Selected GANs

Hao Dong

Peking University

PEKING UNIVERSITY 北京大學

Selected GANs

- Conditional GAN

Find the latent representation
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

Do GANs generate new data?
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

北京大學 PEKING UNIVERSITY

- Conditional GAN

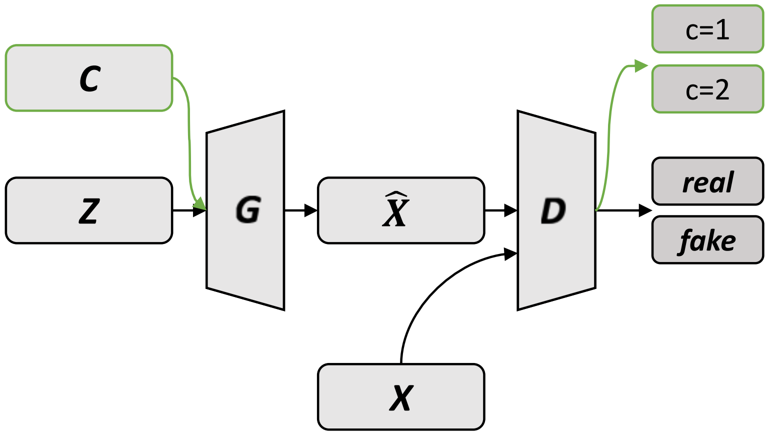
Find the latent representation
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

Do GANs generate new data?
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

Conditional GAN

- A Simple Example: Auxiliary Classifier GANs



$$\mathcal{L}_D = \mathbb{E}_{x\sim p_{data}}[logD_x(x)] + \mathbb{E}_{z\sim p_z}[log(1 - D_x(G(z,c)))]$$

$$\mathbb{E}_{x\sim p_{data}}[logD_c(x)] + \mathbb{E}_{z\sim p_z}[log(1 - D_c(G(z,c)))]$$

$$\mathcal{L}_G = \mathbb{E}_{x\sim p_{data}}[logD_x(G(z,c))] + \mathbb{E}_{z\sim p_z}[logD_c(G(z,c))]$$

monarch butterfly | goldfinch | daisy

Multi-modal problem: one problem has multiple solutions

$p(x|c,z)$

Conditional GAN

“Class” conditional generative models

$P(X =$  $|Y = Cat)$

“Text” conditional generative models

$P(X =$  $|Y = \text{“}a\ flower\ with\ white\ petals\ and\ yellow\ stamen\text{”})$

“Text-image” conditional generative models

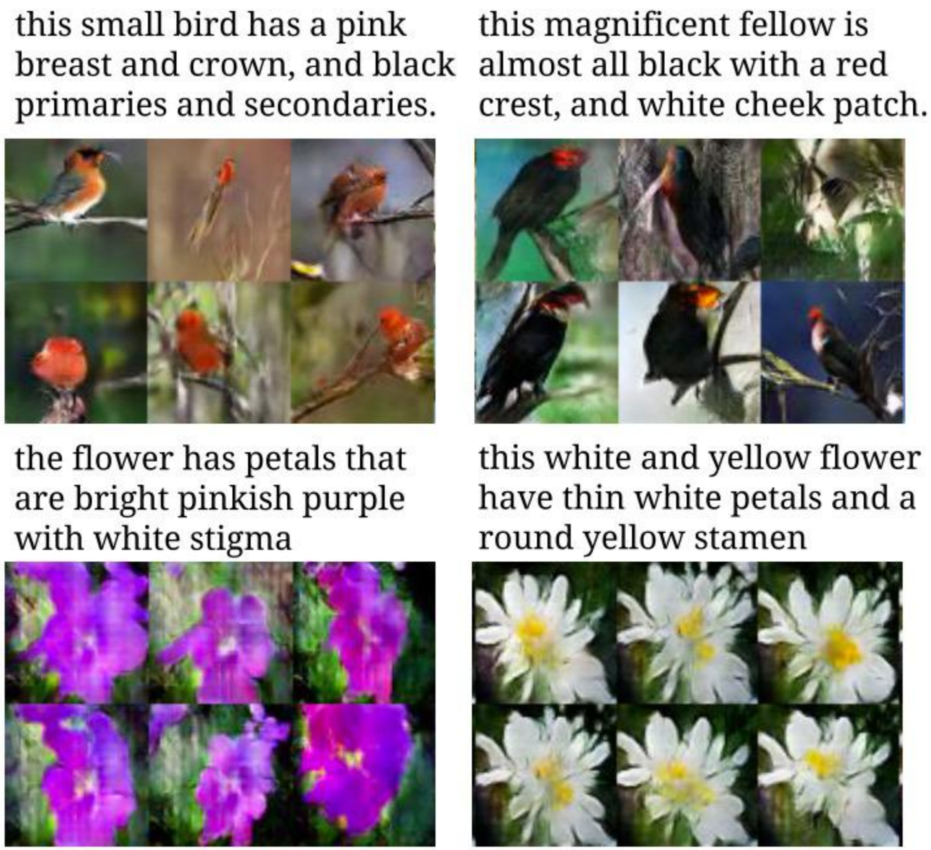
Joint distribution

$P(X =$  $|Y_1 =$  $, Y_2 = \text{“}a\ yellow\ bird\ with\ grey\ wings\text{”})$

Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*
Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

Conditional GAN

- Text-to-image synthesis: Another Multi-modal generation problem



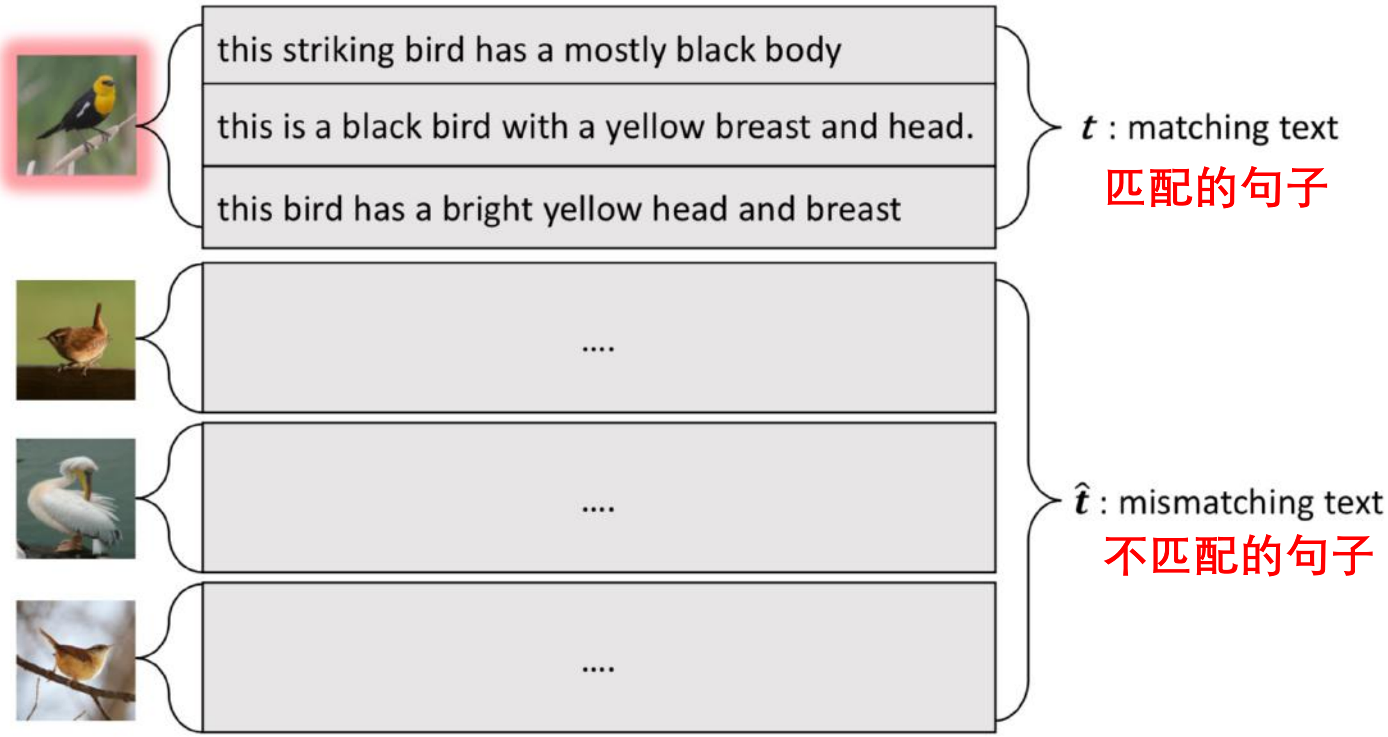
- Classic multi-modal problem

P(t, z)

Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*

Conditional GAN

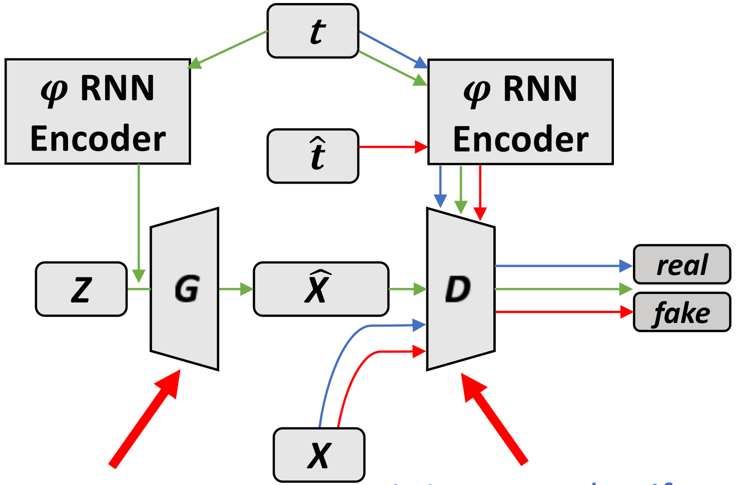
- Text-to-image synthesis



Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*

Conditional GAN

- Text-to-image synthesis

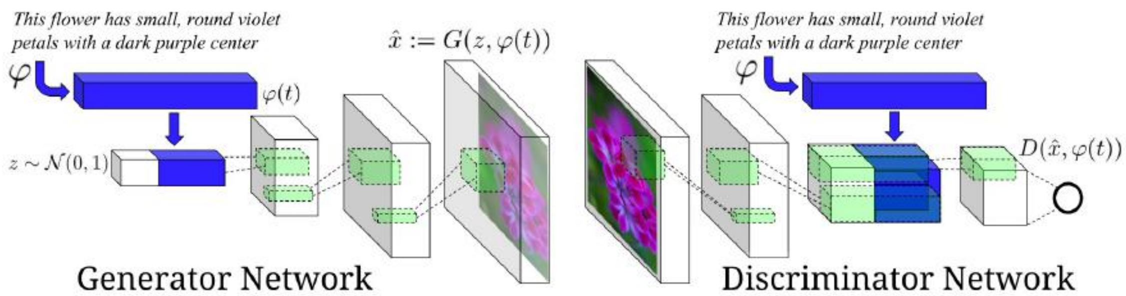


t : matching text

$\hat{t}$: mismatched text

Learn to fool discriminator

1. Learn to classify matching image and text as real sample
2. Learn to classify mismatched image and text as fake sample
3. Learn to classify image from generator as fake sample



PEKING UNIVERSITY 北京大學

Find the latent representation

- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision

- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

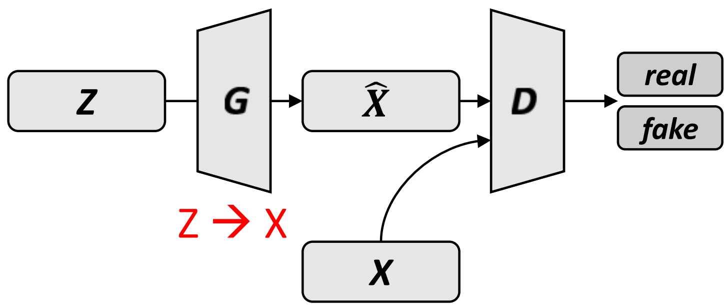
Do GANs generate new data?

- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

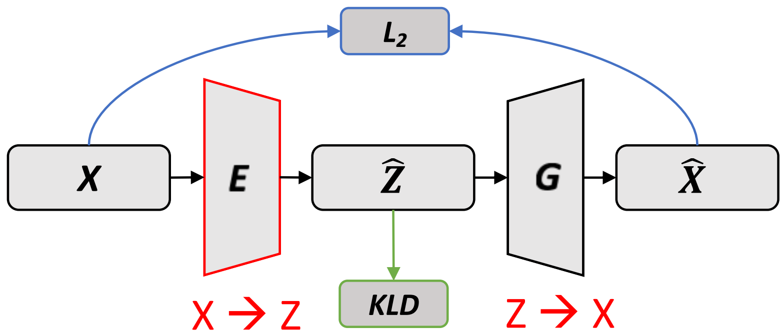
Find Latent Representation by Optimisation

- **Motivation: GAN vs. VAE**

Vanilla GAN



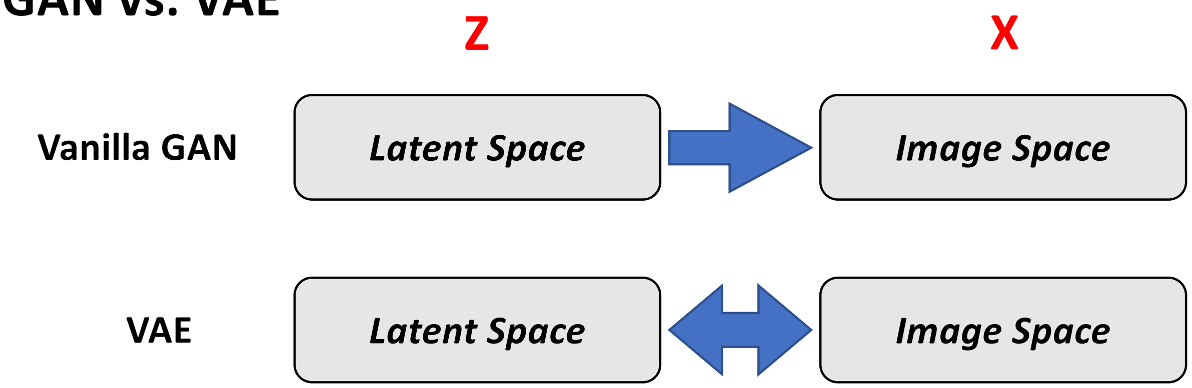
VAE variational autoencoder



VAE has an Encoder that can map x to z

Find Latent Representation by Optimisation

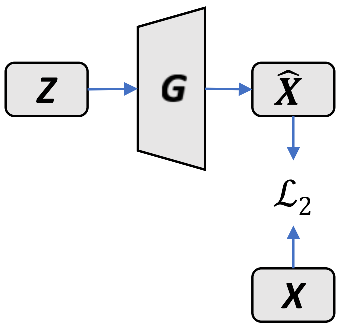
- **Motivation: GAN vs. VAE**



- VAE = **G**enerator + **E**ncoder
- Vanilla GAN = **G**enerator + **D**iscriminator

Find Latent Representation by Optimisation

- **Optimisation-based Method**



- Given a pretrained/fixed G and an image X
- Optimise: $\min_{z} ||x - G(z)||_2^2$
- Limitation: **SLOW**!

- Conditional GAN

Find the latent representation
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

Do GANs generate new data?
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

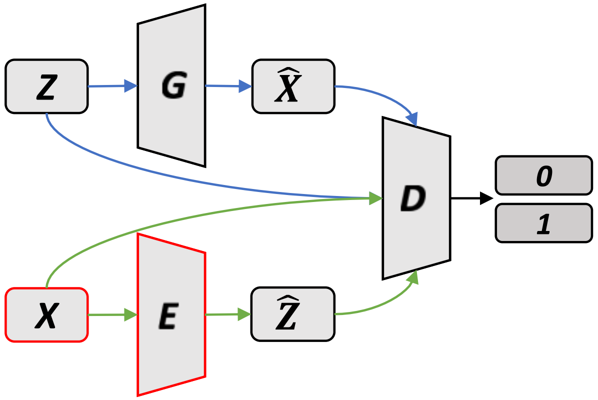
BiGAN: GAN with Encoder

- **Motivation: GAN vs. VAE**

- VAE = **G**enerator + **E**ncoder
- Vanilla GAN = **G**enerator + **D**iscriminator
- Better GAN = **G**enerator + **D**iscriminator + **Encoder**

BiGAN: GAN with Encoder

- **Find the joint distribution of X and Z**



BiGAN

Bidirectional GAN

$$p_G(\hat{X}, Z) = p_G(X|Z)p(Z)$$

$$p_E(X, \hat{Z}) = p_E(Z|X)p(X)$$

$$p_G(\hat{X}, Z) \qquad p_E(X, \hat{Z})$$

Minimise the gap between them

If E and G are optimal, then $E=G^{-1}$ almost everywhere, that is G(E(X))=X and E(G(z))=Z

Adversarial Feature Learning. *Jeff Donahue, Philipp Krahenbuhl, Trevor Darrell. ICLR 2016*

Find the latent representation

- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision

- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

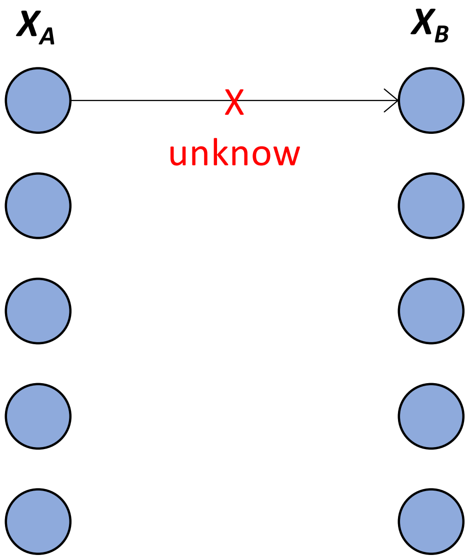
Do GANs generate new data?

- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

CoGAN

- **Learn the joint distribution of two <u>(semantically similar)</u> domains**

Data of two domains without known the mappings
(Learn the joint distribution)

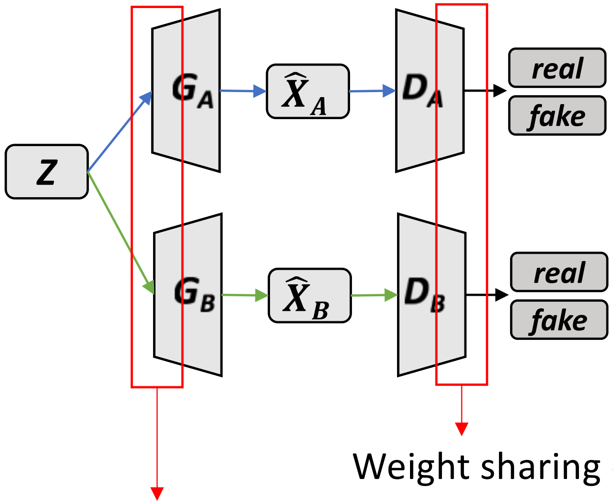


$$p(X_A, X_B) = \sum_{Z} p(X_A, X_B | Z) p(Z)$$

Coupled Generative Adversarial Networks. Liu, Ming-yu. Tuzel, Oncel. NIPS 2016

CoGAN

- **Utilising the model inductive bias**



Weight sharing on the last few layers (responsible for encoding high-level semantics)

Weight sharing on the first few layers (responsible for decoding high-level semantics)

"The weight-sharing constraint allows us to learn a joint distribution of images without correspondence supervision"

(The prior knowledge is from the model inductive bias)

Coupled Generative Adversarial Networks. Liu, Ming-yu. Tuzel, Oncel. NIPS 2016

CoGAN

- **Results**

Blond-hair

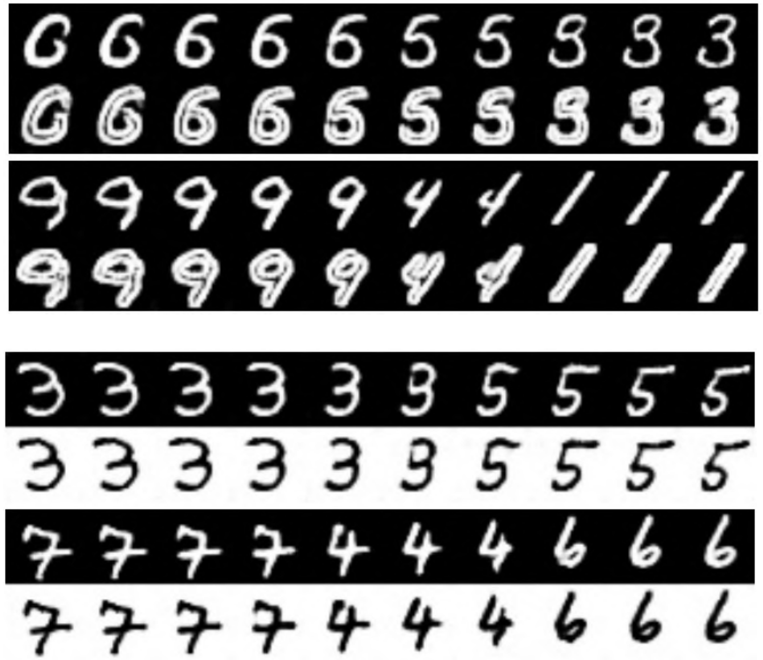
Not blond-hair

Smile

Not smile

Glasses

No glasses



x_A

x_B

x_A

x_B

x_A

x_B

x_A

x_B

Coupled Generative Adversarial Networks. Liu, Ming-yu. Tuzel, Oncel. NIPS 2016

CoGAN

- **Limitation**
  - It learns the joint distribution of two domains without known the mapping,
  - but when given an image, it cannot output the image of the other domain
  - so we need to map the images back to the latent codes for more applications…

Coupled Generative Adversarial Networks. Liu, Ming-yu. Tuzel, Oncel. NIPS 2016

北京大學 PEKING UNIVERSITY

Find the latent representation
- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

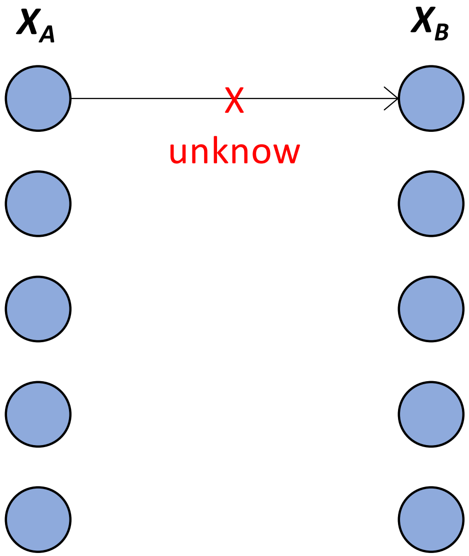
Do GANs generate new data?
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

CycleGAN, DualGAN, DiscoGAN and UNIT

- **<u>Unpaired</u> Image-to-Image Translation**

Data from two domains without known the mappings

(Learn the unknown mappings)



$$X_B = G_{A2B}(X_A), \; X_A = G_{B2A}(X_B)$$

X_A

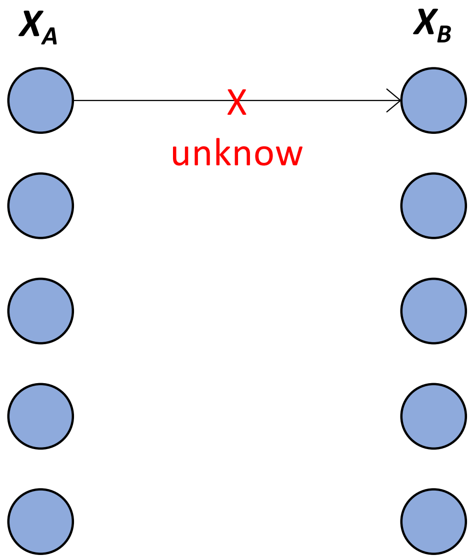
X_B

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

CycleGAN, DualGAN, DiscoGAN and UNIT

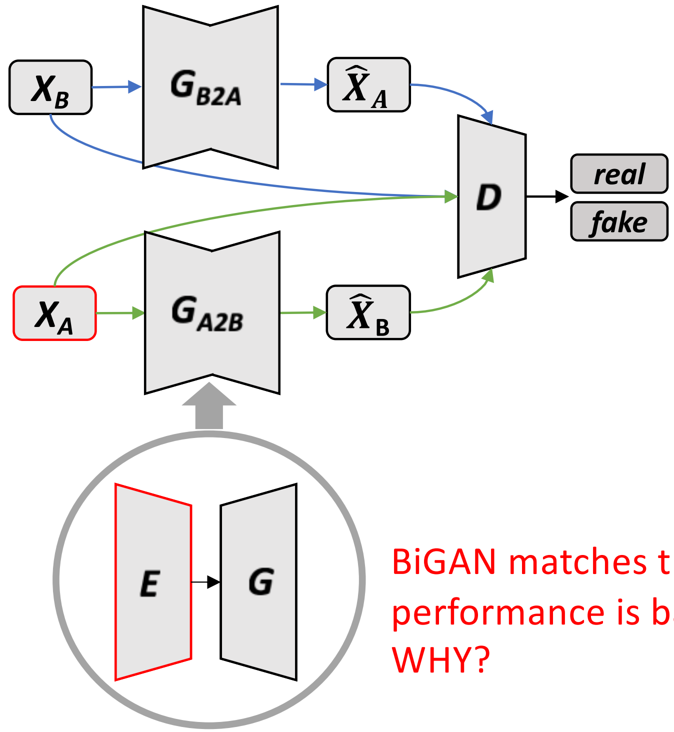
- **<u>Unpaired</u> Image-to-Image Translation**

Data from two domains without known the mappings
(Learn the unknown mappings)



$$X_B = G_{A2B}(X_A), \; X_A = G_{B2A}(X_B)$$

BiGAN with Autoencoders as Generators



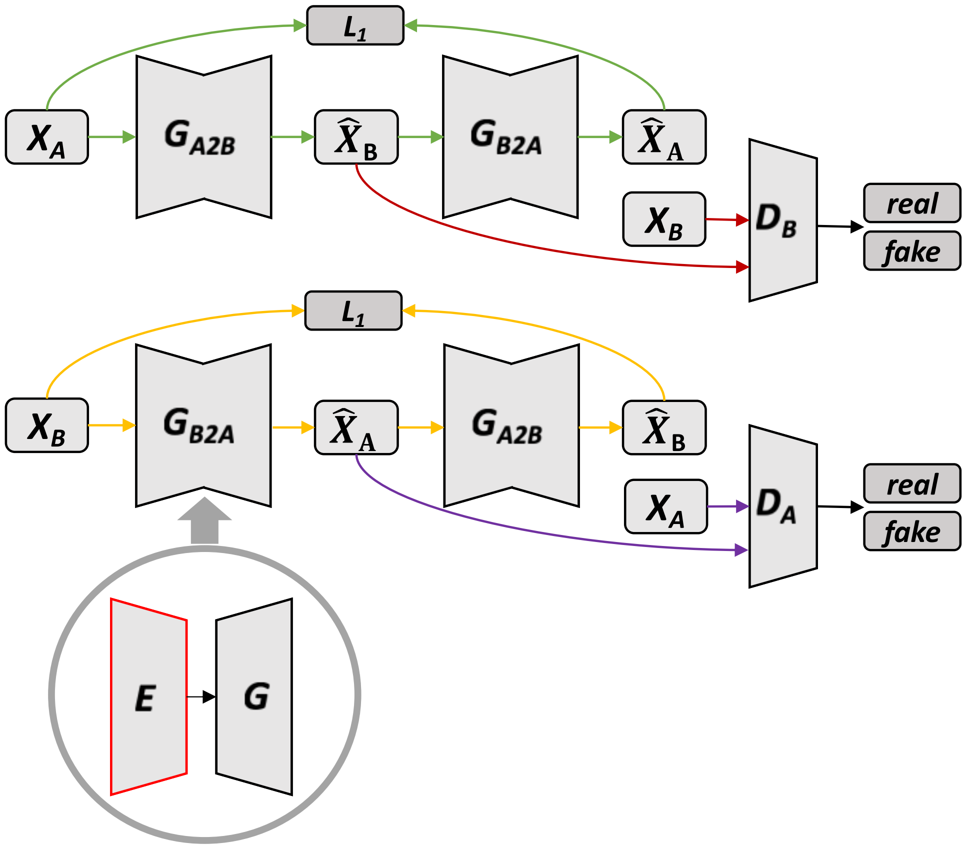
BiGAN matches this setting, but the performance is bad…
WHY?

