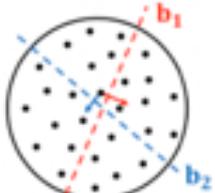


CVPR 2020 Tutorial on
Interpretable Machine Learning for Computer Vision

Anywhere on earth on June 15, 2020

Boundary Search
& Verification

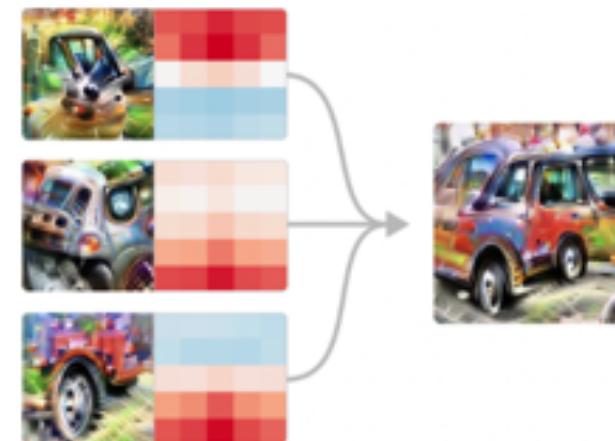
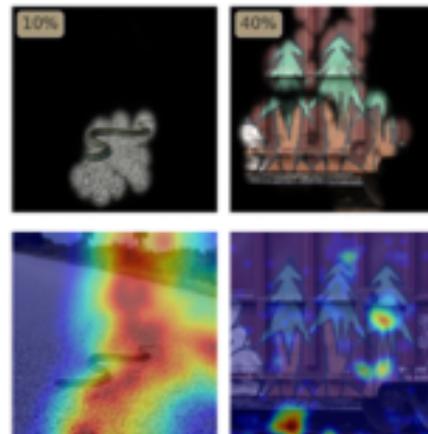


Latent Space Z

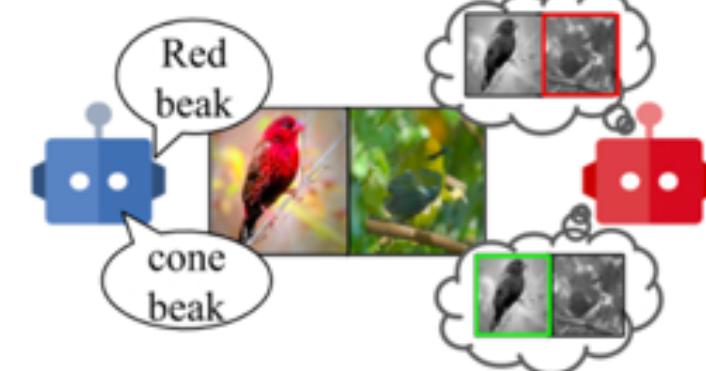
Manipulating Semantics



$$G(\mathbf{z}) \longrightarrow G(\mathbf{z} + \lambda\mathbf{b})$$



CVPR VIRTUAL



Lecturers



Bolei Zhou
CUHK



Ruth C. Fong
Oxford



Christopher Olah
OpenAI



Zeynep Akata
Uni-Tübingen



Exploring and Exploiting Interpretable Semantics in GANs

Bolei Zhou

The Chinese University of Hong Kong (CUHK)

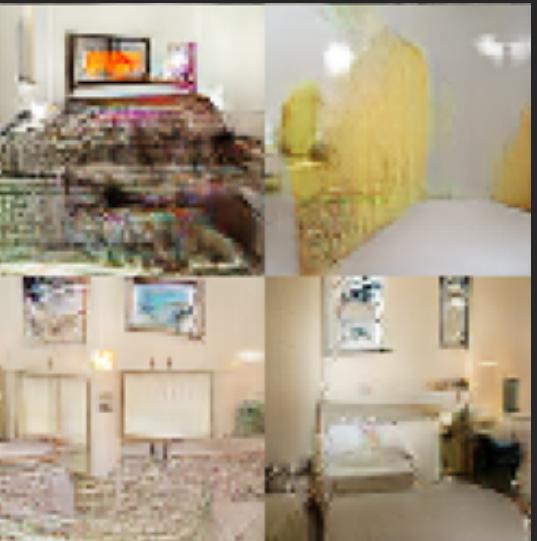
GANs for Synthesizing Images

2014



GANs [Goodfellow et al.]

2015



DCGAN [Radford et al.]

2017



PG-GAN [Karras et al.]

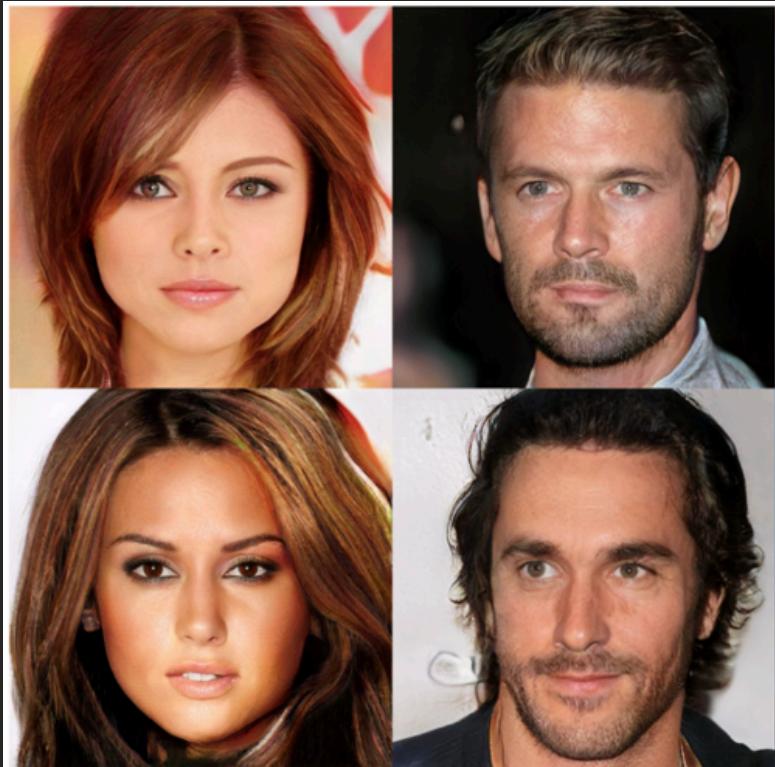
2018



BigGAN [Brock et al.]

...

GANs for Synthesizing Images



PG-GAN

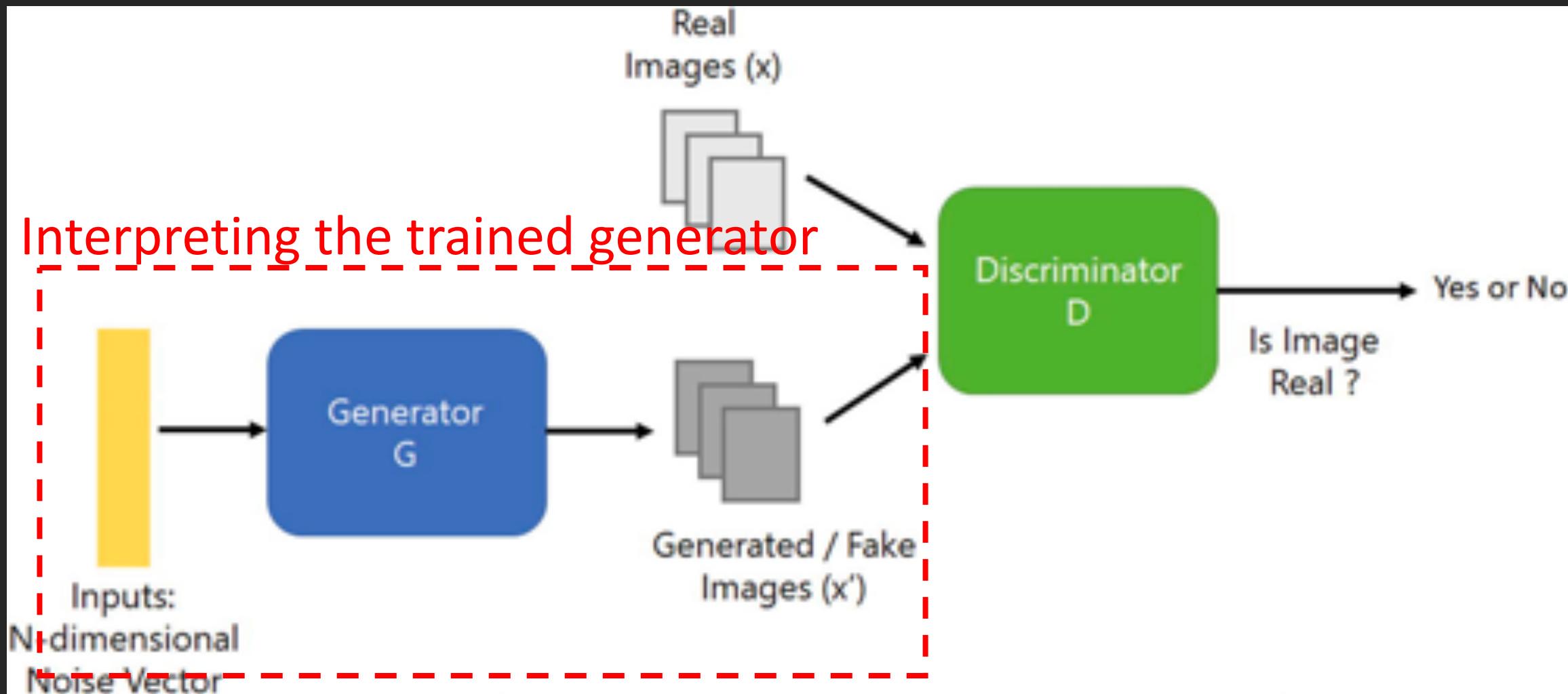


StyleGAN



BigGAN

Generative Adversarial Training



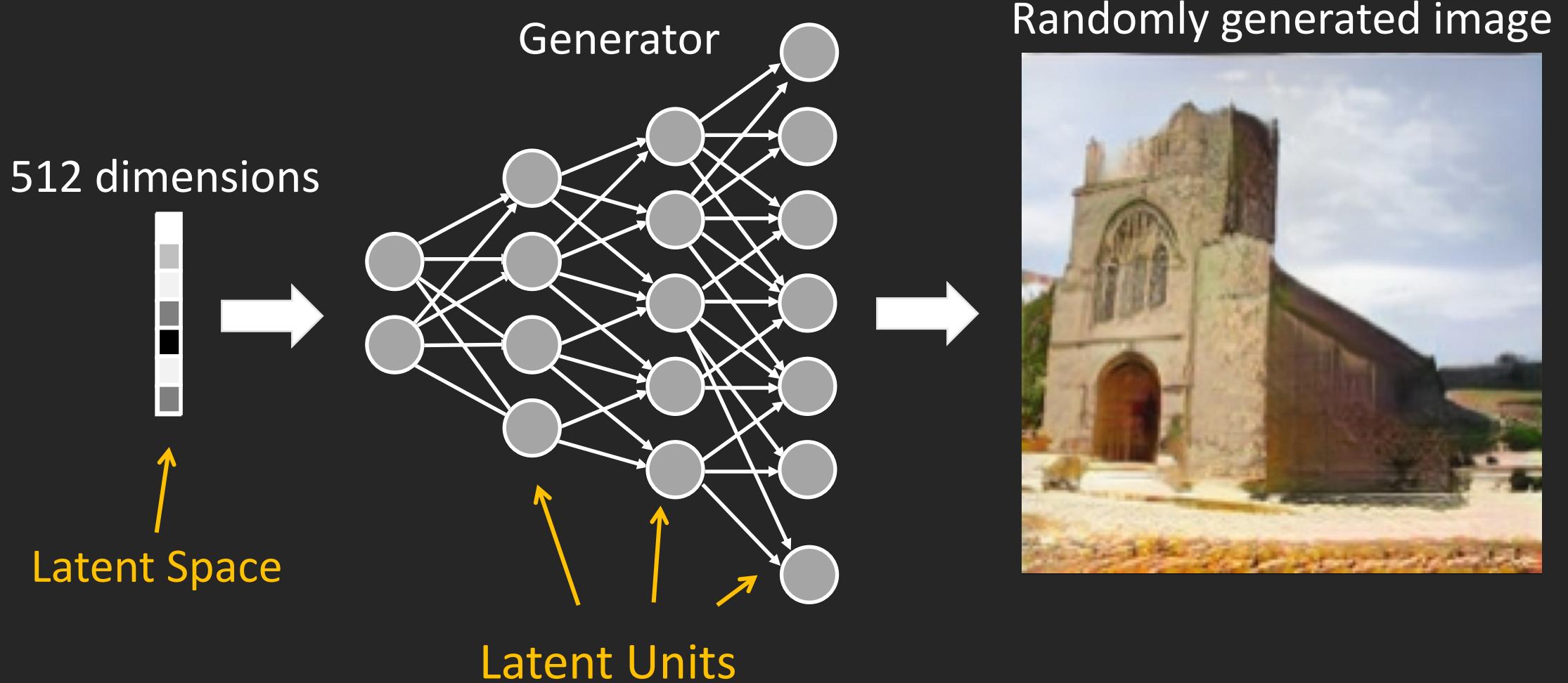
Tutorial Outline

- Semantics in the latent space
 - Supervised approach
 - Unsupervised approach
- Inversion of real images
- Future directions

