Adversarially Learned Inference (ALI)

김경환

모두의연구소

Contents

- Introduction of ALI
- Experimental Results

모두의연구소

Introduction of ALI

모두의연구소

Background

- VAE
  - Learn an approximate inference mechanism
  - Sample from VAE is blurry

- GAN
  - Bypass inference altogether
  - Higher-quality samples than VAE
  - Lack an efficient inference mechanism

모두의연구소

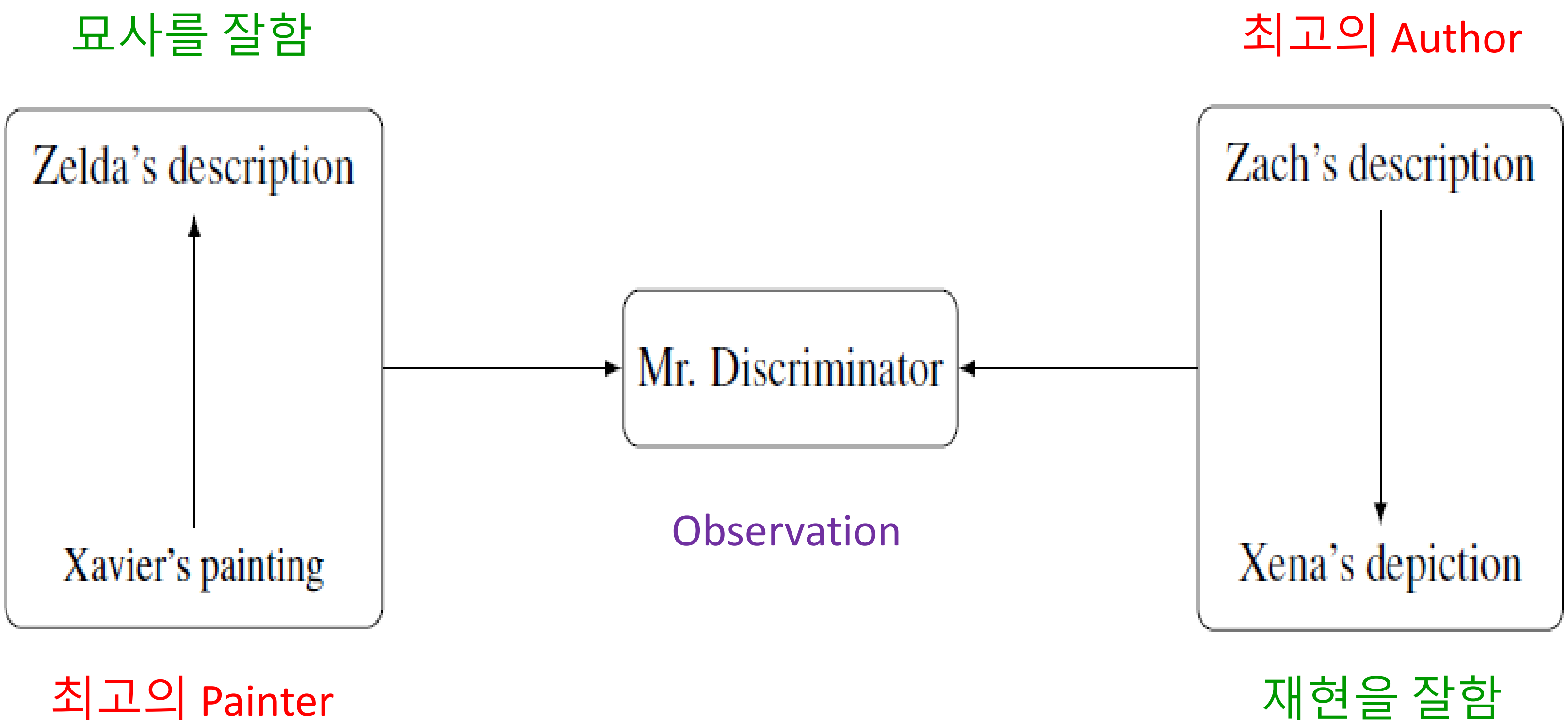
Recently Goal

→ *To bridge the gap between VAEs and GANs*

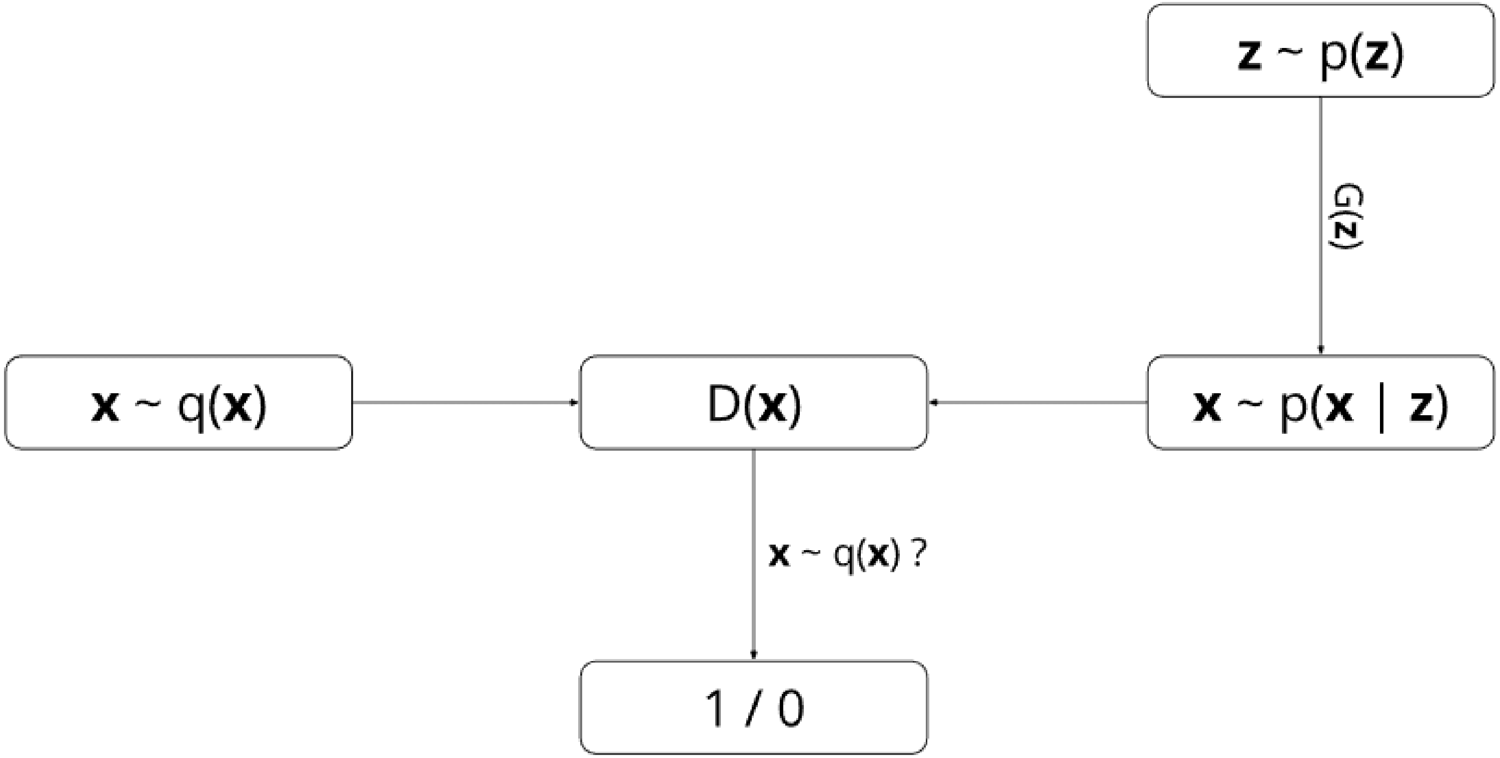
→ *To learn generative models with higher-quality samples learning an efficient inference network*