Encoder is a part of the generator (fully conv nets)

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

CycleGAN, DualGAN, DiscoGAN and UNIT

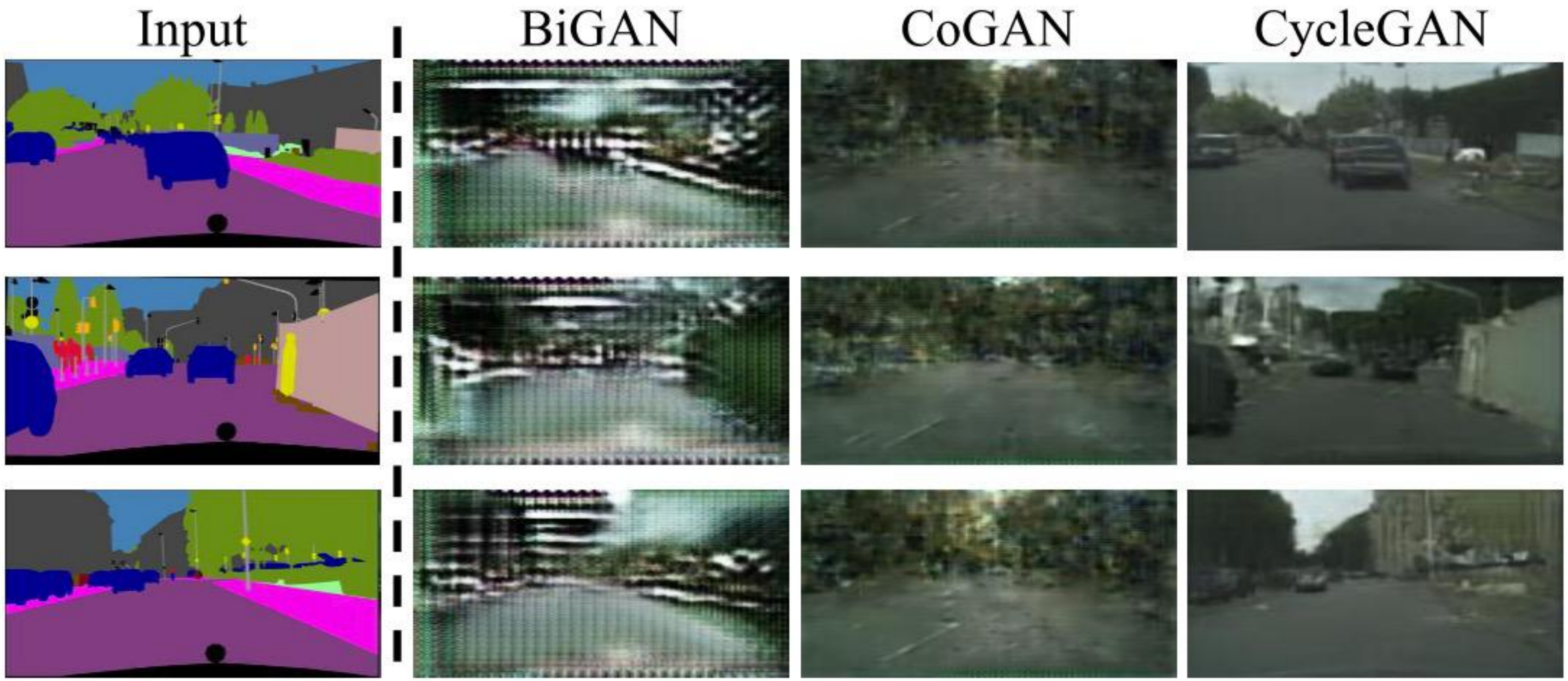
- **Cycle-consistency loss + adversarial loss**