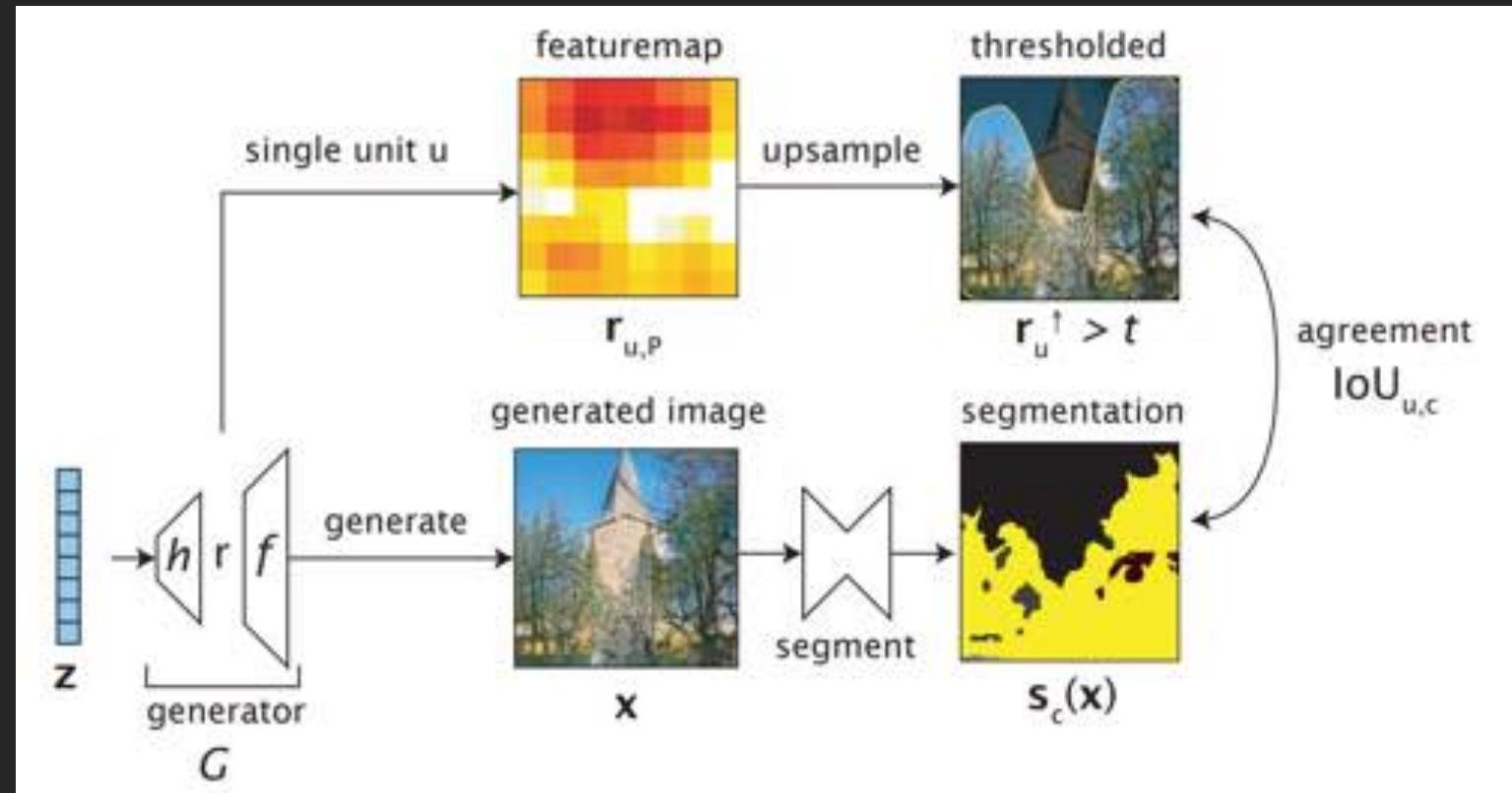
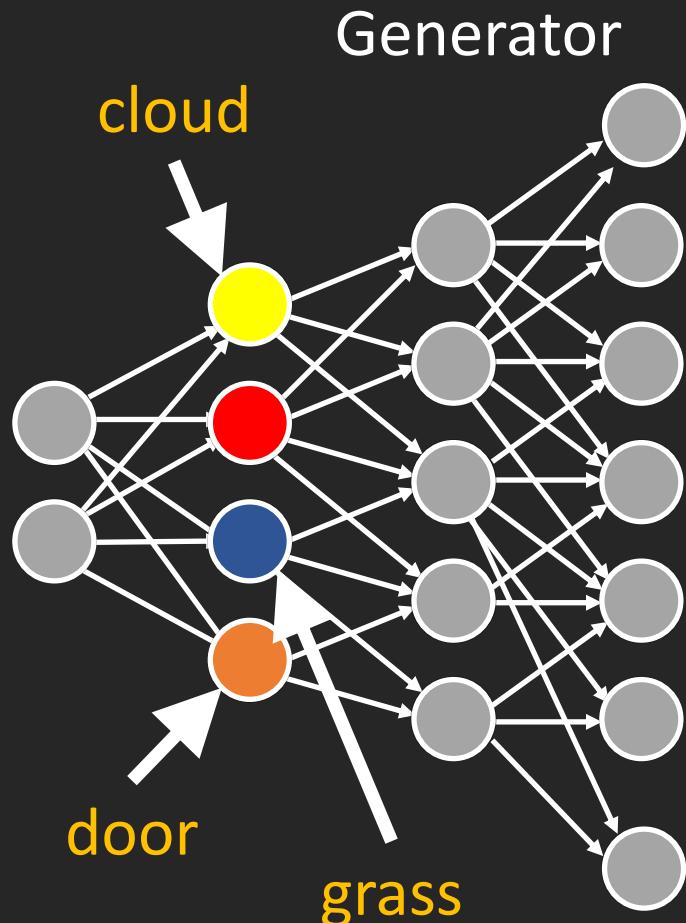
Tutorial Outline

- Semantics in the latent space
 - Supervised approach
 - Unsupervised approach
- Inversion of real images
- Future directions

Deep Generative Representation

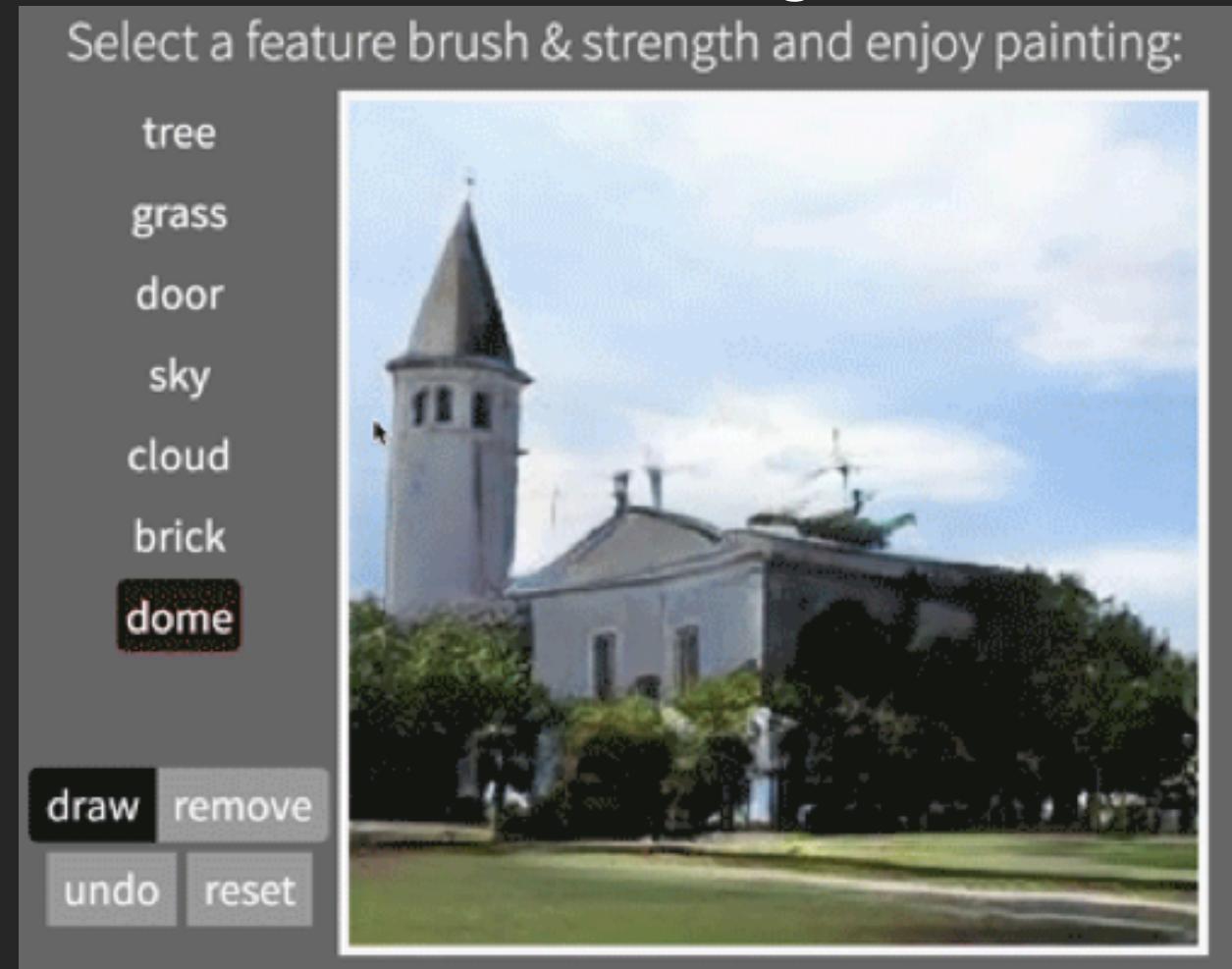
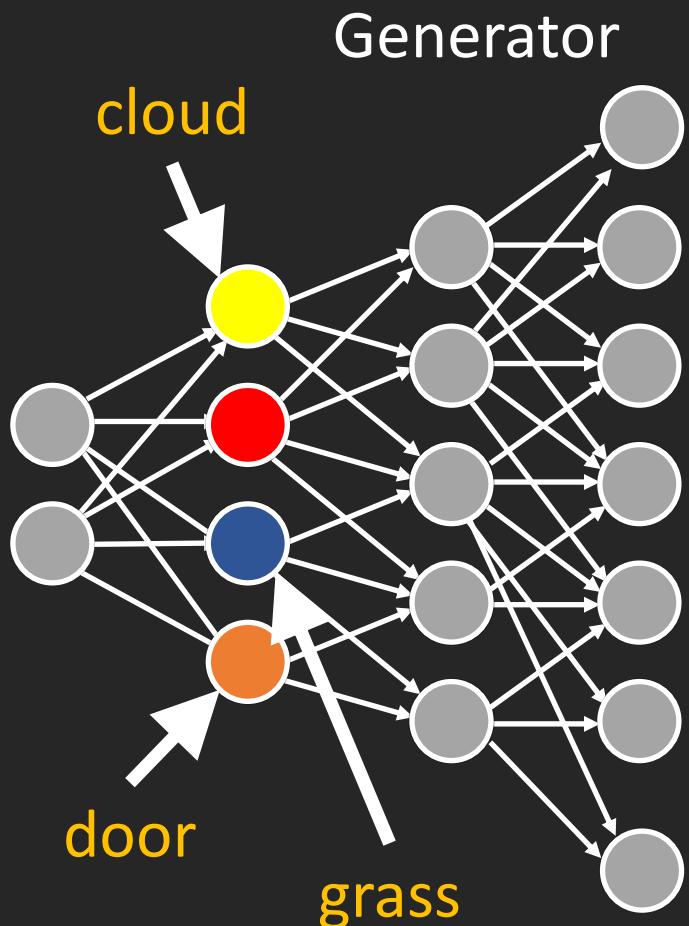


GAN Dissection for Interpreting Latent Units

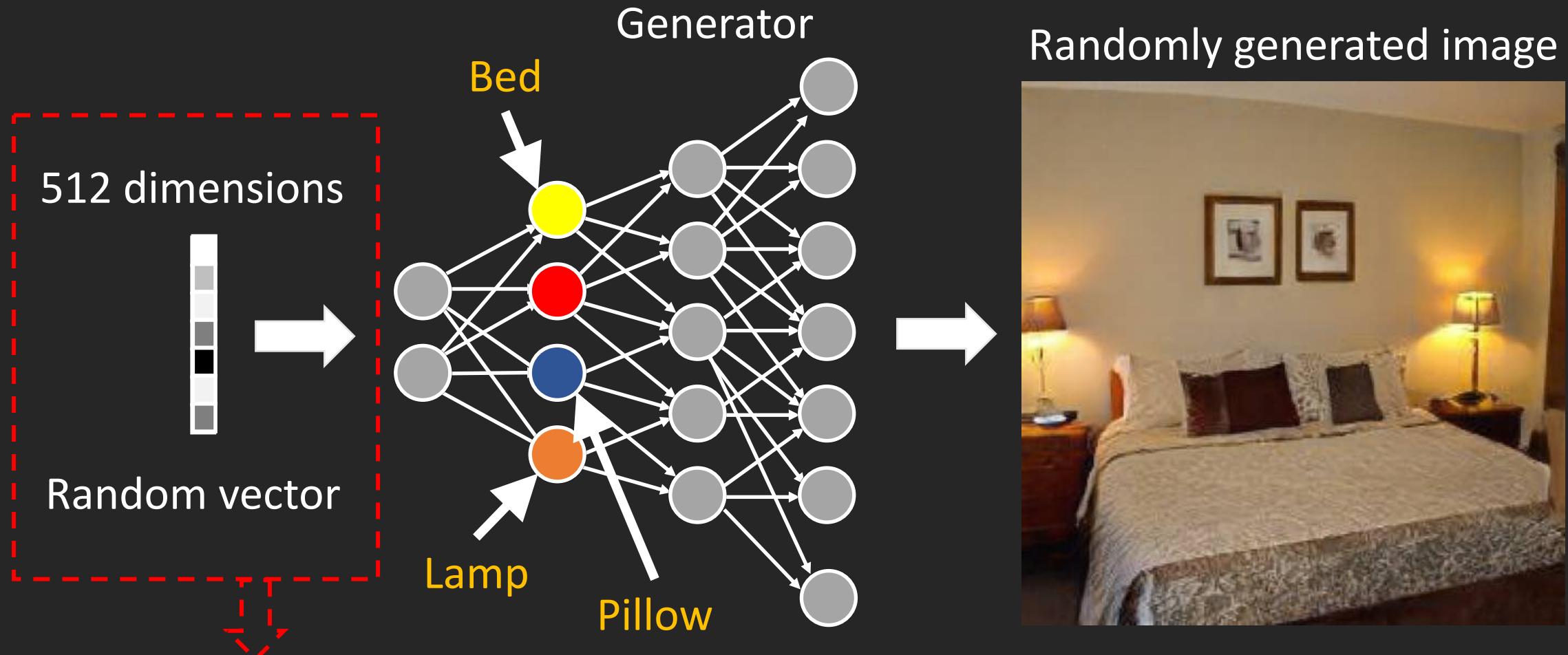


GAN Dissection for Interpreting Latent Units

Interactive editing



Deep Generative Representation



How latent space controls the semantics?

Random Walk in Latent Space of Bedroom



Multiple Levels of Scene Descriptors

Scene category:

bedroom

Scene attributes:

nature lighting

wood

foliage

...

