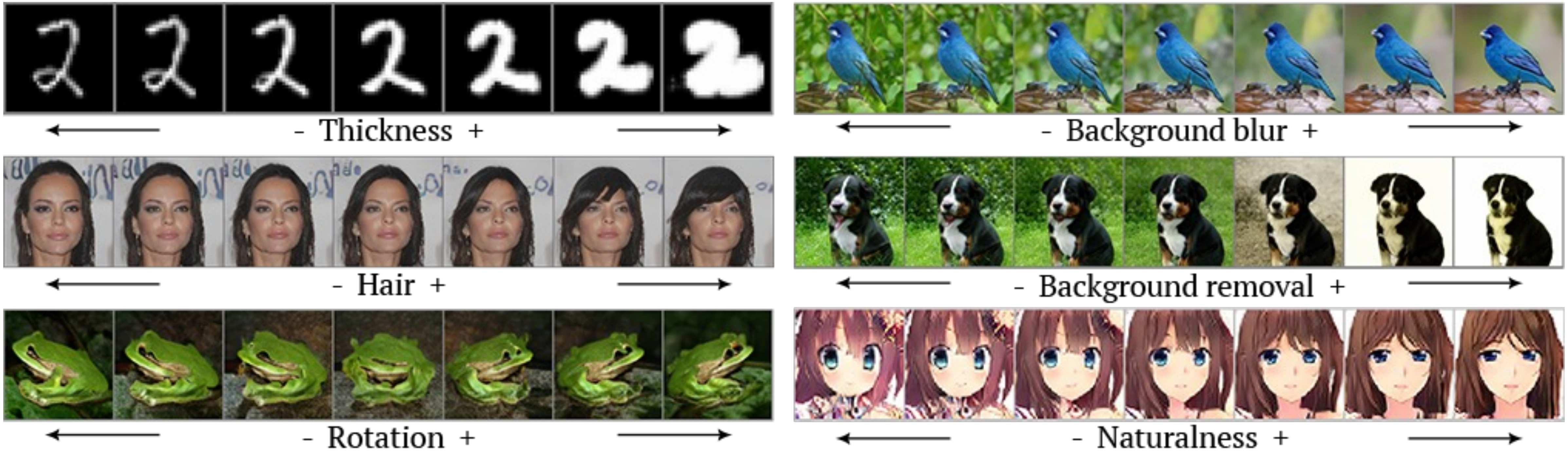


- The latent deformator (A) aims to produce shifts that are easy to distinguish
- The reconstructor (R) aims to reproduce the shift in the latent space that induces a given image transformation.

Andrey Voynov and Artem Babenko. Unsupervised Discovery of Interpretable Directions in the GAN Latent Space.

Unsupervised approach

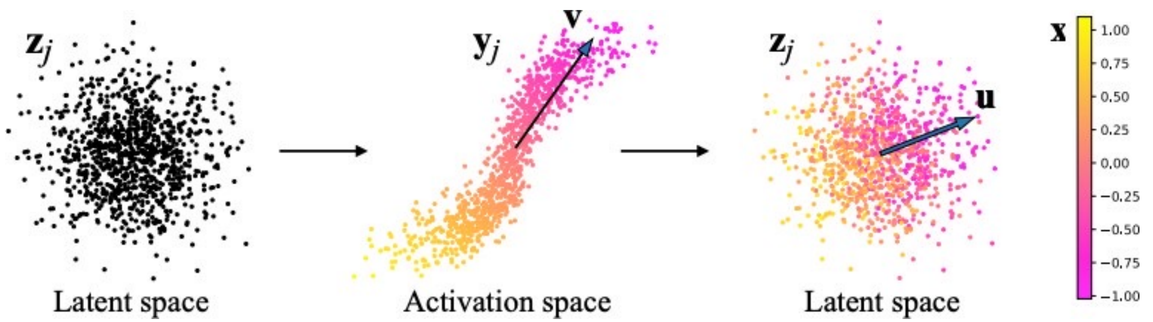
- Unsupervised Discovery of Interpretable Directions



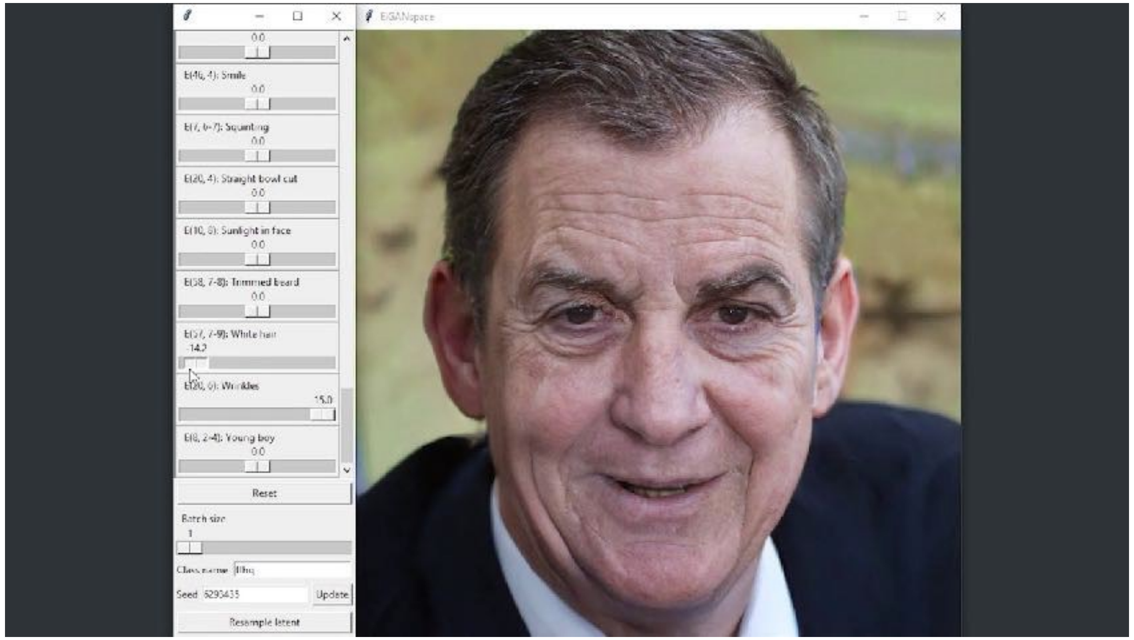
Unsupervised approach

- GANspace: PCA applied to the latent space of StyleGAN

PCA direction
in feature space



Regression from PCA
direction in latent space



Erik Harkonen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: discovering interpretable GAN controls. NeurIPS'20

Unsupervised approach

- Rows 1-3 illustrate the three largest principal components in the intermediate W latent space of StyleGAN2.
- The red square corresponds to location of the original image on each principal axis.
- Rows 4-5 demonstrate the effect of constraining the variation to a subset of the layers.

