

Challenge: Learning Large Encoder

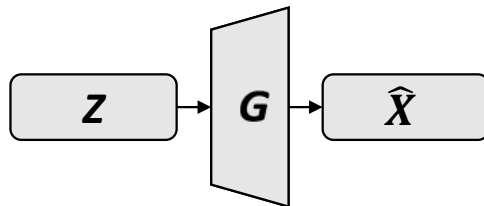
Hao Dong

Peking University

Challenge: Learning Large Encoder

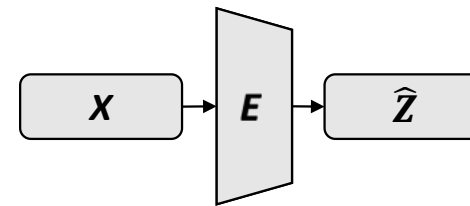
Previous Lecture: Large Image

Scalable



This Lecture: Large Encoder

Reversible



We use images for demonstration

Unsupervised Representation Learning!

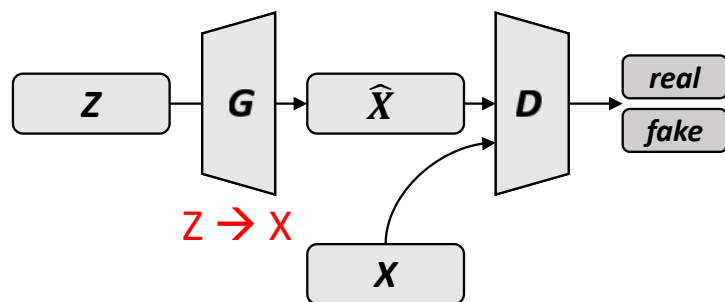
Challenge: Learning Large Encoder

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

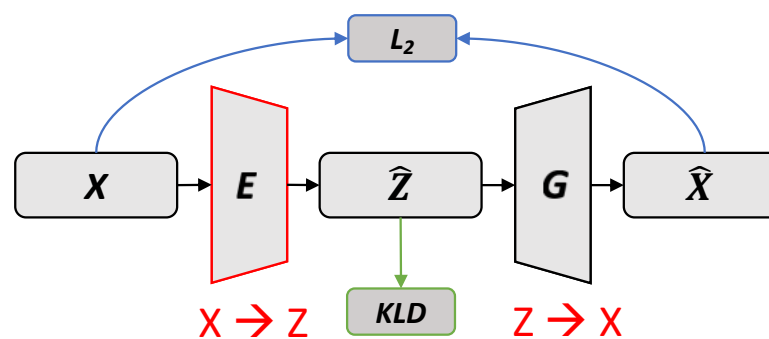
- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

VAE vs. GAN

Vanilla GAN

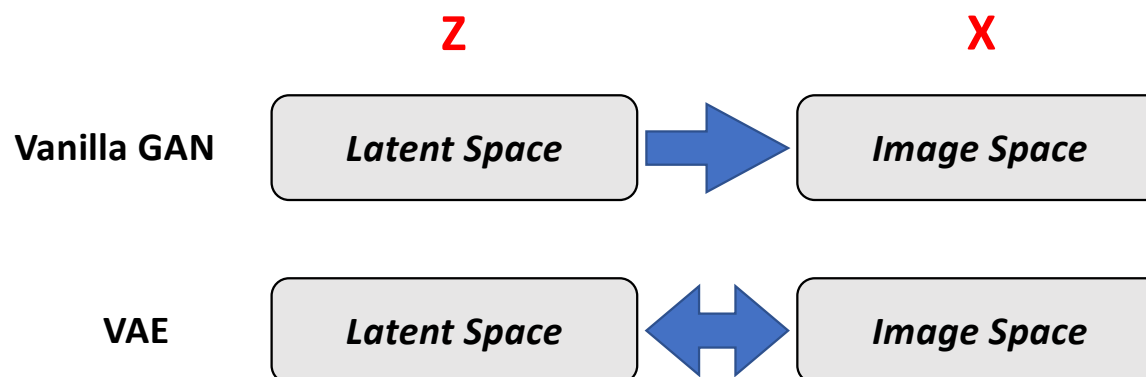


VAE variational autoencoder



VAE has an Encoder that can map x to z

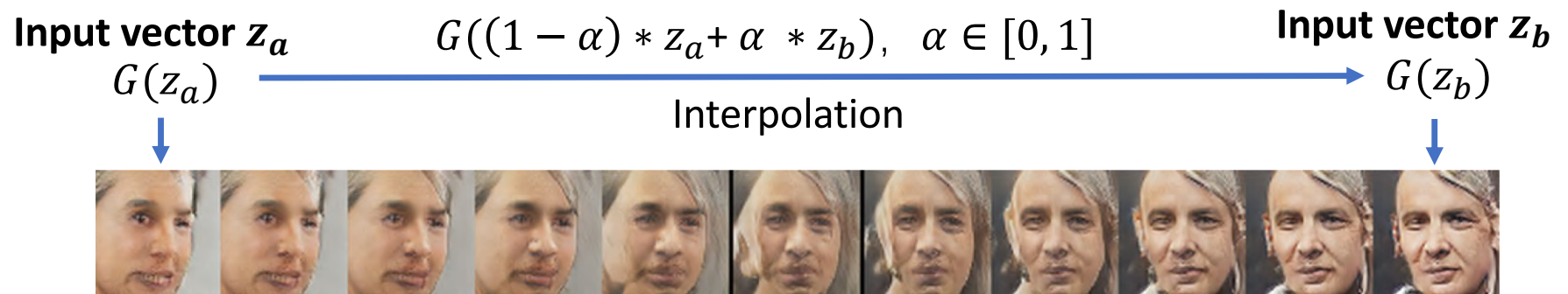
VAE vs. GAN



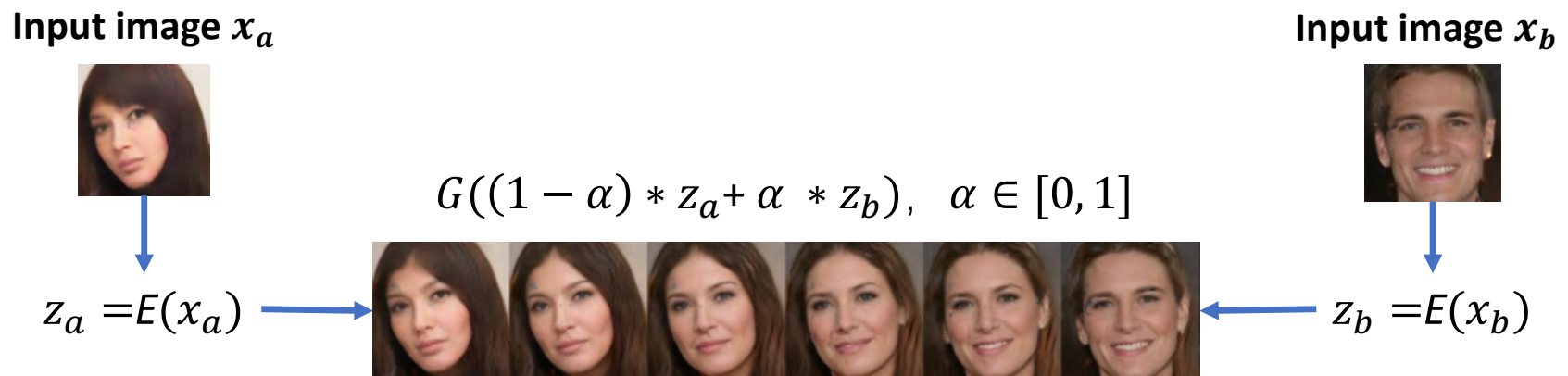
- VAE = **G**enerator + **E**ncoder
- Vanilla GAN = **G**enerator + **D**iscriminator
- Better GAN = **G**enerator + **D**iscriminator + **Encoder**

VAE vs. GAN

- Without Encoder:

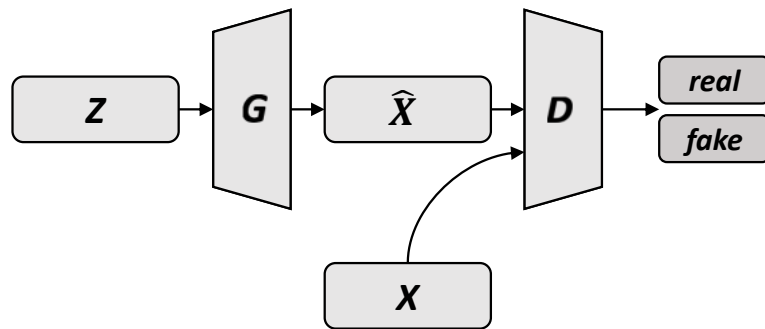


- With Encoder:

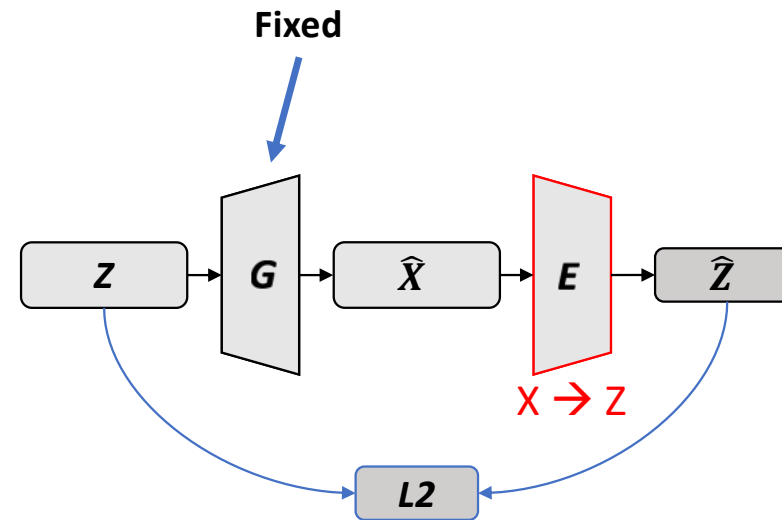


- VAE vs. GAN
- **A Naïve Approach**
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

A Naïve Approach



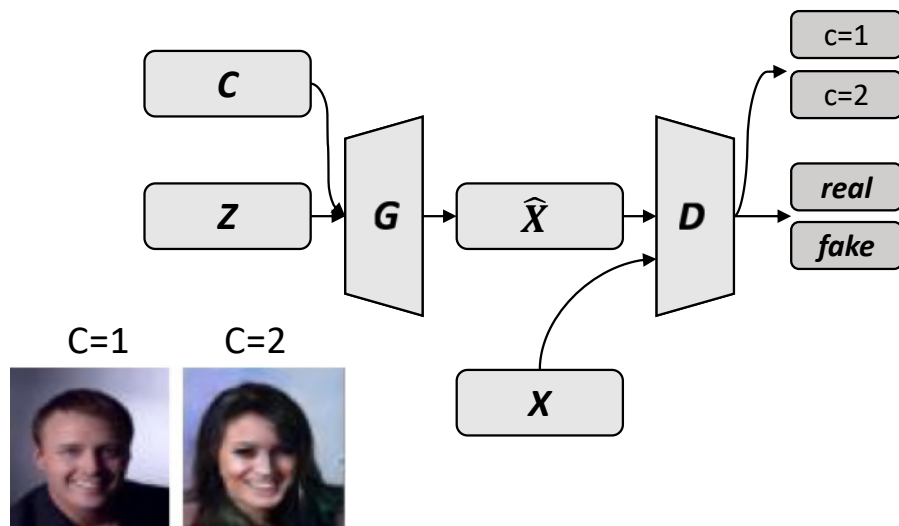
Step 1: Pre-trained G



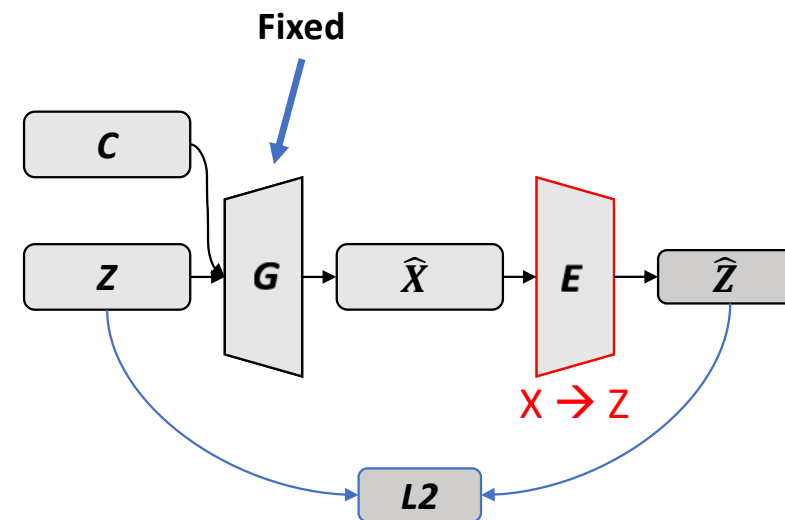
Step 2: Fix G and Train E

A Naïve Approach

- Application: Unsupervised/Unpaired Image-to-Image Translation



Given an ACGAN

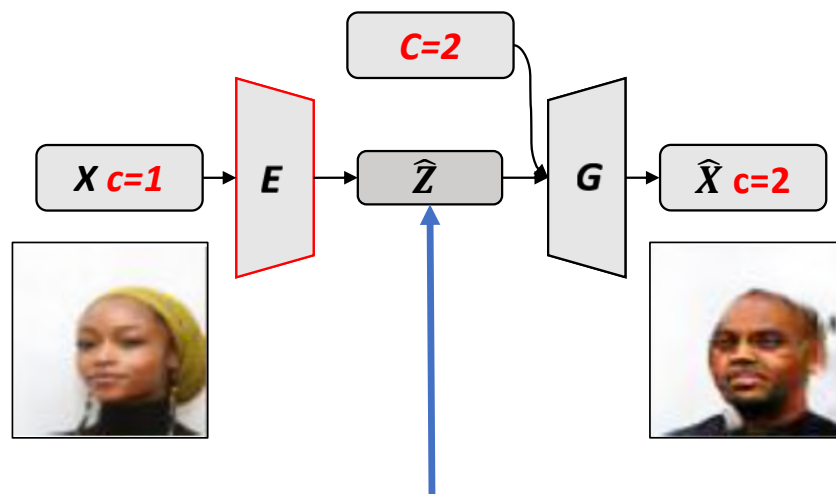


Learning the Encoder in a Brute Force Way

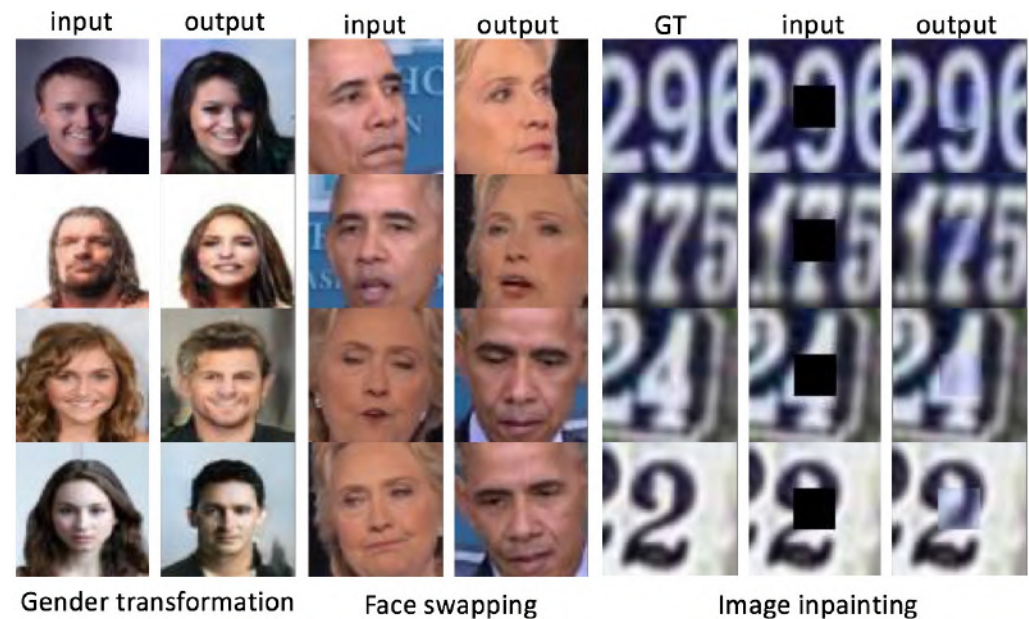
Z : shared latent representation across two domains