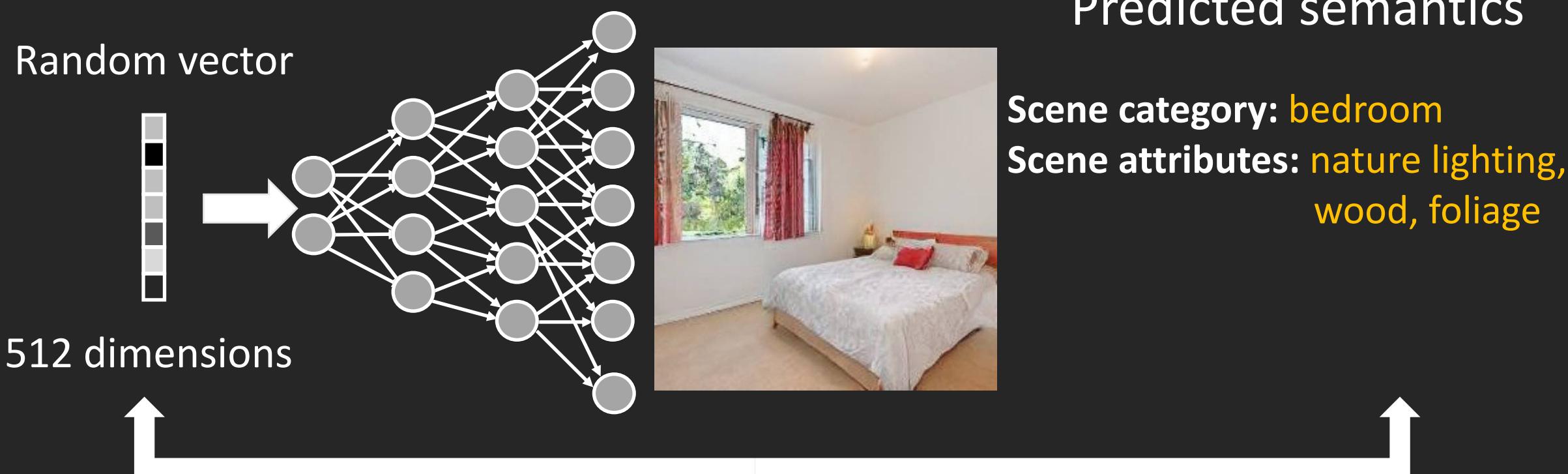
Layout



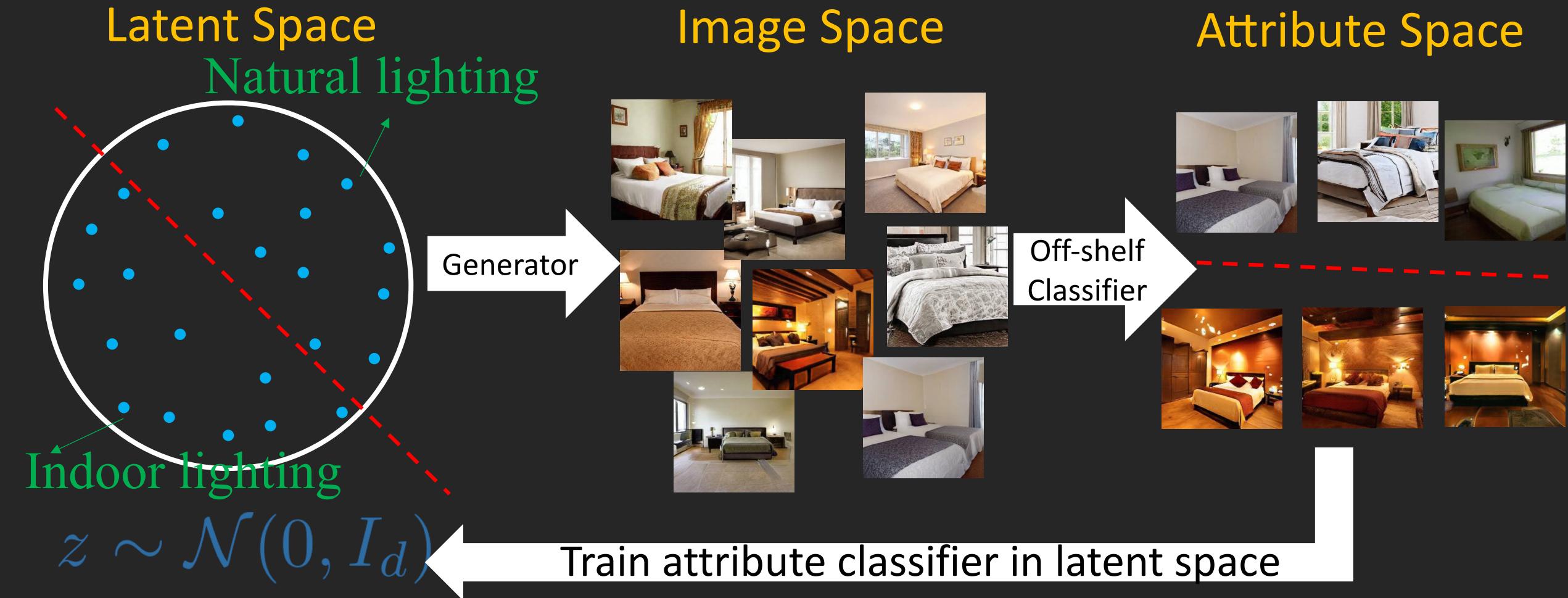
Segmentation



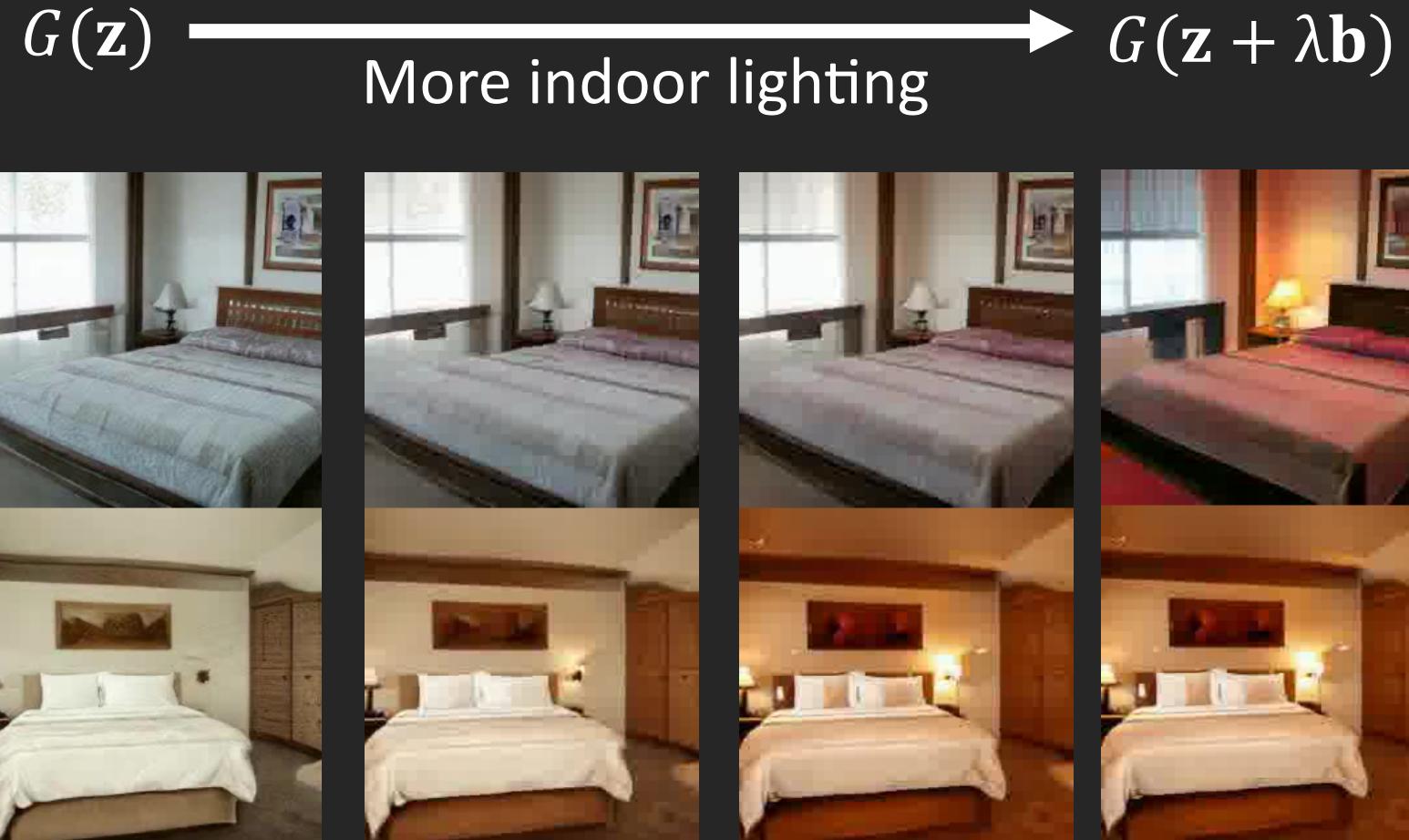
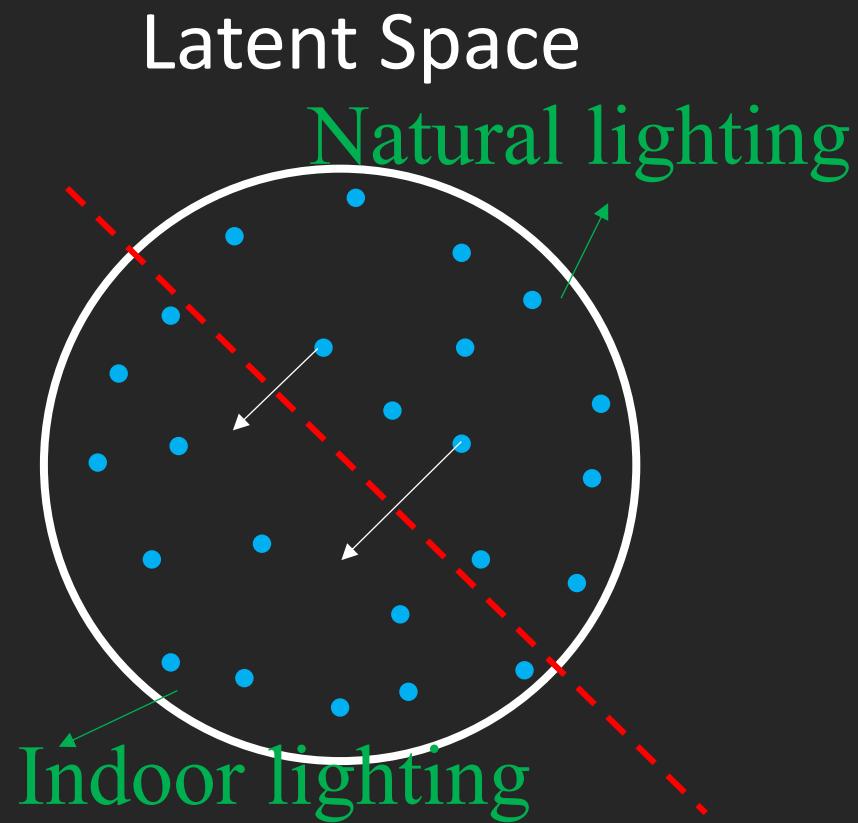
Identifying the Causality in Latent Space



