

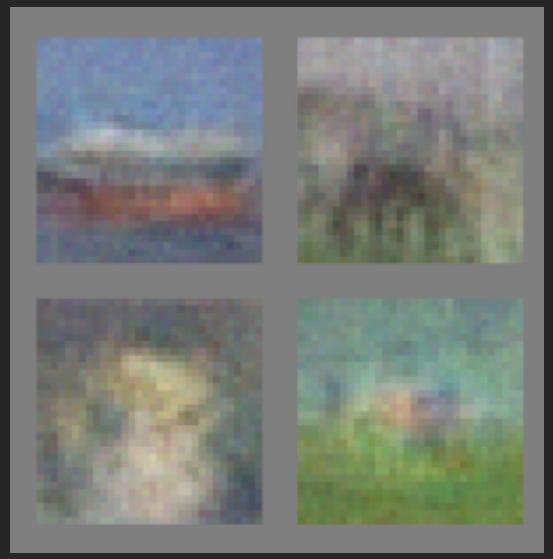

# Understanding Latent Semantics in GANs

Bolei Zhou

The Chinese University of Hong Kong (CUHK)

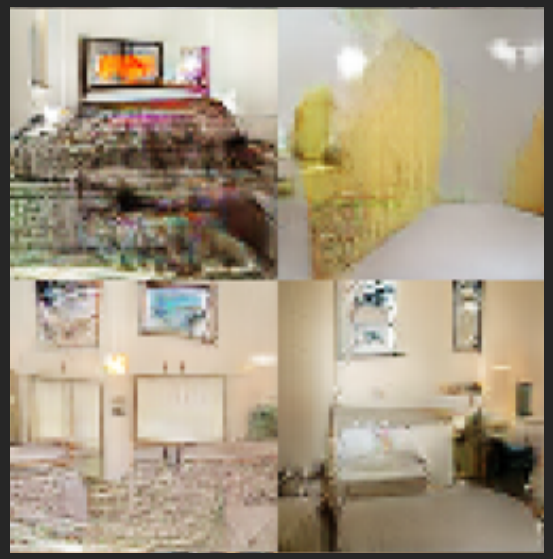
# GANs for Synthesizing Images

2014



GANs [Goodfellow et al.]

2015

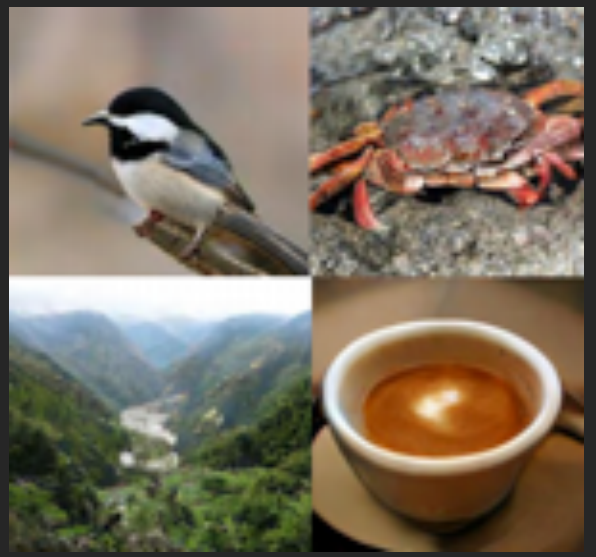


DCGAN [Radford et al.]

2017

PG-GAN [Karras et al.]

2018



BigGAN [Brock et al.]

• • •

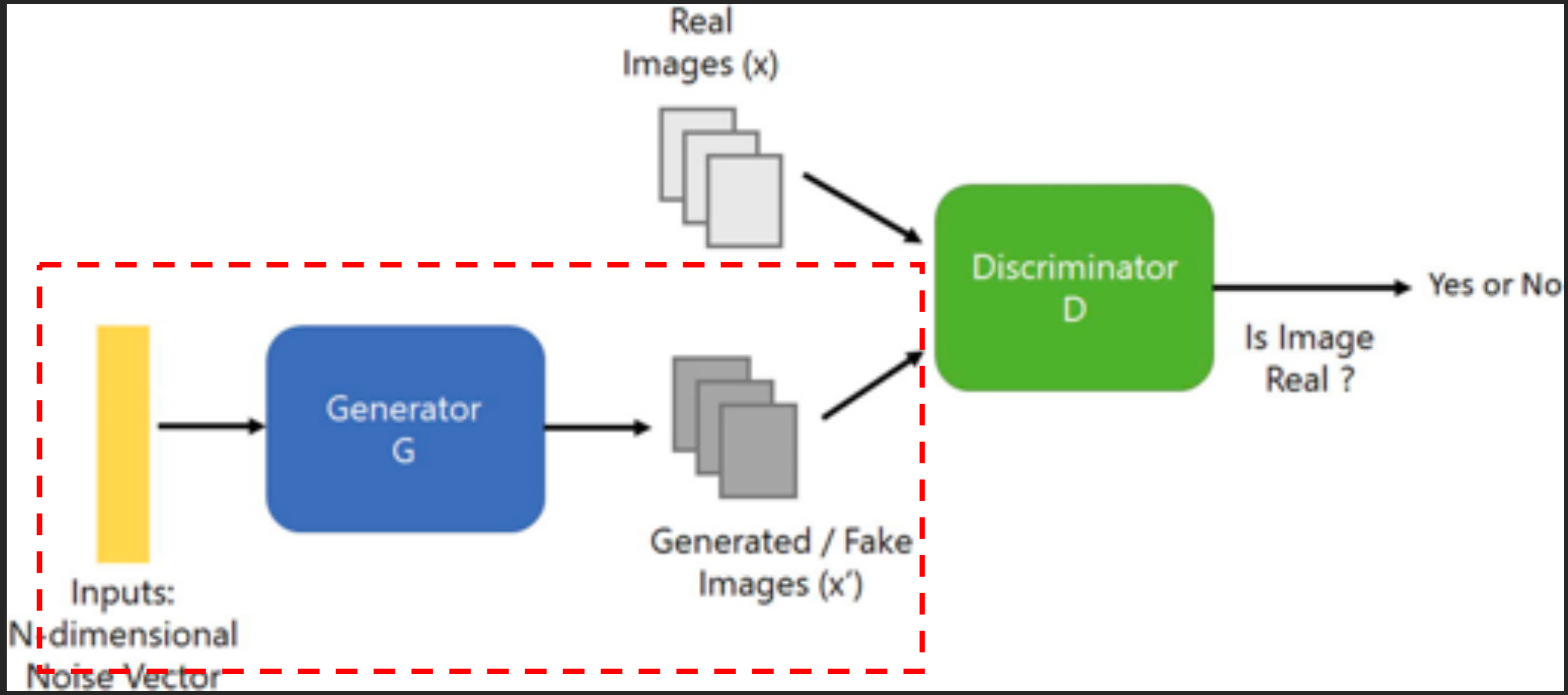
# GANs for Synthesizing Images

PG-GAN

StyleGAN

BigGAN

# Training GANs




Goodfellow et al. NIPS'14

# Tutorial Outline

- Interpretable units in intermediate layers
- Semantics in the latent space
- Inversion of real images
- Future directions

# Tutorial Outline

- Interpretable units in intermediate layers
- Semantics in the latent space
- Inversion of real images
- Future directions

# Deep Generative Representation

Random vector

512 dimensions

Generator

Randomly generated image

Slide from David Bau

Progressive GANs [Karras et al., ICLR 2018]

Slide from David Bau

Progressive GANs [Karras et al., ICLR 2018]

Slide from David Bau



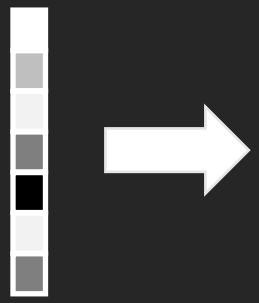




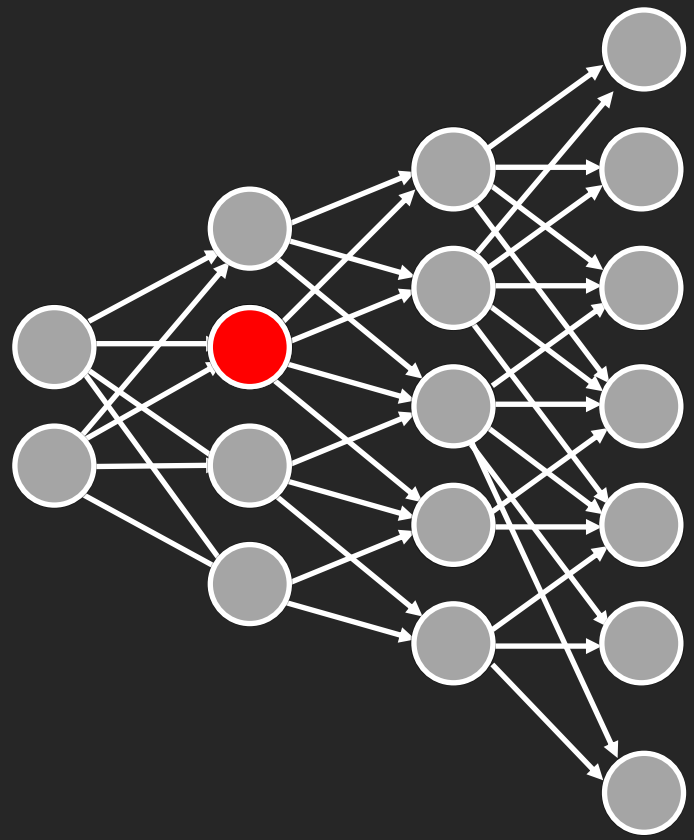

Slide from David Bau

# Are there watermark units?

Random vector



512 dimensions



Randomly generated image

Slide from David Bau








Slide from David Bau

# Deactivating banner units in layer 4








Slide from David Bau

# Deactivating watermark units in layer 4





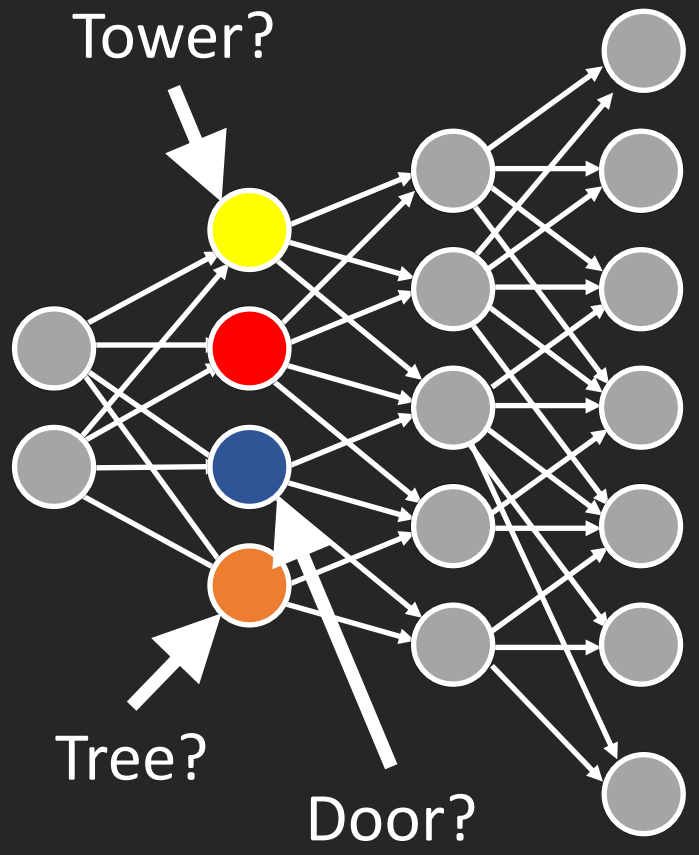



Slide from David Bau

# Are there other objects?

Random vector

512 dimensions




Randomly generated image

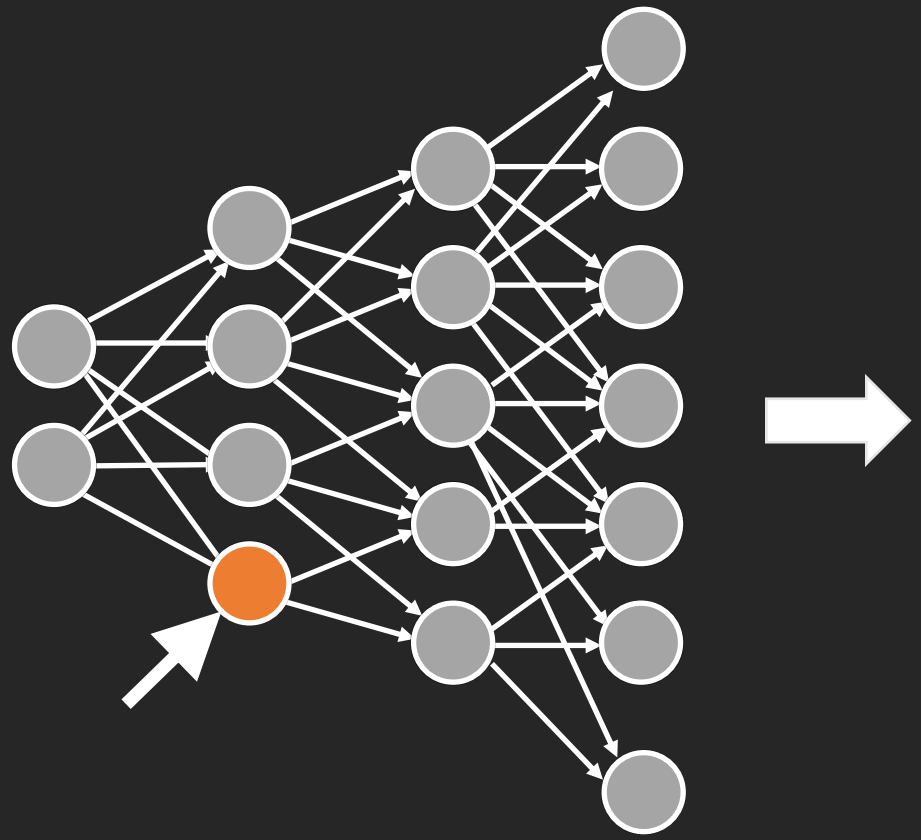
Slide from David Bau

GAN Dissection [Bau et al., ICLR 2019]