A Naïve Approach

- Application: Unsupervised/Unpaired Image-to-Image Translation

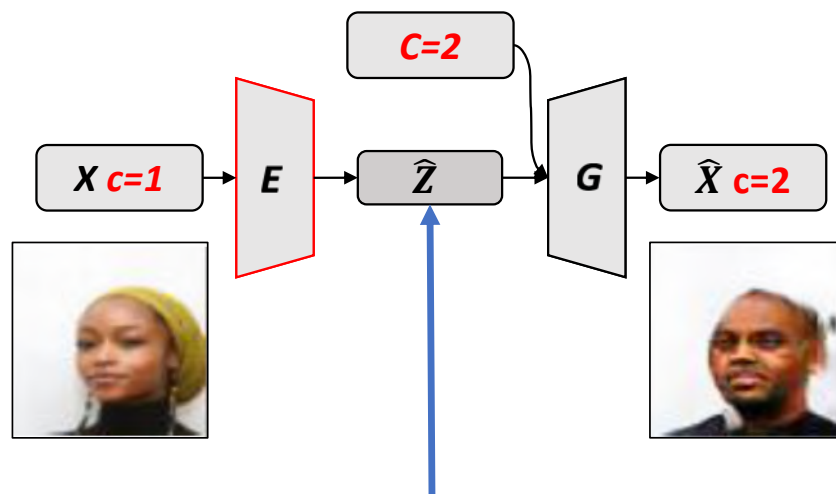


Shared Features Across Domain



A Naïve Approach

- Application: Unsupervised/Unpaired Image-to-Image Translation

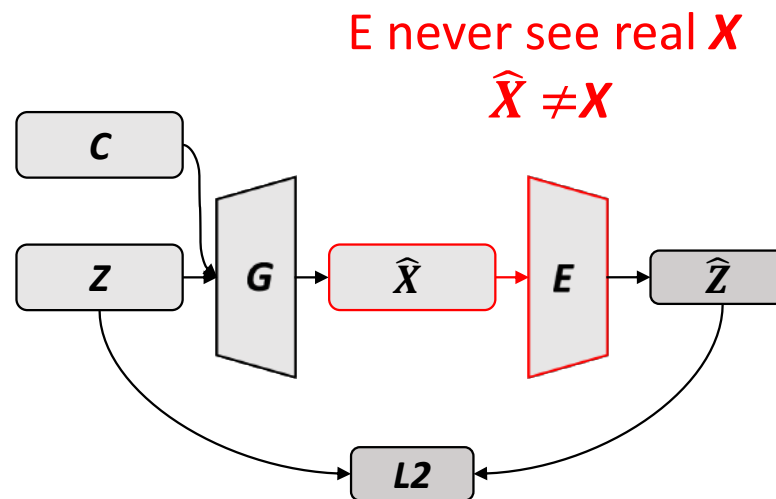


Shared Features Across Domain

Only Work Well for Simple Images with Small Size

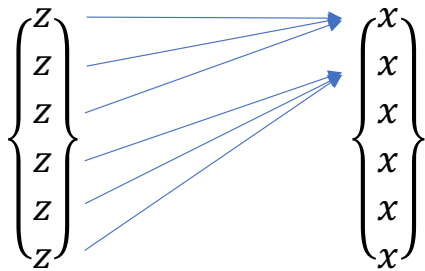
A Naïve Approach

- Limitation: Encoder never see real data sample !

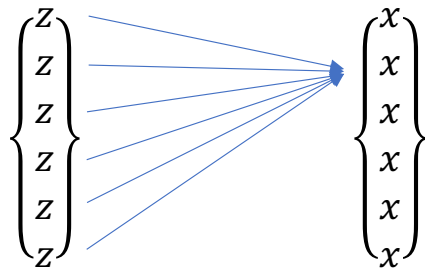


A Naïve Approach

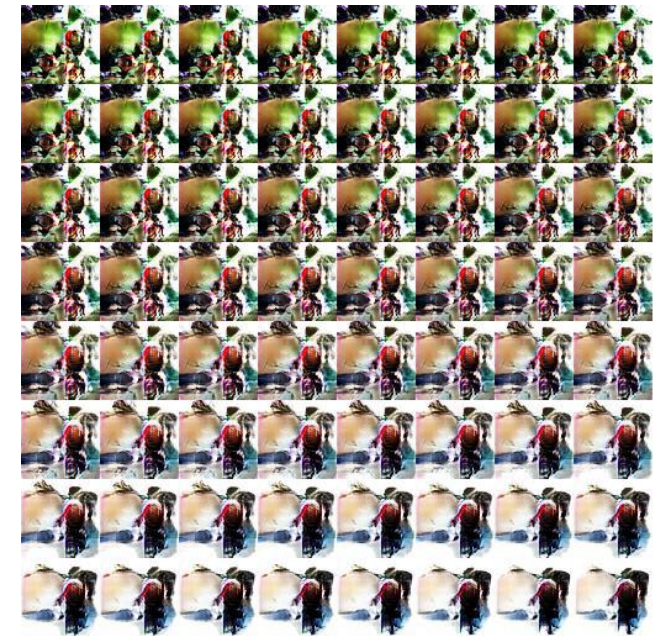
- Limitation: Encoder never see real data sample
and the synthesized data distribution \neq real data distribution
- Mode Collapse



G can only synthesis some part of the dataset x
and can fool D



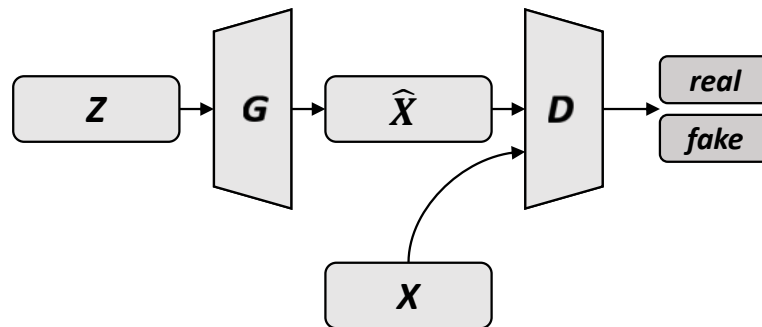
G can even only synthesis one data
and can fool D



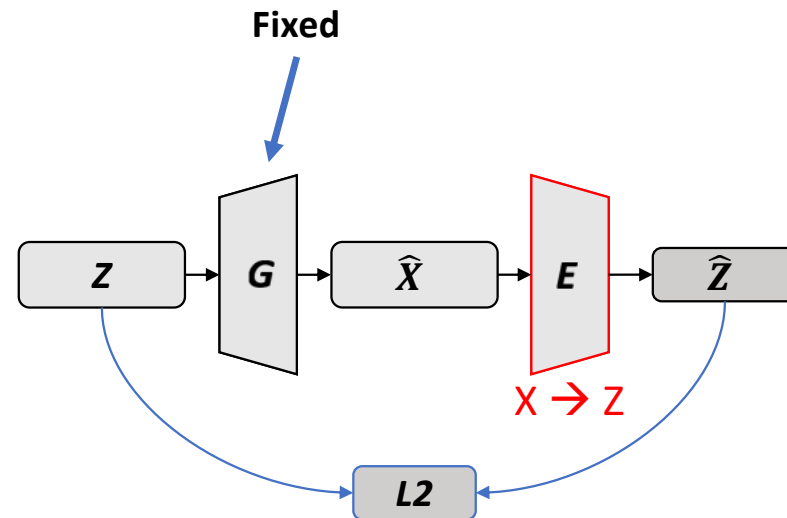
Examples of GAN collapse

A Naïve Approach

- Only work well if only if the fake distribution == the real distribution, but it is impossible in practice.



Step 1: Pre-trained G

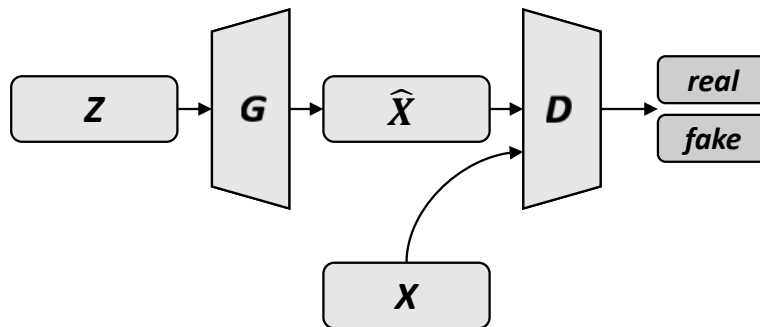


Step 2: Fix G and Train E

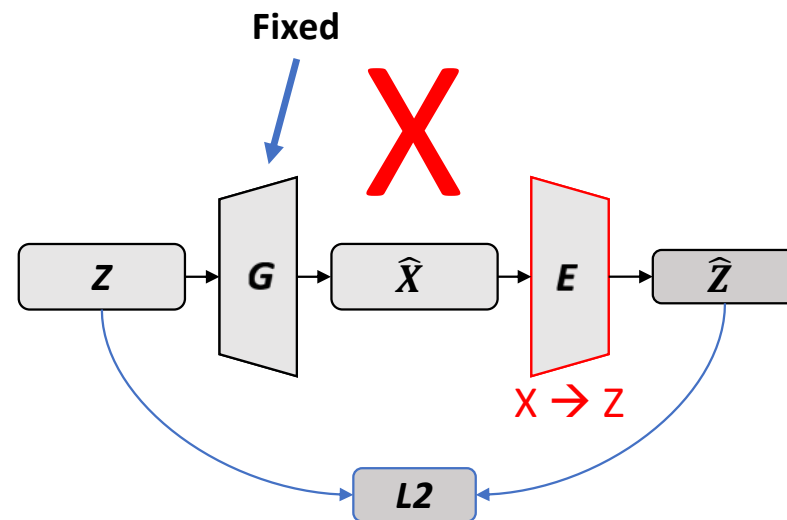
- VAE vs. GAN
- A Naïve Approach
- **Another Naïve Approach**
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

Another Naïve Approach

- Could E see real data sample?