Aligning Latent Space with Attribute Space



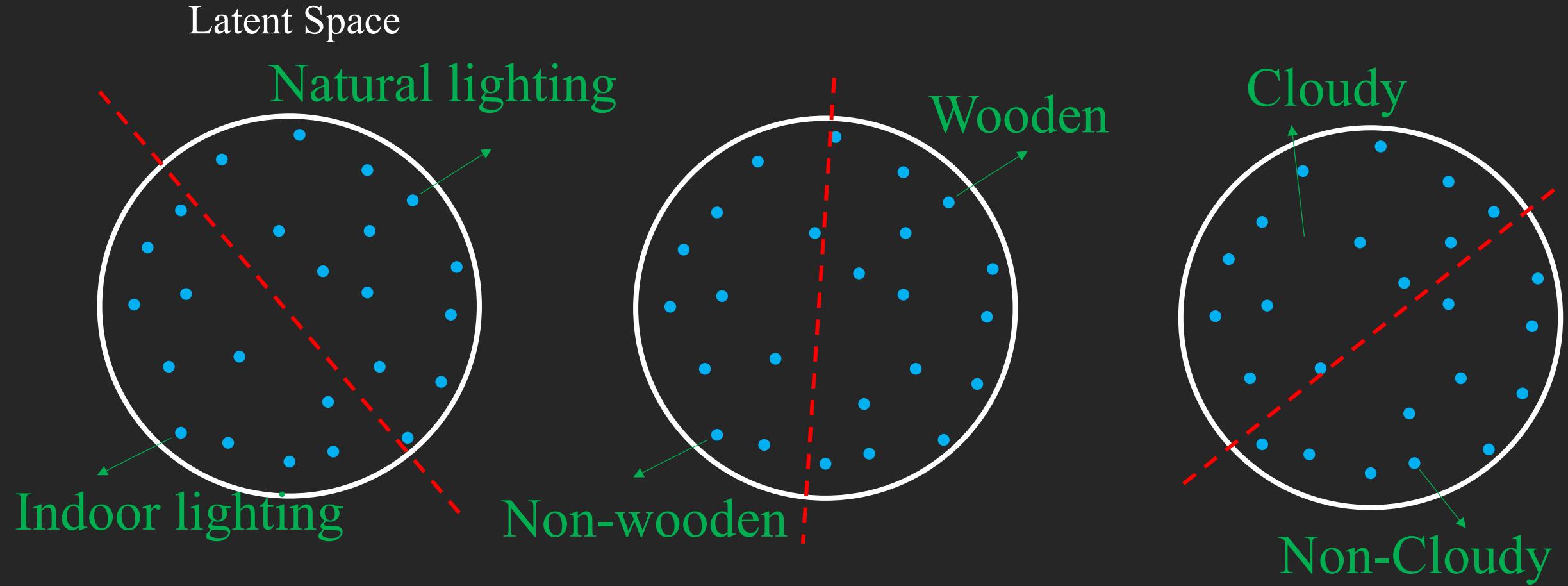
Pushing Latent Code through Boundary



Result on turning up the lights



Various Attribute Boundaries in Latent Space



Ageing the scenes



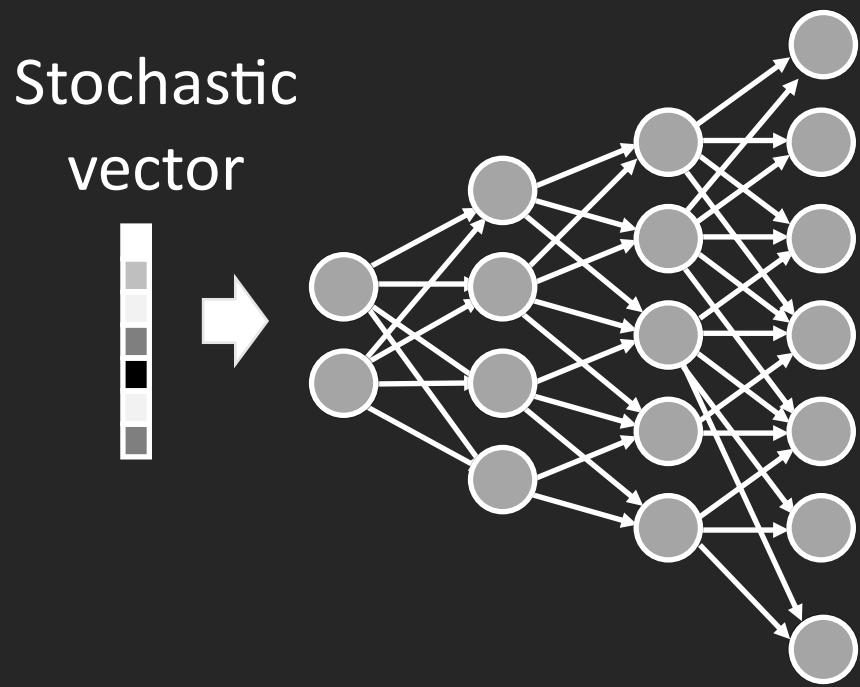
Adding more clouds



Adding more green space

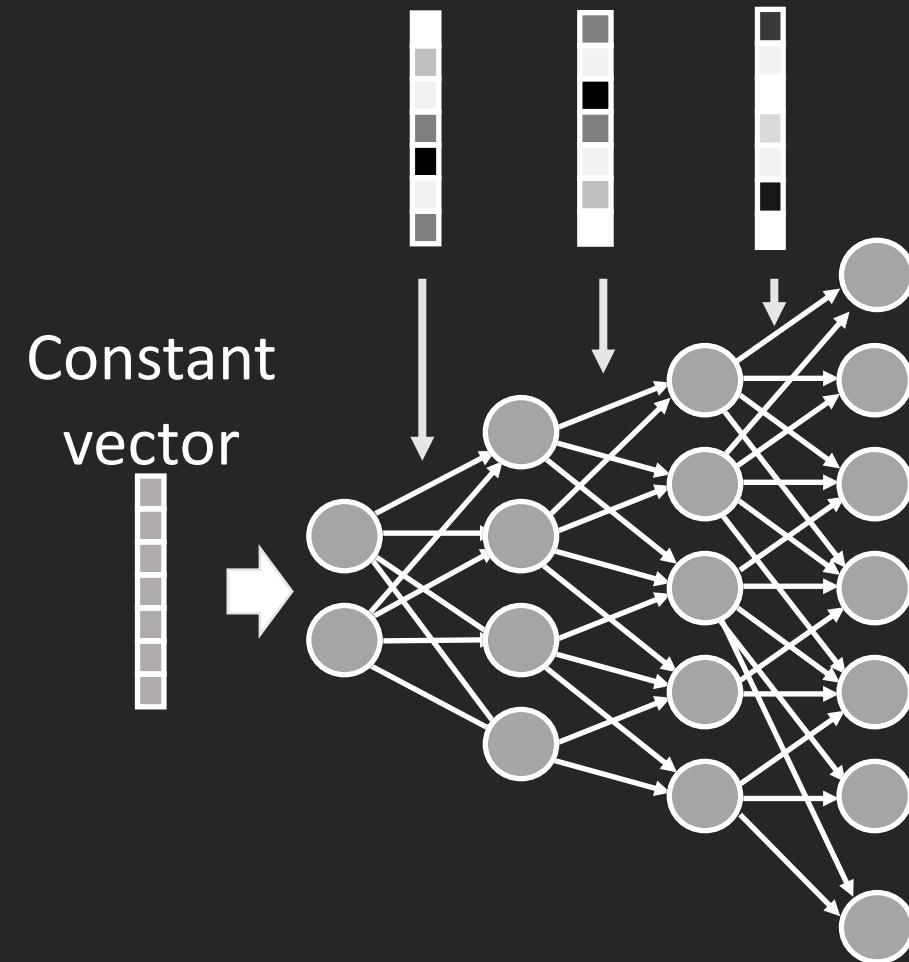


Layer-wise Stochasticity



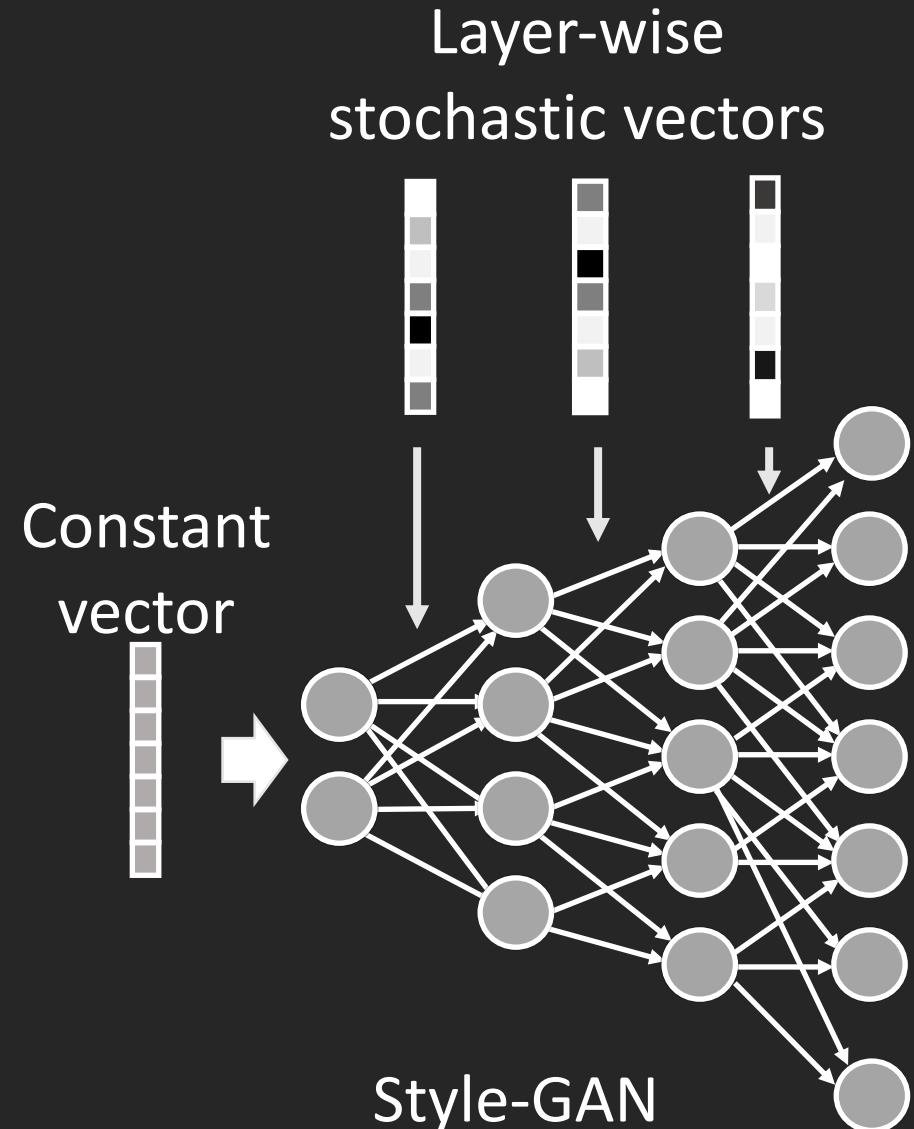
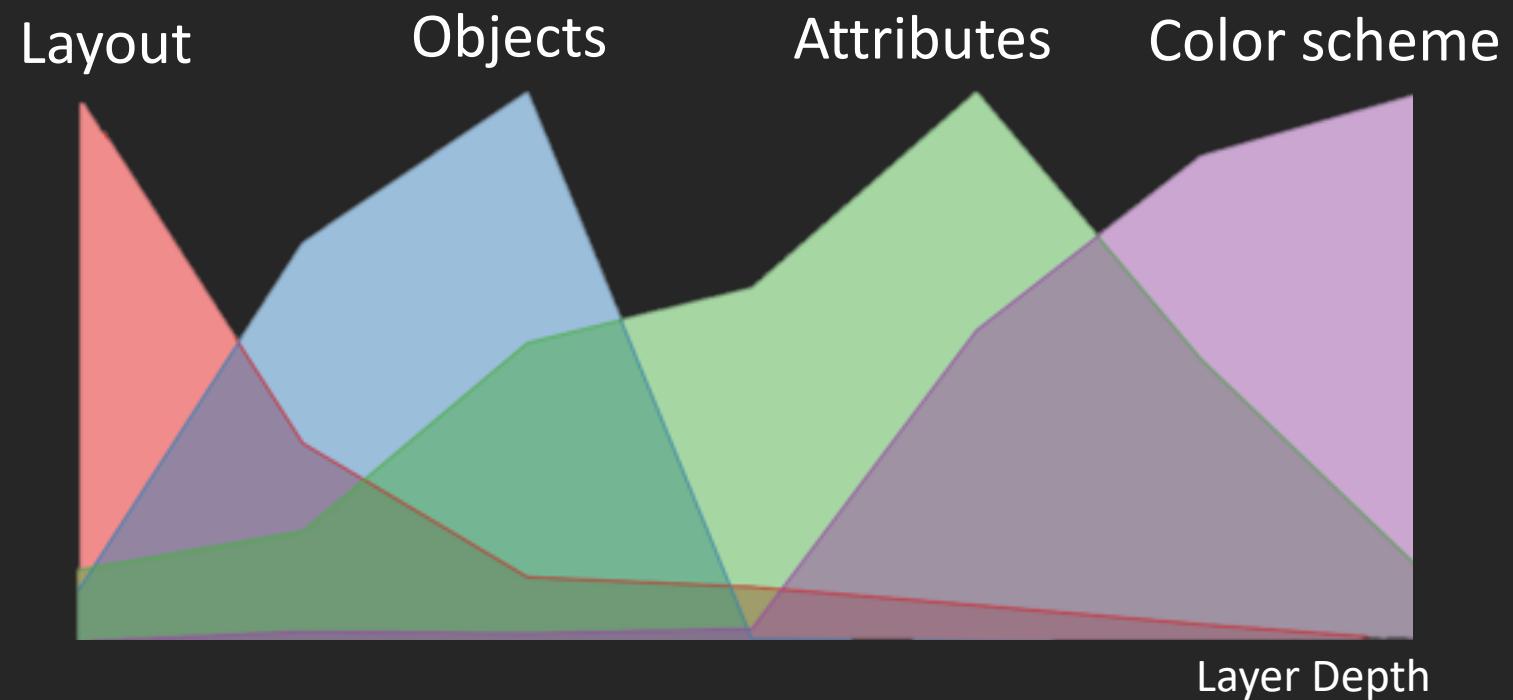
DC-GAN, PG-GAN

Layer-wise
stochastic vectors



Style-GAN [Karras et al]

Semantic hierarchy emerges



Changing layout at Layer0-1



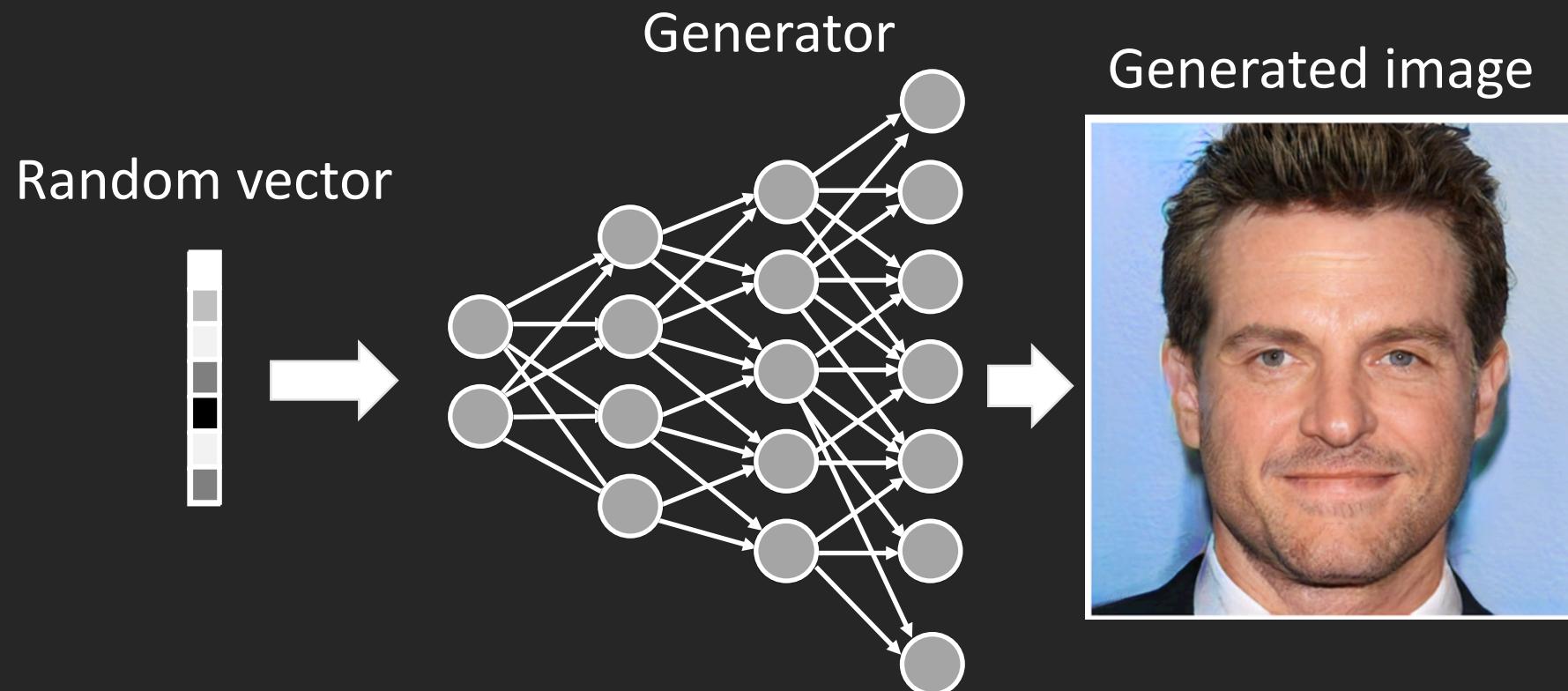
Varying category (Bedroom to Dining Room)
at layers 3-6