Erik Harkonen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: discovering interpretable GAN controls. NeurIPS'20

Challenges for Unsupervised Approach

- How to evaluate the results?
- How to annotate each disentangled dimensions?
- How to improve the disentanglement in GAN training?

Interpretable Latent Space and Inversion Issue in GANs

- Walking on the Latent Space
  - Supervised approach
  - Unsupervised approach
- Inversion of Real Images
  - Optimization-based method
  - Encoder-based method
  - Hybrid approach
  - Pivot-tuning

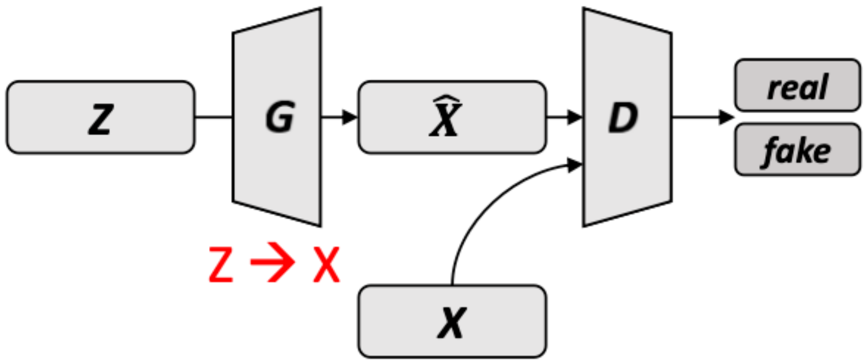
Inversion of real images

- How to edit my own photo?