# Are there other objects?

Random vector

512 dimensions



Randomly generated image

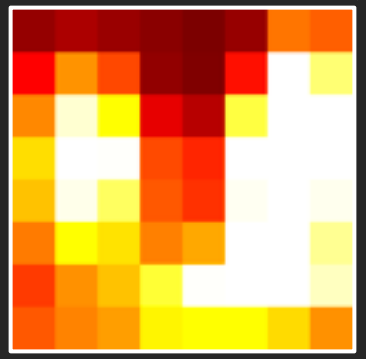
Slide from David Bau

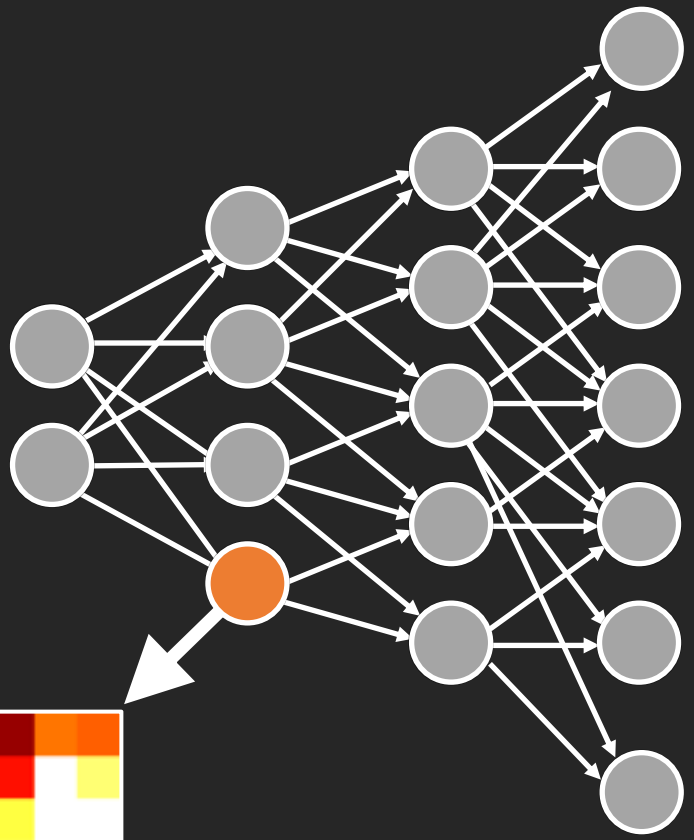
GAN Dissection [Bau et al., ICLR 2019]

# Are there other objects?

Random vector

512 dimensions



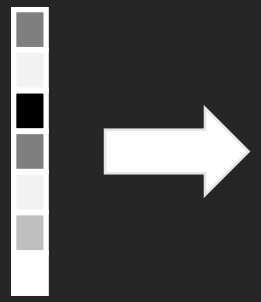


Randomly generated image

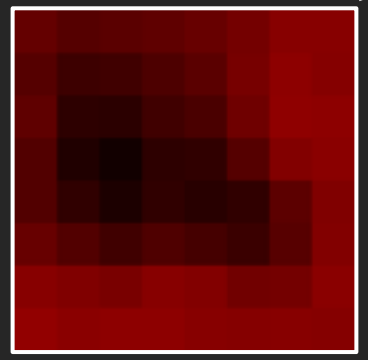
GAN Dissection [Bau et al., ICLR 2019]

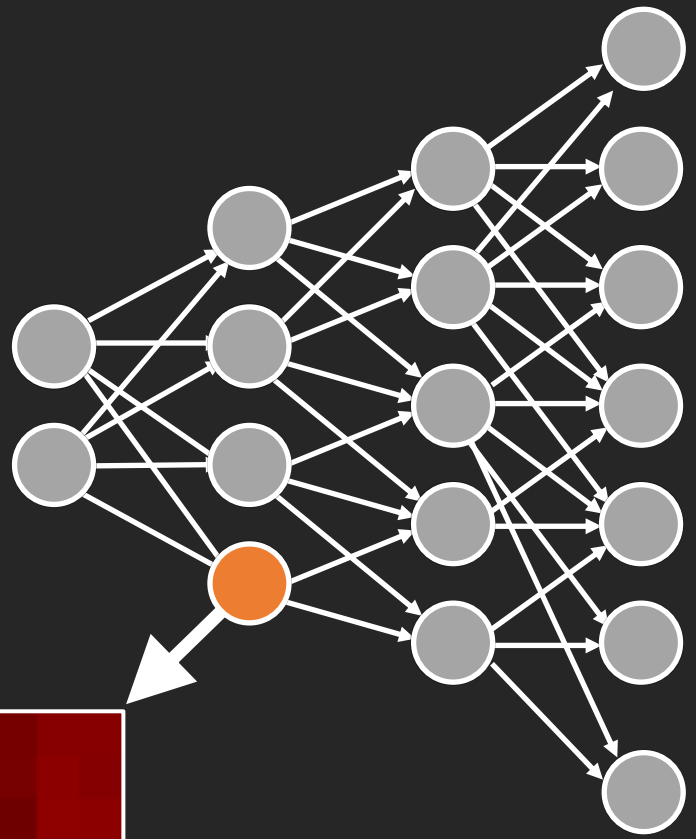
# Are there other objects?

Random vector



512 dimensions



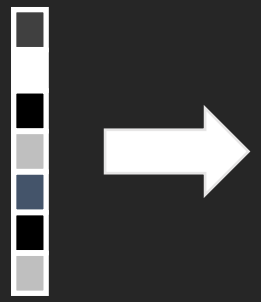


Randomly generated image

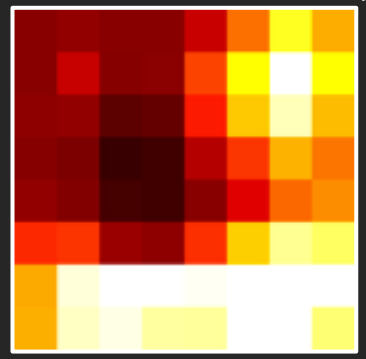
GAN Dissection [Bau et al., ICLR 2019]

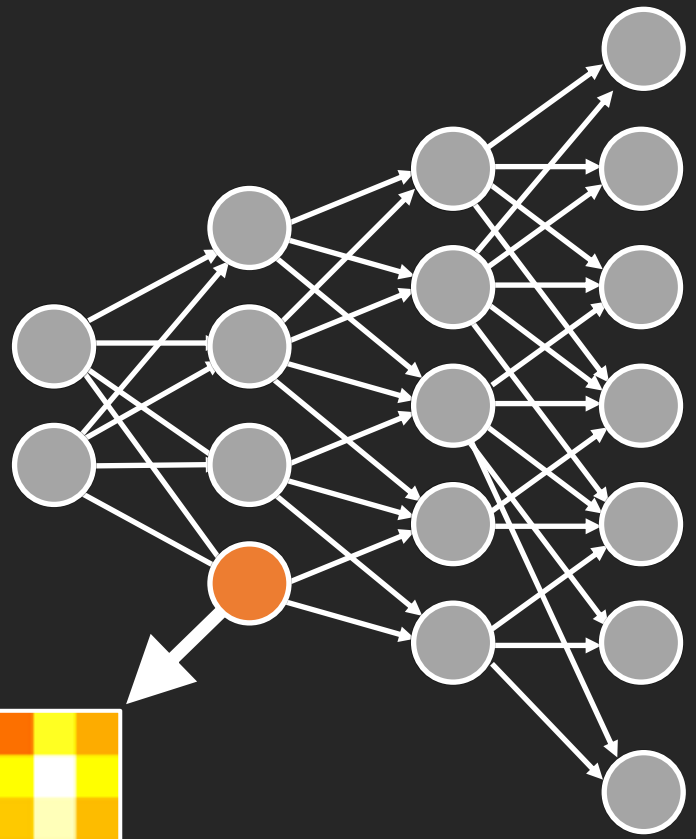
# Are there other objects?

Random vector



512 dimensions





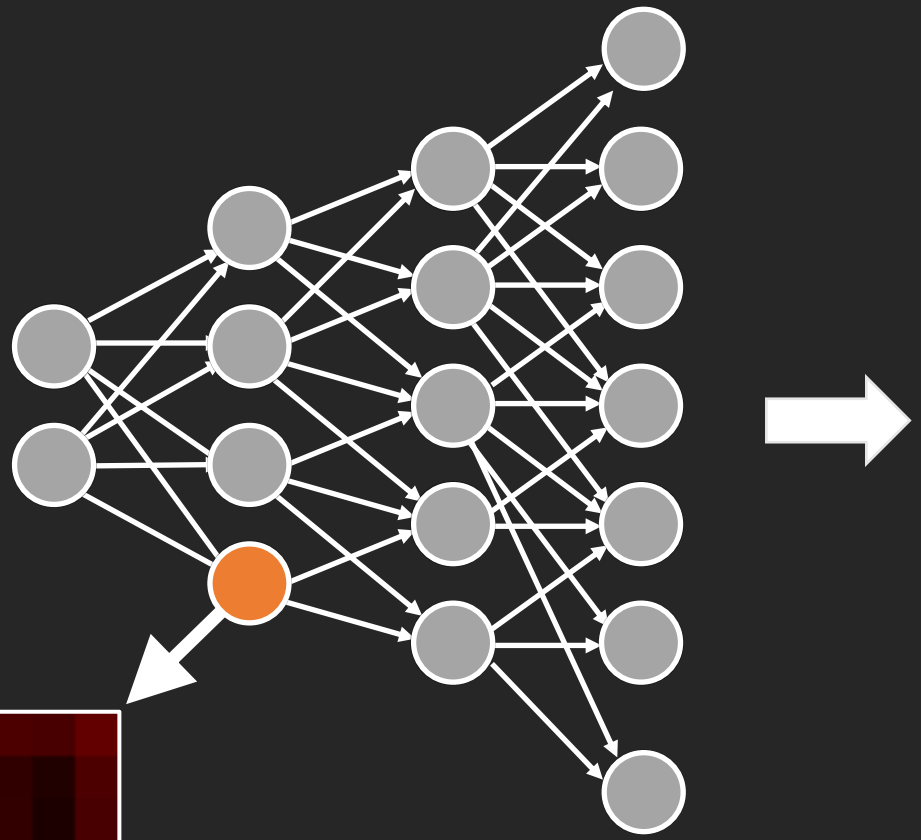
Randomly generated image

GAN Dissection [Bau et al., ICLR 2019]

# Are there other objects?

Random vector

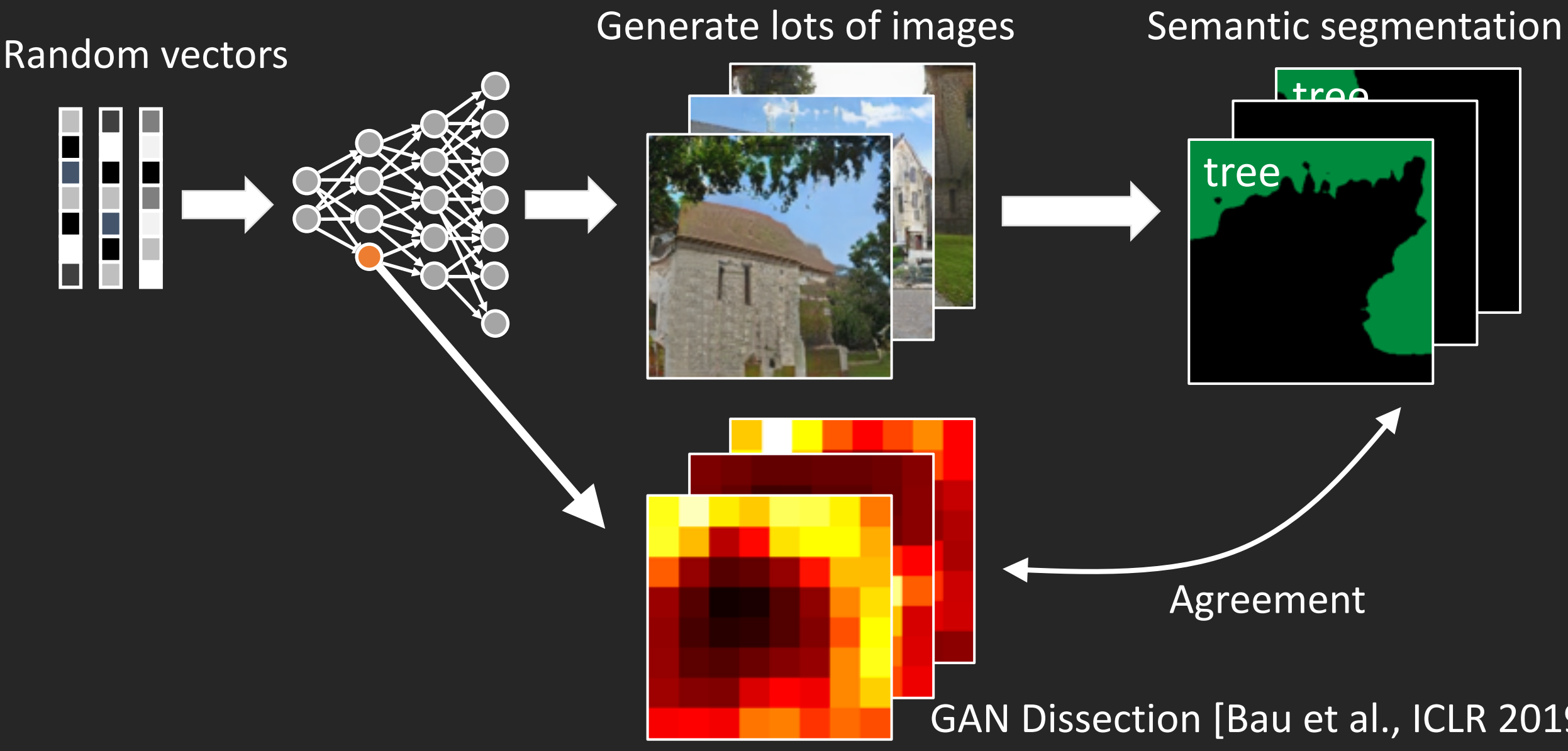
512 dimensions



Randomly generated image

GAN Dissection [Bau et al., ICLR 2019]