Varying category (Bedroom to Living Room) at layers 3-6



Latent Semantics in Face Synthesis GANs



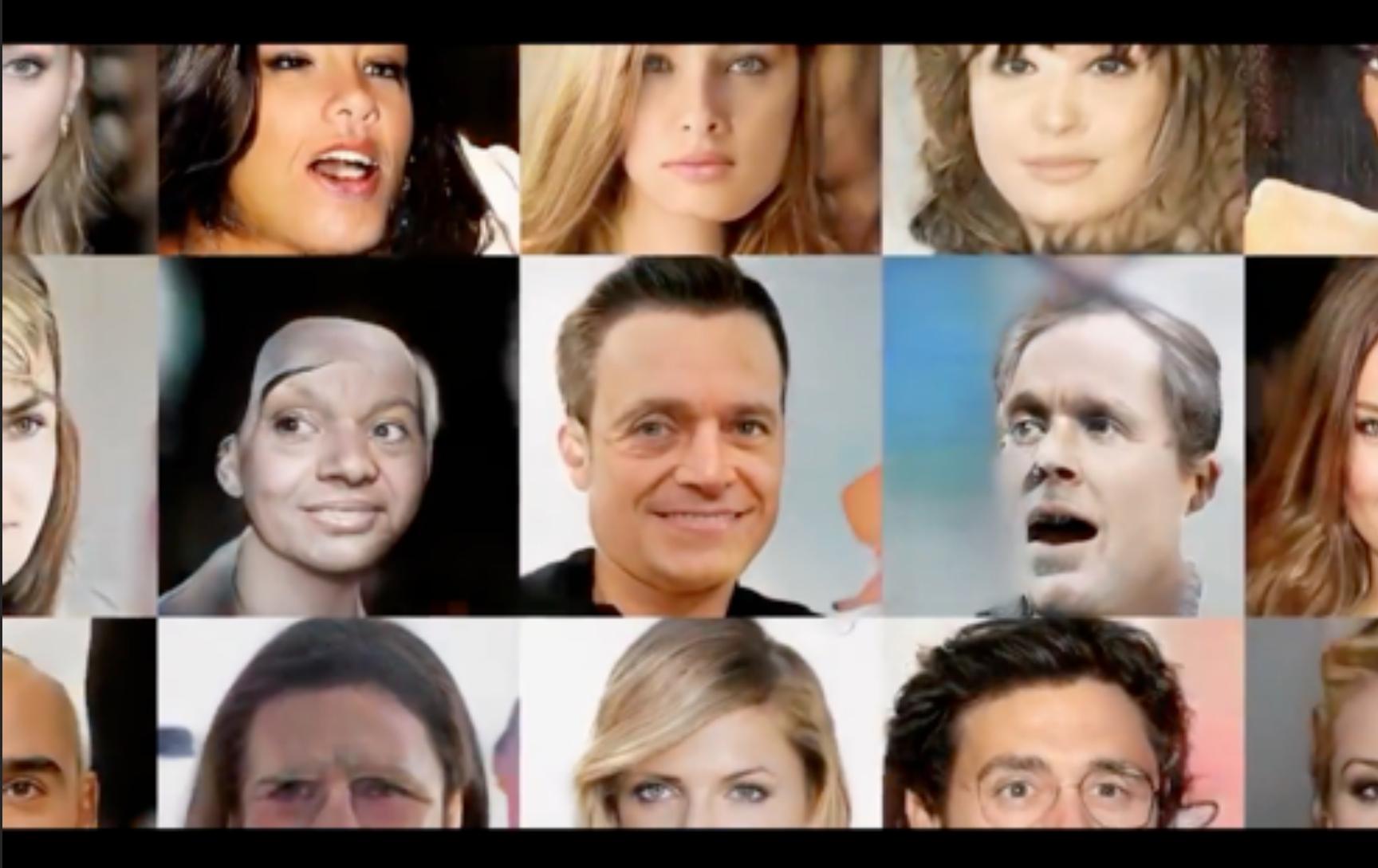
Different angle



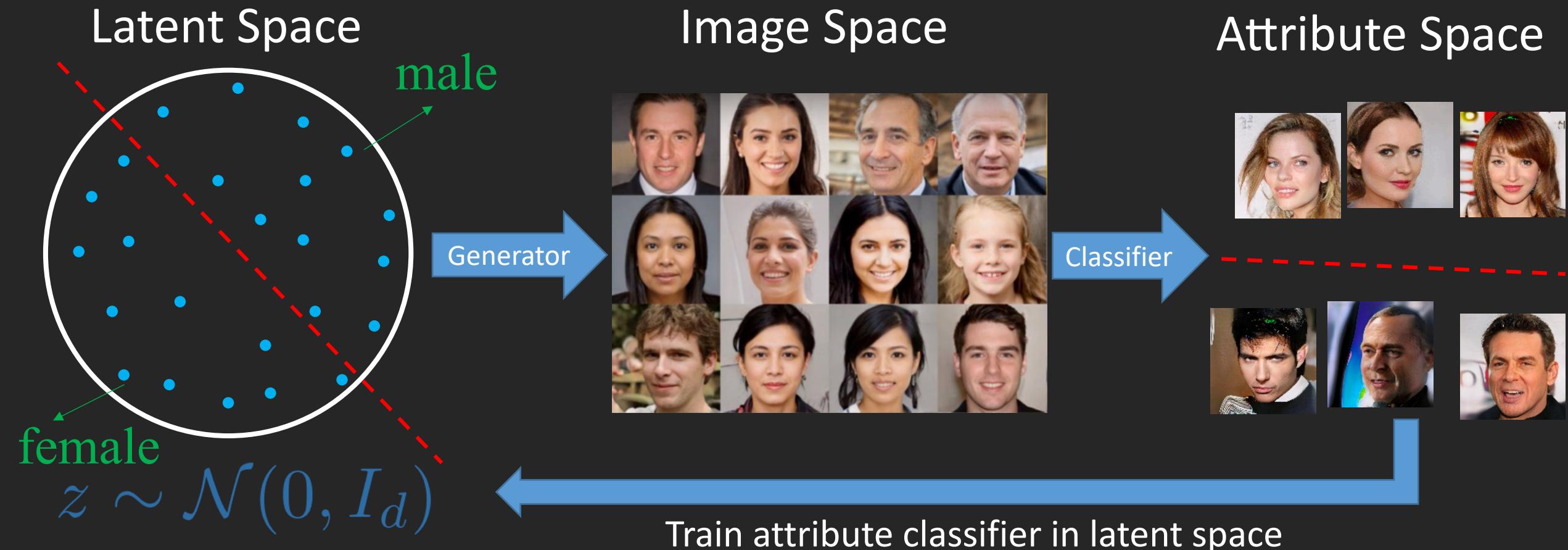
Different age



Interpolation in the Latent Space



InterFaceGAN: Bridging Latent Space to Attribute Space



Make me cooler



Make me younger



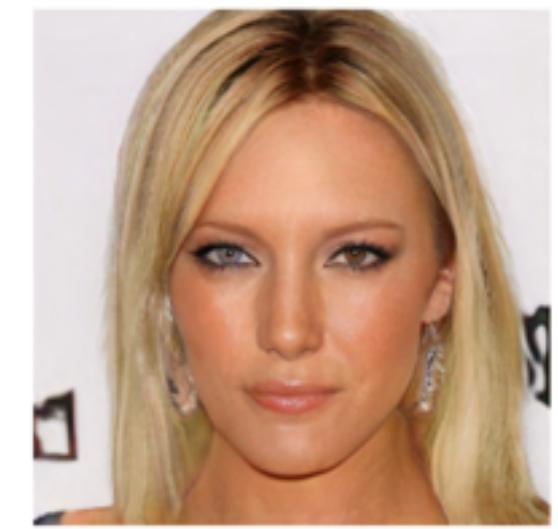
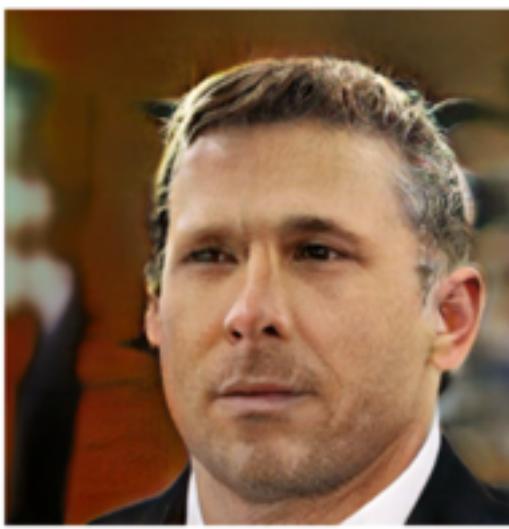
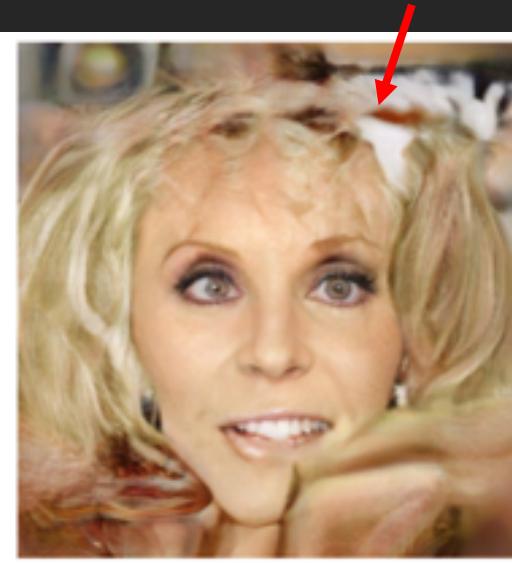
Make me front faced



Make me more man



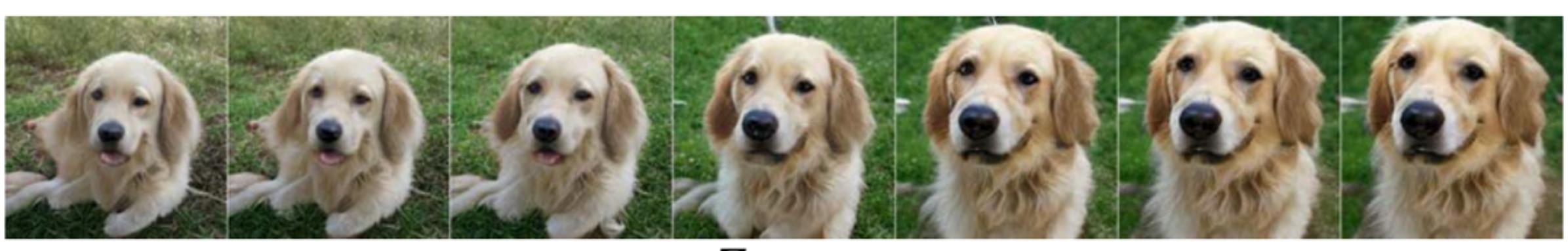
Correcting the Artifacts Made by GAN



Demo Video

Face Manipulation
with InterpretGAN

Steerability of GANs



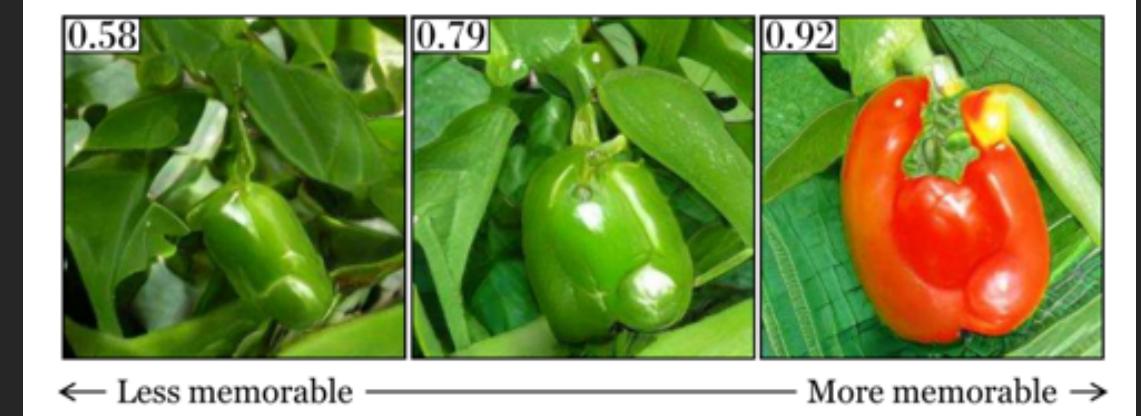
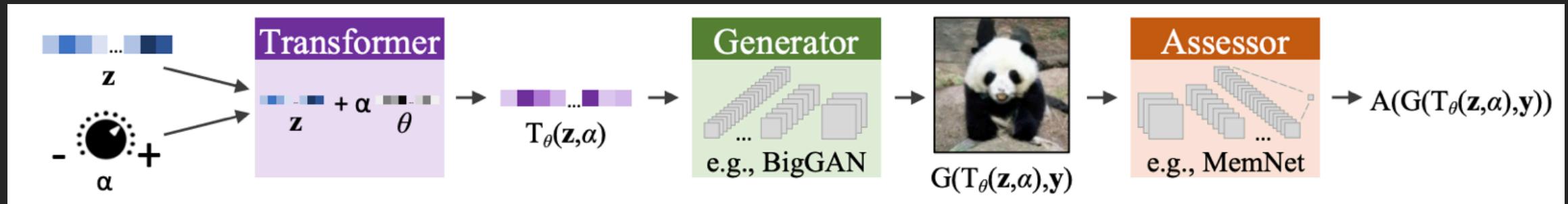
- Zoom +



- Rotate3D +

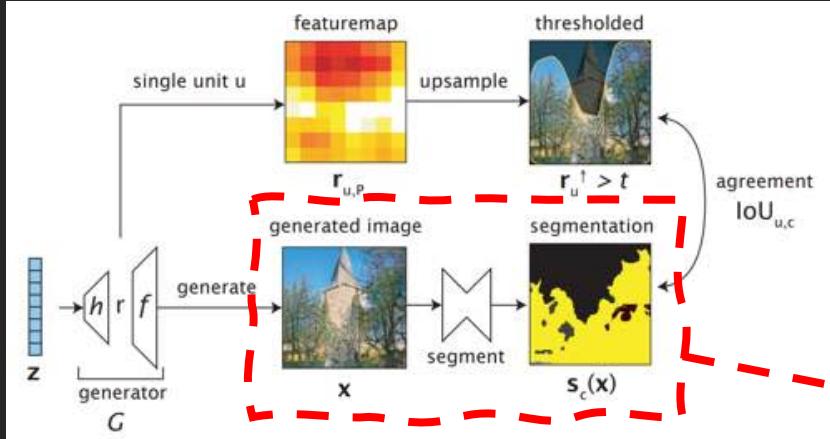
Ali Jahanian, Lucy Chai, Phillip Isola . On the "steerability" of generative adversarial networks. ICLR'20.
<https://arxiv.org/abs/1907.07171>