모두의연구소

Story - A Circle of Infinite Painters

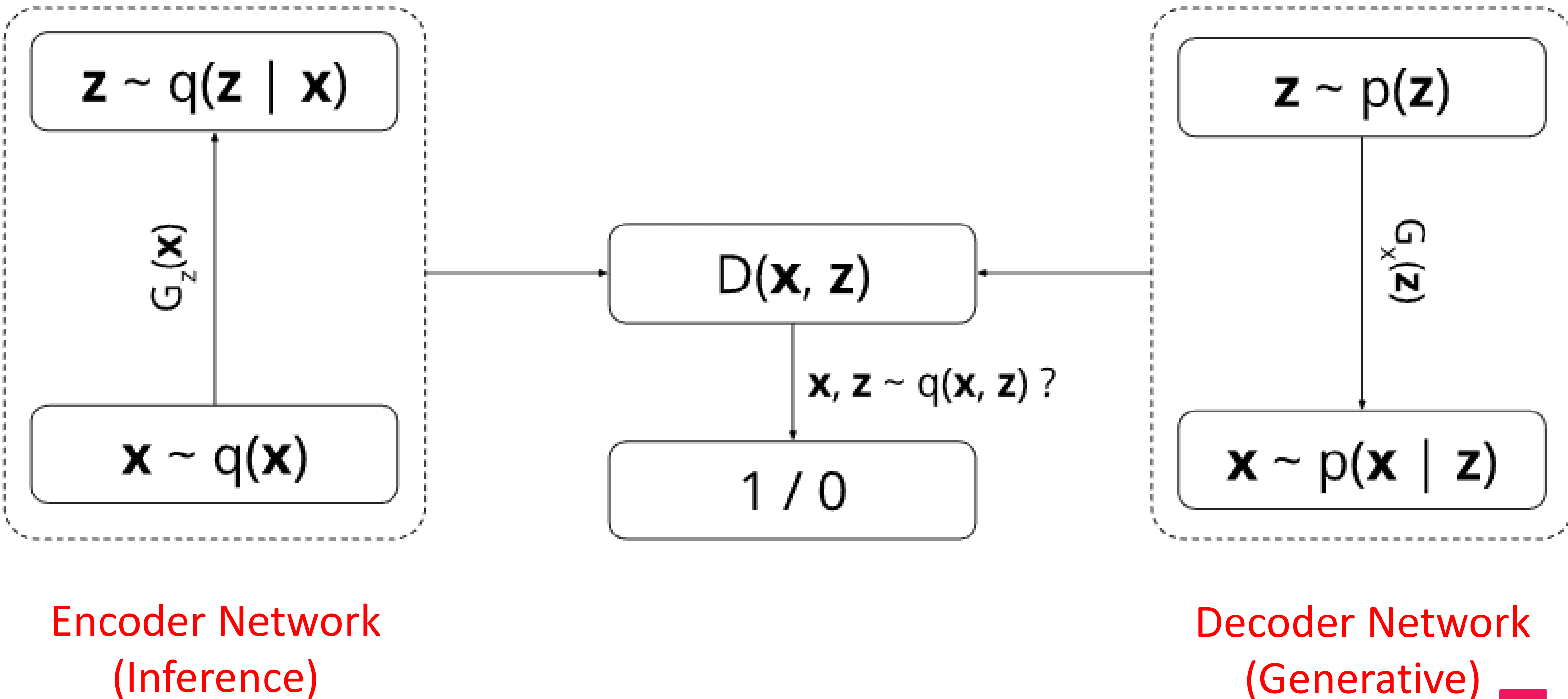


GAN and ALI



< GAN >

GAN and ALI



< ALI >

모두의연구소

Value Function

< GAN >

$$\min_G \max_D V(D,G) = \mathbb{E}_{q(\mathbf{x})}[\log(D(\mathbf{x}))] + \mathbb{E}_{p(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]$$
$$= \int q(\mathbf{x}) \log(D(\mathbf{x})) d\mathbf{x} + \iint p(\mathbf{z}) p(\mathbf{x} \mid \mathbf{z}) \log(1 - D(\mathbf{x})) d\mathbf{x} d\mathbf{z}$$

$\downarrow$

< ALI >

$$\min_G \max_D V(D,G) = \mathbb{E}_{q(\boldsymbol{x})}[\log(D(\boldsymbol{x}, G_z(\boldsymbol{x})))] + \mathbb{E}_{p(\boldsymbol{z})}[\log(1 - D(G_x(\boldsymbol{z}), \boldsymbol{z}))]$$
$$= \iint q(\boldsymbol{x}) q(\boldsymbol{z} \mid \boldsymbol{x}) \log(D(\boldsymbol{x}, \boldsymbol{z})) d\boldsymbol{x} d\boldsymbol{z}$$
$$+ \iint p(\boldsymbol{z}) p(\boldsymbol{x} \mid \boldsymbol{z}) \log(1 - D(\boldsymbol{x}, \boldsymbol{z})) d\boldsymbol{x} d\boldsymbol{z}.$$