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

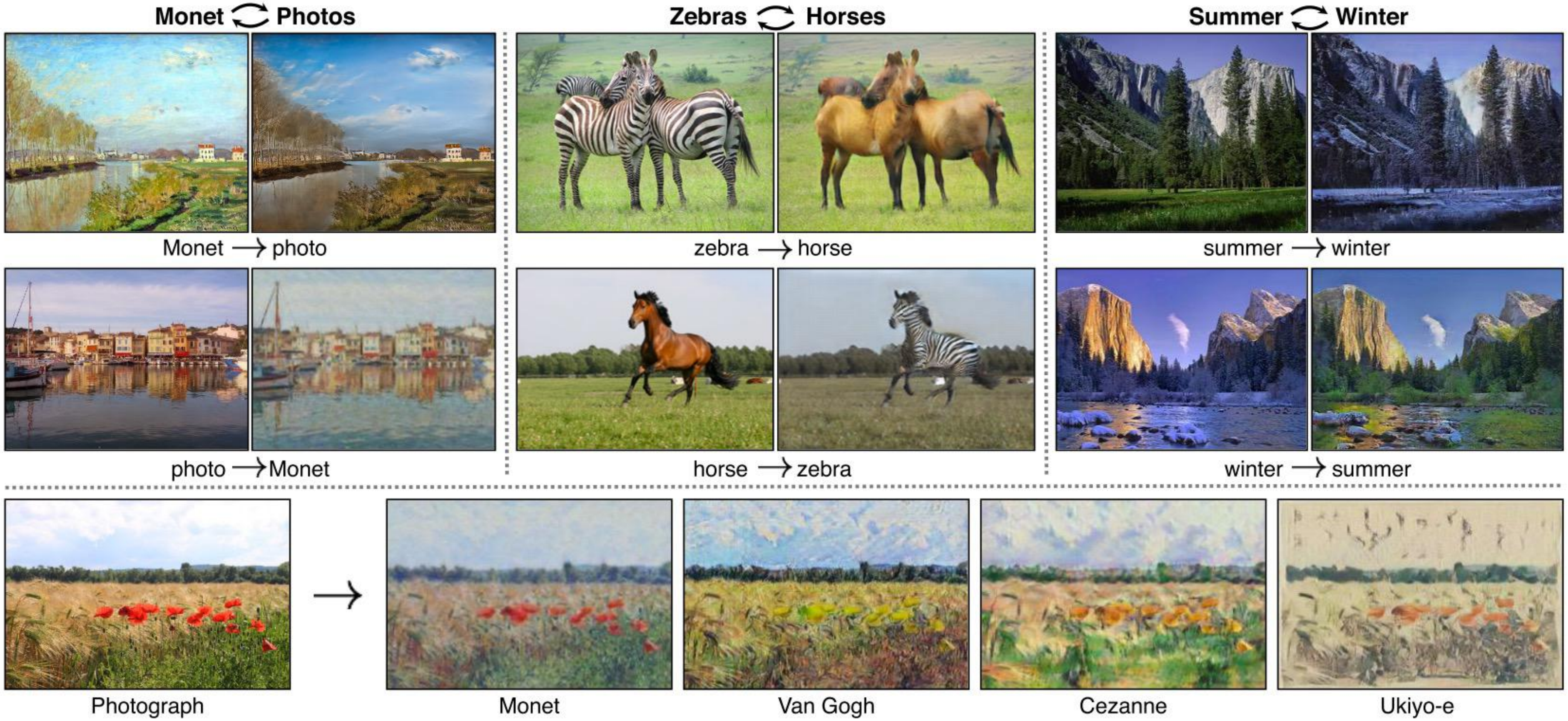
CycleGAN, DualGAN, DiscoGAN and UNIT

- **Importance of cycle-consistency loss**



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

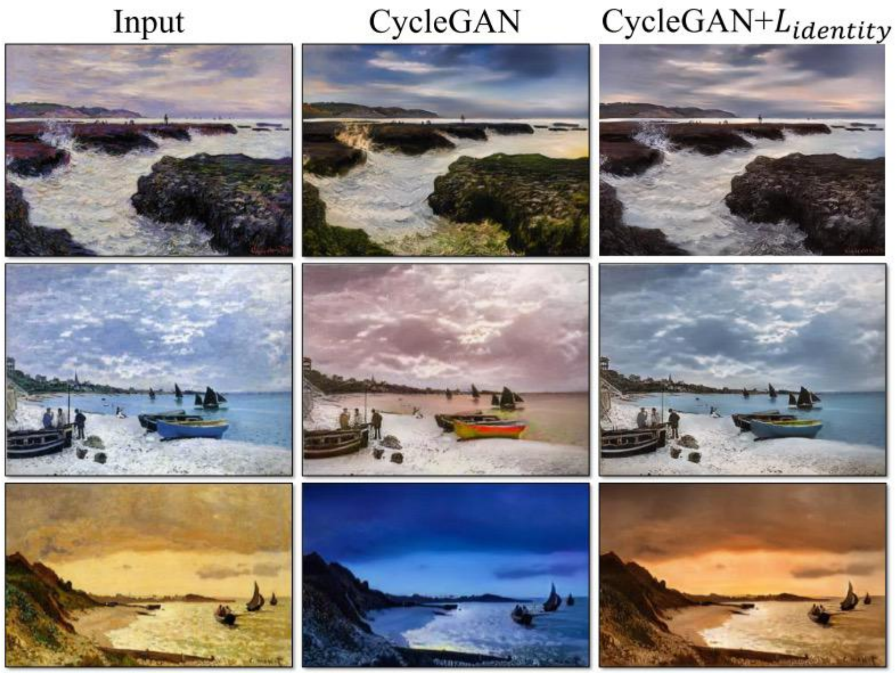
CycleGAN, DualGAN, DiscoGAN and UNIT



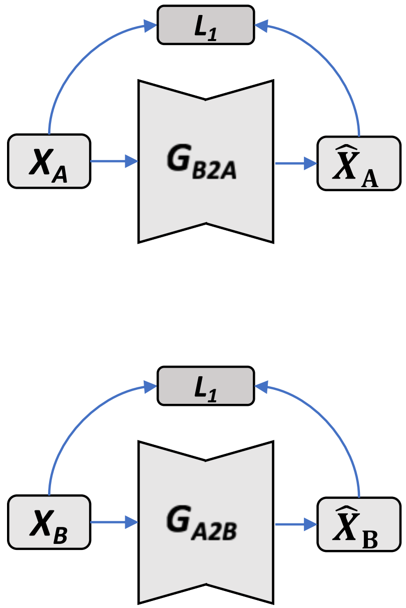
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

CycleGAN, DualGAN, DiscoGAN and UNIT

- **Identity loss**



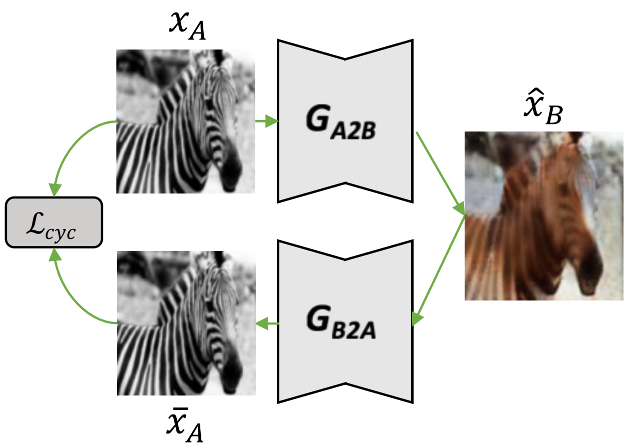
Monet→ Photo

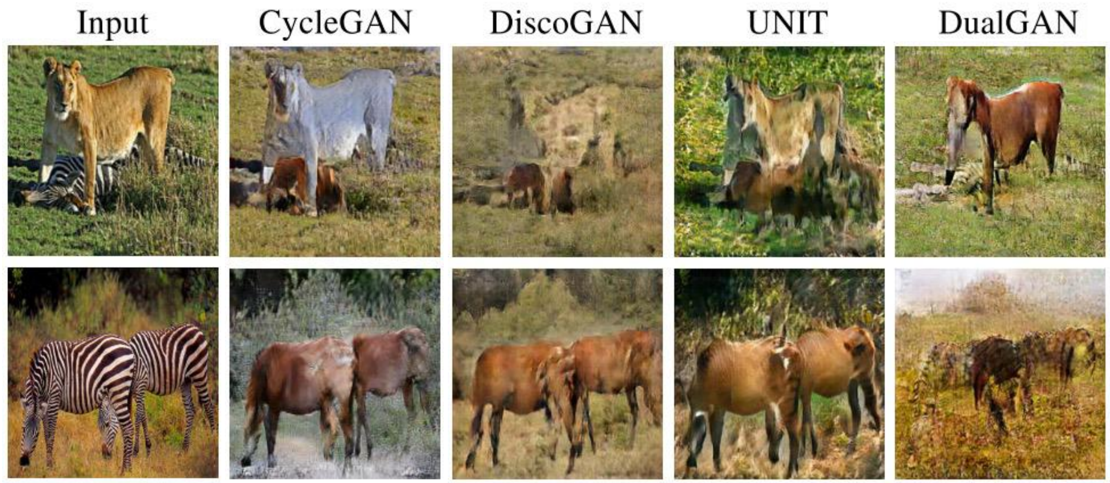


Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

CycleGAN, DualGAN, DiscoGAN and UNIT

- **Limitation of Cycle Consistency Loss**





Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

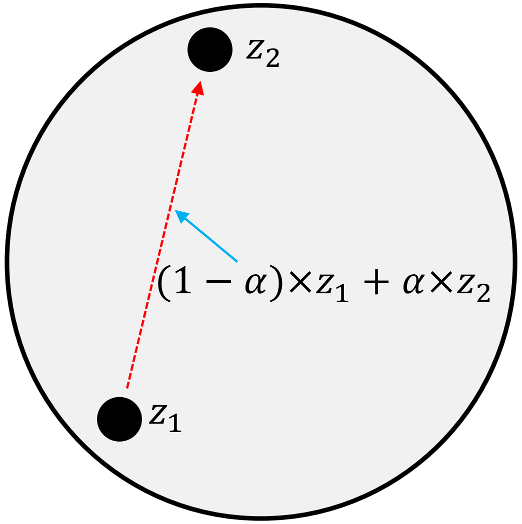
Find the latent representation

Find the mapping without supervision

Do GANs generate new data?

Walking on the Latent Space

- **Linear Interpolation**



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)**

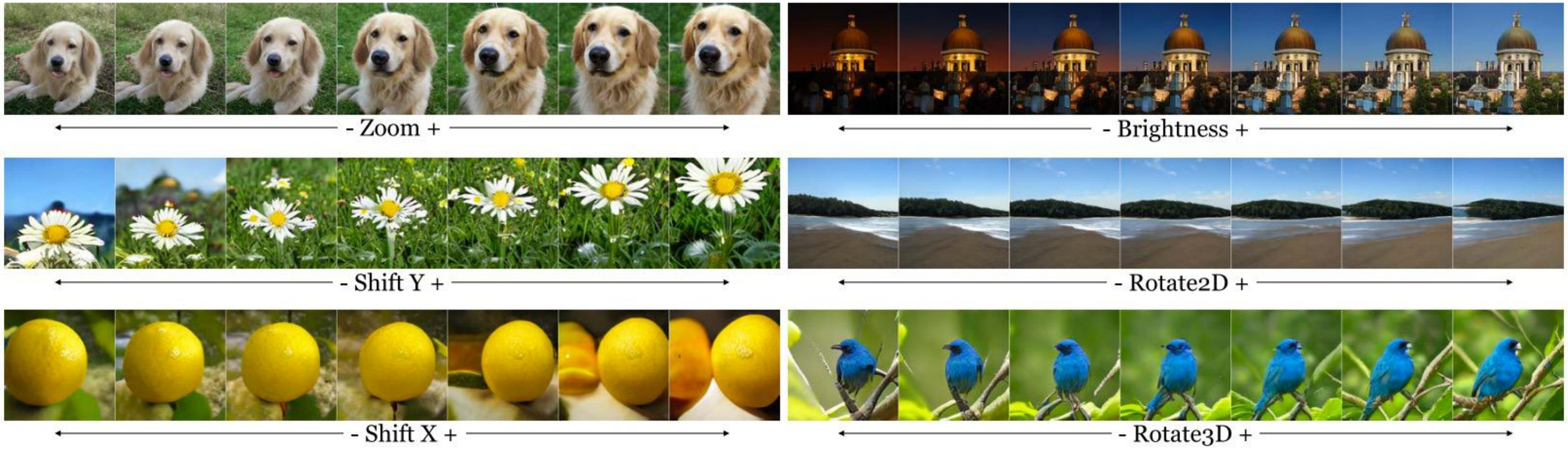
$\hat{x} = G(z)$

artifact in the space

$\hat{x} = G(z)$

Walking on the Latent Space

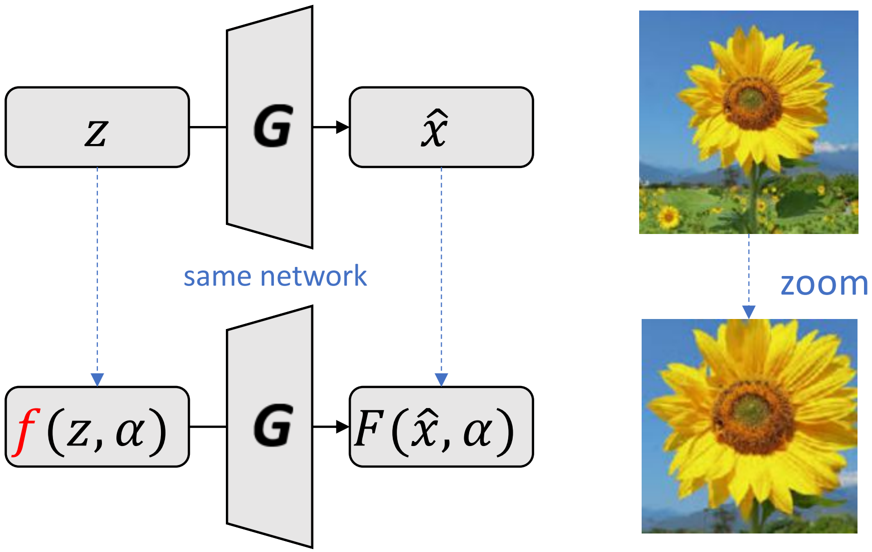
- **Beyond Random Walk: How to Control the Walking on the Latent Space?**
  - Given the prior knowledge: the transformation functions (zoom, shift ..) on image space
  - Find the corresponding function on the latent space