GANalyze for changing image memorability



All the approaches below need supervision

GAN Dissection

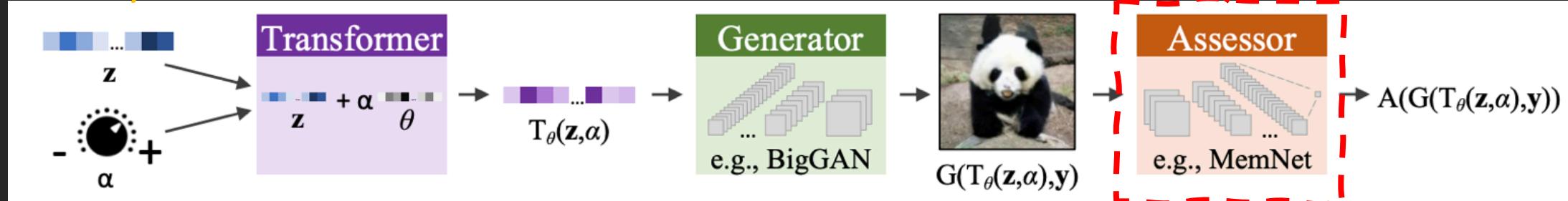


InterFaceGAN & GAN Hierarchy



Off-the-shelf supervised classifiers

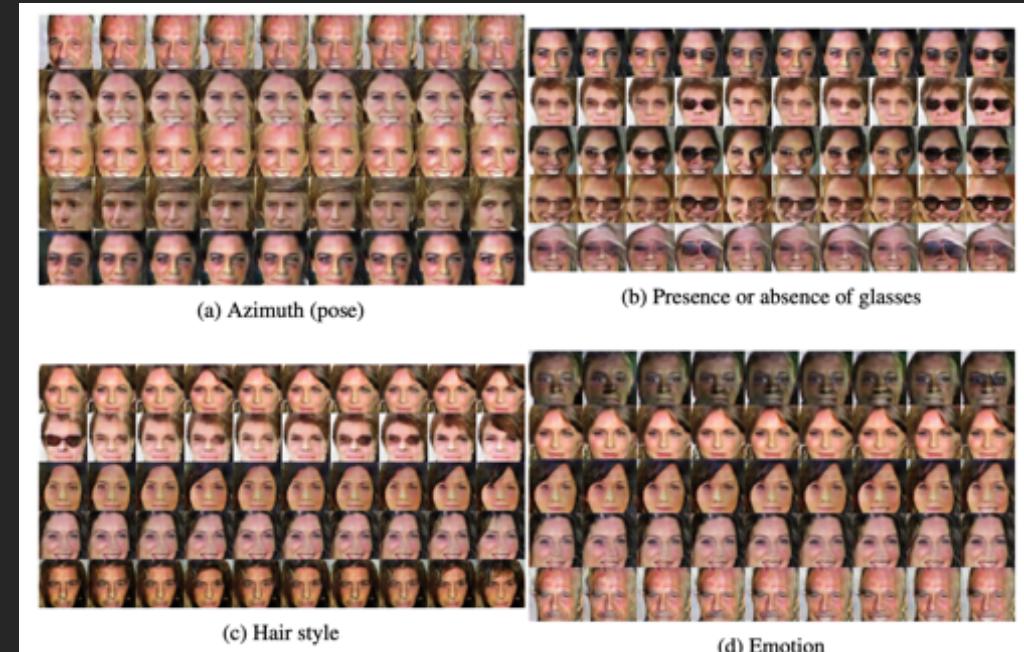
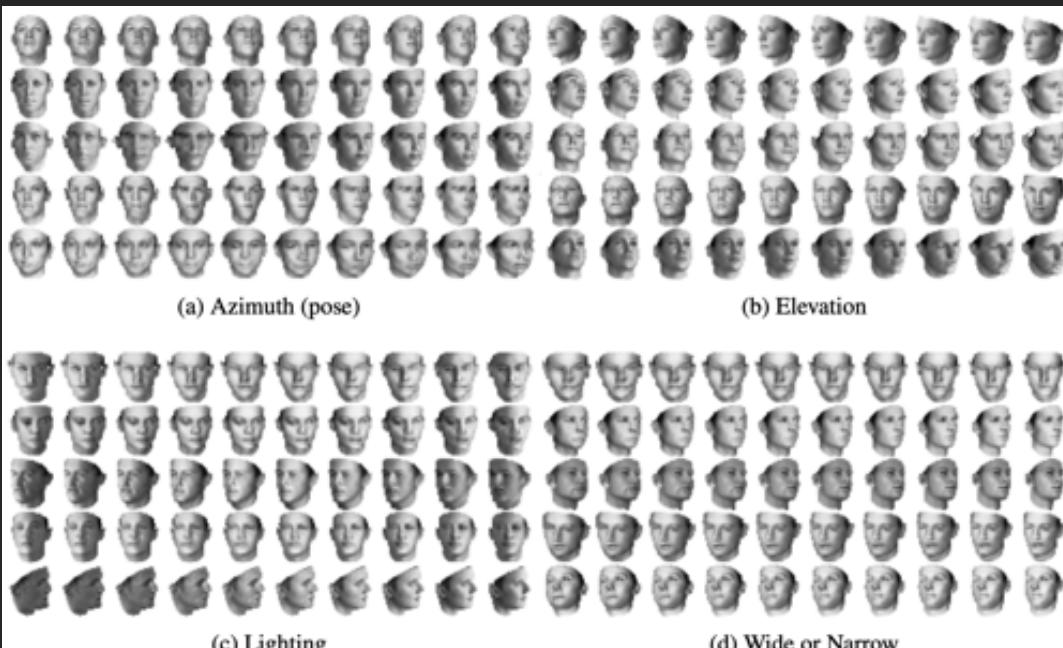
GANalyze



Unsupervised Attribute Discovery in GANs

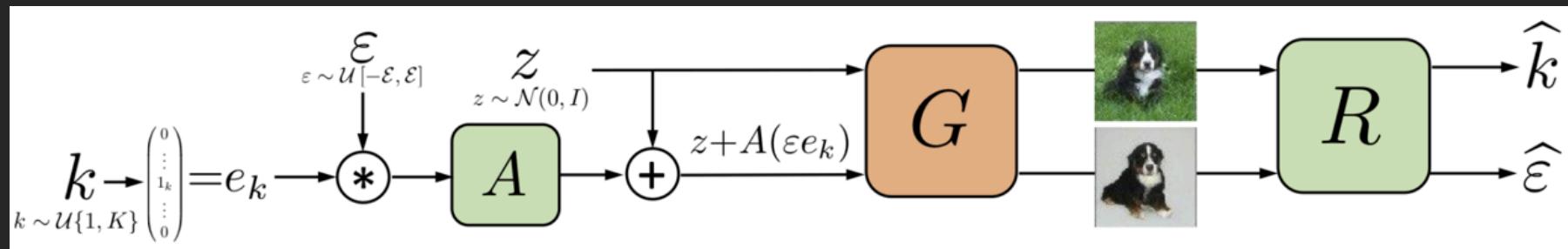
InfoGAN: Mutual information maximization is added as a training regularizer

$$\min_G \max_D V_I(D, G) = V(D, G) - \lambda I(c; G(z, c))$$



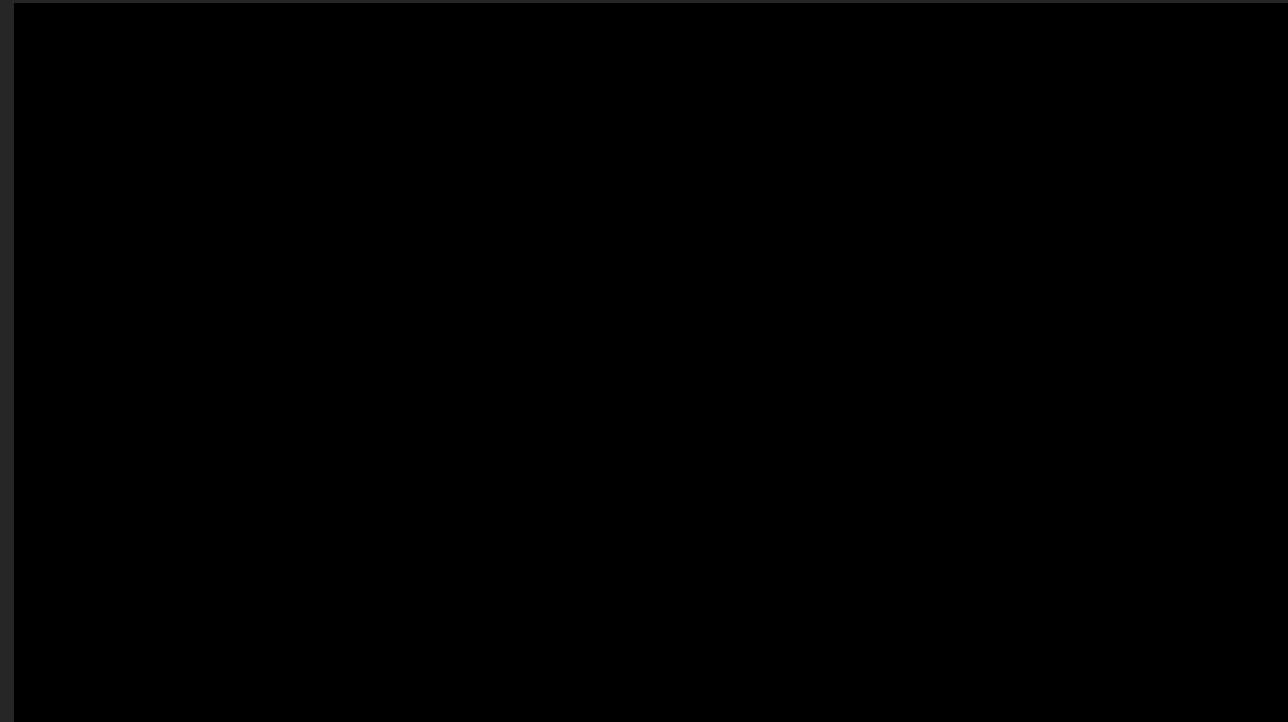
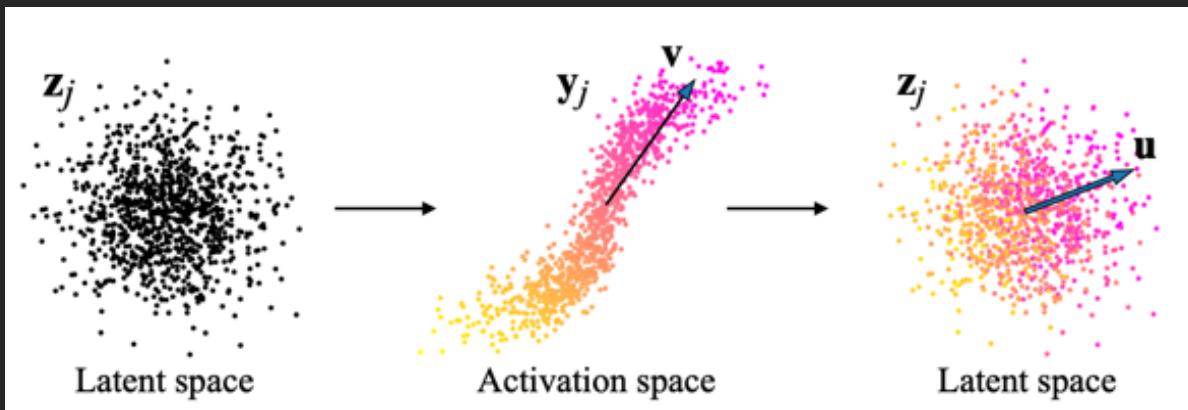
Unsupervised Attribute Discovery in GANs

Identifying the variation factors



Unsupervised Attribute Discovery in GANs

PCA on the latent activations



Unsupervised Attribute Discovery in GANs

Disentanglement based on the learned representations



Issues for unsupervised approaches

- How to evaluate the discovered interpretable dimensions?
- How to measure the disentanglement?
- How to compare different methods?

Tutorial Outline

- Semantics in the latent space
 - Supervised approach
 - Unsupervised approach
- Inversion of real images
- Future directions

How to edit my own photo?