# Dissecting a GAN

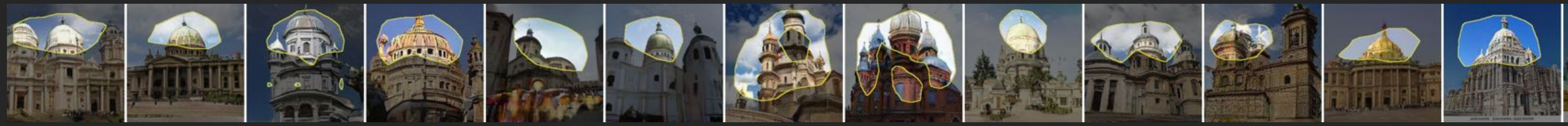



GAN Dissection [Bau et al., ICLR 2019]

# Units Emerge as Drawing Objects

Unit 365 draws trees

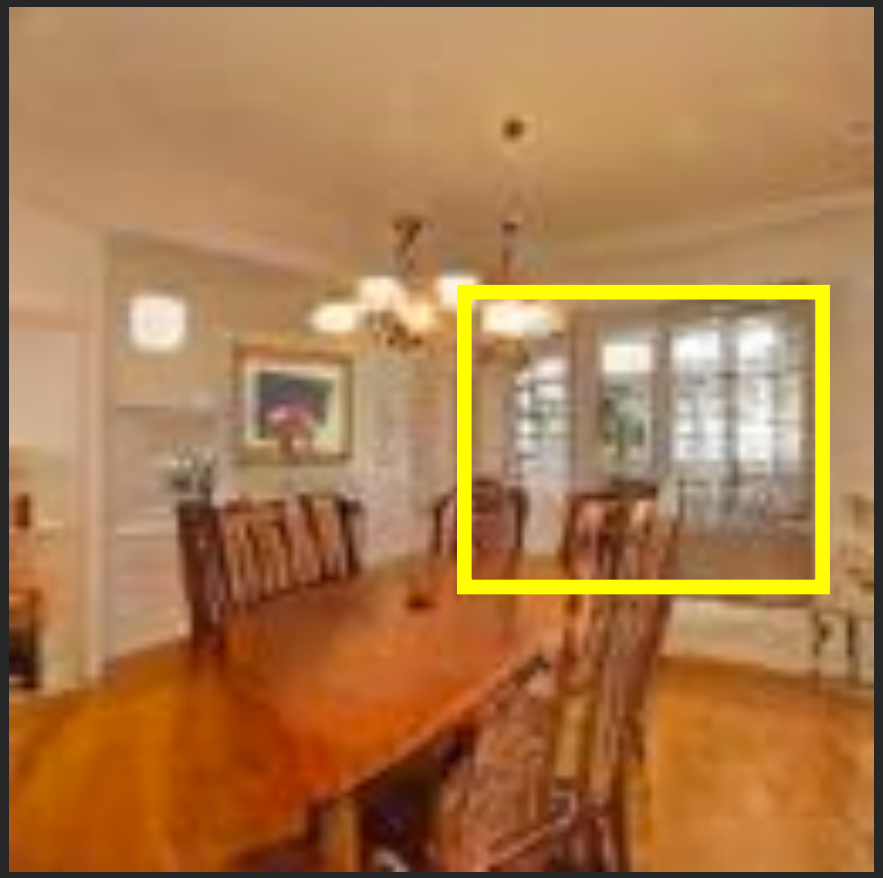
Unit 43 draws domes



Unit 14 draws grass

Unit 276 draws towers

# Turning off window units



# Turning on door units

# Turning on door units

# Turning on door units

# Turning on door units

# Turning on door units

# Turning on door units

# Interactive Object Editing


Online demo:
**ganpaint.io**

Semantic Photo Manipulation with a Generative Image Prior. SIGGRAPH 2019

# Editing Objects vs. Global Attributes

# Manipulating Global Attributes

Indoor lighting

Wooden style

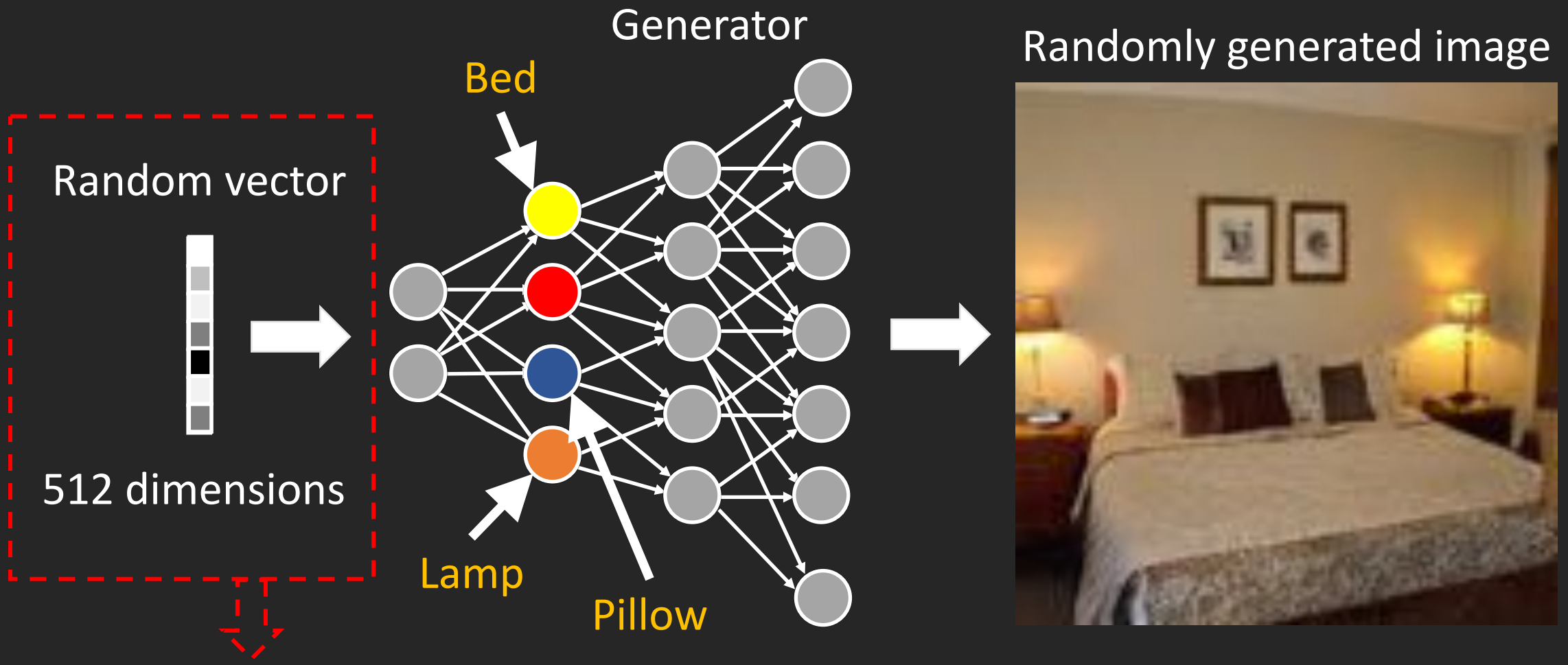
Different layout

Living room

# Tutorial Outline

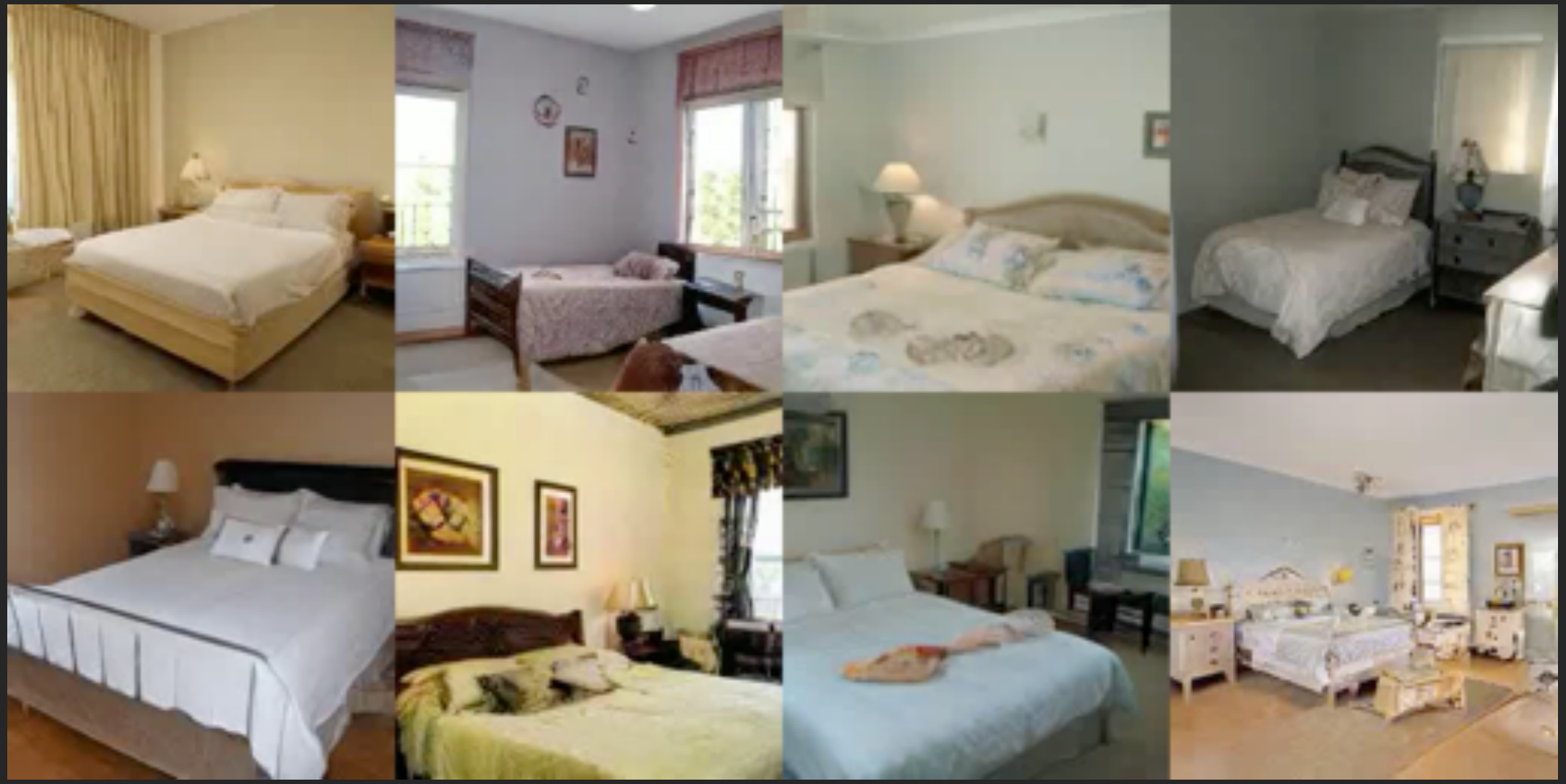
- Interpretable units in intermediate layers
- Semantics in the latent space
- Inversion of real images
- Future directions

# Deep Generative Representation




How latent code affects the output?

# Random Walk in Latent Space of Bedroom



# Multiple Levels of Abstractions for Scenes

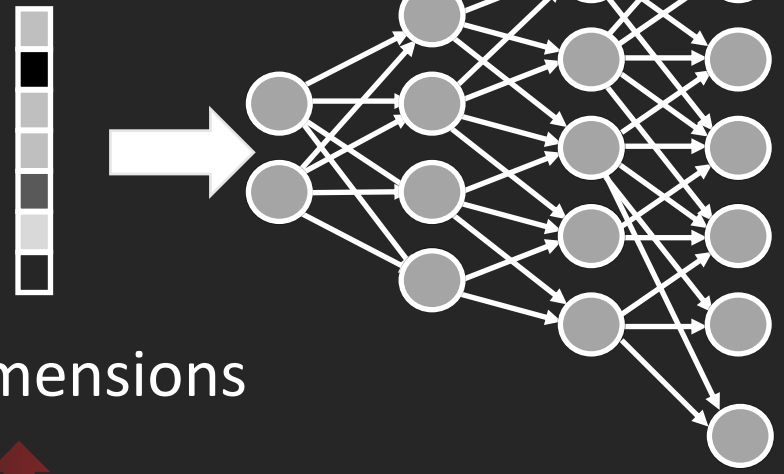
**Scene category:**
bedroom
**Scene attributes:**
nature lighting
wood
foliage
...