Generator Value Function

: *when D gets too far ahead,*
it is too hard to minimize the value function of G

< GAN >

$$\log(1 - \left(D\left(G(z)\right)\right) \quad \rightarrow \quad \text{minimize}$$

$$\log\left(D\left(G(z)\right)\right) \quad \rightarrow \quad \text{maximize}$$

< ALI >

$$V'(D, G) = \mathbb{E}_{q(\boldsymbol{x})}[\log(1 - D(\boldsymbol{x}, G_{\boldsymbol{z}}(\boldsymbol{x})))] + \mathbb{E}_{p(\boldsymbol{z})}[\log(D(G_{\boldsymbol{x}}(\boldsymbol{z}), \boldsymbol{z}))]$$

Value Function

< Discriminator >

$$D(x, E(x)) \rightarrow 1, \qquad D(G(z),\ z) \rightarrow 0$$

< Generator >

$$D(x, E(x)) \rightarrow 0, \qquad D(G(z),\ z) \rightarrow 1$$

Proposition

1. Given a fixed generator G, the optimal discriminator is given by

$$D^*(x,z) = \frac{q(x,z)}{q(x,z) + p(x,z)}$$

2. Under an optimal discriminator D, the generator minimizes the Jensen-Shanon divergence which attains its minimum if and only if q(x; z) = p(x; z).

→ The proof in *GAN(Goodfellow et al. 2014).*

모두의연구소

Discriminator Optimality

< GAN >

$$D^*(z) = \frac{q(x)}{q(x) + p(z)}$$

< ALI >

$$D^*(x,z) = \frac{q(x,z)}{q(x,z) + p(x,z)}$$

모두의연구소

Invertibility

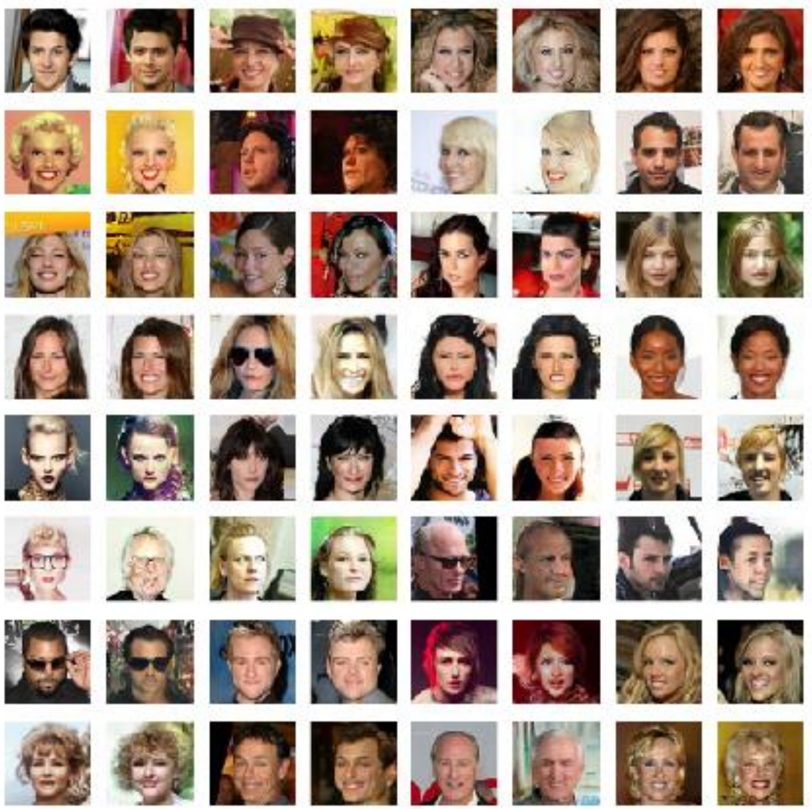
3. Assuming optimal discriminator D and generator G. If the encoder Gx is deterministic, then $G_X = G_Z^{-1}$ *and* $G_z = G_x^{-1}$ *almost everywhere.*