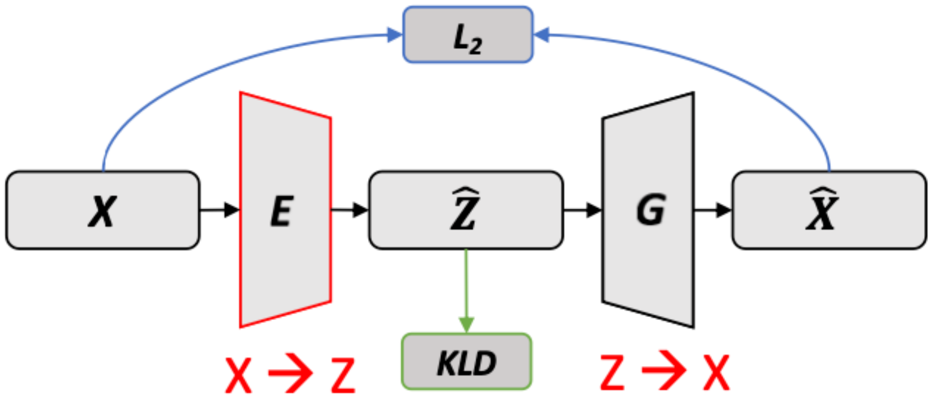
Inversion of real images

Motivation: GAN vs. VAE

GAN



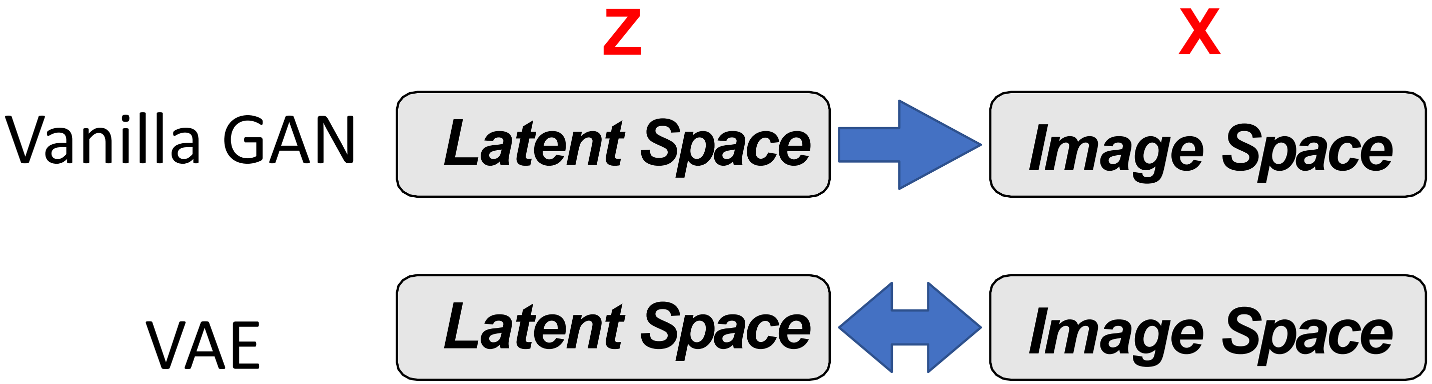
VAE



VAE has an Encoder that can map x to z

Inversion of real images

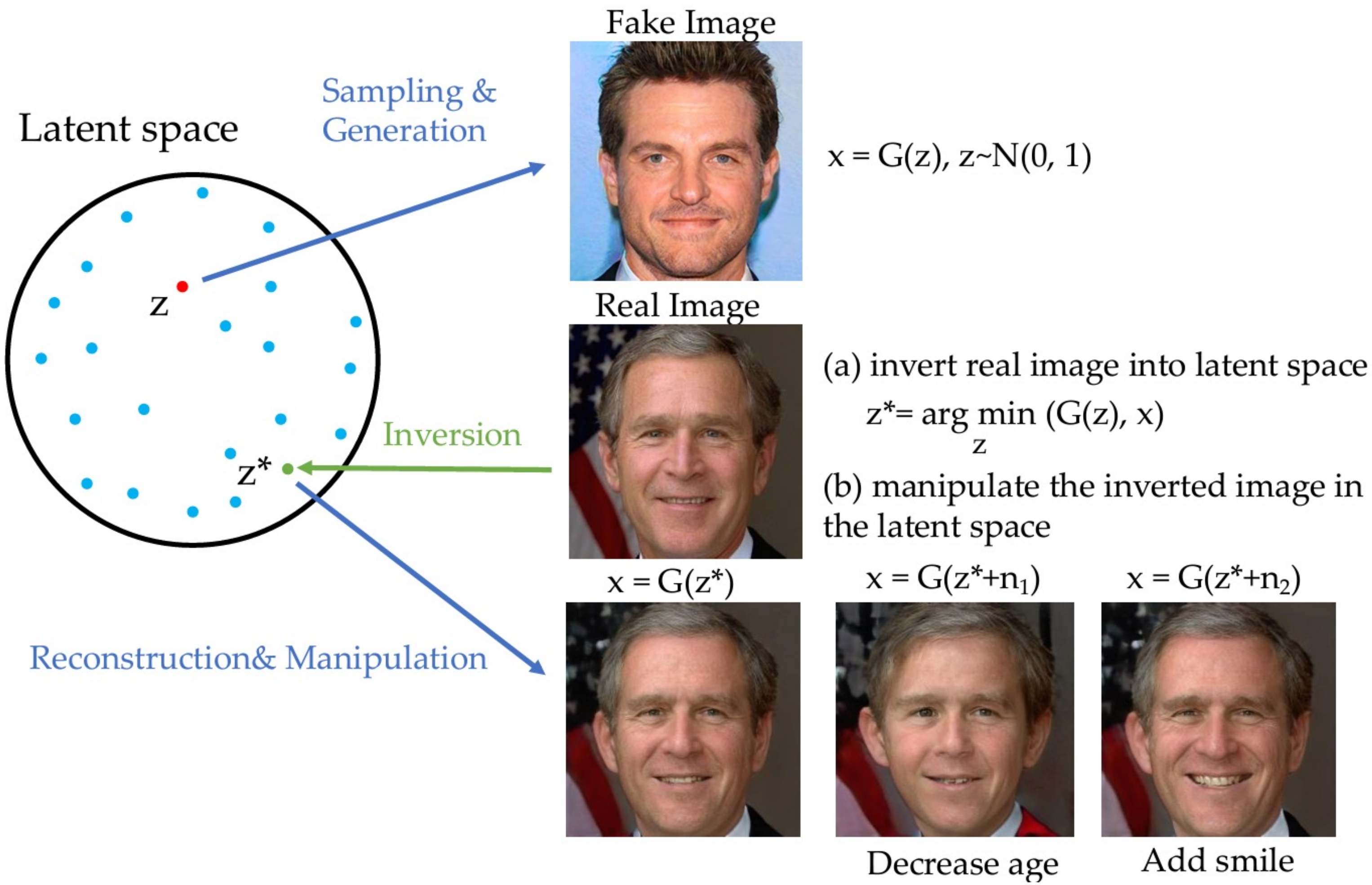
Motivation: GAN vs. VAE



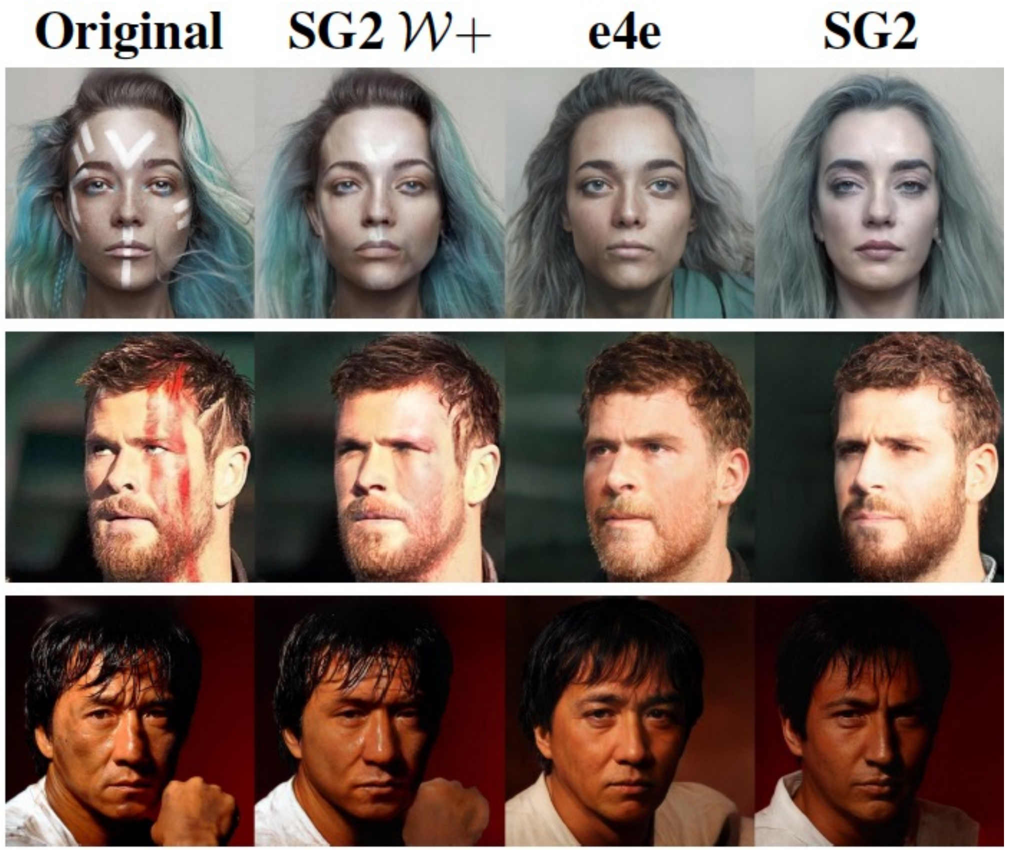
VAE = Generator + Encoder

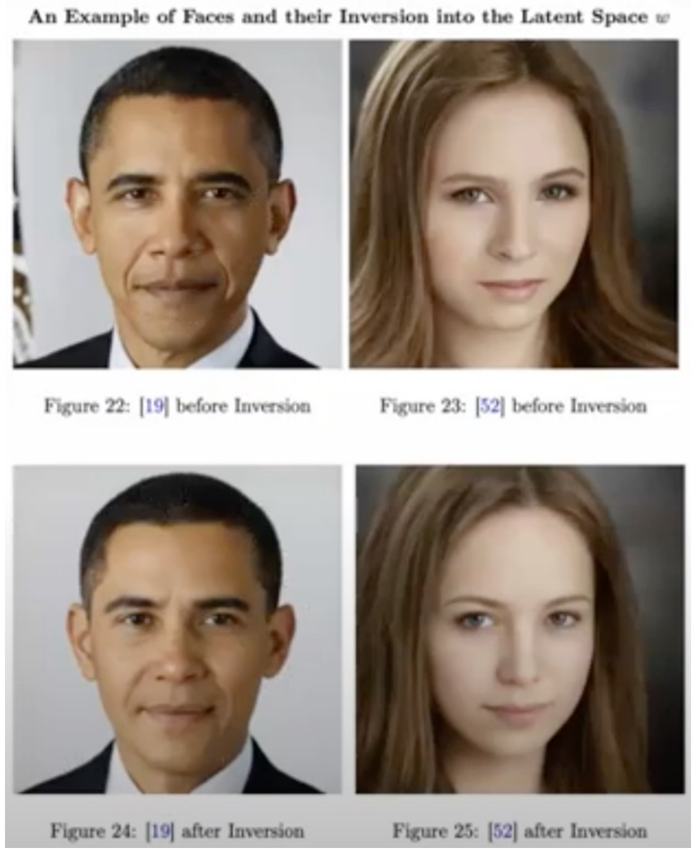
Vanilla GAN = Generator + Discriminator

GAN Inversion:
Inverting Real Faces to Latent Code



Inversion issue of GANs



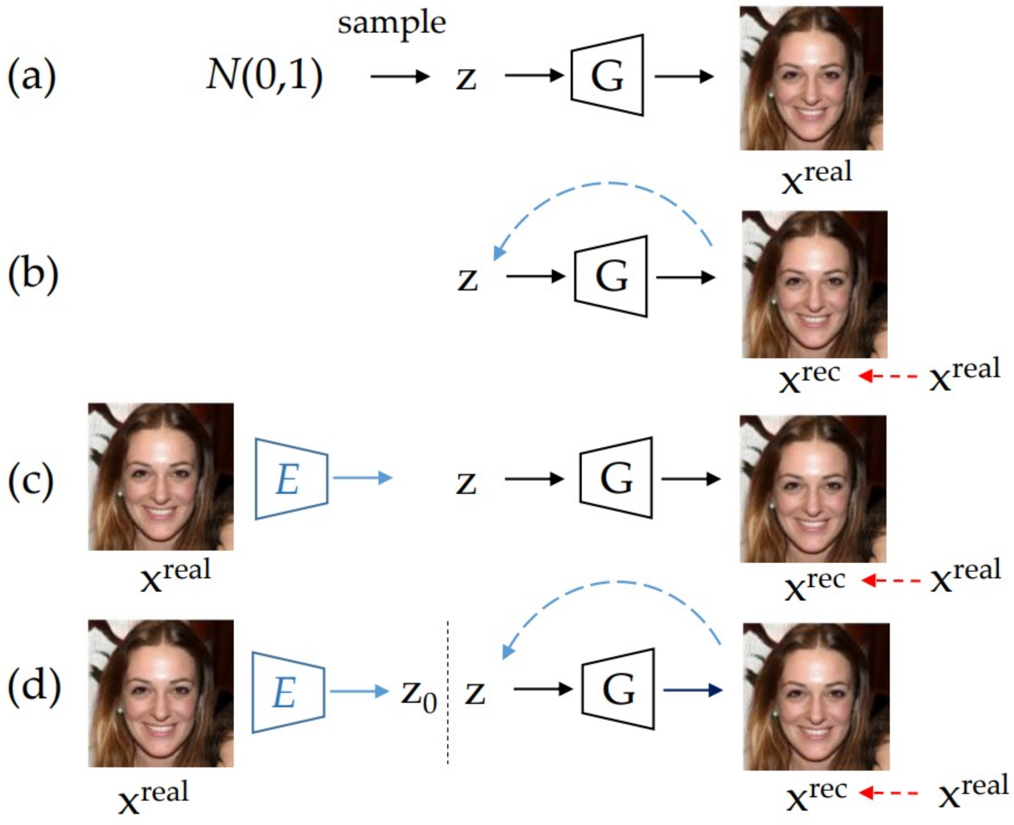


Roich, Daniel, et al. "Pivotal tuning for latent-based editing of real images." *ACM Transactions on Graphics (TOG)* 42.1 (2022): 1-13.

[MyStyle: A Personalized Generative Prior]

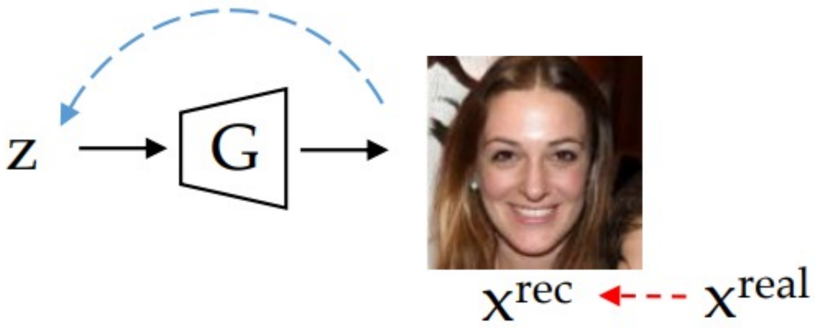
GAN Inversion methods

- **Optimization-based** inversion uses an optimization algorithm to iteratively optimize the latent code to minimize the pixel-wise reconstruction loss.
- **Learning-based** inversion builds an encoder network that maps an image into the latent space.
- **Hybrid approach** uses the encoder to generate an initialization for optimization, i.e., an encoder network is first used to obtain an approximate embed



Optimization-based method

Image2StyleGAN (ICCV'2019)



- Given a pretrained/fixed G and an image X, optimize Z for each image：

$$z^* = \arg\min_z ||x^{rec} - x^{real}||^2$$

- Limitation: SLOW!

Background -- StyleGAN

- Structure of StyleGAN
- Which latent space (z, w or w+) to use?
- Tradeoff between distortion and editability