Layout

Segmentation

# Causal Relations between Latent Code and Semantics

Random vector



512 dimensions

Predicted semantics

**Scene category:** bedroom

**Scene attributes:** nature lighting, wood, foliage

To identify the cause-effect relations

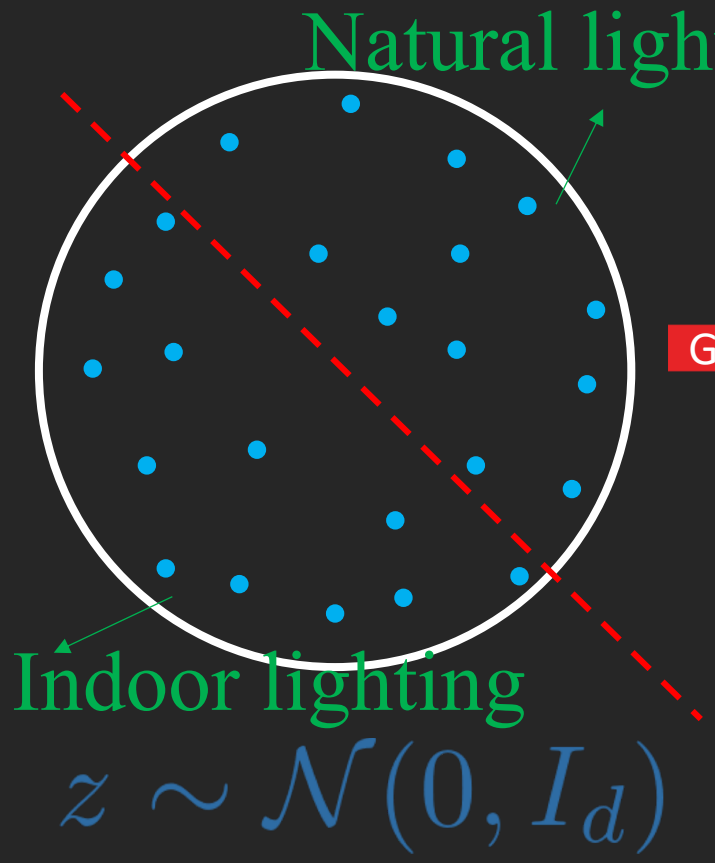
Ceyuan Yang, Yujun Shen, Bolei Zhou. Work in submission

# Aligning Latent Space with Attribute Space

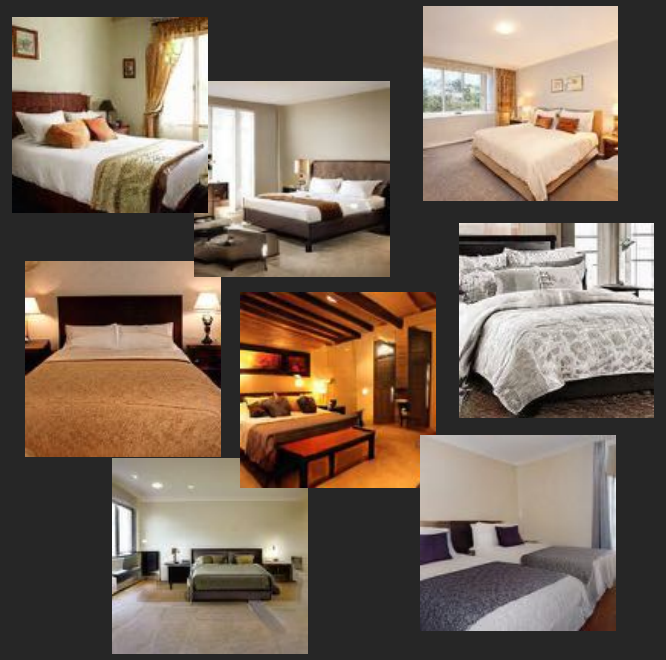
Latent Space

Image Space

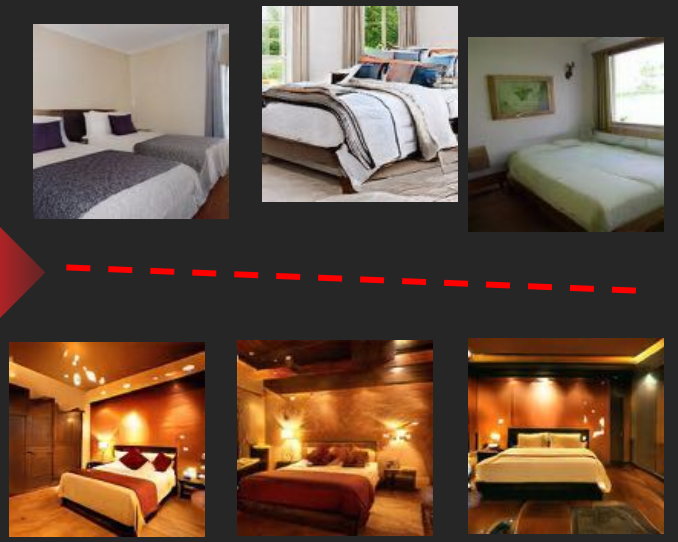
Attribute Space




Generator

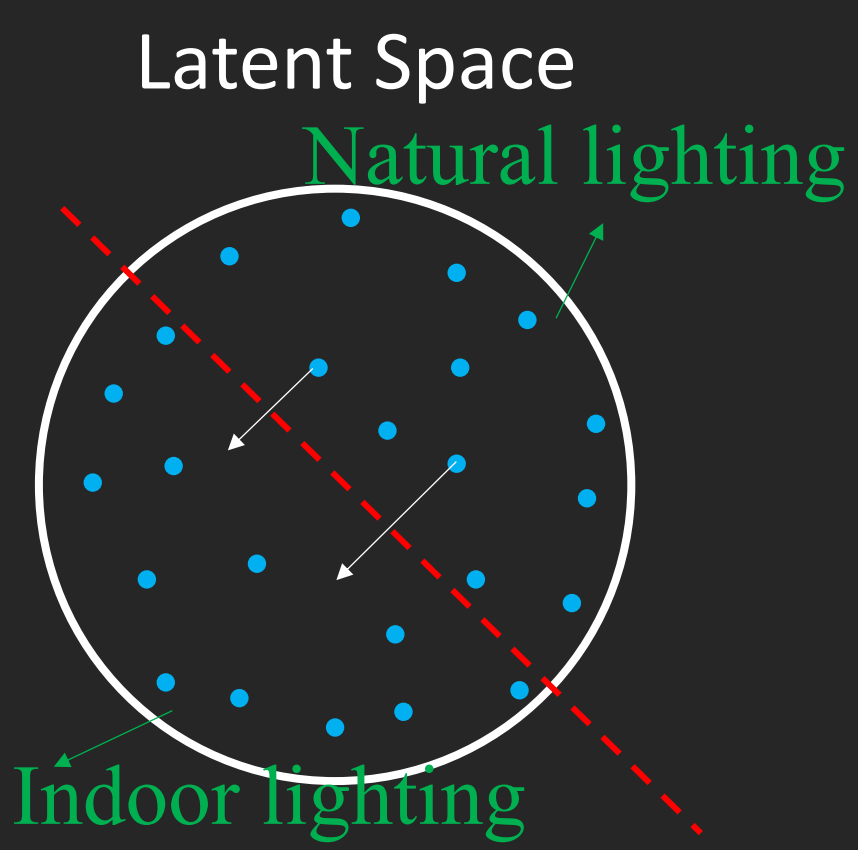


Predictor



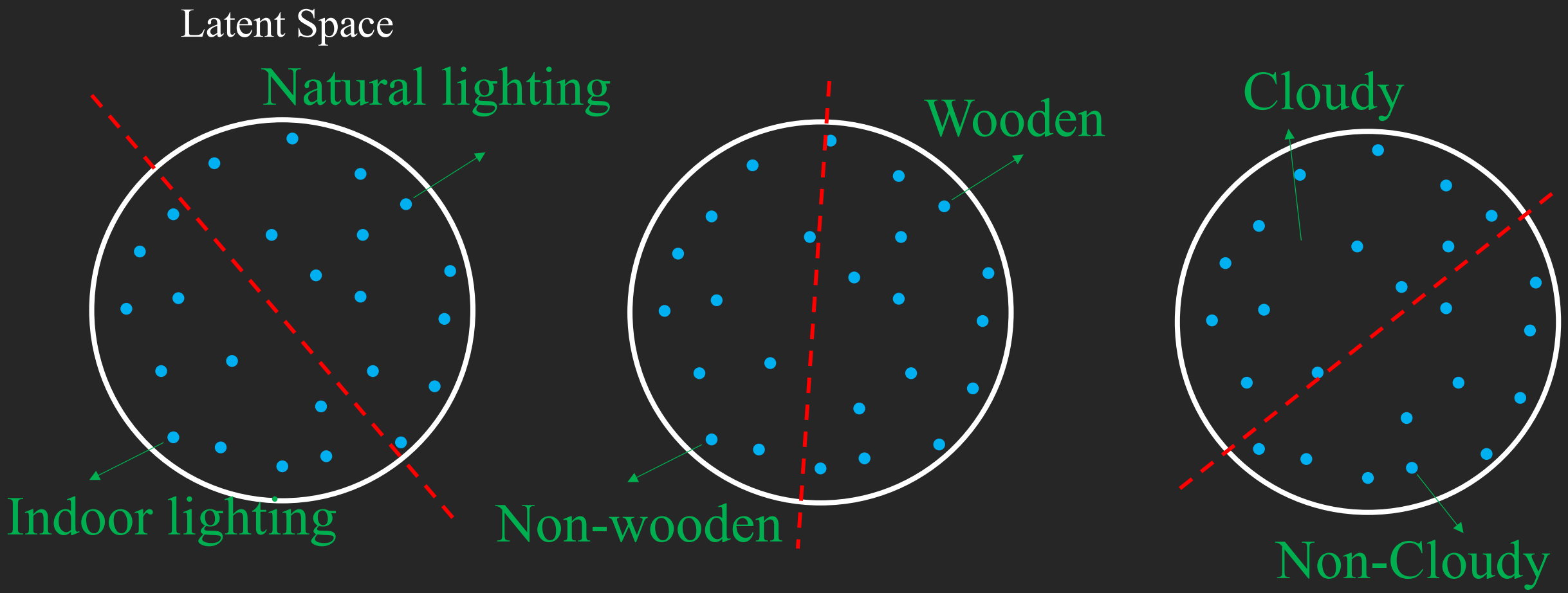
Train attribute classifier in latent space

# Pushing Latent Code through Boundary




$G(\mathbf{z})$ → More indoor lighting → $G(\mathbf{z} + \lambda\mathbf{b})$

# Various Attribute Boundaries in Latent Space



# Turning up the light

# How to edit my own photo?

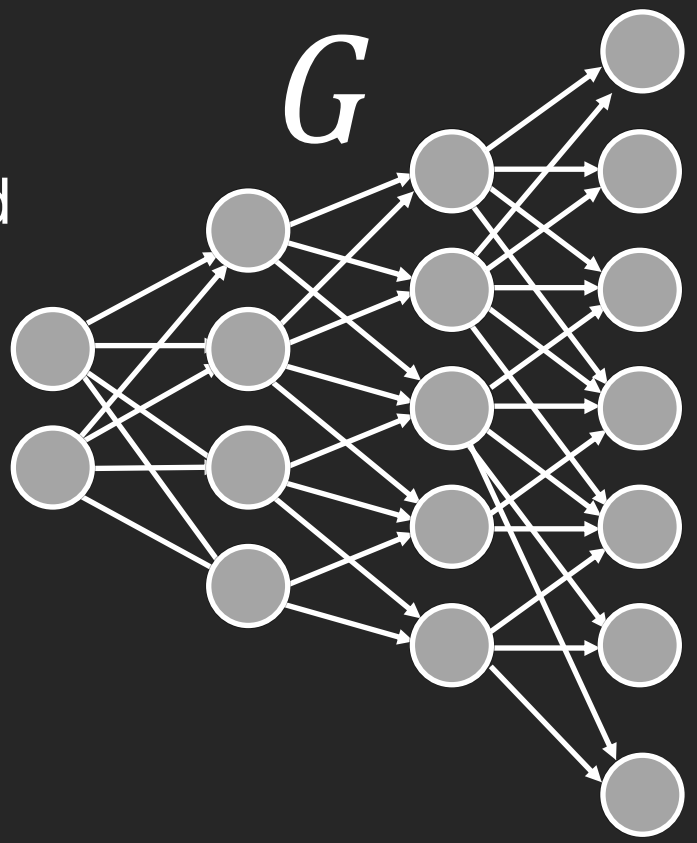
GAN-Synthesized Kitchen

My Kitchen Photo

# Tutorial Outline

- Interpretable units in intermediate layers
- Semantics in the latent space
- Inversion of real images
- Future directions