→ The proof in *BiGAN(Donahue et al., 2016)*

Experimental Results

Sample and Reconstructions

< SVHN >

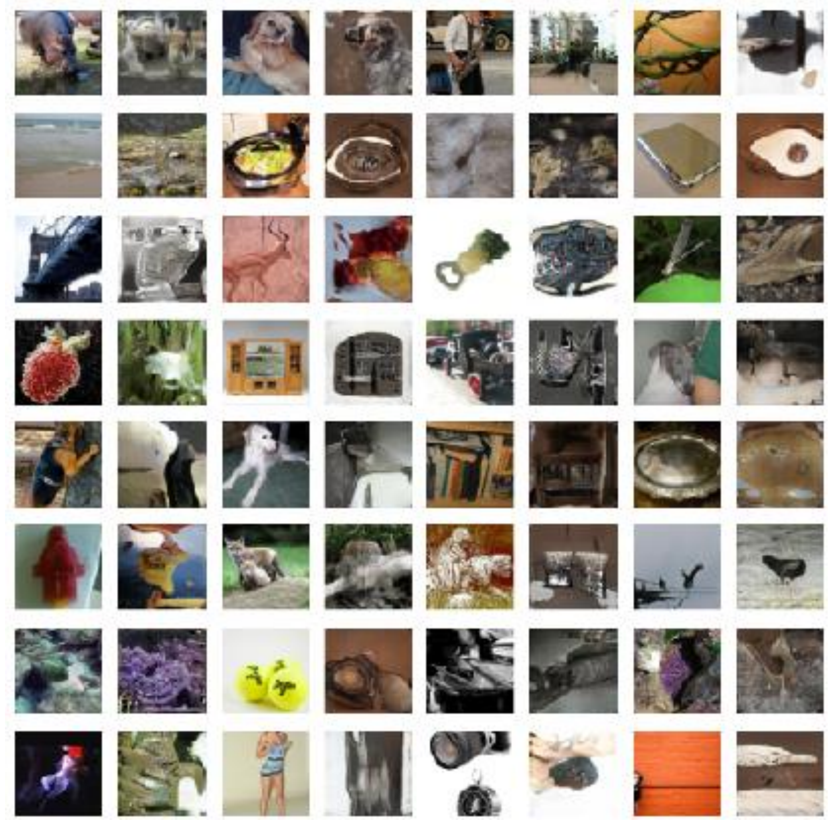


< CelebA >

모두의연구소

Sample and Reconstructions

< CIFAR 10 >



< ImageNet >

모두의연구소

Sample and Reconstructions

- To observe that reconstructions are not always faithful reproductions of the inputs

- a more complex input distribution, the model exhibits less faithful reconstructions