Make me cooler



Make me younger



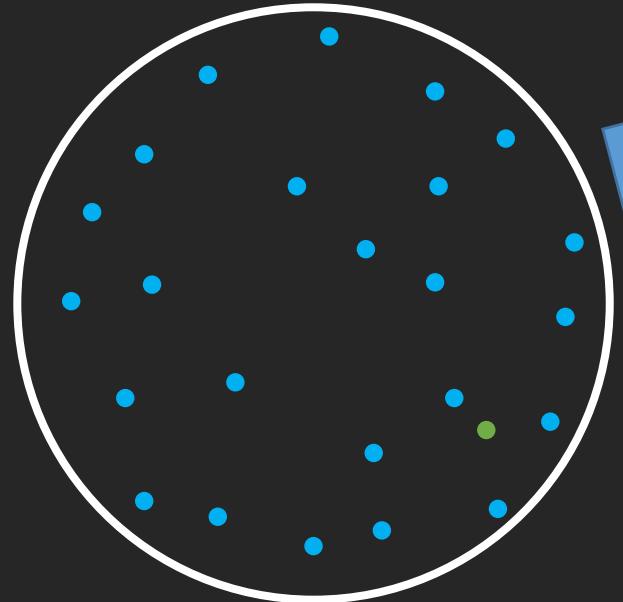
My Profile photo

GAN-Synthesized Image

GAN Inversion: Inverting Real Faces to Latent Code

Synthesized Image $\mathbf{x} = G(\mathbf{z})$

Latent Space



Generation

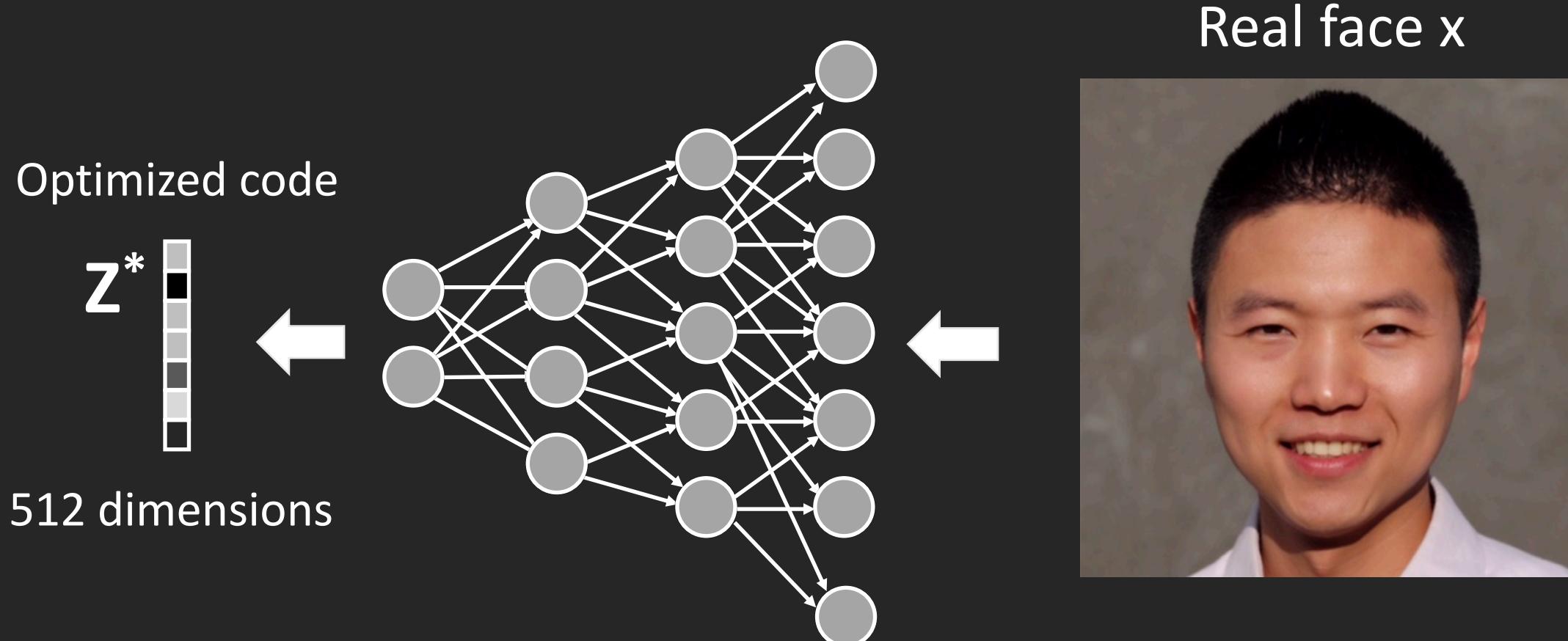


Real Image \mathbf{x}



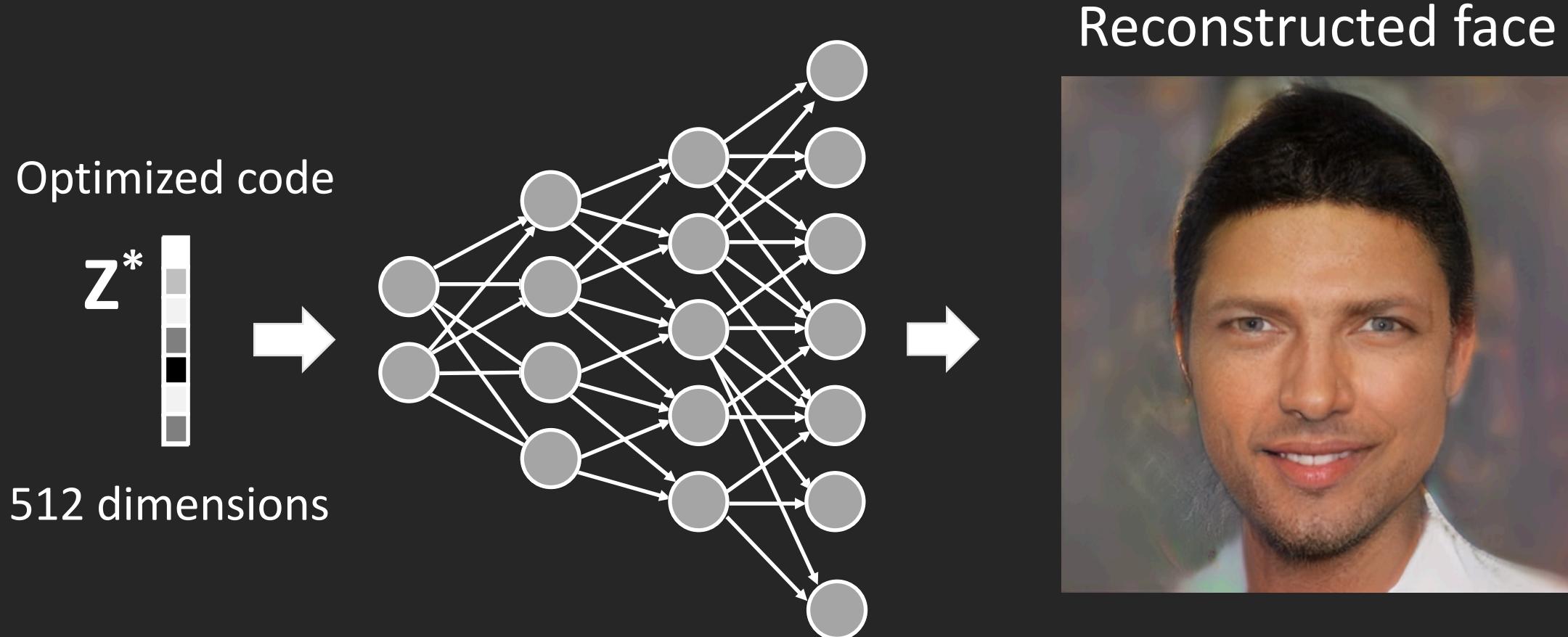
$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|G(\mathbf{z}) - \mathbf{x}\|^2$$

GAN Inversion for Faces



$$z^* = \arg \min_z ||G(z) - x||^2$$

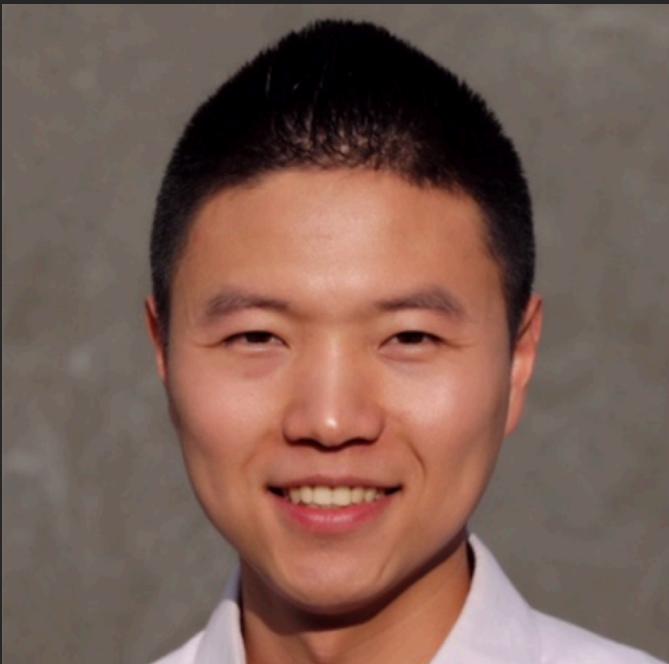
GAN Inversion for Faces



GAN inversion is challenging!

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} ||G(\mathbf{z}) - \mathbf{x}||^2$$

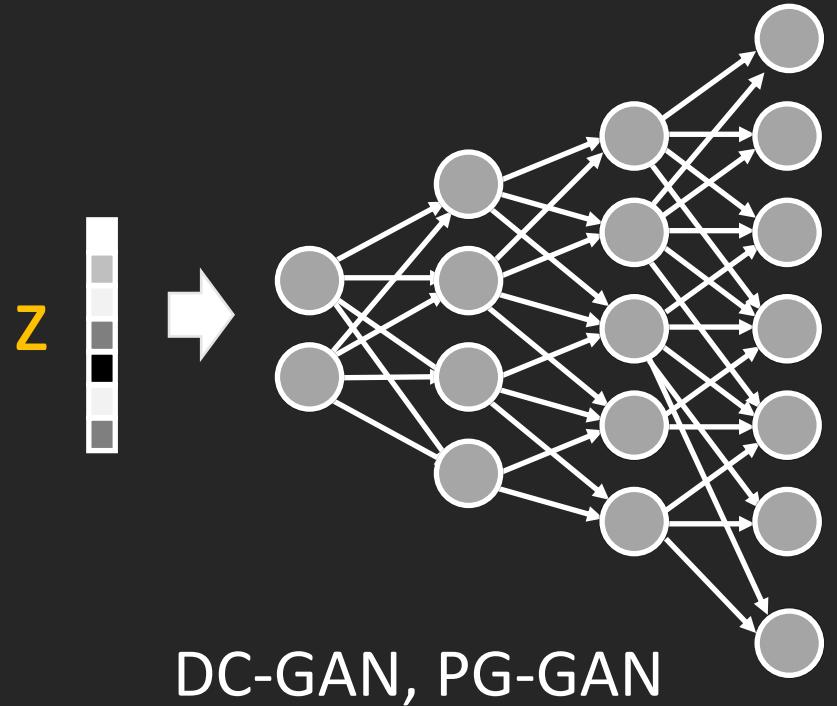
Inversion



Different initialization



Extended latent codes



$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|G(\mathbf{z}) - \mathbf{x}\|^2$$

$$\mathbf{z}_1^*, \mathbf{z}_2^*, \dots = \arg \min_{\mathbf{z}_1, \mathbf{z}_2, \dots} \|G(\mathbf{z}_1, \mathbf{z}_2, \dots) - \mathbf{x}\|^2$$

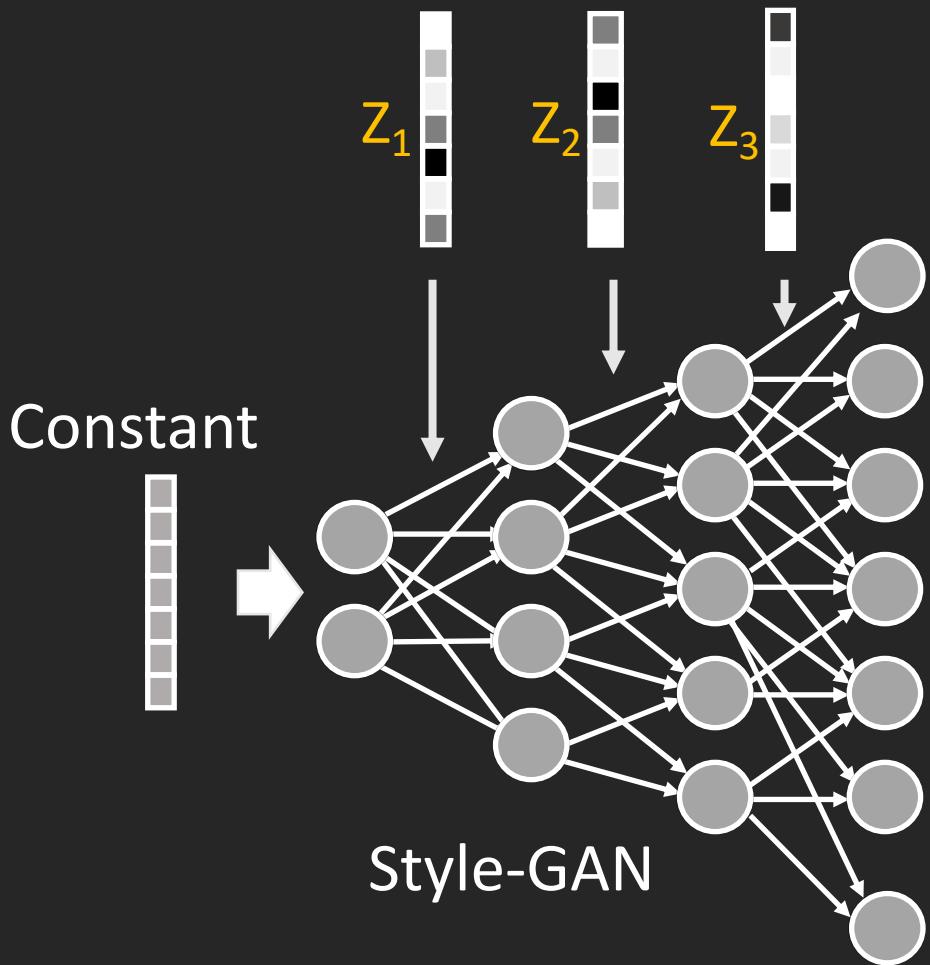


Image2StyleGAN inversion: it kind of works!

Dimension of the latent space is increased from 512 to 14x512

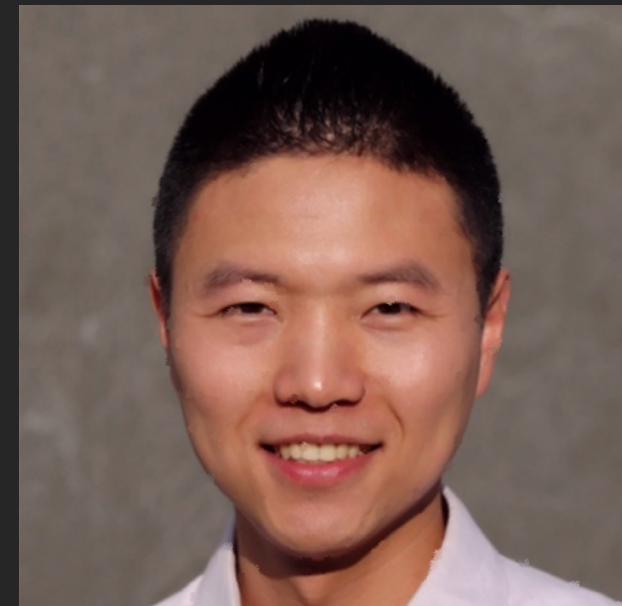
Target



Inversion from
single code



Inversion from
Image2StyleGAN (multiple codes)



But it seems overfitting the given image

StyleGAN generator is trained on human faces only

Target



Image2StyleGAN



Target

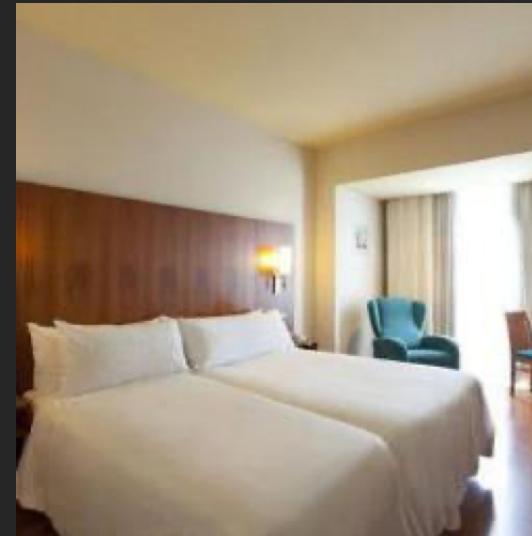


Image2StyleGAN



But it seems overfitting the given image

Inverted codes don't very well support the manipulation

Target

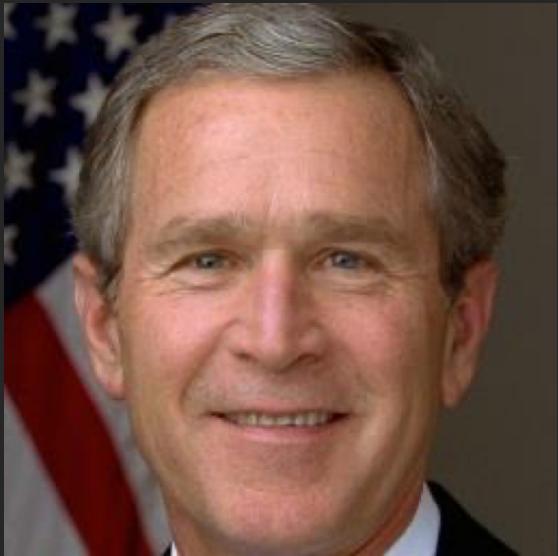


Image2StyleGAN



Add smile

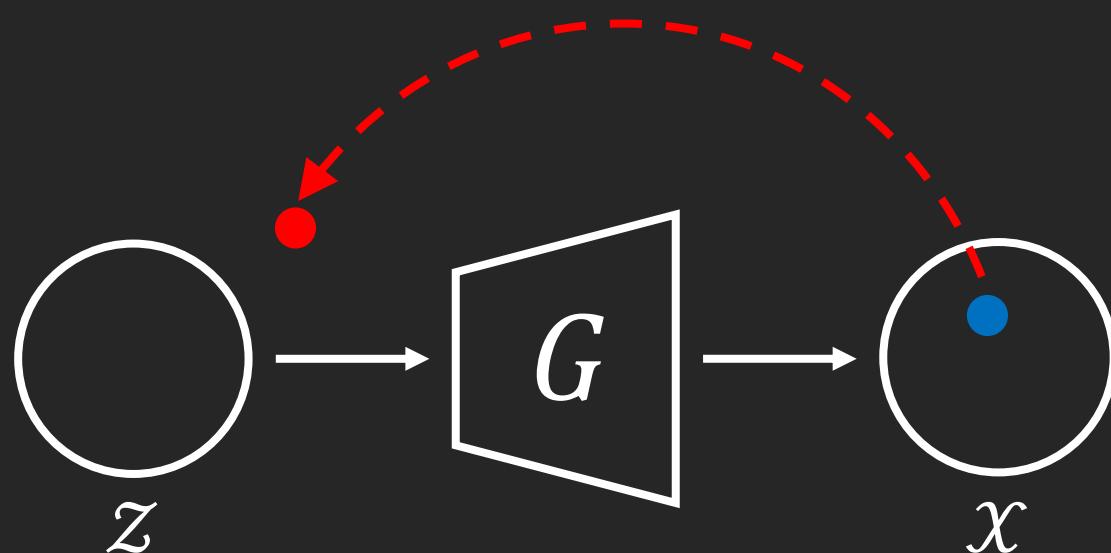


Add glasses



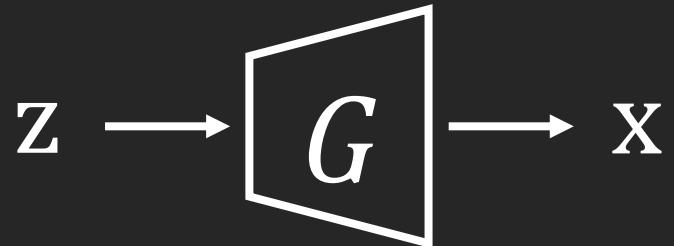
Issue: resulting code might be out of the original latent domain