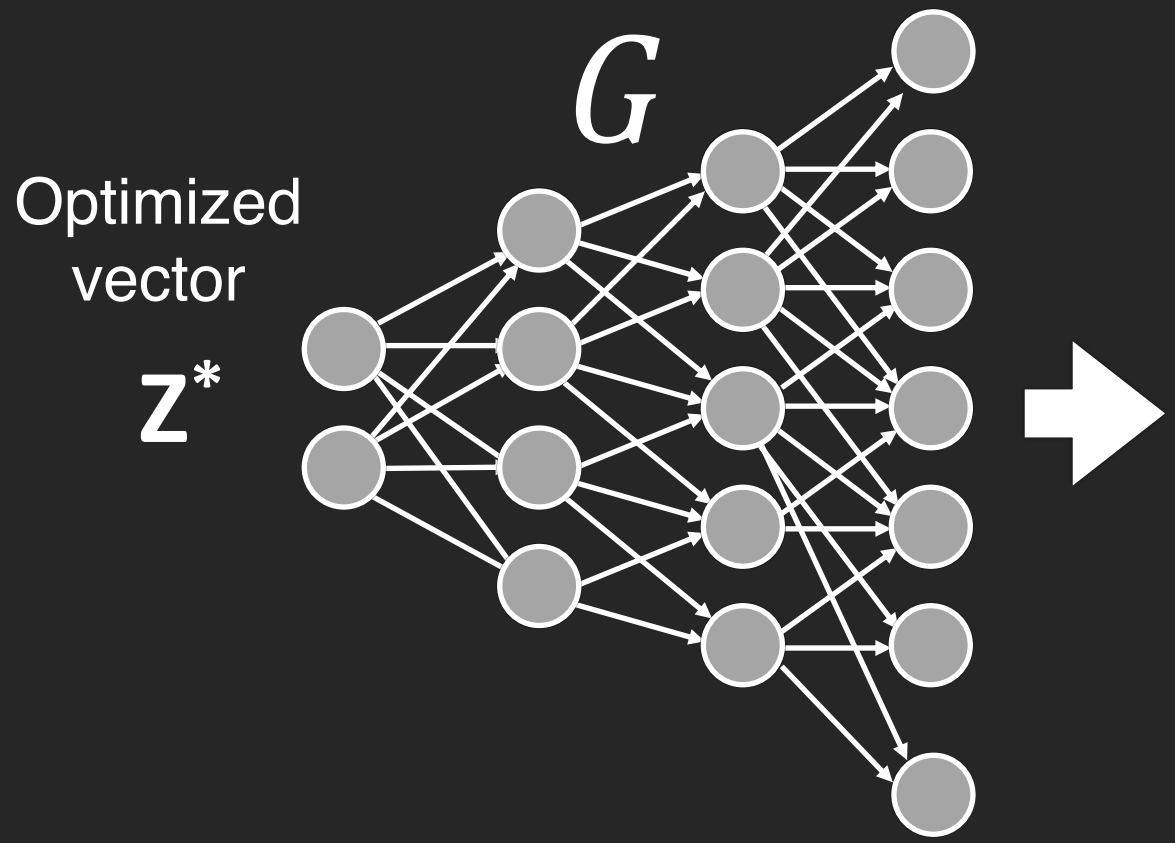
# Inverting image to GAN latent code




Real image **x**

$$\mathbf{z}^* = \arg\min_{\mathbf{z}} \frac{1}{2}||G(\mathbf{z}) - \mathbf{x}||^2$$

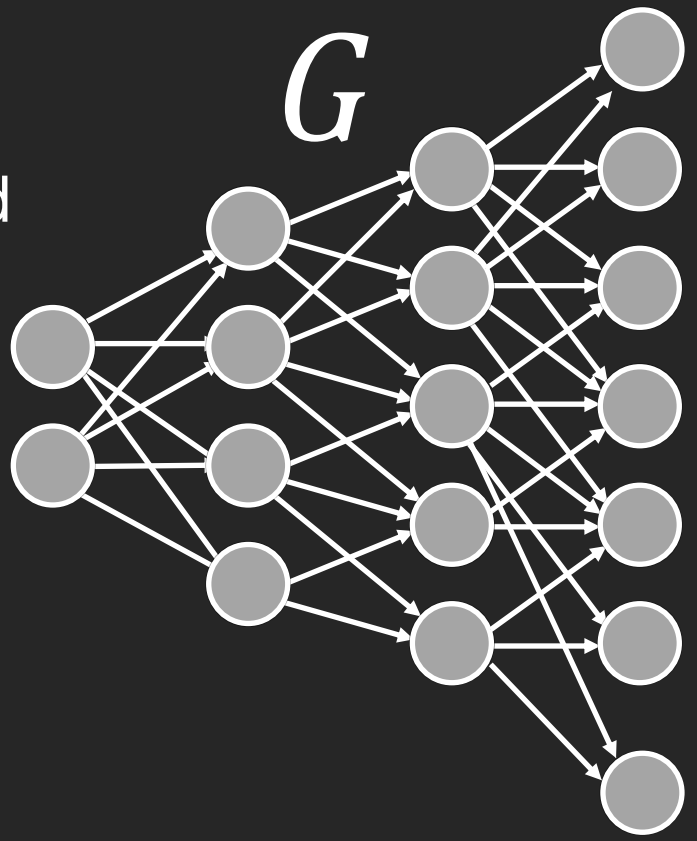
# Inverting image to GAN latent code



# Inverting image to GAN latent code

Optimized vector

$\mathbf{z}^*$




Reconstructed image $G(\mathbf{z}^*)$

# Find the differences…

Original image


GAN reconstructed image

# Fact:
# GANs don’t like a lot of things!

# GANs don’t like people







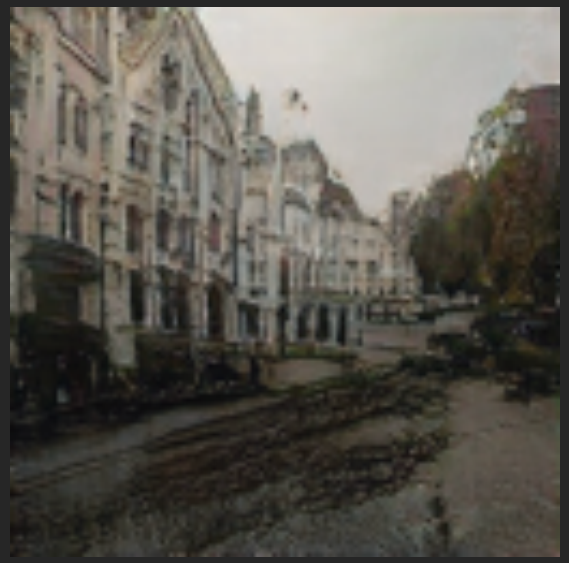
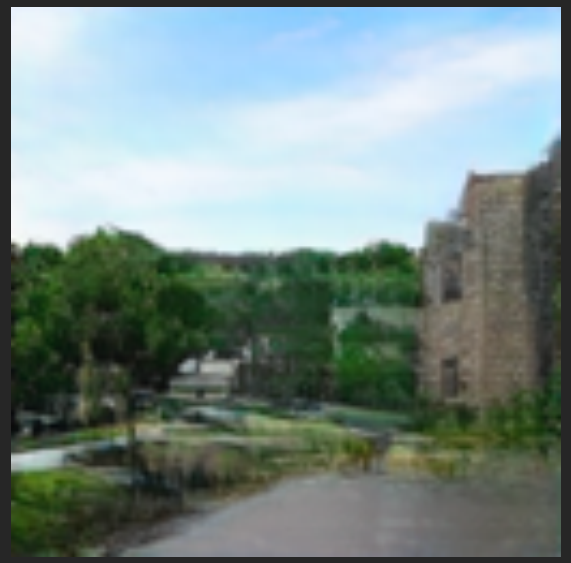

# GANs don’t like people

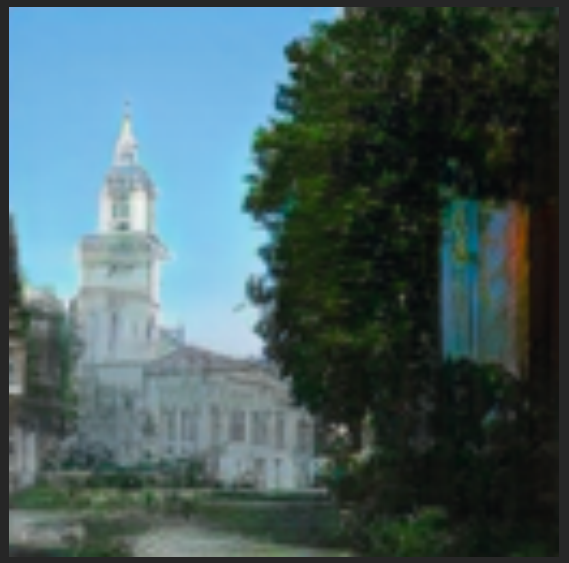
BigGAN’s synthesis of Bar

# GANs don’t like vehicles







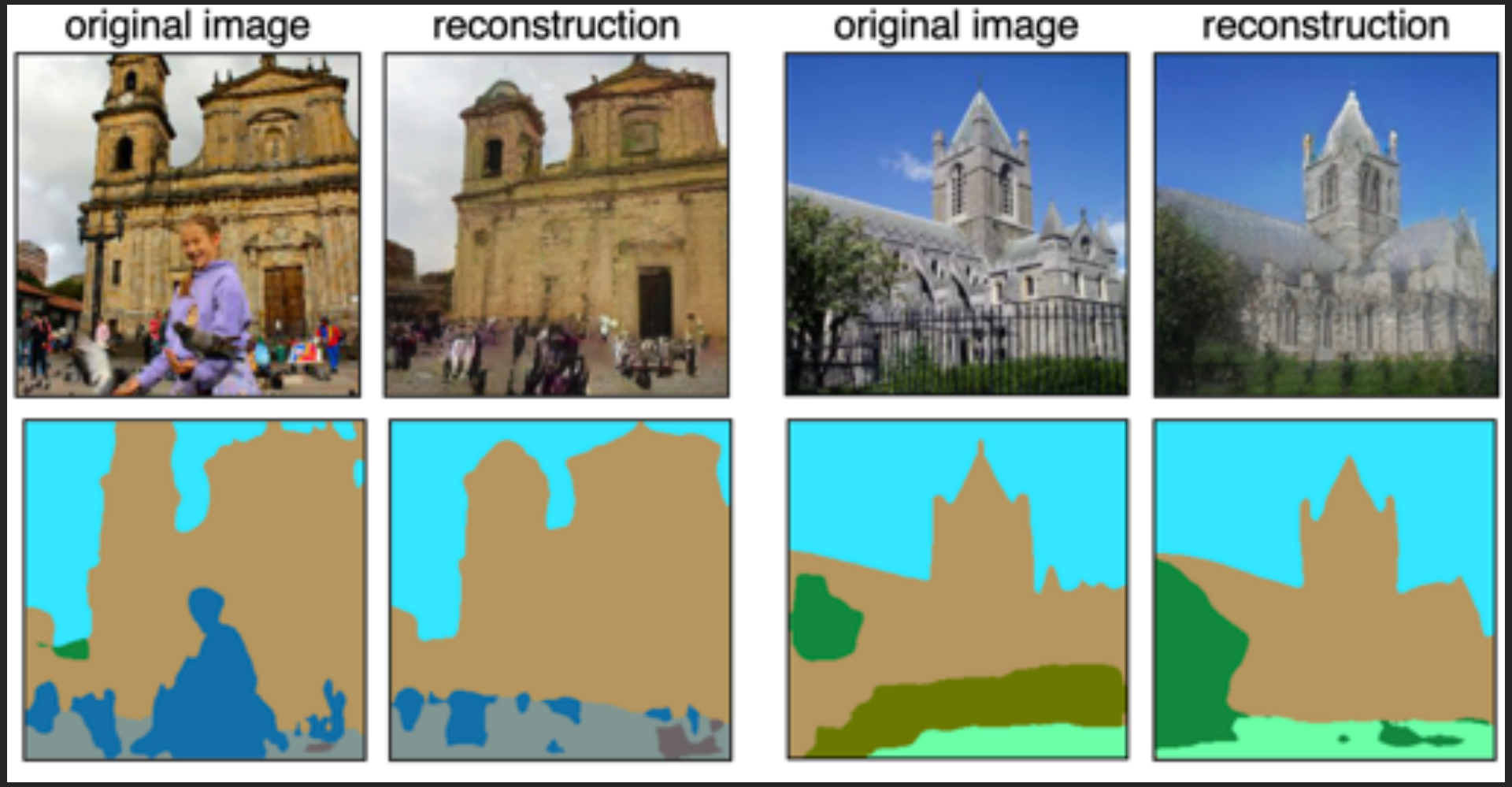
# GANs don’t like signs







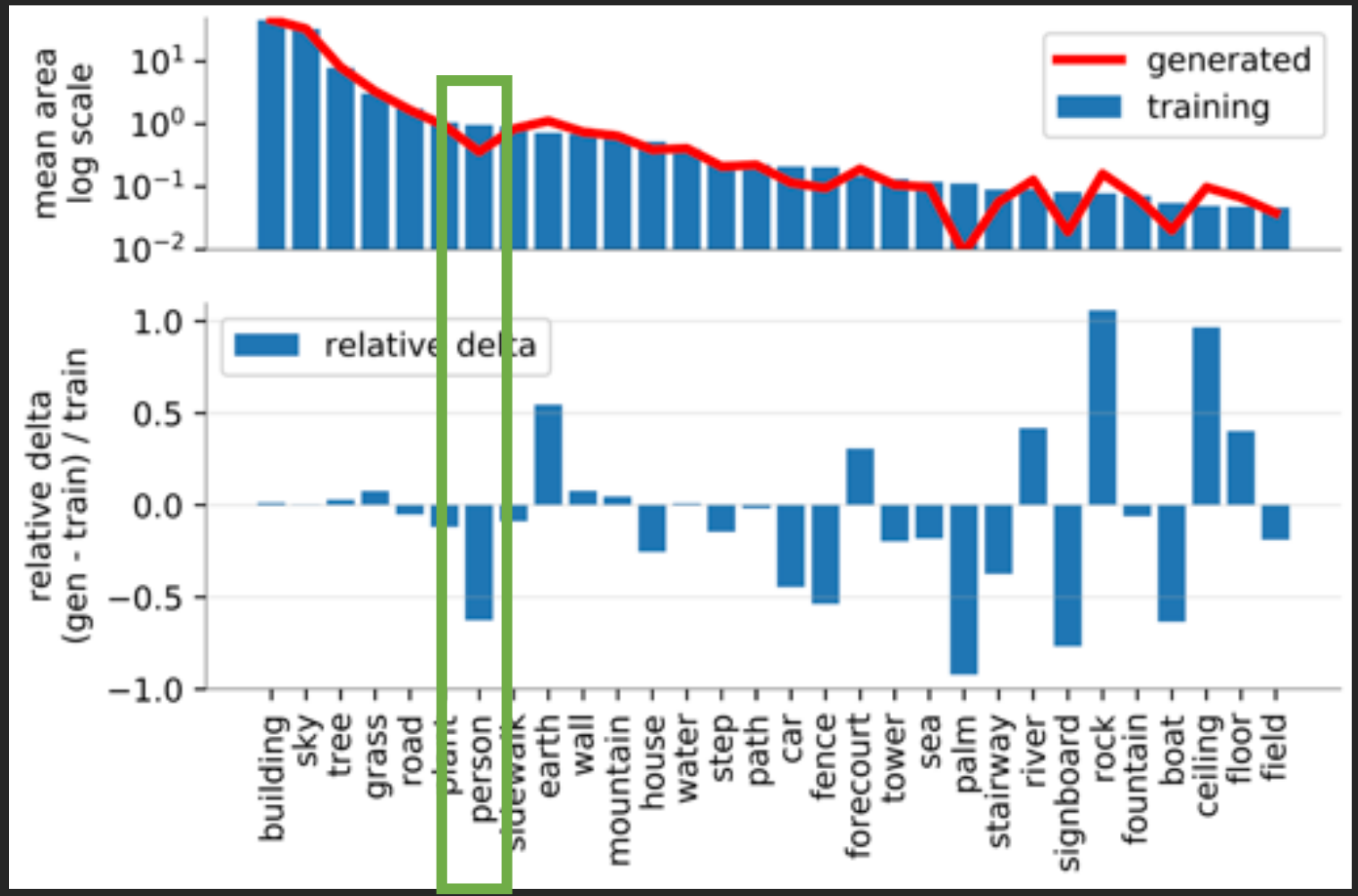
# Identify what's missing by comparing masks