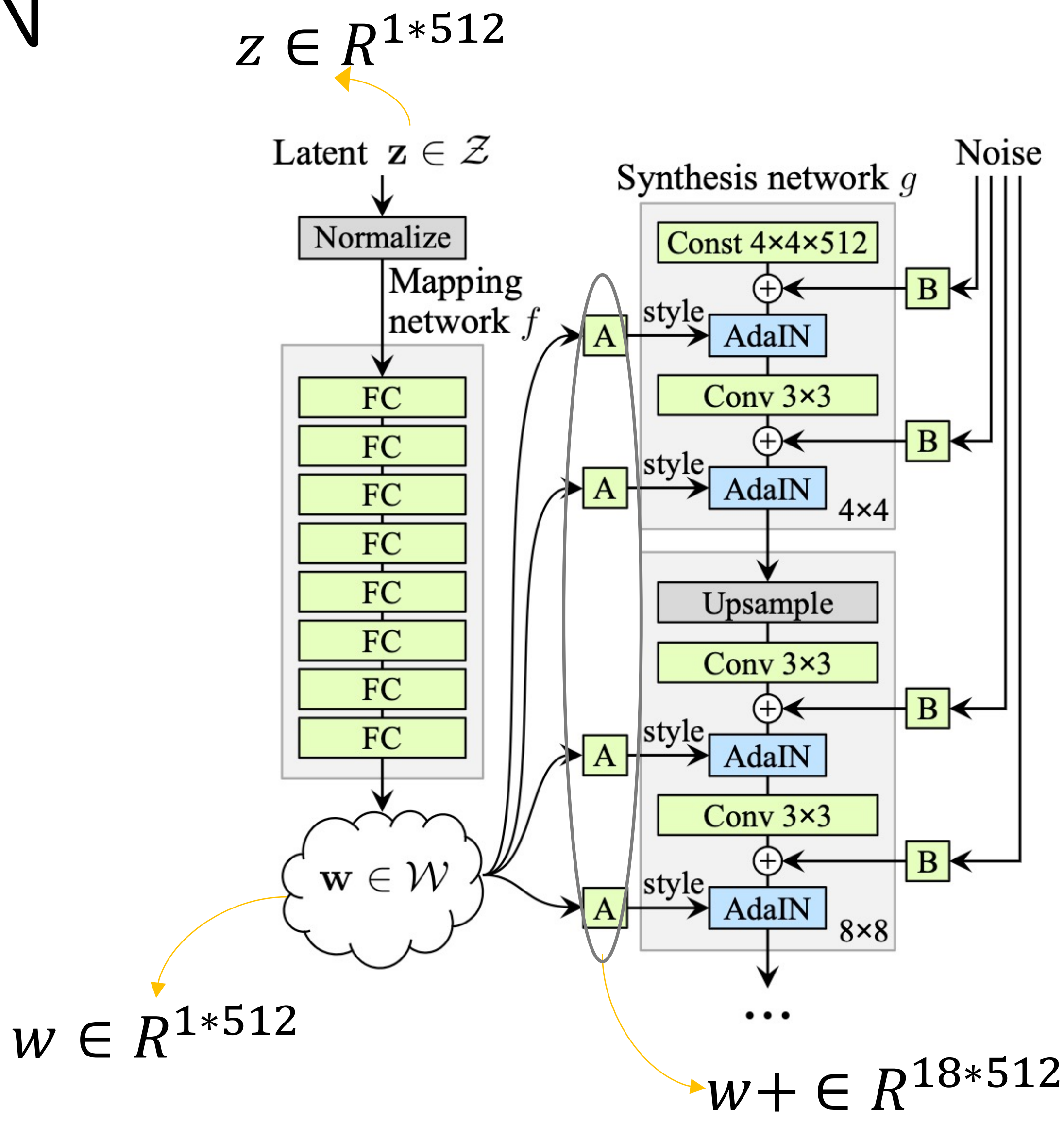
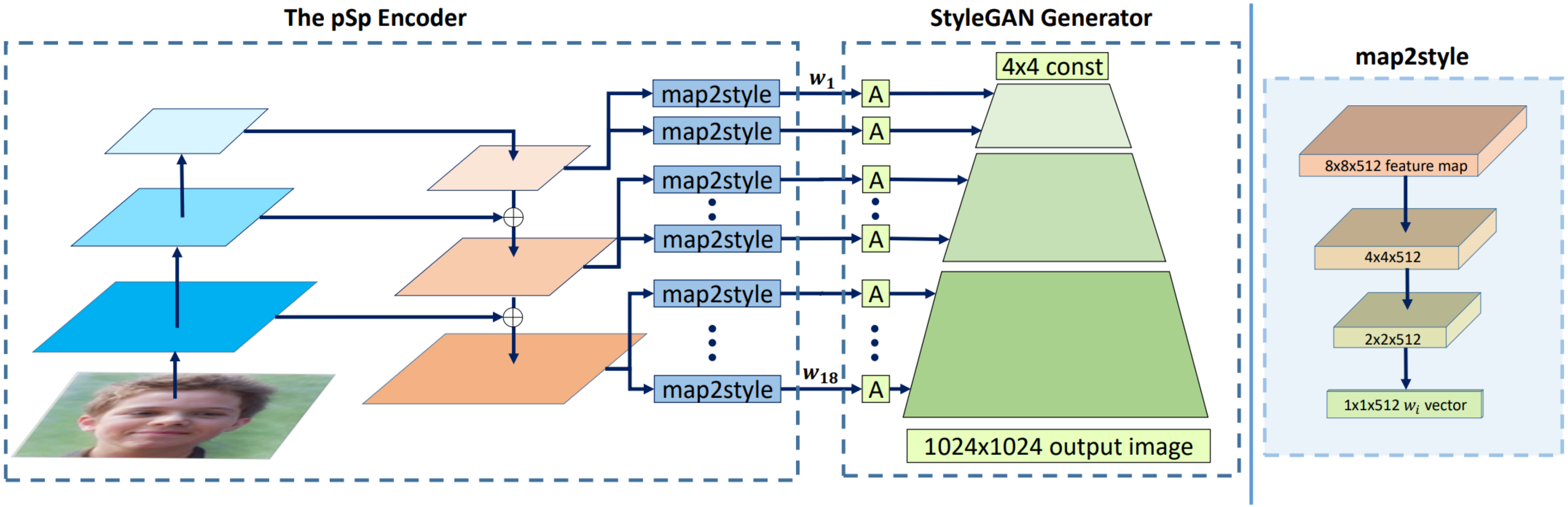


Image borrowed from StyleGAN (CVPR'19)

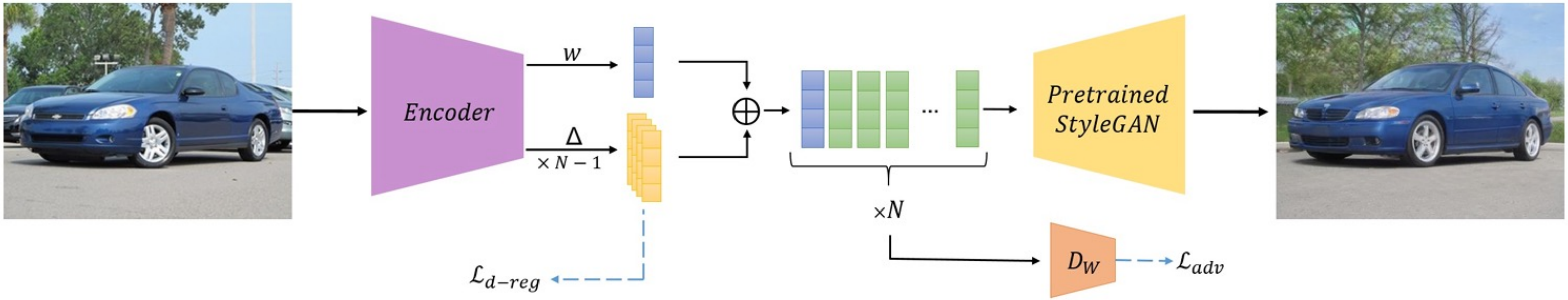
Encoder-based method

- Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation (CVPR'21)
  - Directly train an encoder to the latent space W
  - Main novelty: a novel encoder architecture
  - Loss: pixel-wise L2 loss + perceptual loss (LPIPS) + Identity loss