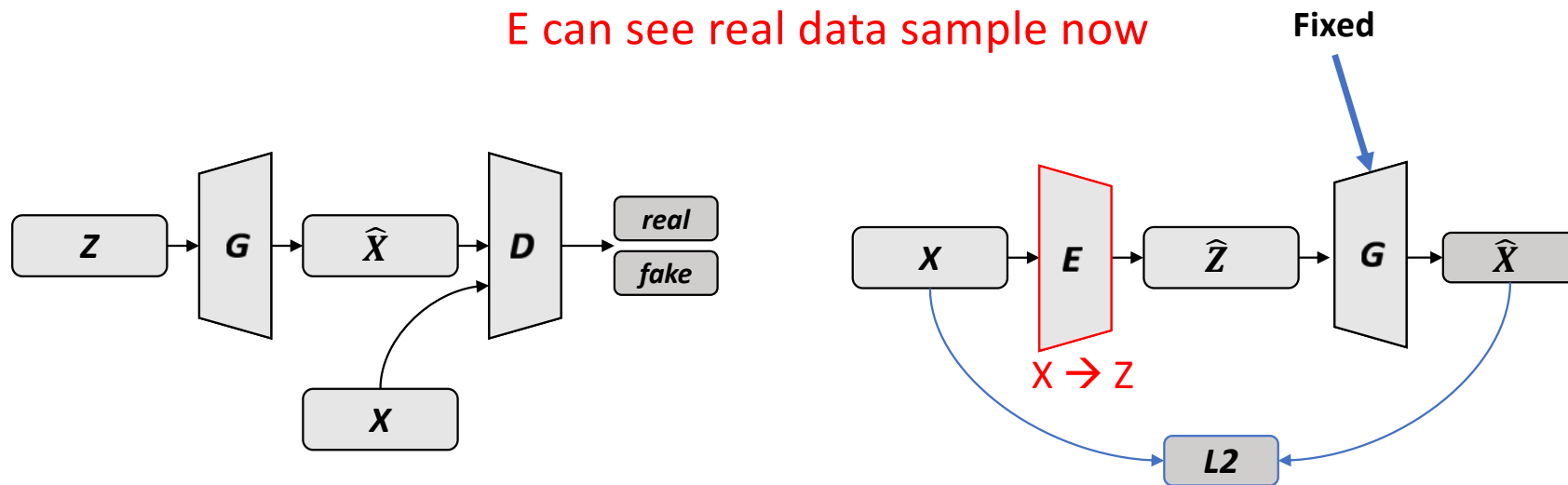
Step 1: Pre-trained G



Step 2: Fix G and Train E

Another Naïve Approach

- Could E see real data sample?



Another Naïve Approach

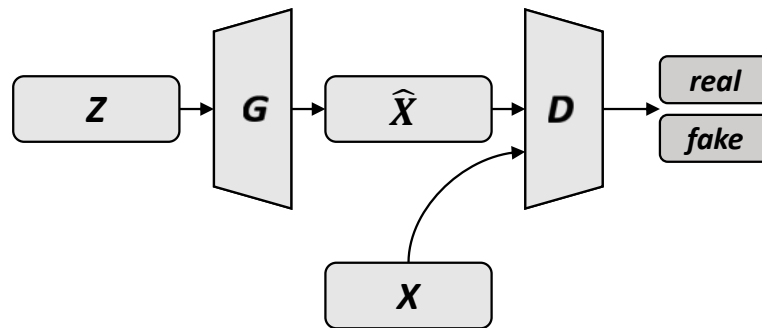
- Problem:
Difficult to converge (even using a super-deep E)
- Reason:
Model Collapse: G cannot synthesise the input image,
so the loss cannot be reduced

The quality of synthesised images \neq real images,
so the loss cannot be reduced

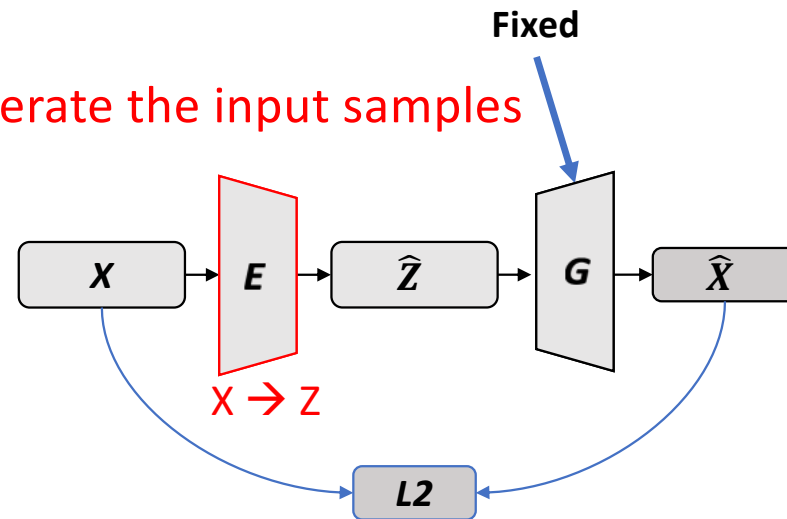
Another Naïve Approach

- Only work well if only if the fake distribution == the real distribution, but it is impossible in practice.

E can see real data sample now,
but G cannot always generate the input samples



Step 1: Pre-trained G

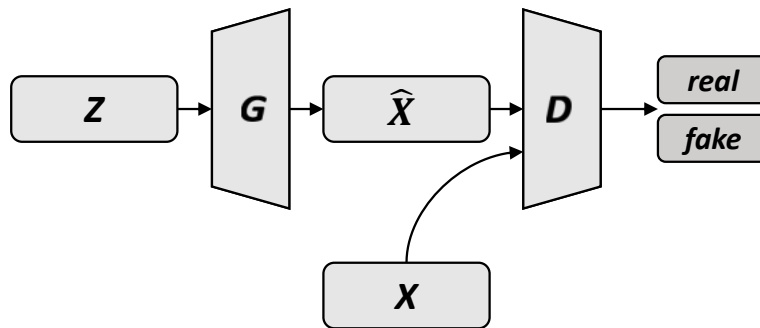


Step 2: Fix G and Train E

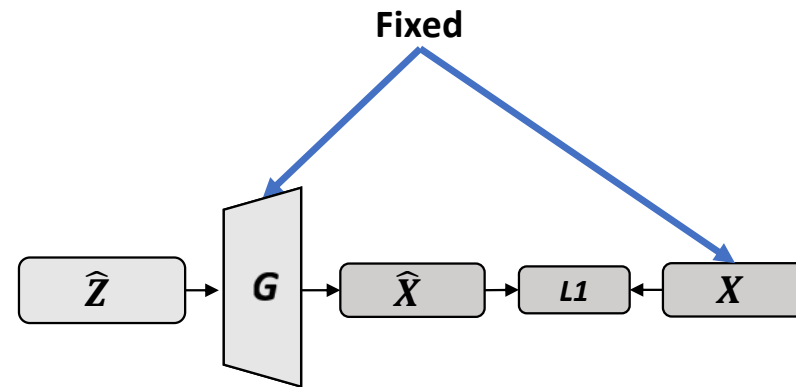
- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- **Without Encoder**
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

Without Encoder

- Optimisation-based method: find the optimal z **iteratively**



Step 1: Pre-trained G



Step 2: Fix G and X, Train Z

Without Encoder

- Limitation?