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**

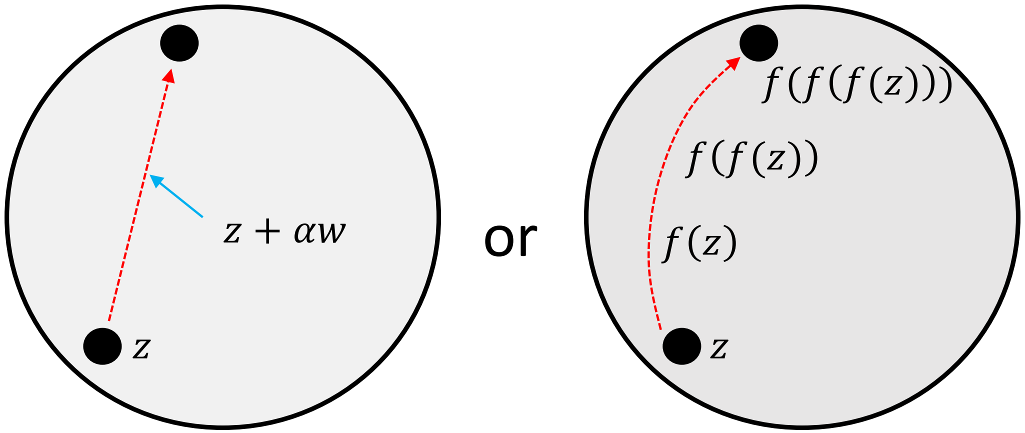


- Given
  - 1. image transformation function F
    - shifting, zooming, brightness …
    - α controls the degree
  - 2. pre-trained generator G
- Find latent transformation f

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

Walking on the Latent Space

- **Latent Transformation Function**



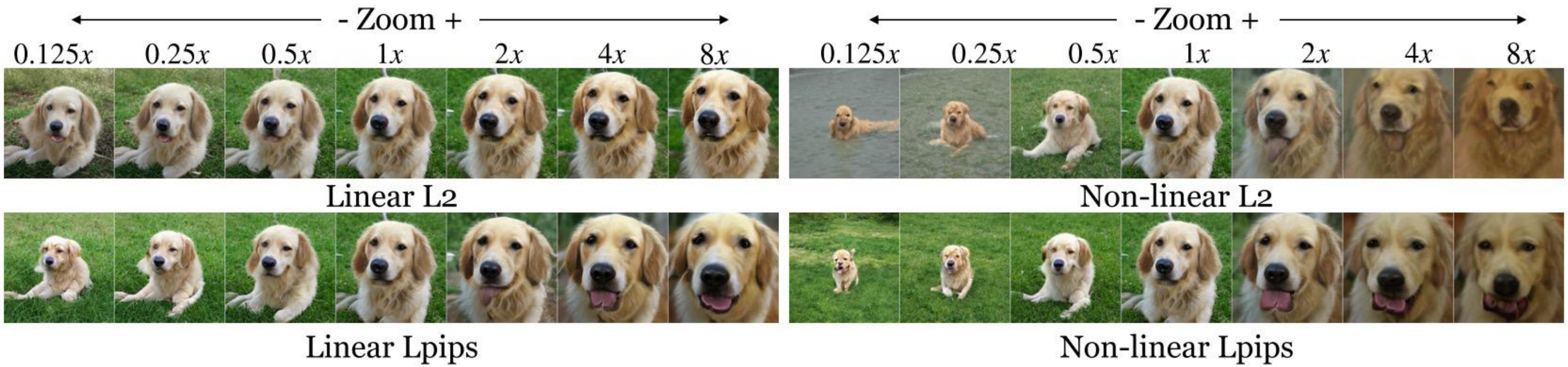
Linear walk on latent space

Non-linear walk on latent space
Function is a neural networks

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

Walking on the Latent Space

- **Latent Transformation Function**

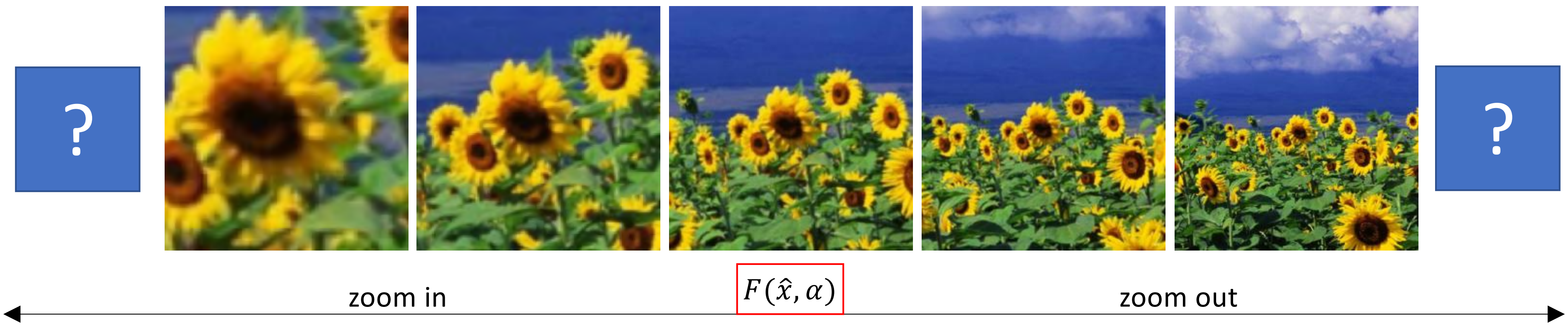


Comparison of linear and nonlinear walks for the zoom operation. The linear walk undershoots the targeted level of transformation, but maintains more realistic output.

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

Walking on the Latent Space

- **Discussion: Can We Zoom In/Out an Object Infinitely?**

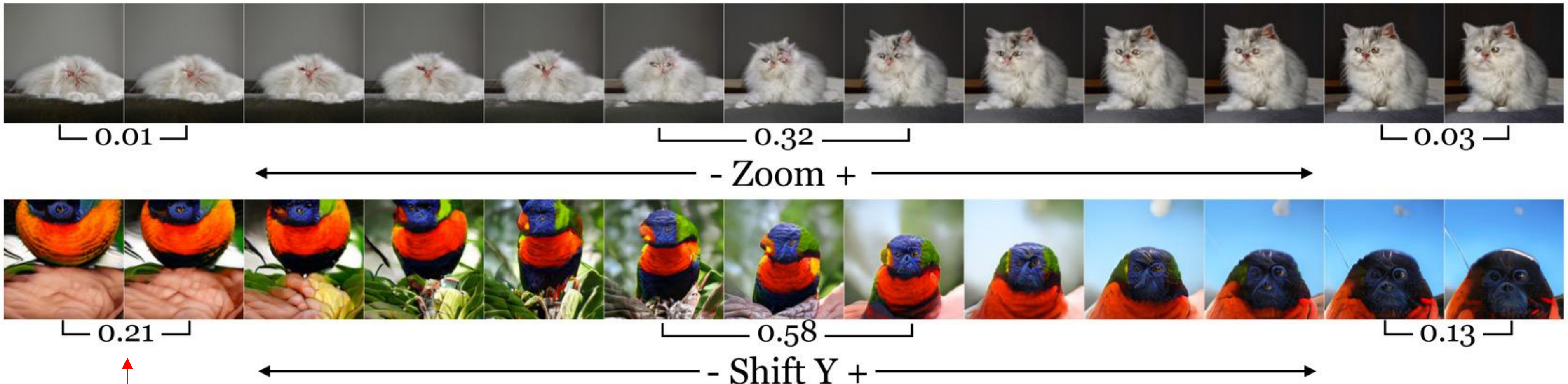


On the “steerability” of generative adversarial networks. Jahanian, Ali. Chai, Lucy. Isola, Phillip. ICLR 2020

Walking on the Latent Space

- **Latent Transformation Limits**

$\boxed{G(f(z,\alpha))}$



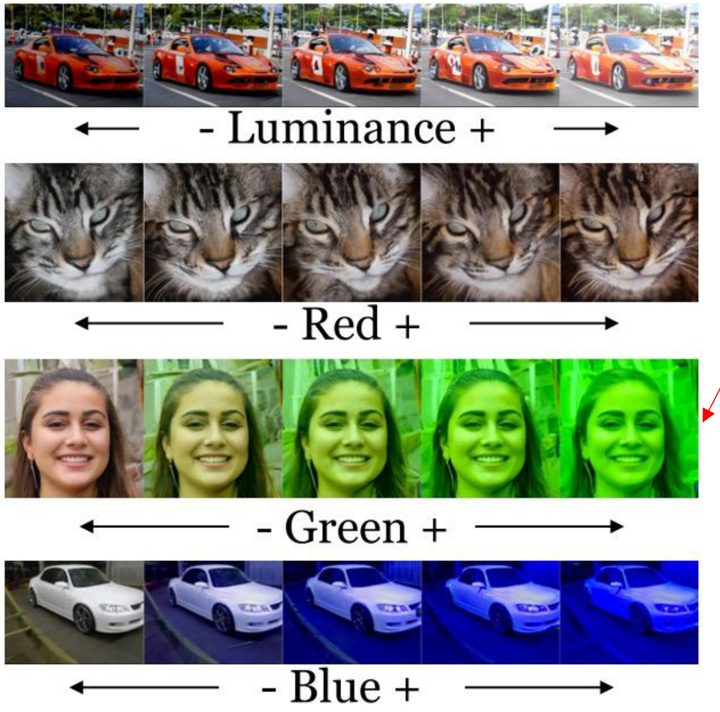
Reach the limit, the image starts to deviate from the natural image manifold

we can transform images to some degree but cannot extrapolate entirely outside the support of the training data.