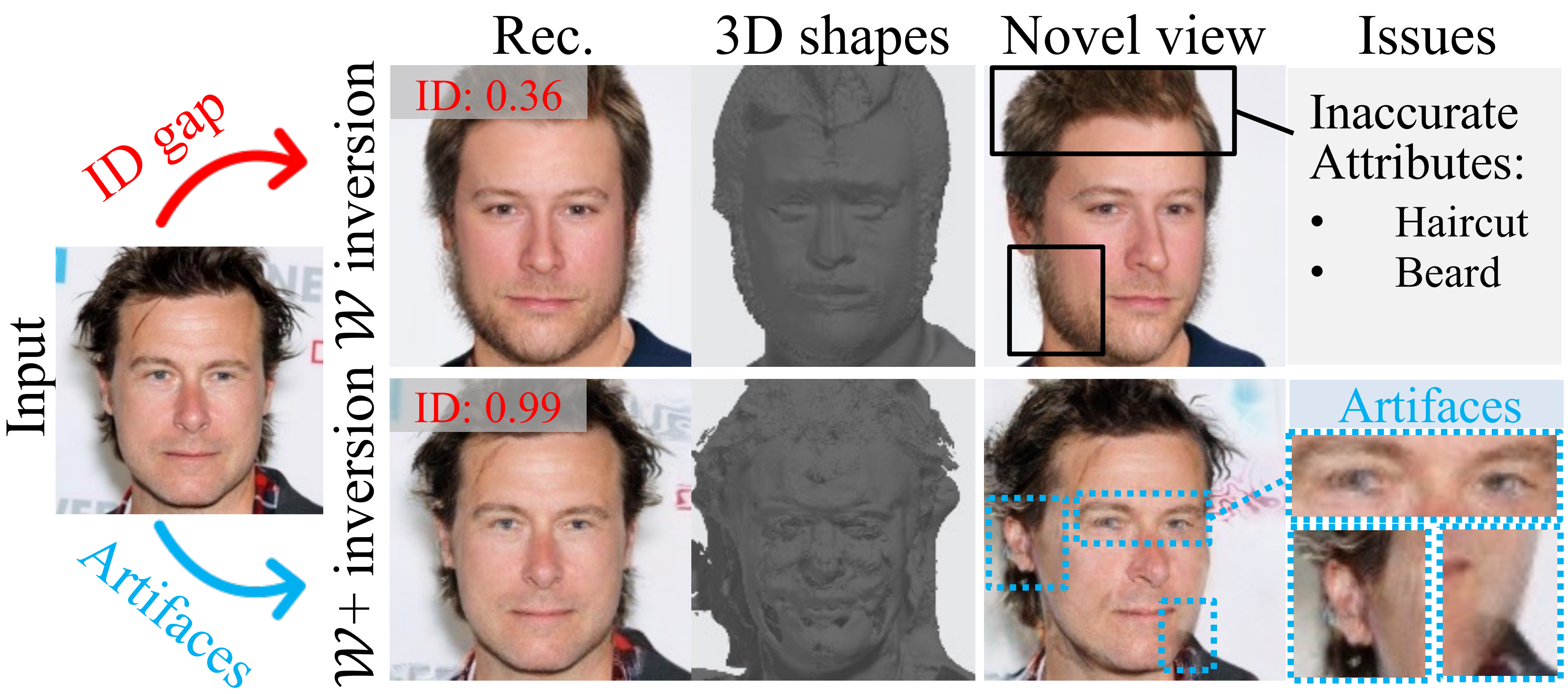


Encoder-based method

- Designing an Encoder for StyleGAN Image Manipulation (ACM TOG'21)
  - Trade-offs among distortion, perception, and editability
  - a new encoder, which is explicitly encouraged to invert images close to W space



Inversion trade-off in 3D-GANs



Yin, Yu, et al. "NeRFInvertor: High Fidelity NeRF-GAN Inversion for Single-shot Real Image Animation." *CVPR*. 2023.

Hybrid method

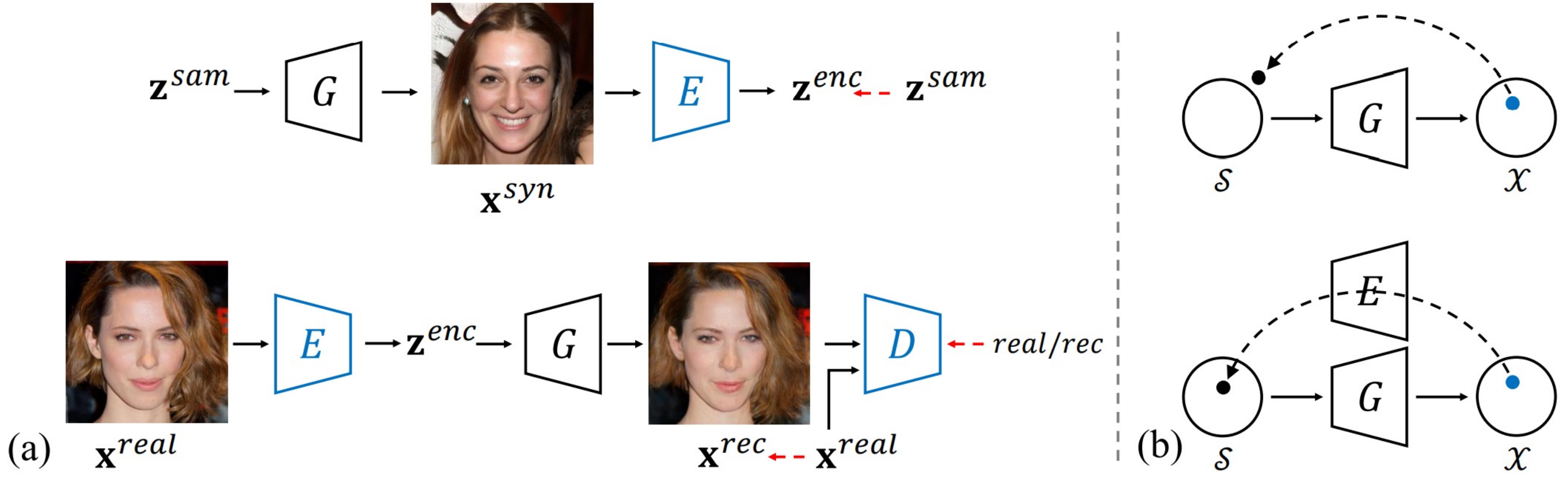
In-domain gan inversion for real image editing (ECCV'20)

- Init using a trained encoder --> optimize Z for each image
- Loss: pixel-wise L2 loss + perceptual loss + adversarial loss
- Optimization: (F is perceptual loss)

$$\arg\min_{z} ||x^{real} - G(z)|| + ||F(x^{real}) - F(G(z))|| + ||z - E(G(z))||$$

Others:

Proposed:



Comparisons of SOTA works

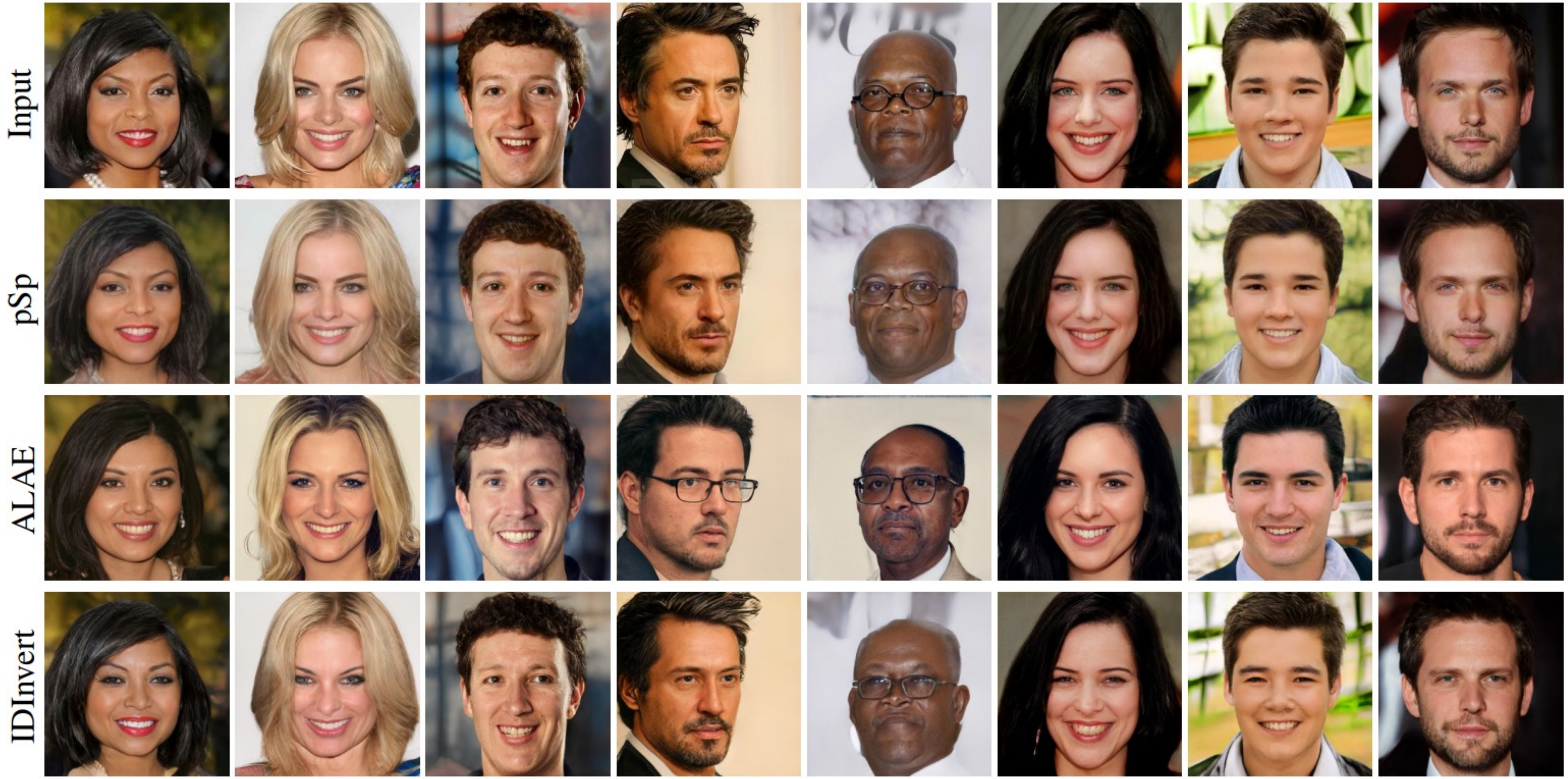
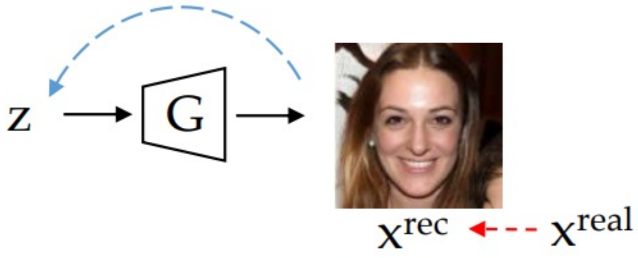


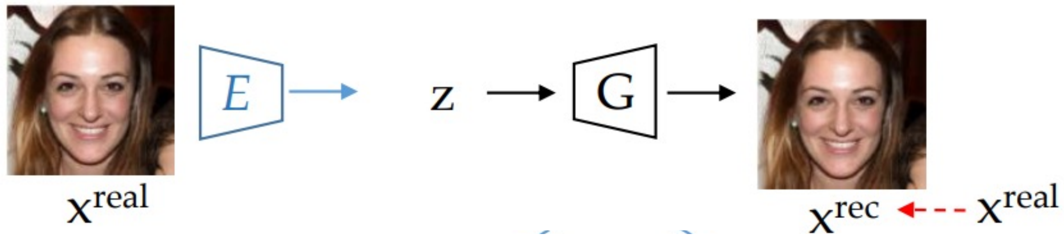
Figure 4: Results of our pSp framework for StyleGAN inversion compared to other approaches on CelebA-HQ.

Drawbacks of current methods



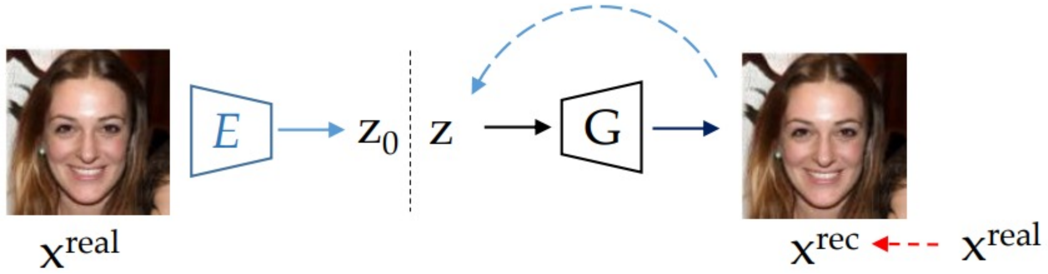
Optimization-based

- Time consuming
- Stuck at local minimum
- Performs badly on real images (large domain gap between real & synthesized data)