SLOW

A naïve way to speed up this method is to:

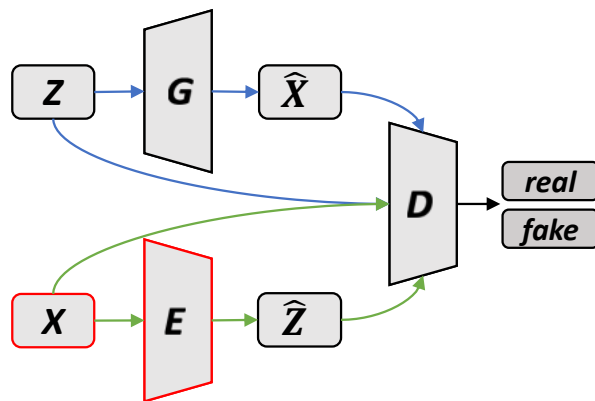
use one of the previous naïve way to pretrain an encoder, then

step 1: use the encoder to initialize the latent code z when given an image x

step 2: find the optimal z iteratively

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- **Recap: BiGAN**
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

Recap: Bidirectional GAN



BiGAN
Bidirectional GAN

$$\{X, \hat{Z}\} - \{\hat{X}, Z\}$$

Consider a BiGAN discriminator input pair (\mathbf{x}, \mathbf{z}) . Due to the sampling procedure, (\mathbf{x}, \mathbf{z}) must satisfy at least one of the following two properties:

$$(a) \mathbf{x} \in \hat{\Omega}_{\mathbf{X}} \wedge E(\mathbf{x}) = \mathbf{z}$$

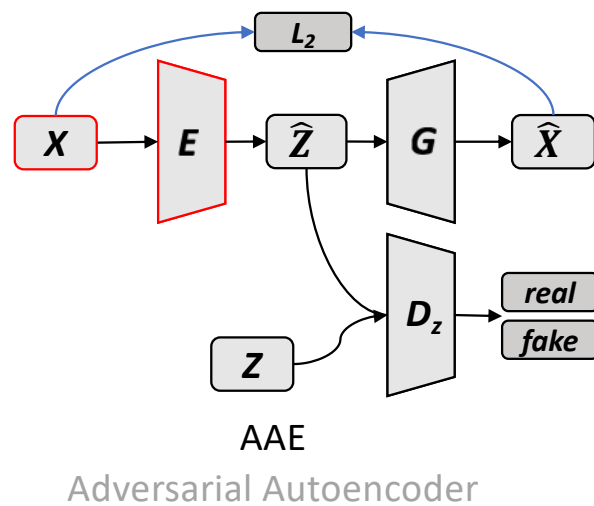
$$(b) \mathbf{z} \in \hat{\Omega}_{\mathbf{Z}} \wedge G(\mathbf{z}) = \mathbf{x}$$

If *only* one of these properties is satisfied, a perfect discriminator can infer the source of (\mathbf{x}, \mathbf{z}) with certainty: if only (a) is satisfied, (\mathbf{x}, \mathbf{z}) must be an encoder pair $(\mathbf{x}, E(\mathbf{x}))$ and $D_{EG}^*(\mathbf{x}, \mathbf{z}) = 1$; if only (b) is satisfied, (\mathbf{x}, \mathbf{z}) must be a generator pair $(G(\mathbf{z}), \mathbf{z})$ and $D_{EG}^*(\mathbf{x}, \mathbf{z}) = 0$.

Therefore, in order to fool a perfect discriminator at (\mathbf{x}, \mathbf{z}) (so that $0 < D_{EG}^*(\mathbf{x}, \mathbf{z}) < 1$), E and G must satisfy *both* (a) and (b). In this case, we can substitute the equality $E(\mathbf{x}) = \mathbf{z}$ required by (a) into the equality $G(\mathbf{z}) = \mathbf{x}$ required by (b), and vice versa, giving the inversion properties $\mathbf{x} = G(E(\mathbf{x}))$ and $\mathbf{z} = E(G(\mathbf{z}))$.

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- **Adversarial Autoencoder**
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

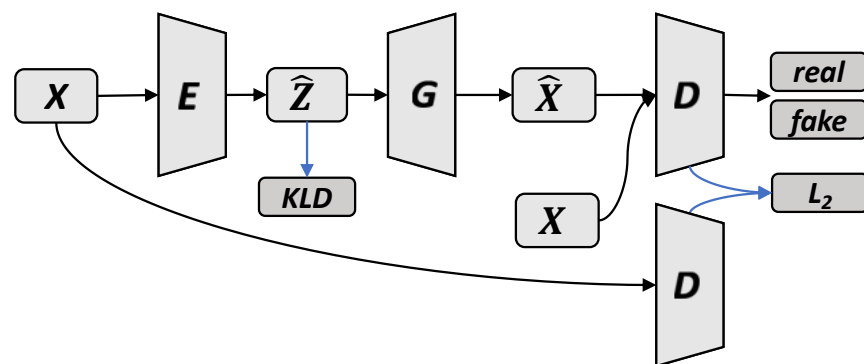
Adversarial Autoencoder



$$\{\hat{Z}\} - \{Z\}$$

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- **VAE+GAN**
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

VAE+GAN



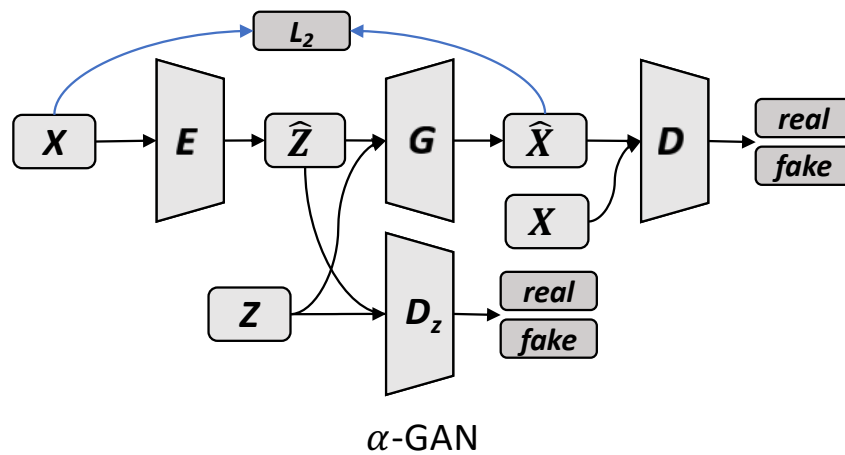
VAE+GAN

Discriminator as the feature extractor

$$\{\hat{X}\} - \{X\}$$

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- **α -GAN**
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

α -GAN



$$\{\hat{X}\} - \{X\}$$
$$\{\hat{Z}\} - \{Z\}$$

- Training the G and E in Autoencoder way can force the G to be able to generate all X, **avoiding GAN mode collapse**

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- **BigBiGAN**
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- Summary

BigBiGAN

- Work on large images
- Combine BigGAN and BiGAN

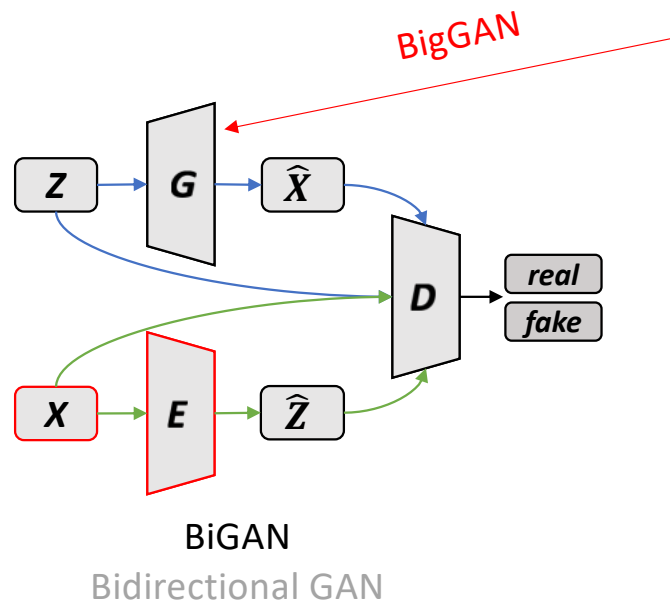


Figure 4: Samples from our BigGAN model with truncation threshold 0.5 (a-c) and an example of class leakage in a partially trained model (d).

BigBiGAN

- Limitation

image size of 512x512x3 → Latent code with size of 1x512

$$\frac{512}{512 \times 512 \times 3} = 0.000651$$

Difficult to be **lossless**

BigBiGAN

- Limitation



Figure 2: Selected reconstructions from an unsupervised BigBiGAN model (Section 3.3). Top row images are real data $\mathbf{x} \sim P_{\mathbf{x}}$; bottom row images are generated reconstructions of the above image \mathbf{x} computed by $\mathcal{G}(\mathcal{E}(\mathbf{x}))$. Unlike most explicit reconstruction costs (e.g., pixel-wise), the reconstruction cost implicitly minimized by a (Big)BiGAN [4, 7] tends to emphasize more semantic, high-level details. Additional reconstructions are presented in Appendix B.