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|\mathbf{G}(\mathbf{z}) - \mathbf{x}\|^2$$

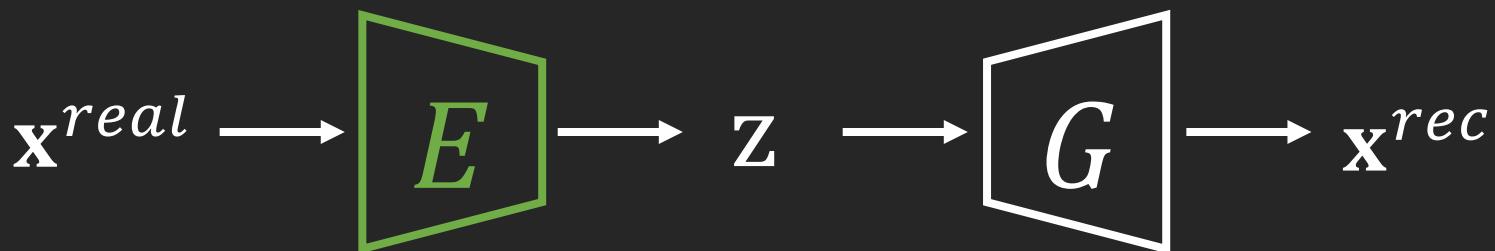


Out-of-Domain Inversion

GAN generator lacks the inference ability:



Many recent works on adding encoder to GAN generator:



BigBiGAN (NeurIPS'19)



Input

Reconstruction

Adversarial Latent Autoencoders (CVPR'20)



Input

Reconstruction

In-Domain Inversion (arXiv)

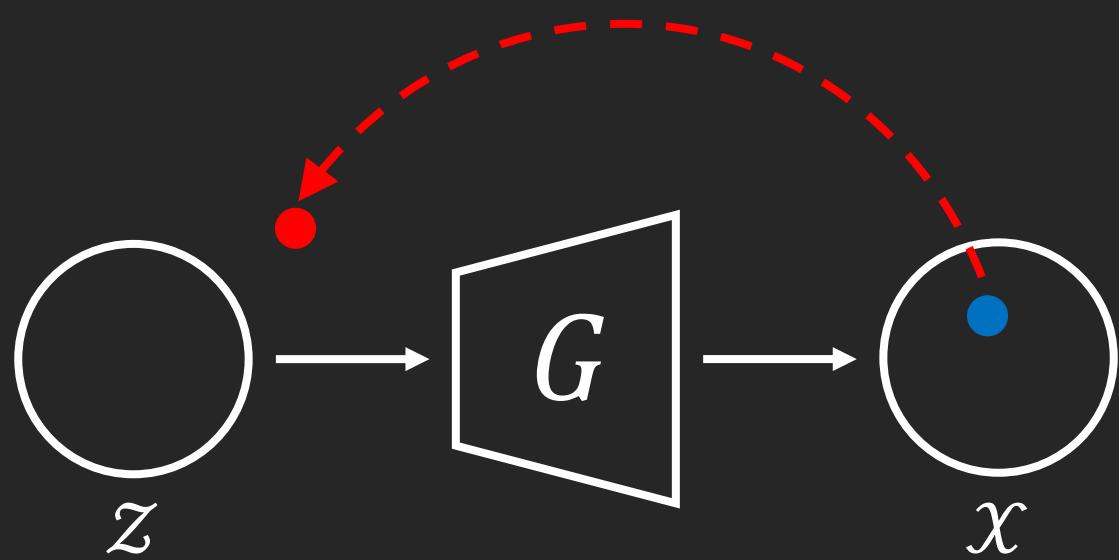


Input

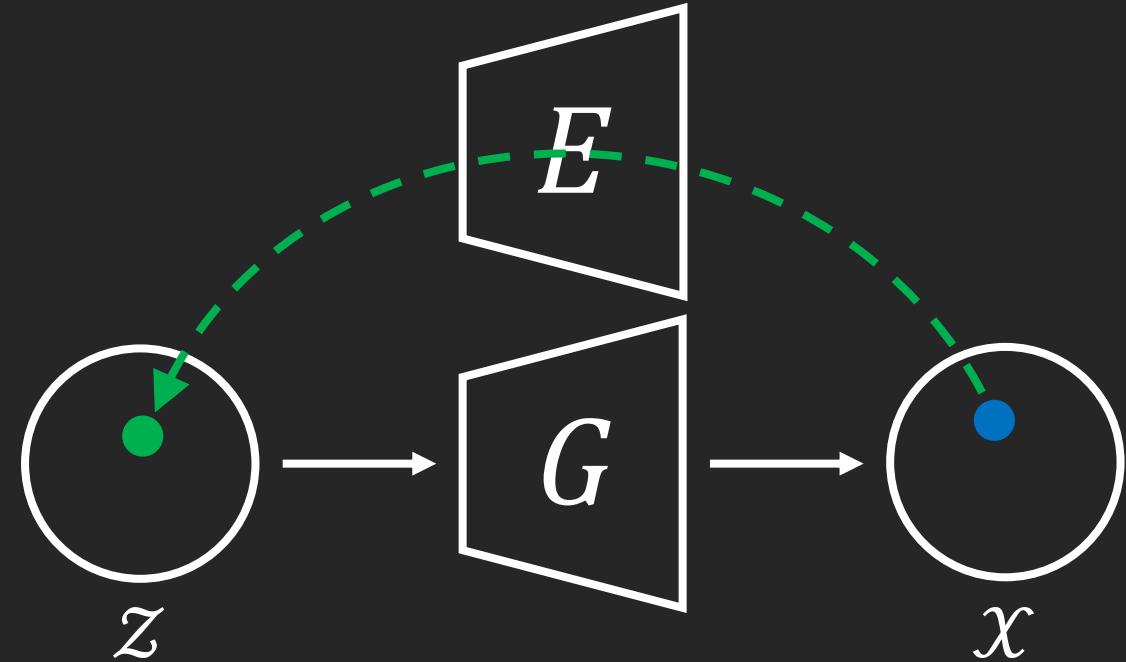
Reconstruction

Encoder-constrained Optimization

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|\mathbf{G}(\mathbf{z}) - \mathbf{x}\|^2$$



Unconstrained Inversion

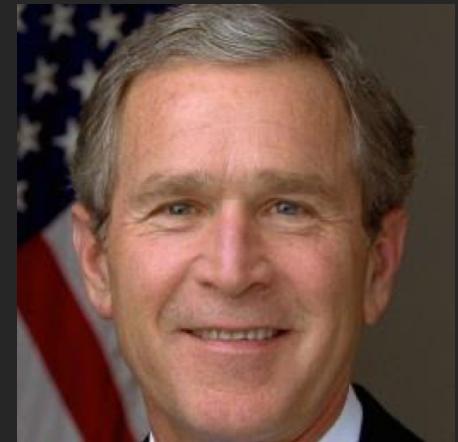


In-Domain Inversion

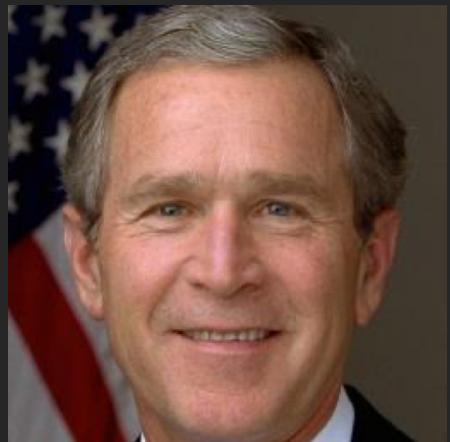
$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \|\mathbf{G}(\mathbf{z}) - \mathbf{x}\|^2 + \|\mathbf{z} - E(\mathbf{G}(\mathbf{z}))\|^2$$

Comparison with Image2StyleGAN

Image2StyleGAN



In-Domain Inversion



Target

Reconstruction

Decrease age

Add smile

Add glasses

Demo of image manipulation



Expression

Demo of image interpolation



Real Image Interpolation
Face

Fun Application: Semantic Diffusion

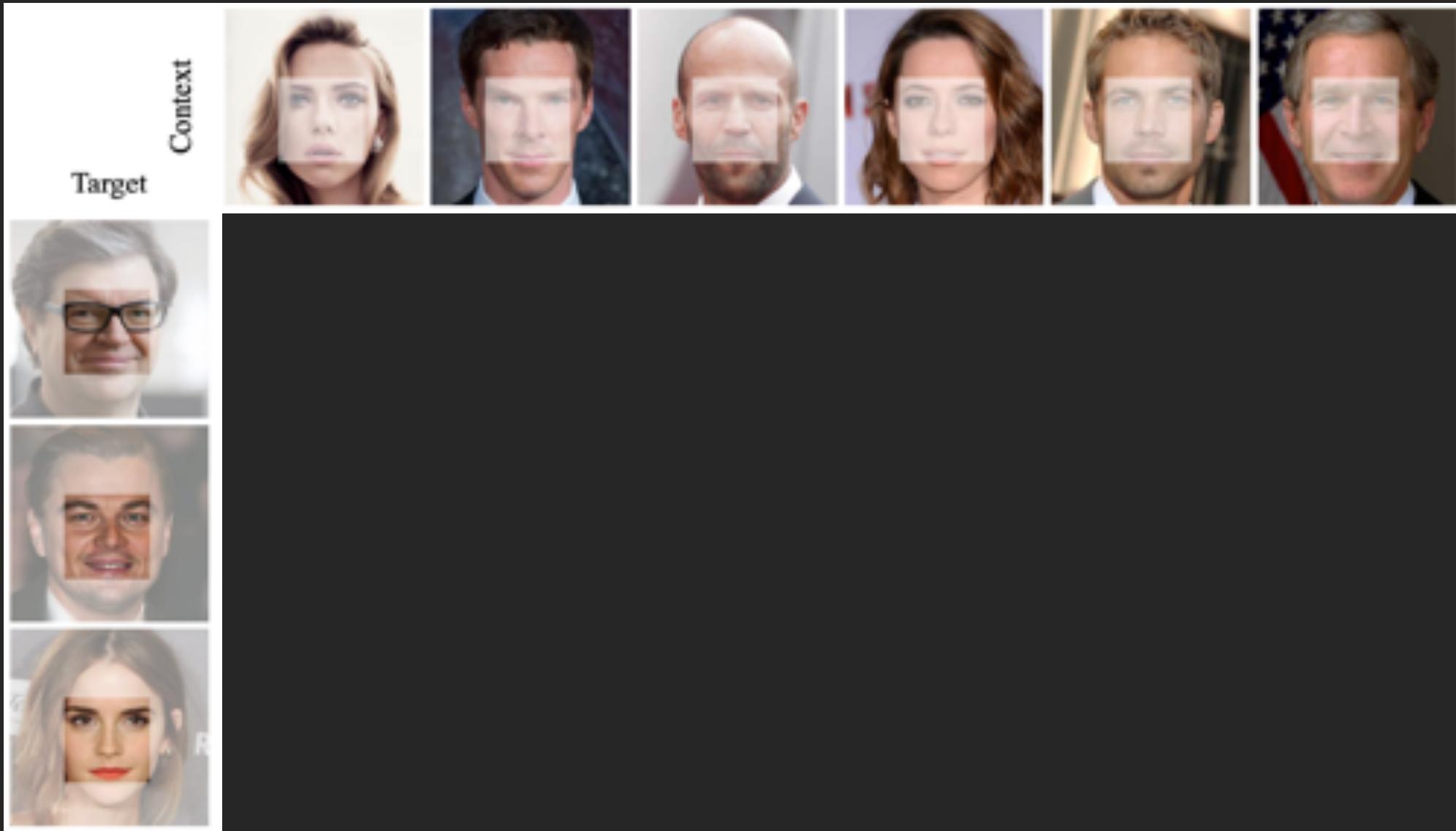


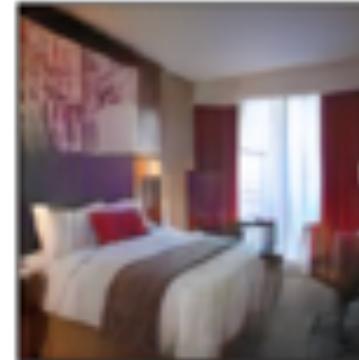
Image Processing with GAN Prior



(a) GAN inversion result



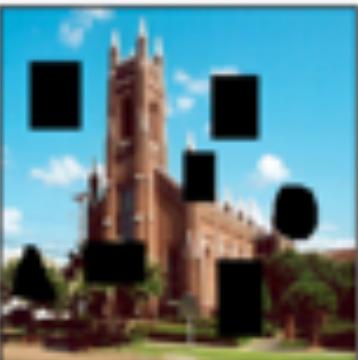
(b) Colorization



(c) Super-Resolution



(d) Image reconstruction



(e) Image inpainting



(f) Semantic image manipulation

Acknowledgement

My research team at CUHK

Yujun Shen



Ceyuan Yang



Jinjin Gu



Jiapeng Zhu



Code and papers are at
<https://genforce.github.io/>



Research Initiative on Generative Modeling

Projects



InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs

Yujun Shen, Ceyuan Yang, Xiaou Tang, Bolei Zhou
arXiv.2005.09635 preprint

[Paper] [Project Page] [Code] [Demo Video]



In-Domain GAN Inversion for Real Image Editing

Jiapeng Zhu*, Yujun Shen*, Deli Zhao, Bolei Zhou
arXiv.2004.00049 preprint

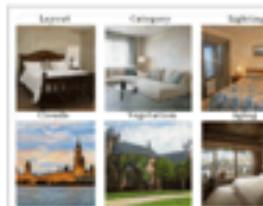
[Paper] [Project Page] [Code] [Demo Video]



Disentangled Inference for GANs with Latently Invertible Autoencoder

Jiapeng Zhu*, Deli Zhao*, Bo Zhang, Bolei Zhou
arXiv.1906.08090 preprint

[Paper] [Code]



Semantic Hierarchy Emerges in Deep Generative Representations for Scene Synthesis

Ceyuan Yang*, Yujun Shen*, Bolei Zhou
arXiv.1911.09267 preprint

[Paper] [Project Page] [Code] [Demo Video]



Image Processing Using Multi-Code GAN Prior

Jinjin Gu, Yujun Shen, Bolei Zhou
Computer Vision and Pattern Recognition (CVPR), 2020

[Paper] [Project Page] [Code]

May Generative Force be with You

- Interpretability of generative models is a rising topic
- Please leave comment below for any questions