  → poor reconstructions are a sign of underfitting

Latent space Interpolations

→ ALI is not overfitting learning

→ it rather has learned latent features that generalize well

모두의연구소

Semi-Supervised Learning

Table 1: SVHN test set missclassification rate

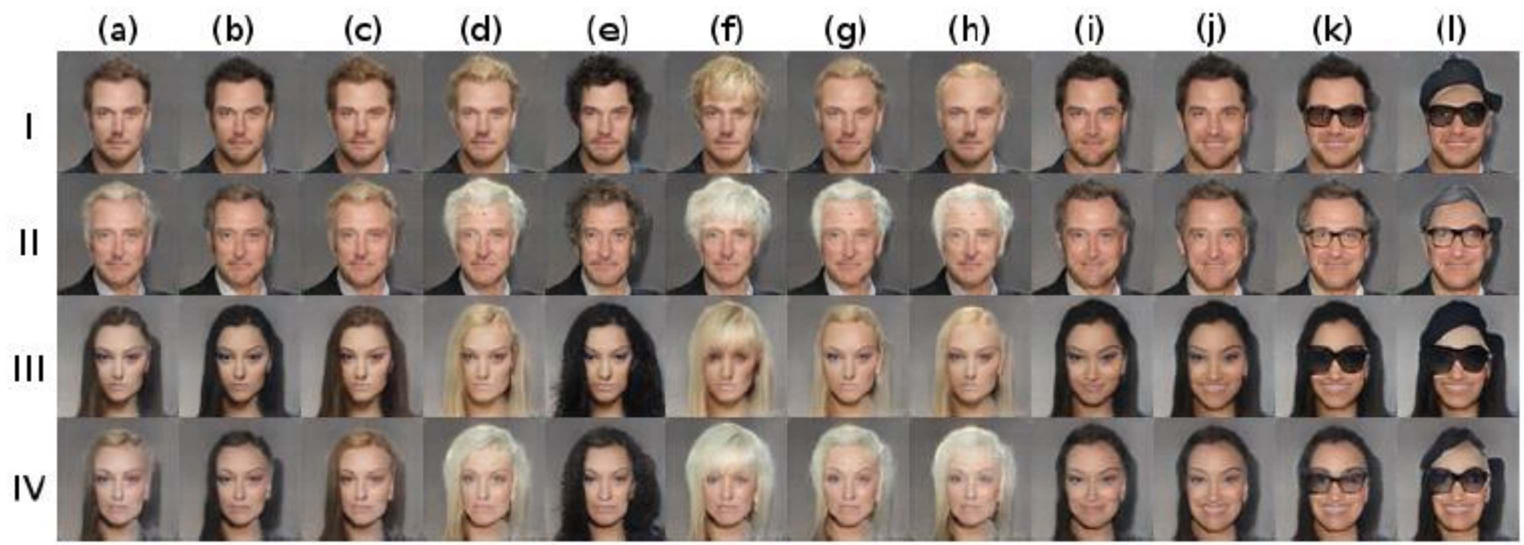
Model	Misclassification rate
VAE (M1 + M2) (Kingma et al., 2014)	36.02
SWWAE with dropout (Zhao et al., 2015)	23.56
DCGAN + L2-SVM (Radford et al., 2015)	22.18
SDGM (Maaløe et al., 2016)	16.61
GAN (feature matching) (Salimans et al., 2016)	**8.11 ± 1.3**
ALI (ours, L2-SVM)	19.14 ± 0.50
ALI (ours, no feature matching)	**7.42 ± 0.65**

Table 2: CIFAR10 test set missclassification rate for semi-supervised learning using different numbers of trained labeled examples. For ALI, error bars correspond to 3 times the standard deviation.

Number of labeled examples	1000	2000	4000	8000
Model		Misclassification rate		
Ladder network (Rasmus et al., 2015)			20.40	
CatGAN (Springenberg, 2015)			19.58	
GAN (feature matching) (Salimans et al., 2016)	**21.83 ± 2.01**	**19.61 ± 2.09**	**18.63 ± 2.32**	**17.72 ± 1.82**
ALI (ours, no feature matching)	**19.98 ± 0.89**	**19.09 ± 0.44**	**17.99 ± 1.62**	**17.05 ± 1.49**

Conditional Generation

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{q(\boldsymbol{x})\,p(\boldsymbol{y})}[\log(D(x, G_z(x,y), y))] + \mathbb{E}_{p(\boldsymbol{z})\,p(\boldsymbol{y})}[\log(1 - D(G_x(z,y), z, y))]$$



Importance of Learning Inference Jointly with Generation

- 2D gaussian 25 mixture components laid out on a grid
- 100,000 q(x) samples
- Measuring the extends that the trained models covered all 25 modes by drawing 10,000 samples from their p(x) distribution
- and assigning each sample to a q(x) mixture component according to the mixture responsibilities.
- ALI models covered 13.4 ± 5.8 modes on average (min: 8, max:25)
- GAN models covered 10.4 ± 9.2 modes on average (min: 1, max: 22)