BigBiGAN

- Limitation

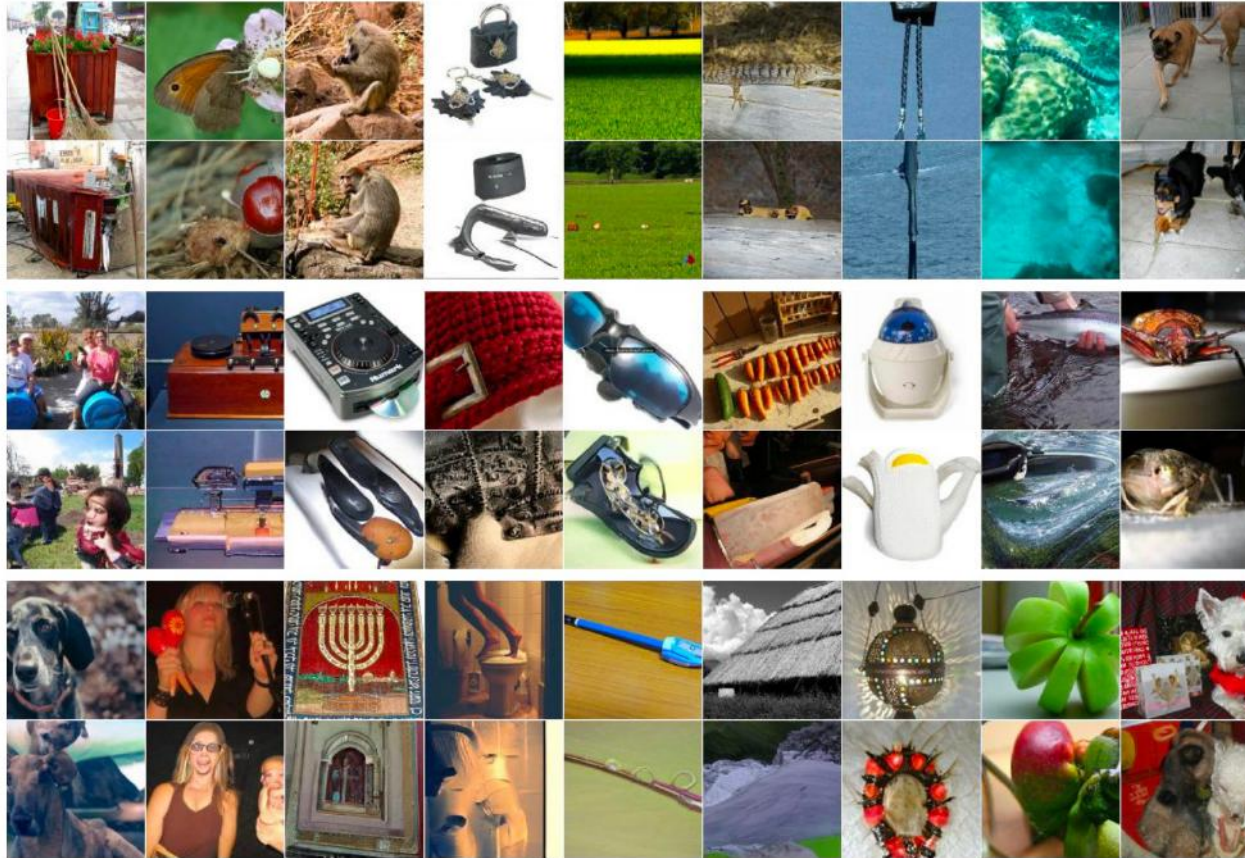


Figure 7: 128×128 reconstructions from an unsupervised BigBiGAN model, trained using the lighter augmentation from [24] with generation results reported in Table 3. The top rows of each pair are real data $\mathbf{x} \sim P_{\mathbf{x}}$, and bottom rows are generated reconstructions computed by $\mathcal{G}(\mathcal{E}(\mathbf{x}))$.

BigBiGAN

- Main Goal: Large Scale Adversarial Representation Learning

Metric	Top-1 / Top-5 Acc. (%)			
	$k = 1$	$k = 5$	$k = 25$	$k = 50$
D_1	38.09 / -	41.28 / 58.56	43.32 / 65.12	42.73 / 66.22
D_2	35.68 / -	38.61 / 55.59	40.65 / 62.23	40.15 / 63.42

Table 6: Accuracy of k nearest neighbors classifiers in BigBiGAN feature space on the ImageNet validation set. We report results under the normalized ℓ_1 distance D_1 as well as the normalized ℓ_2 (cosine) distance D_2 .

BigBiGAN

- Main Goal: Large Scale Adversarial Representation Learning

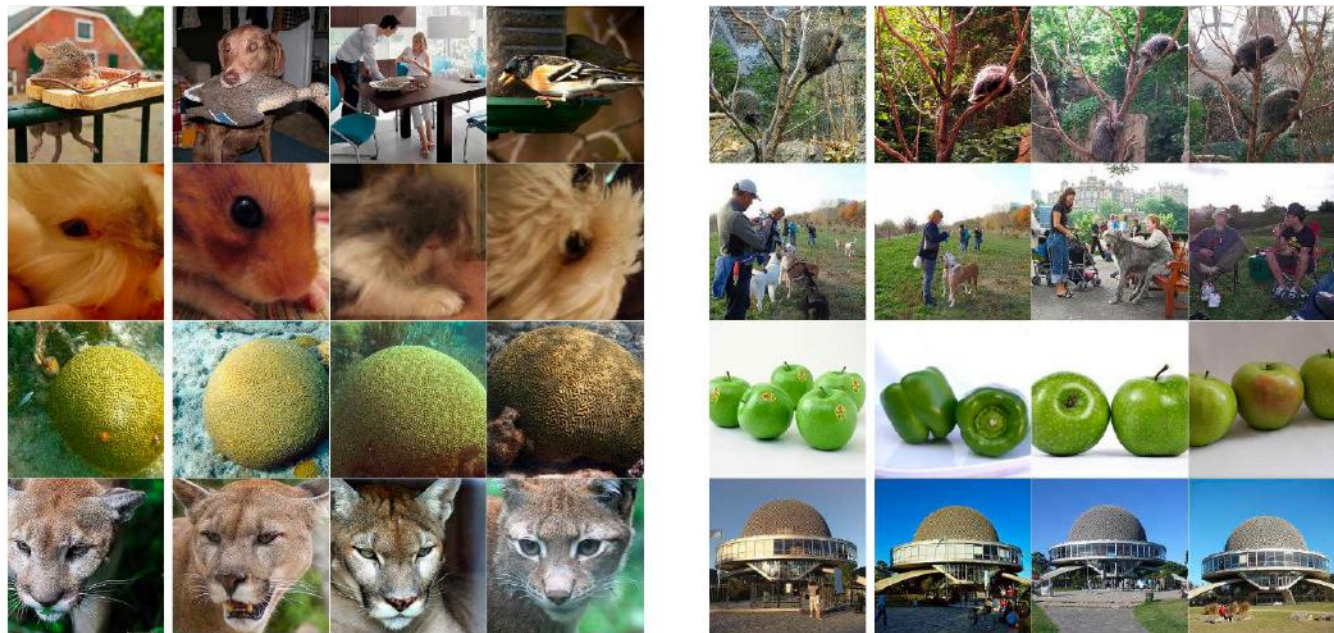


Figure 12: Nearest neighbors in BigBiGAN \mathcal{E} feature space, from our best performing model (*RevNet* $\times 4$, $\uparrow \mathcal{E} LR$). In each row, the first (left) column is a query image, and the remaining columns are its three nearest neighbors from the training set (the leftmost being the nearest, next being the second nearest, etc.). The query images above are the first 24 images in the ImageNet validation set.