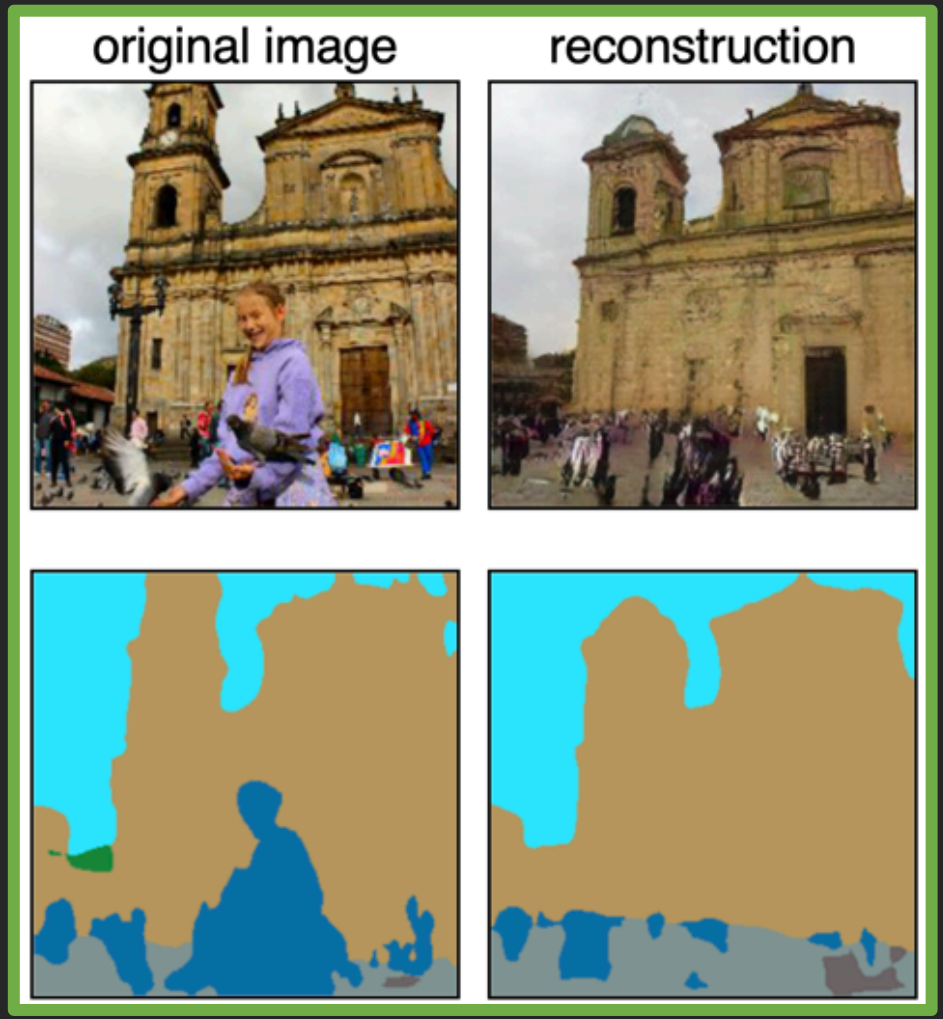



Thursday, October 31, 0900–1030 Oral 3.1A

Bau, Zhu, Wulff, Beebles. Strobelt, Zhou, Torralba. Seeing What GAN Cannot Generate. ICCV'19

# Statistics tells what GANs cannot generate

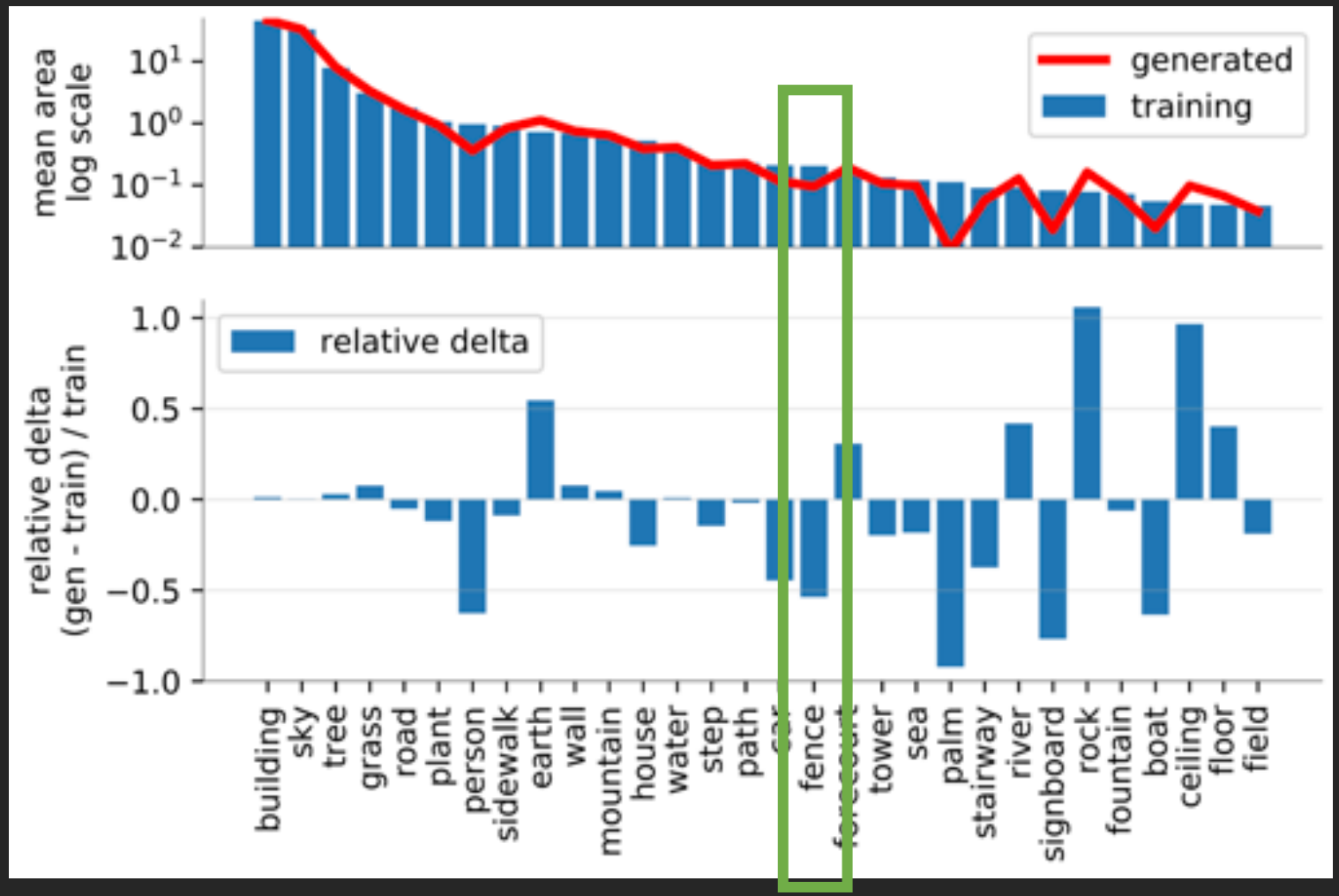


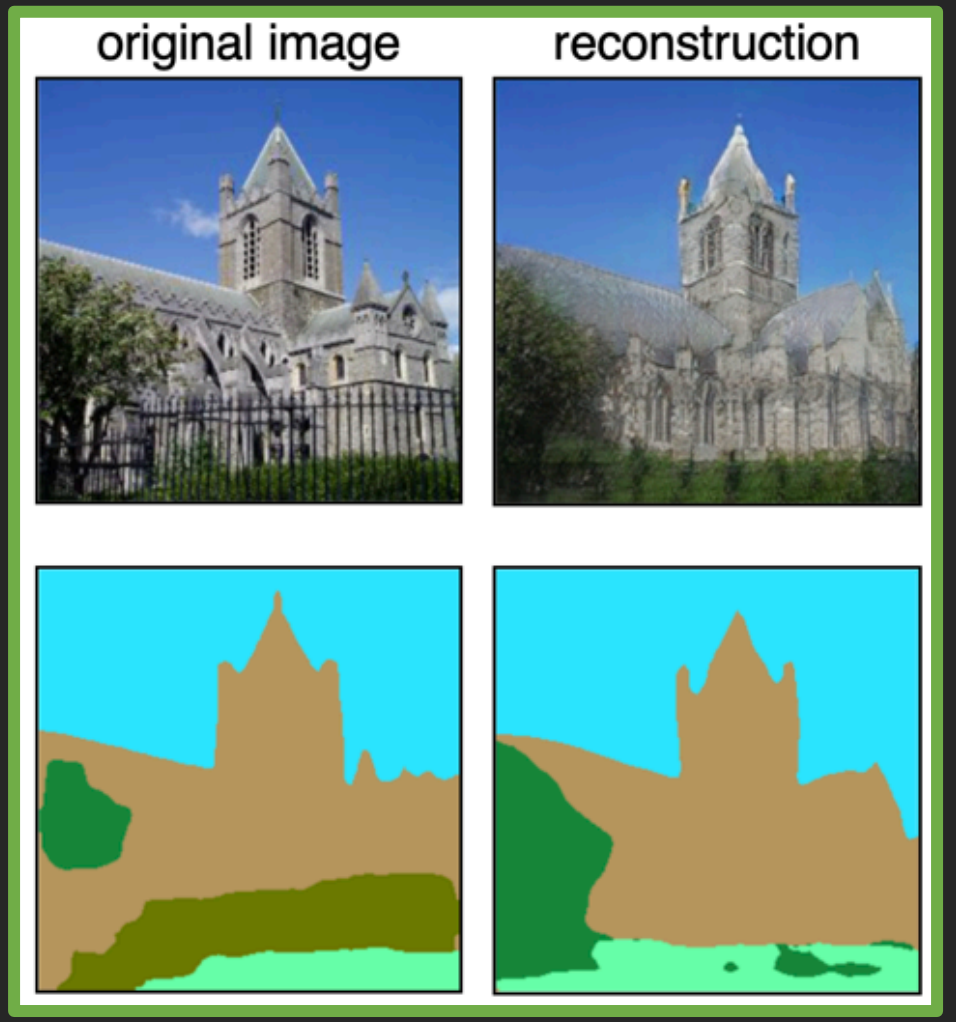




Thursday, October 31, 0900–1030 Oral 3.1A

Bau, Zhu, Wulff, Beebles. Strobelt, Zhou, Torralba. Seeing What GAN Cannot Generate. ICCV'19

# Statistics tells what GANs cannot generate



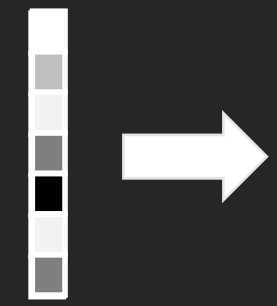




Thursday, October 31, 0900–1030 Oral 3.1A

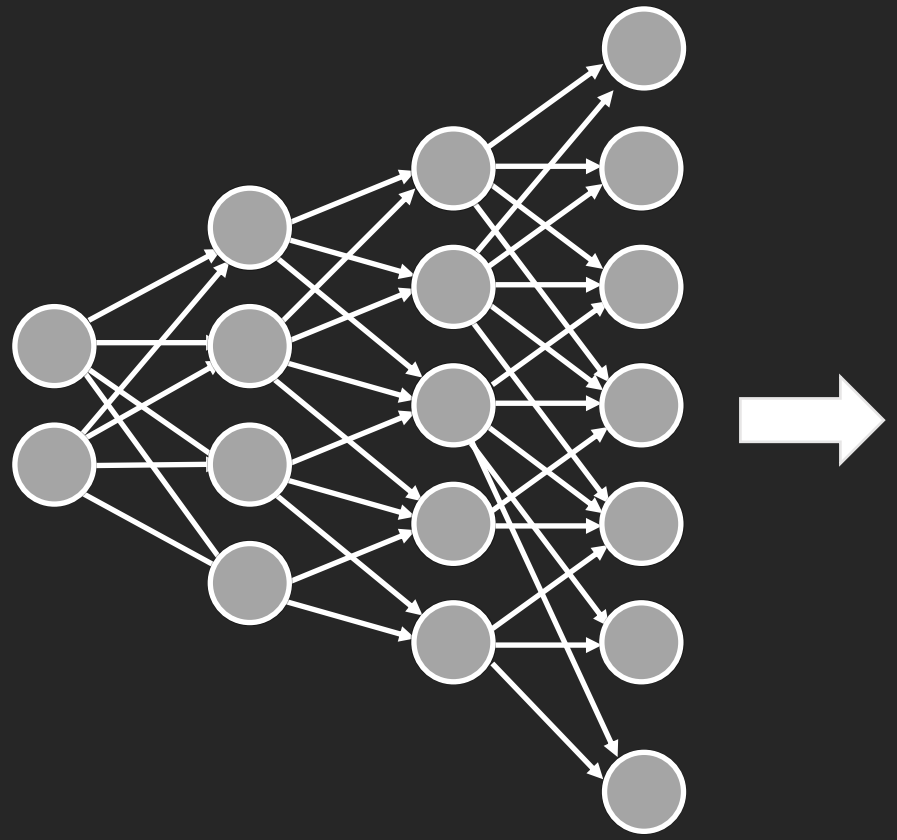
Bau, Zhu, Wulff, Beebles. Strobelt, Zhou, Torralba. Seeing What GAN Cannot Generate. ICCV'19