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

Walking on the Latent Space

- **Discussion: Did GANs really learn to generate new data?**
  - After learning the concepts of luminance, color, brightness?

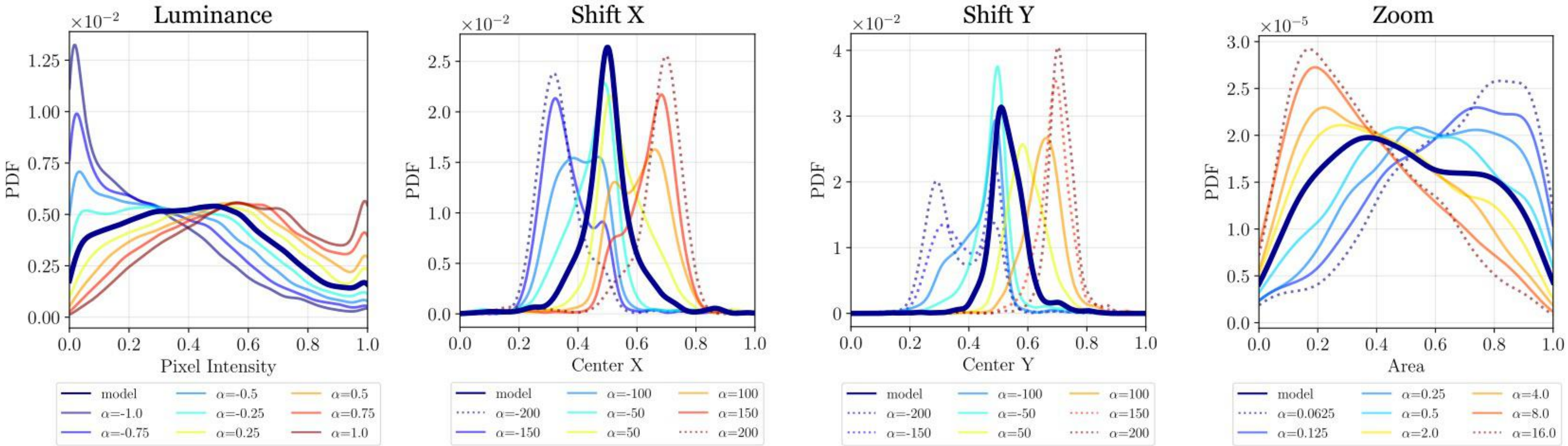


Does not exist on the original dataset

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

Walking on the Latent Space

- **Discussion: Did GANs really learn to generate new data?**
  - After learning the concepts of zooming, shifting …. ?



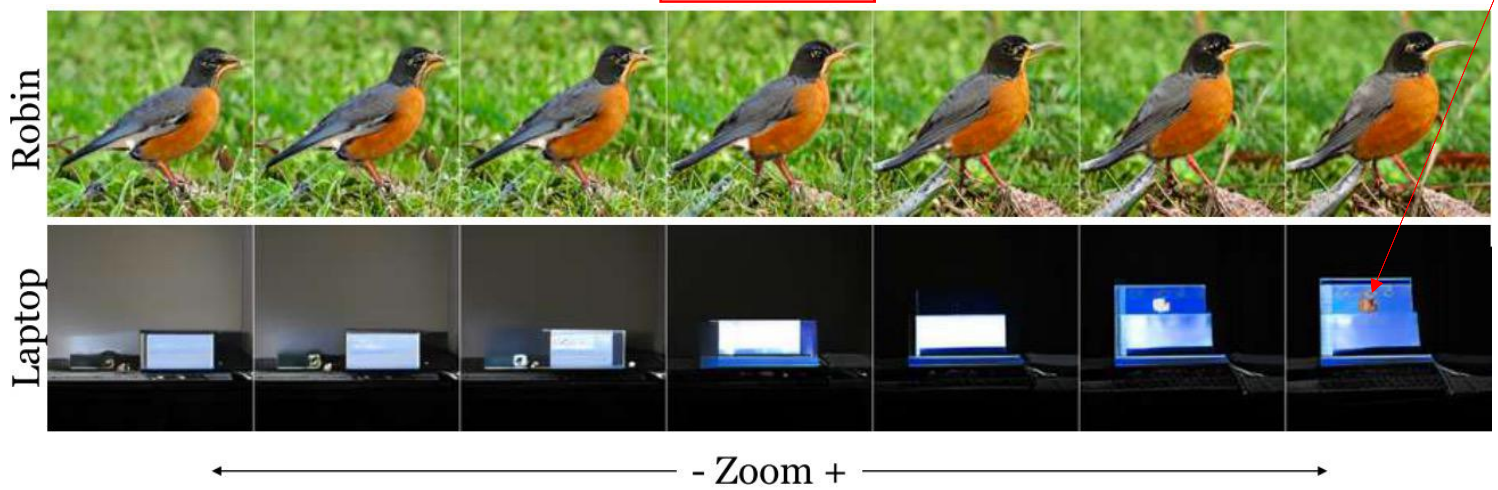
On the “steerability” of generative adversarial networks. Jahanian, Ali. Chai, Lucy. Isola, Phillip. ICLR 2020

Walking on the Latent Space

- **Discussion: Image Transformation vs. Latent Transformation**
  - Latent space is a prior distribution,
    so data points in latent space always generate "plausible" images
    i.e., the prior distribution is a constraint
  - Latent transformation uses the generator as a "memory":

Generate sth on the screen which does not exist on the original image

$\boxed{G(f(z, \alpha))}$



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

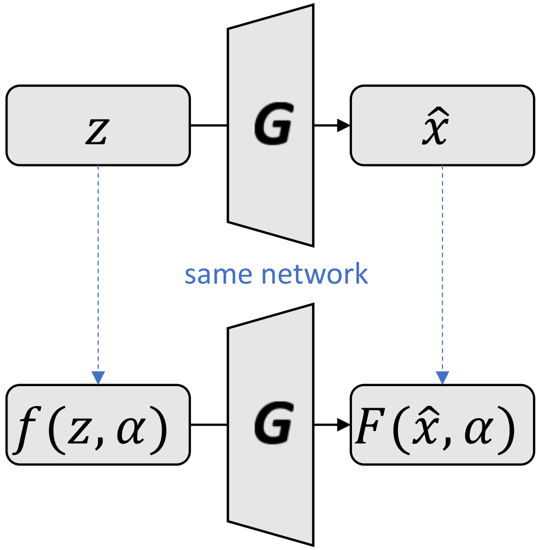
Walking on the Latent Space

GANs learn to generate new data, they able to generalise outside of the training distribution in some degree

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

Walking on the Latent Space

- **A "Steerability" Application: GANalyze**
  - Previous: From Image Transformation to Latent Transformation
  - Now: Differentiable Assessor (e.g., classifier/regressor) to Latent Transformation



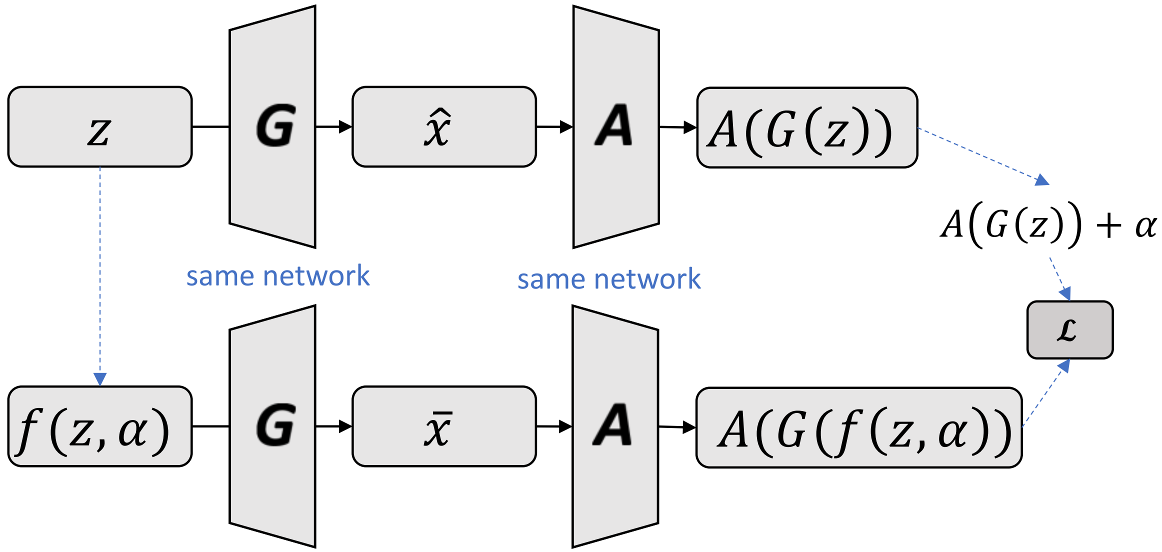
Previous

GANalyze: Toward Visual Definitions of Cognitive Image Properties. Andonian, Alex. Isola, Phillip. arXiv 2019