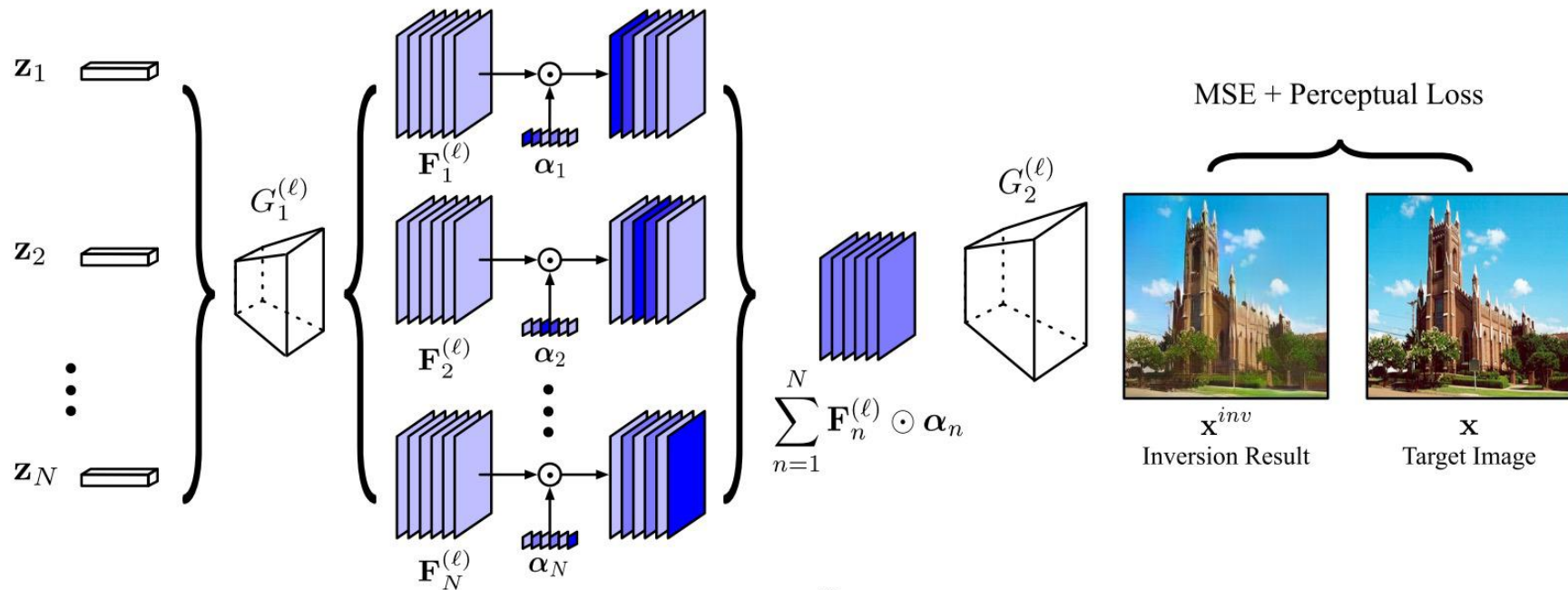
BigBiGAN

- Summary
 - A single latent code cannot represent a high-resolution image
 - Other information inside the generator
 - High compression rate
 - Next: any solutions?

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- **Multi-code GAN prior**
- Implicit vs. Explicit Encoder
- Summary

Multi-code GAN prior

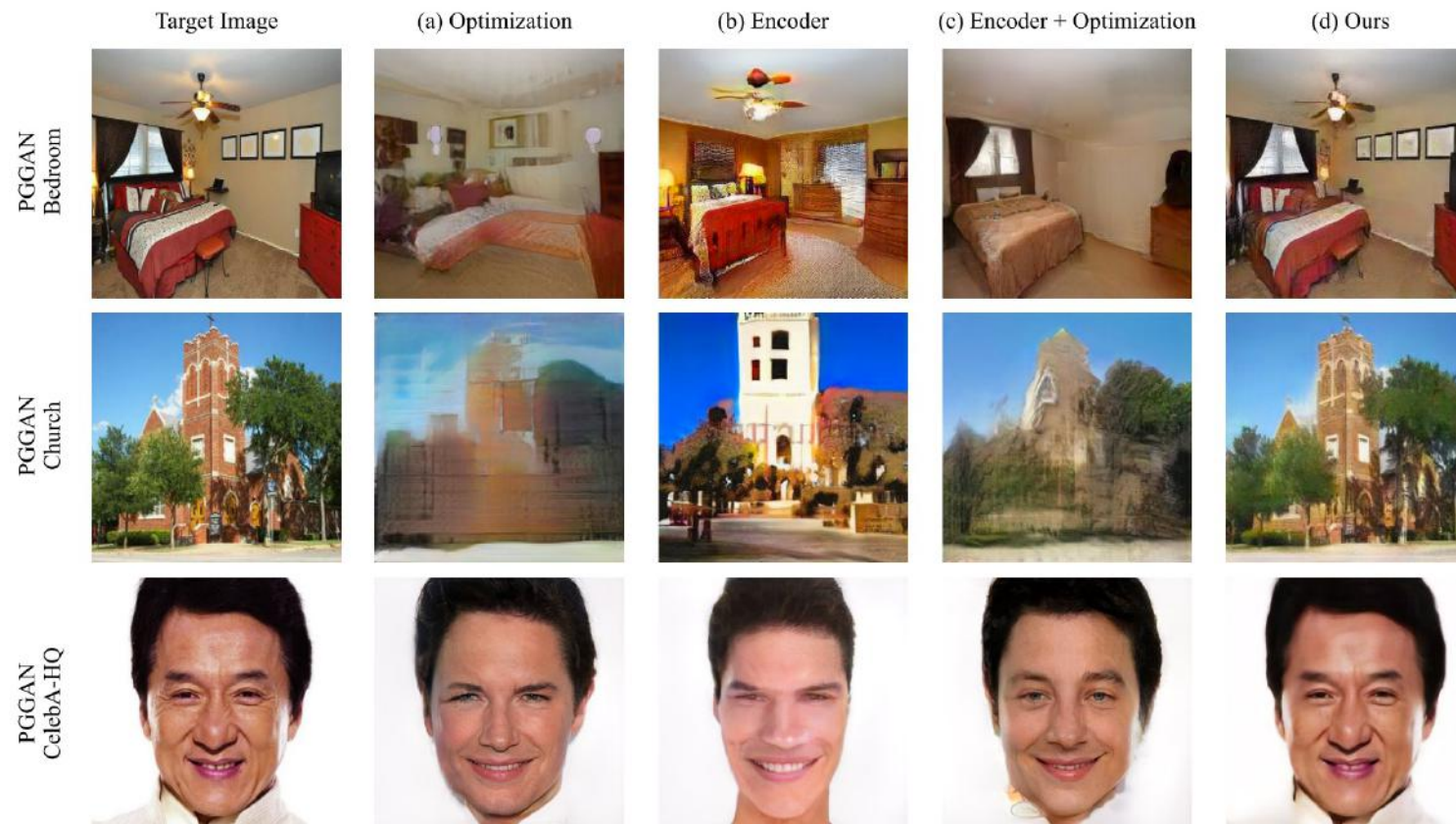
- An Optimisation-based Method



A single latent code is not enough to recover all detailed information.
We can use multiple latent codes to recover different feature maps.

Multi-code GAN prior

- Reconstruction



Multi-code GAN prior

- Inpainting



Multi-code GAN prior

- More



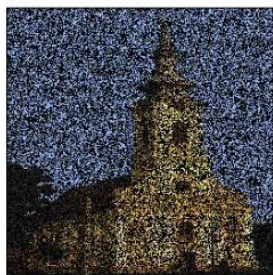
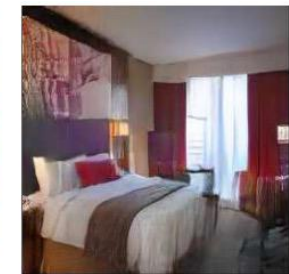
(a) Image Reconstruction



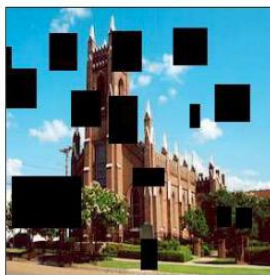
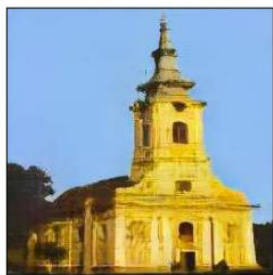
(b) Image Colorization