# GAN Inversion for Faces

Random vector



512 dimensions

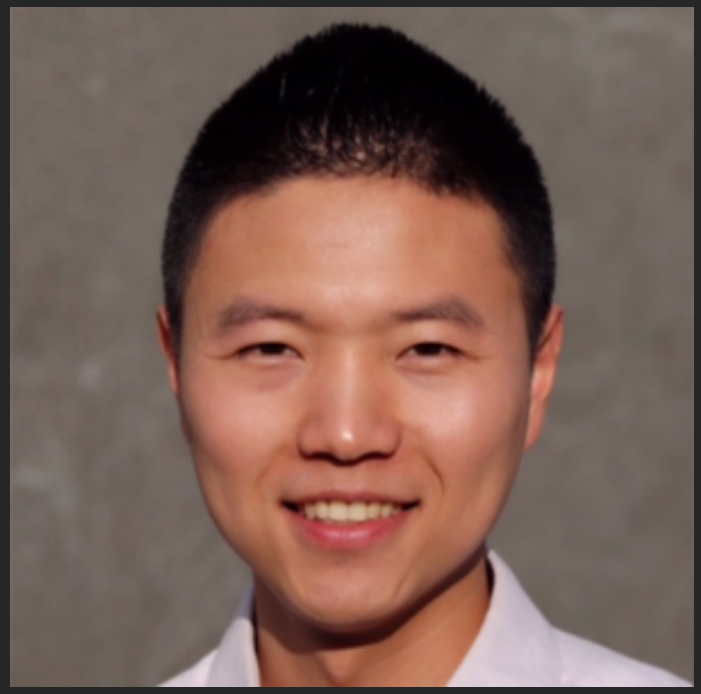


GAN trained on CelebHD

Randomly generated image

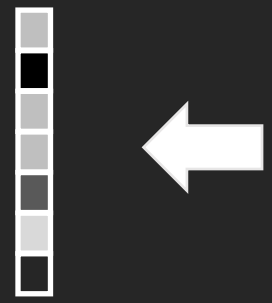
# GAN Inversion for Faces

Real face x

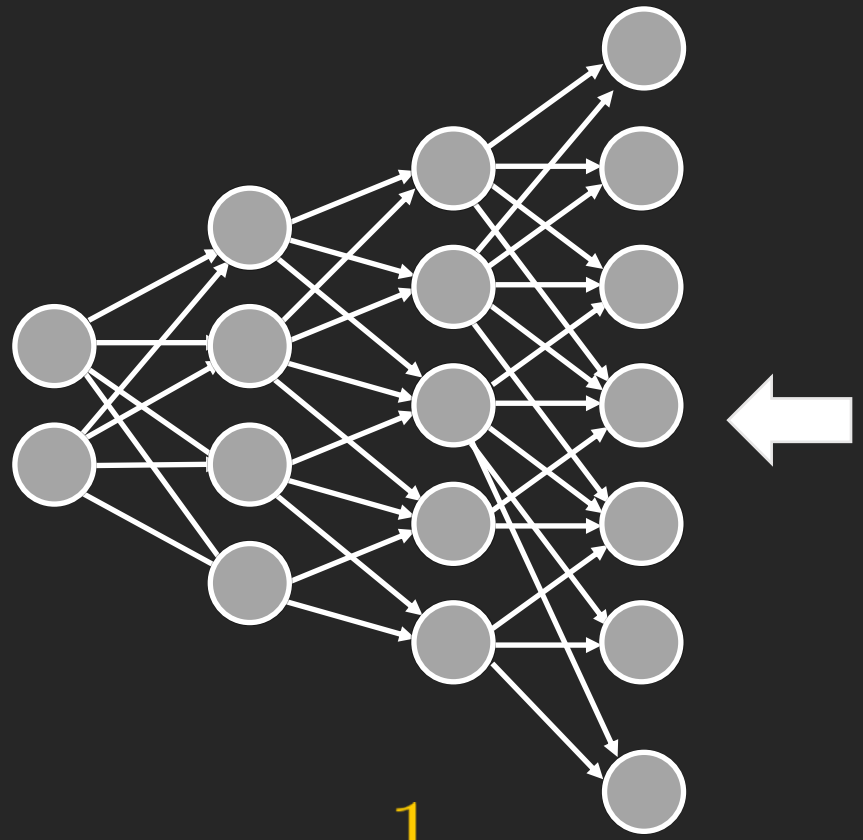


Optimized code

$\mathbf{z}^*$



512 dimensions

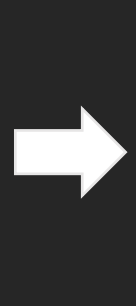


$$\mathbf{z}^* = \arg\min_{\mathbf{z}} \frac{1}{2}||G(\mathbf{z}) - \mathbf{x}||^2$$

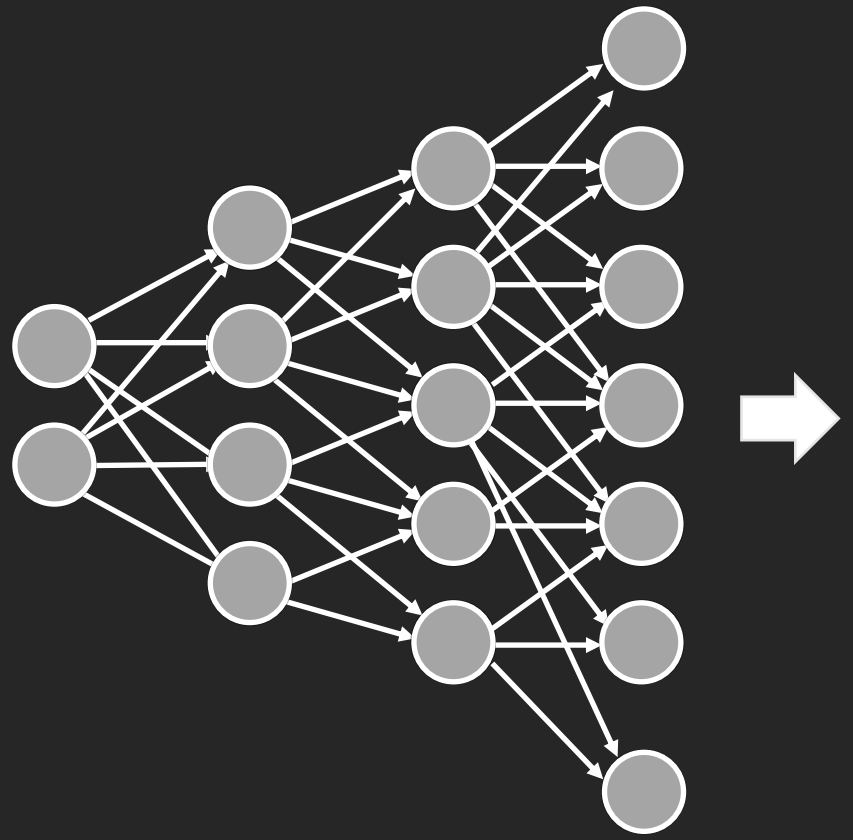
# GAN Inversion for Faces

Optimized code

$z^*$



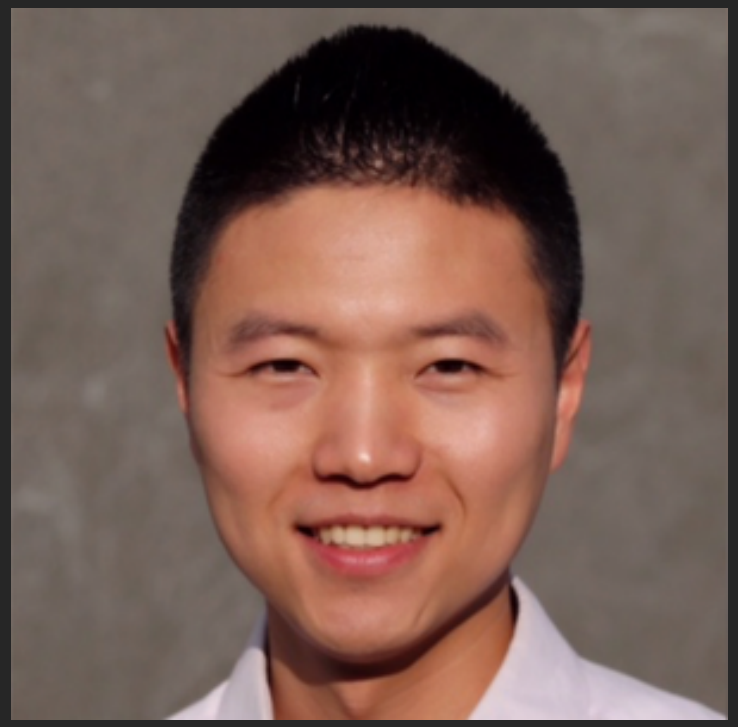
512 dimensions



Reconstructed face

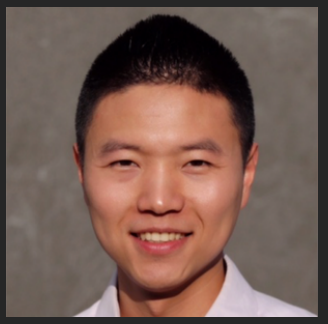
# Bias in GANs

Real face

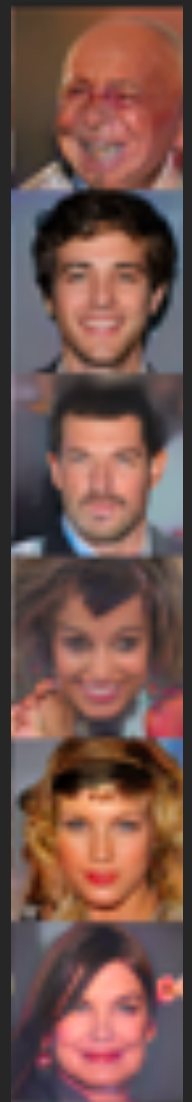


Reconstructed face

# Different random initializations



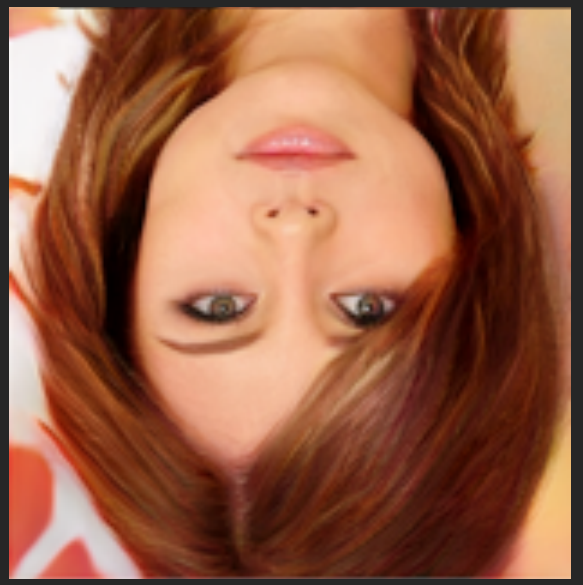
Initializations

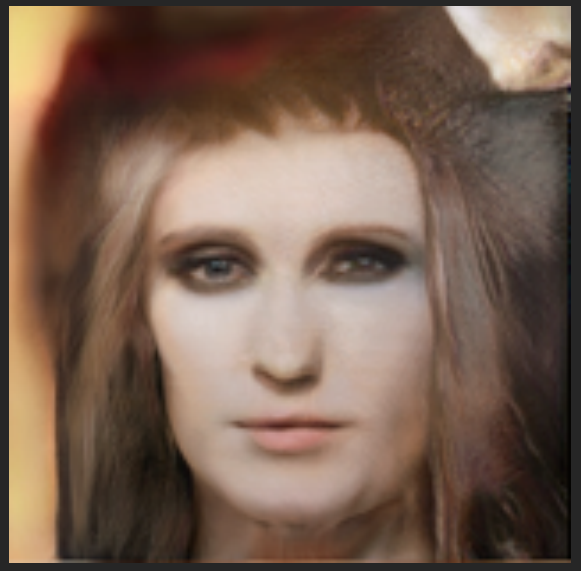


Iteration 1 -> 500

# GAN Inversion of Non-face Images





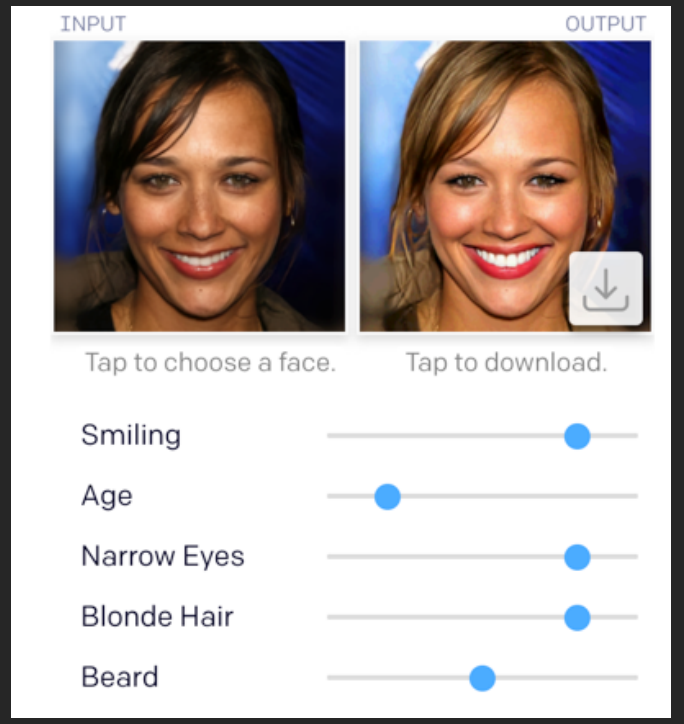



# Tutorial Outline

- Interpretable units in intermediate layers
- Semantics in the latent space
- Inversion of real images
- Future directions