Walking on the Latent Space

- **A "Steerability" Application: GANalyze**
  - Given:
    - a pretrained MemNet as the Accessor A (tell how memorable the images are)
    - a pretrained Generator G
    - α controls how to change the memorable score
  - Find latent transformation f

Learn to change the score $A(G(z))$ by learning to change the latent space

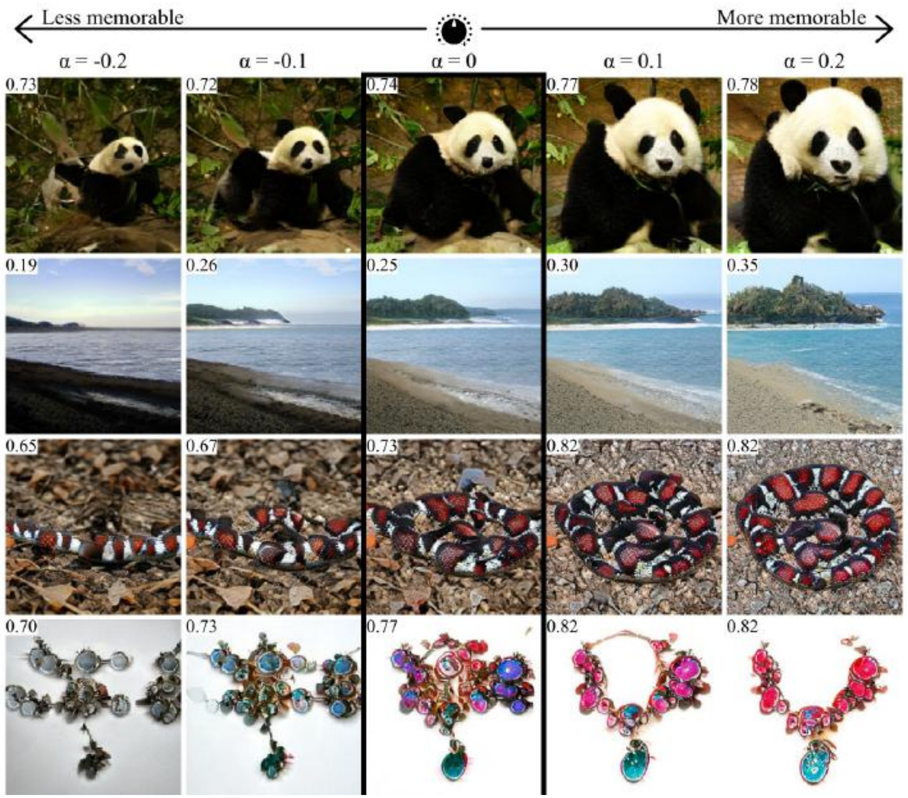


GANalyze: Toward Visual Definitions of Cognitive Image Properties. Andonian, Alex. Isola, Phillip. arXiv 2019

Walking on the Latent Space

- **A “Steerability” Application: GANalyze**

initial score



GANalyze: Toward Visual Definitions of Cognitive Image Properties. Andonian, Alex. Isola, Phillip. arXiv 2019

Walking on the Latent Space

- **A “Steerability” Application: GANalyze**

  - Discussion: Why it works?
    - The accessor network is differentiable, so the latent transformer can “feel” how to change the latent code
    - Similar to “steerability”, the latent space is a prior distribution, which is a constraint to ensure that the generator always generate “plausible” image. The generator would not generate a strange image to fool the accessor.

  - Prior knowledge is always required
    - “steerability”: image transformation functions
    - GANalyze: a pretrained accessor
    - … more applications … face editing with facial feature scorer …

GANalyze: Toward Visual Definitions of Cognitive Image Properties. Andonian, Alex. Isola, Phillip. arXiv 2019

北京大學 PEKING UNIVERSITY

Find the latent representation

Find the mapping without supervision

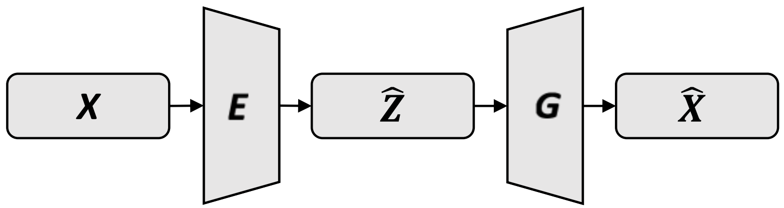
Do GANs generate new data?

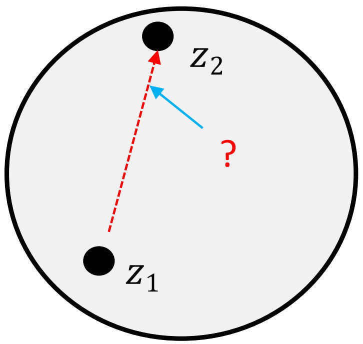
- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

Improving Interpolation via Adversarial Regularisation

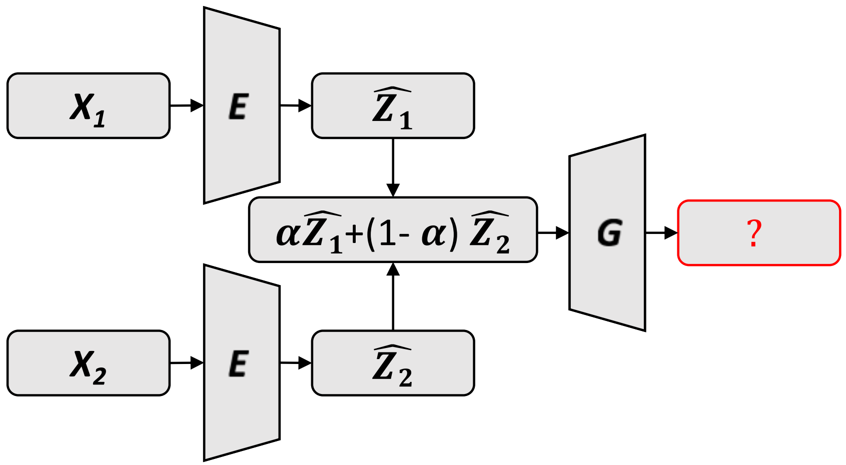
- **Beyond data: GAN learns concepts?**

VAE



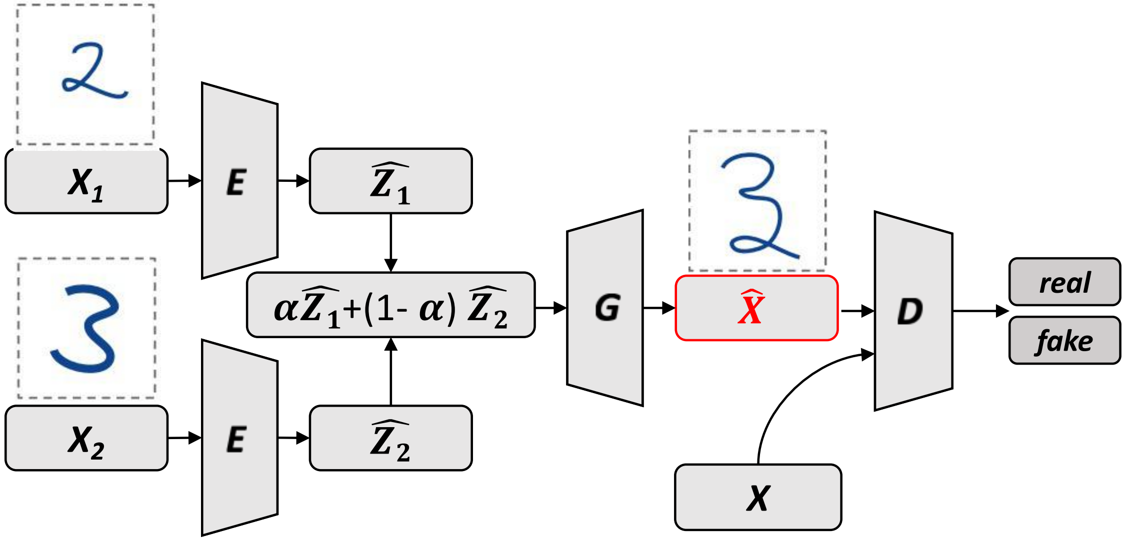


What is the GT?



Improving Interpolation via Adversarial Regularisation

- **Beyond data: GAN learns concepts?**



Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer.
David, Berthelot. Colin, Raffel. Aurko, Roy. Ian, Goodfellow. arXiv 2019

Improving Interpolation via Adversarial Regularisation

- **Experiments**

Training data: random clock

Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer.
David, Berthelot. Colin, Raffel. Aurko, Roy. Ian, Goodfellow. arXiv 2019

Improving Interpolation via Adversarial Regularisation

- **Results**

Alpha Blending
(Interpolation on data space)

Autoencoder
(Interpolation on latent space)

Autoencoder +
Adversarial Regularisation
(Interpolation on latent space)

Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer.
David, Berthelot. Colin, Raffel. Aurko, Roy. Ian, Goodfellow. arXiv 2019

Summary: Selected GANs

北京大學 PEKING UNIVERSITY

Find the latent representation
- Conditional GAN
- Find Latent Representation by Optimisation
- BiGAN: GAN with Encoder

Find the mapping without supervision
- CoGAN
- CycleGAN, DualGAN, DiscoGAN and UNIT

Do GANs generate new data?
- Walking on the Latent Space
- Improving Interpolation via Adversarial Regularisation

北京大學
PEKING UNIVERSITY

Thanks