(c) Image Super-Resolution



(d) Image Denoising



(e) Image Inpainting



(f) Semantic Manipulation

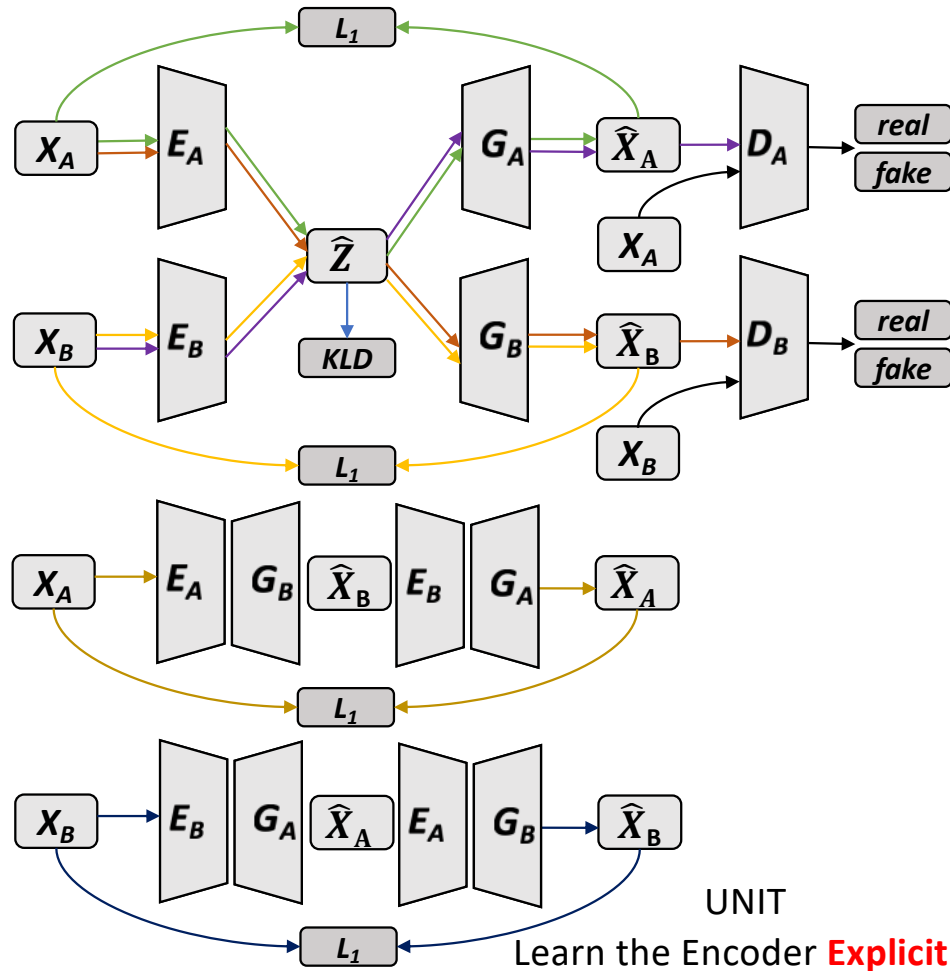


Multi-code GAN prior

- Discussion
 - Why it works?
 - Limitations?

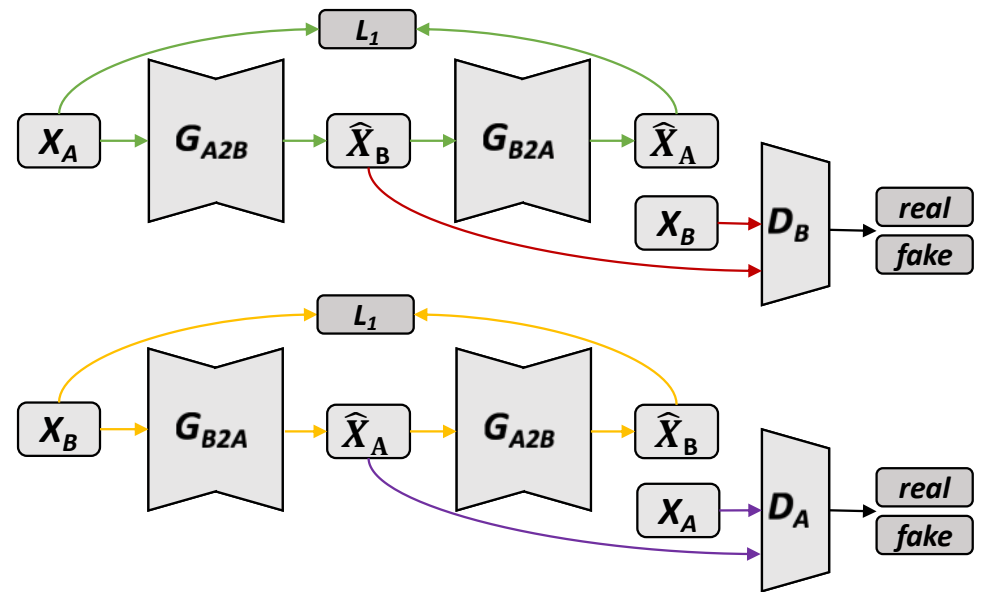
- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- **Implicit vs. Explicit Encoder**
- Summary

Implicit vs. Explicit Encoder



Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*

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



CycleGAN

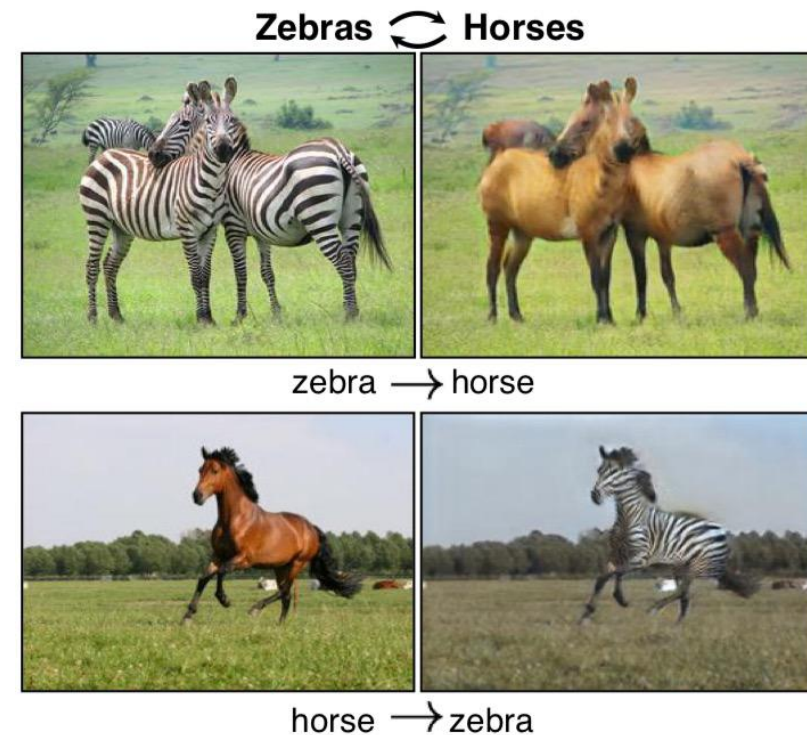
Learn the Encoder **Implicitly**

Implicit vs. Explicit Encoder



Liu et al.

Learn the Encoder **Explicitly**



CycleGAN

Learn the Encoder **Implicitly**

Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*

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

Implicit vs. Explicit Encoder



Input GTA5 CG

<https://blog.csdn.net/gdymind>



Output image with German street view style

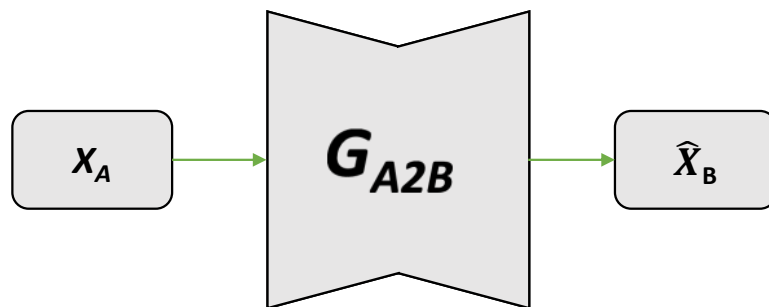
blog.csdn.net/gdymind

Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*

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

Implicit vs. Explicit Encoder

- Simple normal distribution is difficult to model complex images
- 3D tensors can contain more spatial information than vectors
- Many applications do not need interpolation



- Image inpainting
- Image super resolution
- Image-to-image translation
-

- VAE vs. GAN
- A Naïve Approach
- Another Naïve Approach
- Without Encoder
- Recap: BiGAN
- Adversarial Autoencoder
- VAE+GAN
- α -GAN
- BigBiGAN
- Multi-code GAN prior
- Implicit vs. Explicit Encoder
- **Summary**

Summary

- GAN : $G + D \rightarrow G + D + E$
- Learning E from real data is important
- GAN mode collapse
- BiGAN, AAE, VAE+GAN, α -GAN, BigBiGAN
- Autoencoder can help to avoid mode collapse
- Learning E implicitly
- The E can be extended to text and any other data type
- Still on the way ...

Thanks