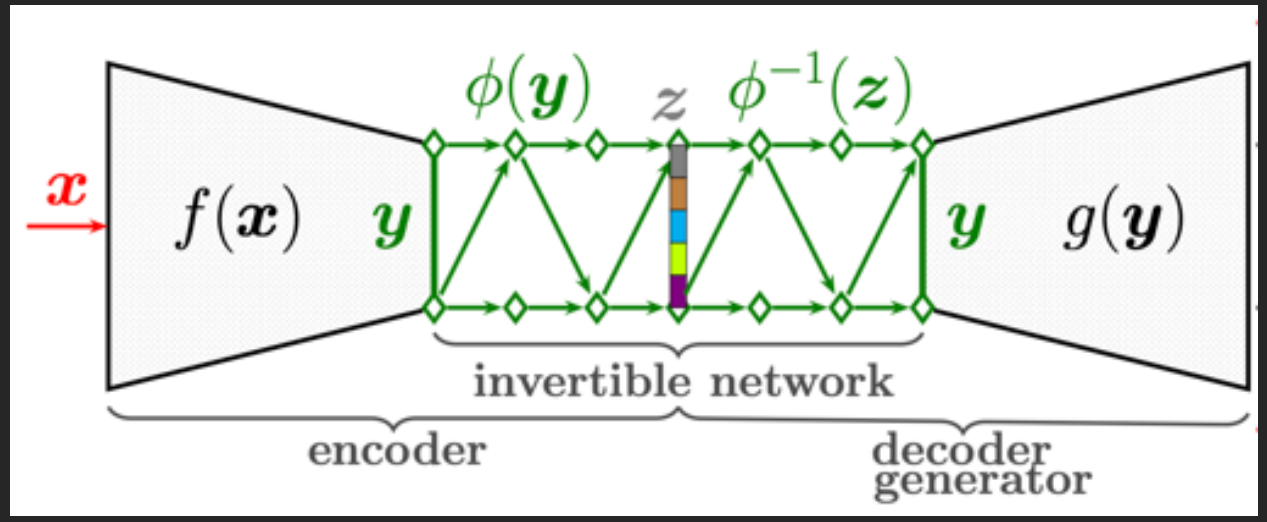
# Invertible Generative Models

- Explicitly design reversible generative models
- Nice (Dinh et al, ICLR'15), RealNVP (Dinh et al, ICLR'17), Glow (Kingma et al, NIPS'18)

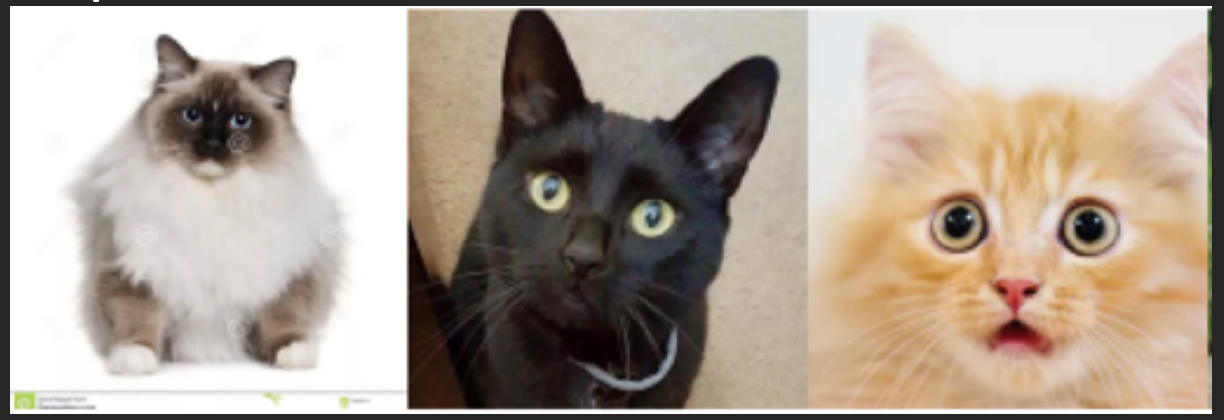



https://arxiv.org/pdf/1410.8516.pdf https://arxiv.org/pdf/1807.03039.pdf https://openai.com/blog/glow/

# Invertible Network + VAEs



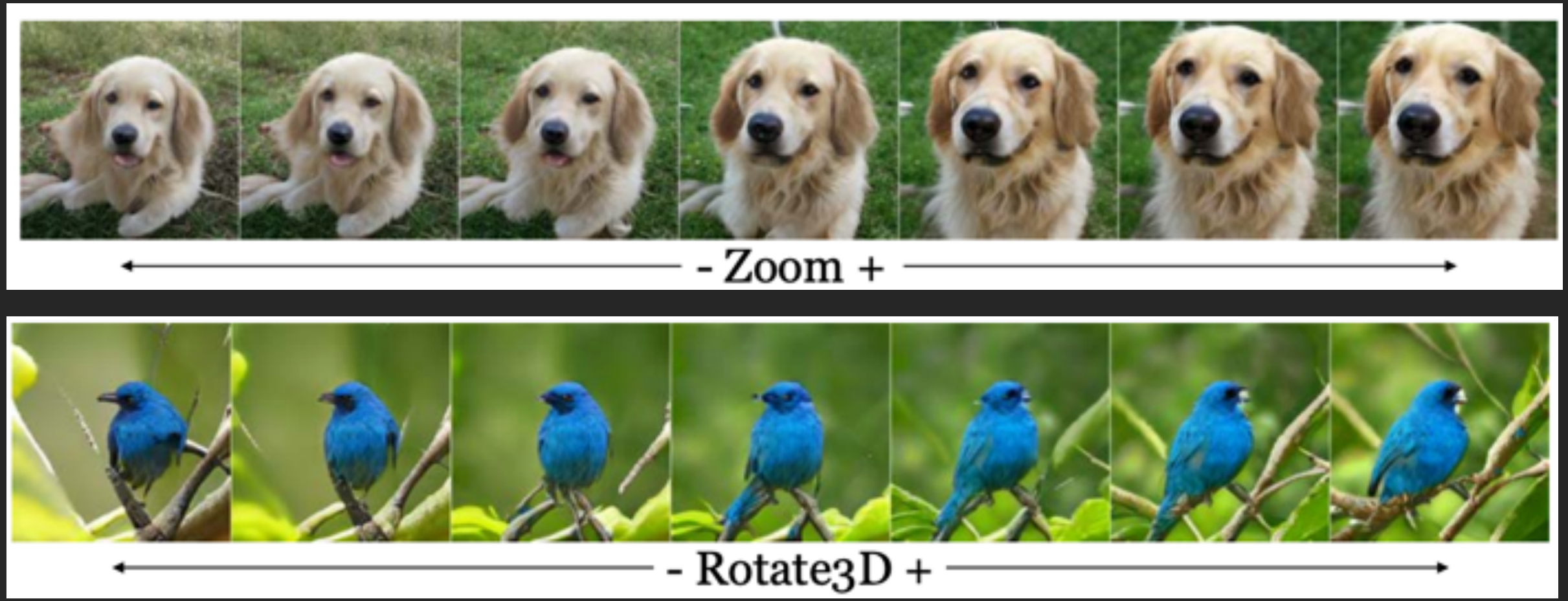

Input



Reconstruction

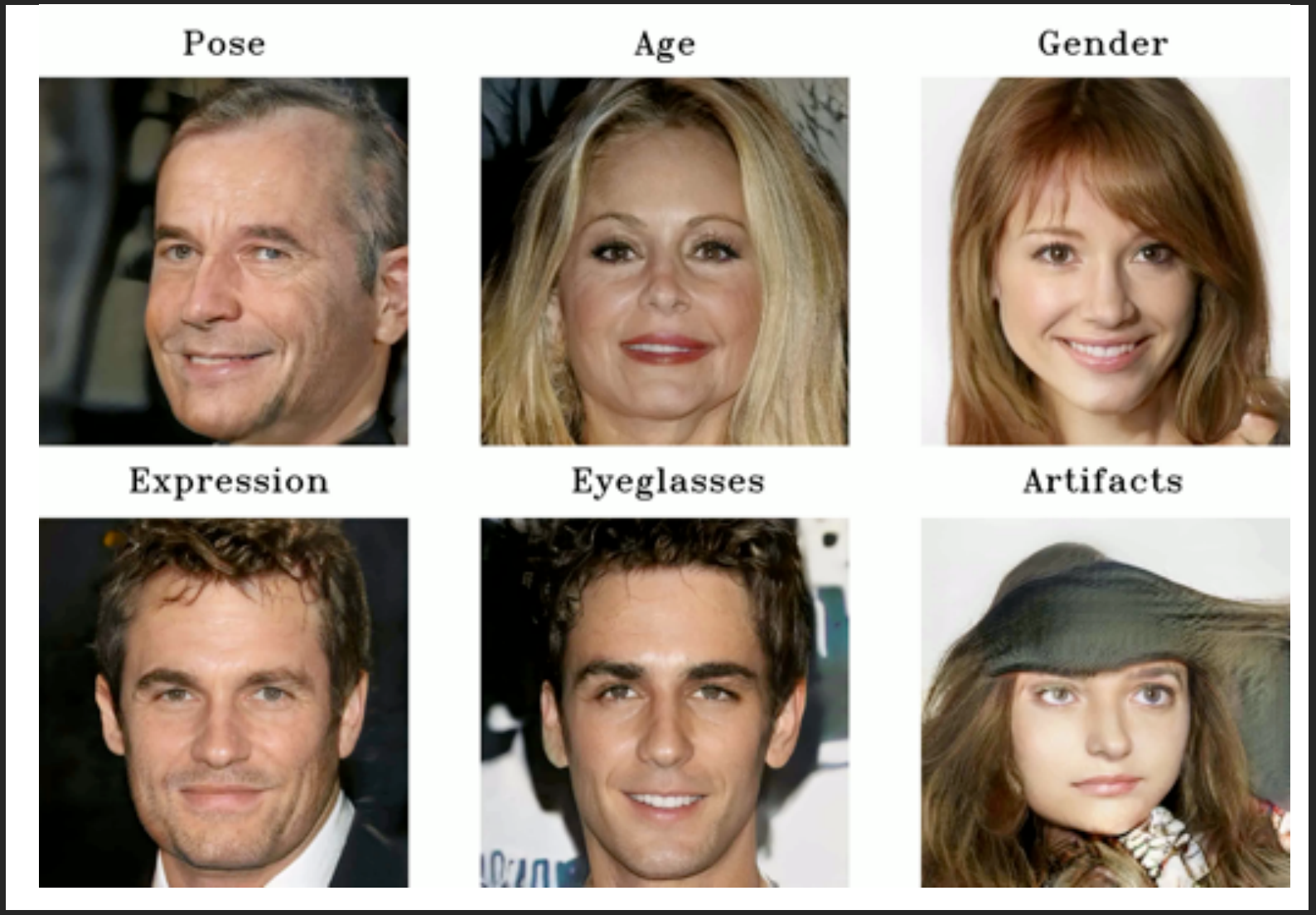
Zhu, Zhao, Zhang. LIA: Latently Invertible Autoencoder with Adversarial Learning. arXiv:1906.08090

# Steerability of GANs




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

# Facial Attributes Disentangled in Latent Space




Shen, Jinjin, Tang, Zhou. arXiv:1907.10786

https://shenyujun.github.io/InterFaceGAN/

# Acknowledgement

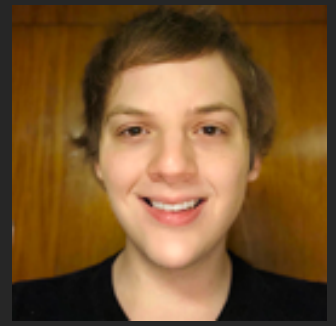

MIT team Key player

David Bau

Jun-Yan Zhu

Hendrik Strobelt

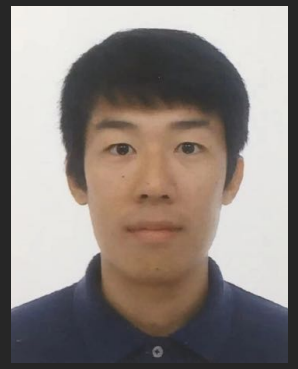
Bill Peebles



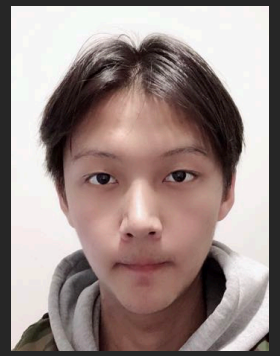
Antonio Torralba

CUHK team

Yujun Shen



Ceyuan Yang



Jinjin Gu

# Conclusion

- Latent semantics emerge in GANs
- Code and papers at http://bzhou.ie.cuhk.edu.hk