모두의연구소

Alternative Approaches to Inference in GANs

- InfoGAN
- The inverse mapping from GAN samples
- Post-hoc learned inference

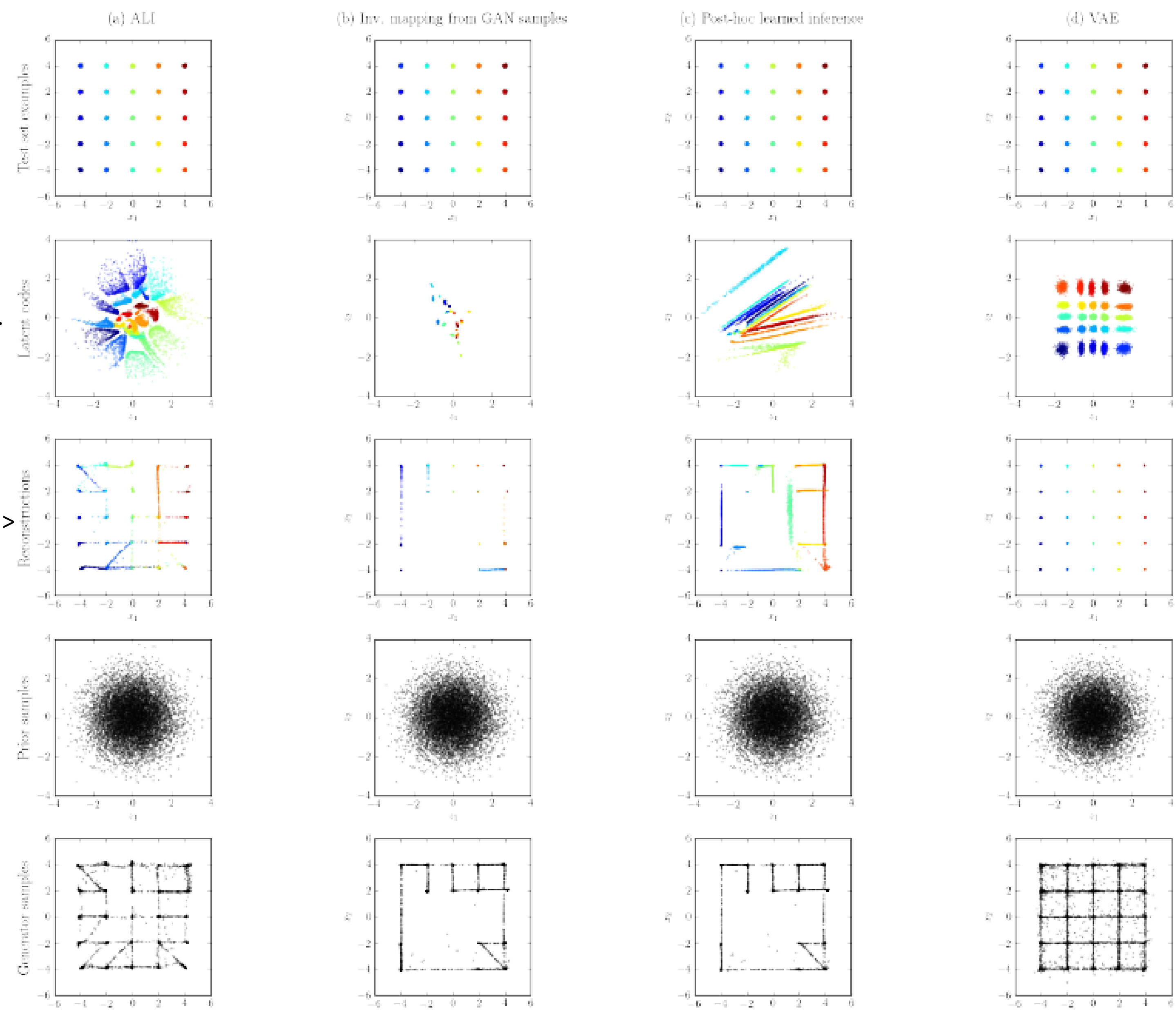
< Test set >

< Latent codes >
Q(z)

< Reconstruction >

< Z >
P(z)

< Generator samples >



< ALI >

< GAN >
(inverse mapping)

< GAN >
(post-hoc learned)

< VAE >

Importance of Learning Inference Jointly with Generation

- The ALI encoder models q(z) that matched p(z) fairly well
- GAN's generator has more trouble reaching all the modes than ALI generator
- Learning an inverse mapping from GAN samples does not work very well : the generator dropping modes
- Learning inference post-hoc does not work as well : negatively impacts how the latent space
- VAE covers all modes easily, but they have tendency to smear out probably density and leave holes in latent space

Reference

- https://ishmaelbelghazi.github.io/ALI/
- http://jaejunyoo.blogspot.com/2017/01/generative-adversarial-nets-1.html

모두의연구소