Learning-based

- less accurate, larger content loss

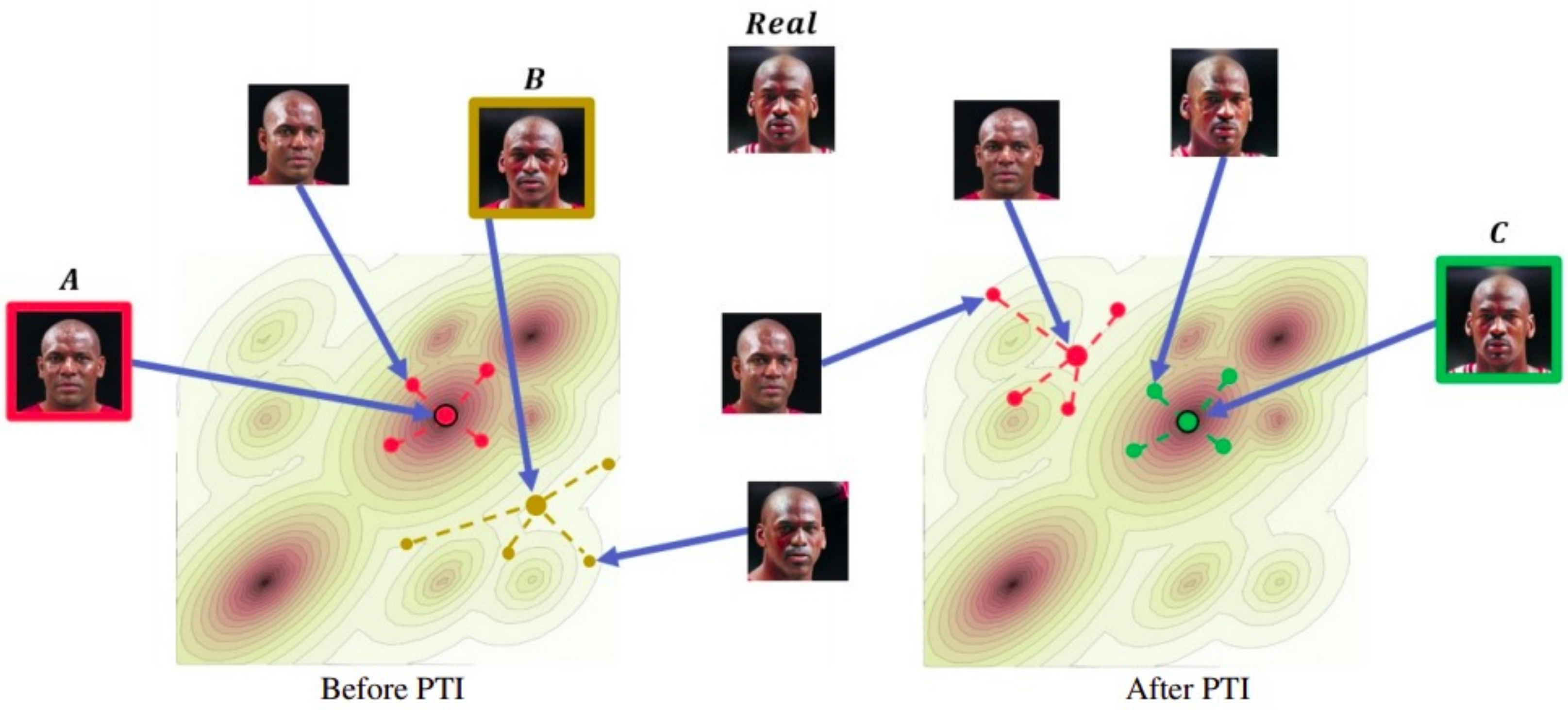


Hybrid approach

- Time consuming

PTI: Pivotal Tuning for Latent-based editing of Real Images (ACM TOG'22)

PTI

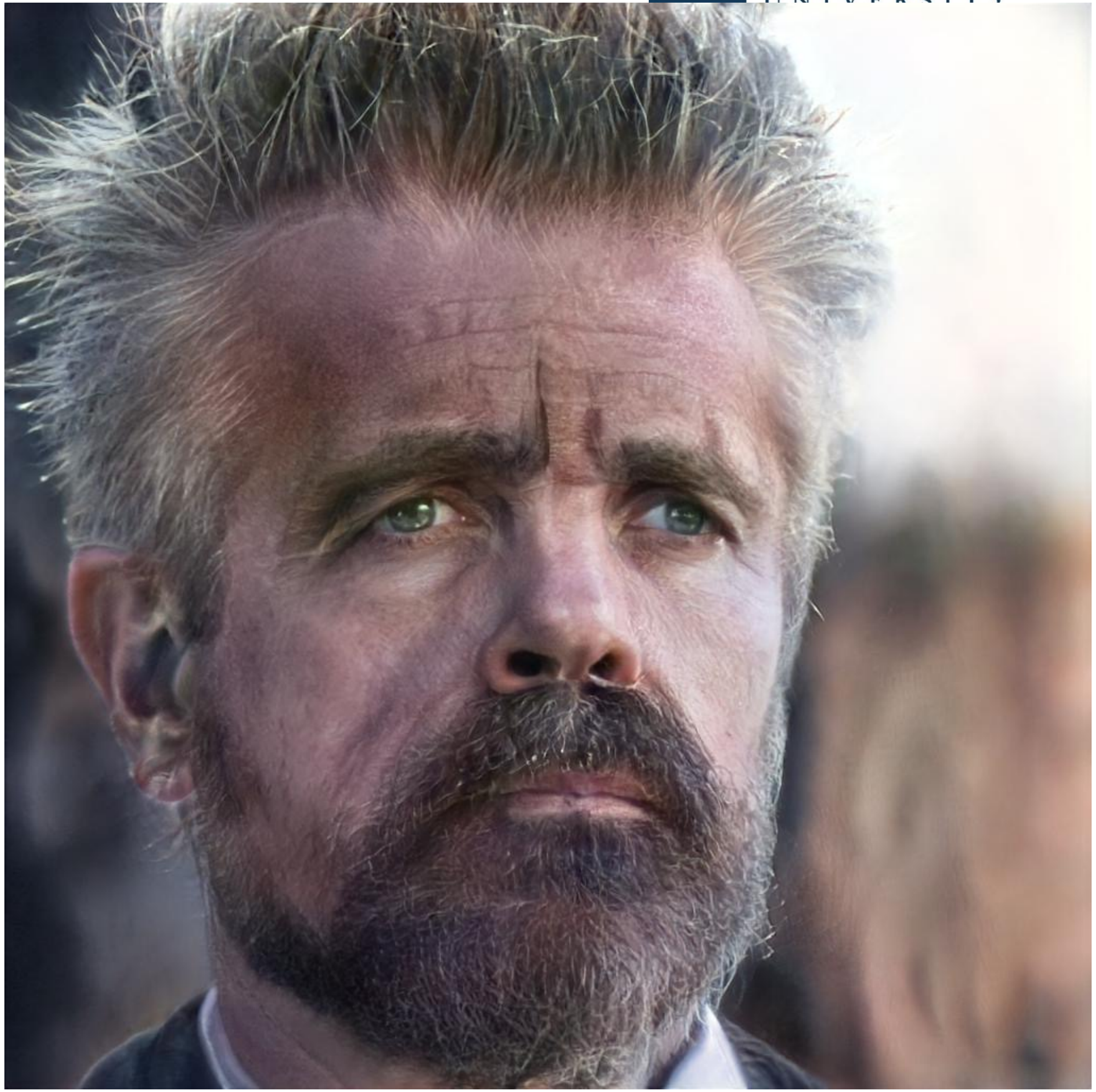


1. Perform optimization-based inversion method on the input image X and obtain the latent vector w_{pivot}.
2. Obtain a new image by feeding the w_{pivot} into the generator G called image X_{pivot}.
3. The generator G 's weights are tuned in such that w_{pivot} generates X and not X_{pivot}.
4. The Pivotal Tuning feedback is guided by two losses: LPIPS and L2 distance.

PTI results

Reconstruction. The images order is: Original image, W+ inversion, e4e inversion, W inversion, PTI inversion

InterfaceGAN pose edit comparison between different methods. The images order is: Original, W+, e4e, W, PTI



Next

- Diffusion Models

Thank You

- Questions?

- Email: yu.yin@case.edu

CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Reference slides and papers

- Hao Dong. Deep Generative Models
- Bolei Zhou. Exploring and Exploiting Interpretable Semantics in GANs
- Weihao Xia, et al. GAN Inversion: A Survey