Improved GAN and Applications

ISTD 50.035

Computer Vision

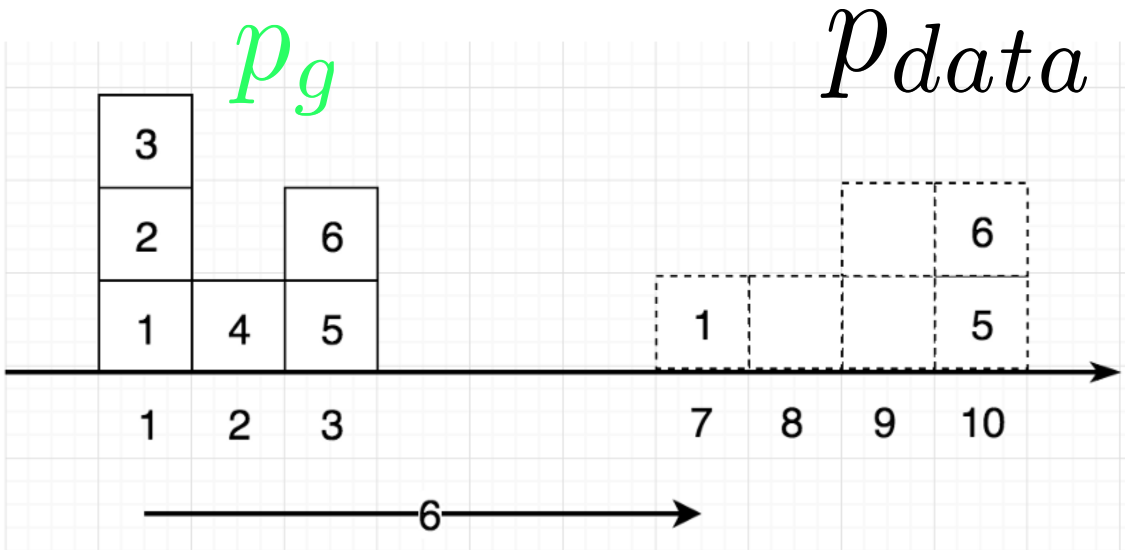
Acknowledgement: Some images are from various sources: UCF, Stanford cs231n, National Taiwan University, etc.

Earth-Mover distance / Wasserstein distance

Cost of transport plan to move the distribution from one to the other: weight times the distance

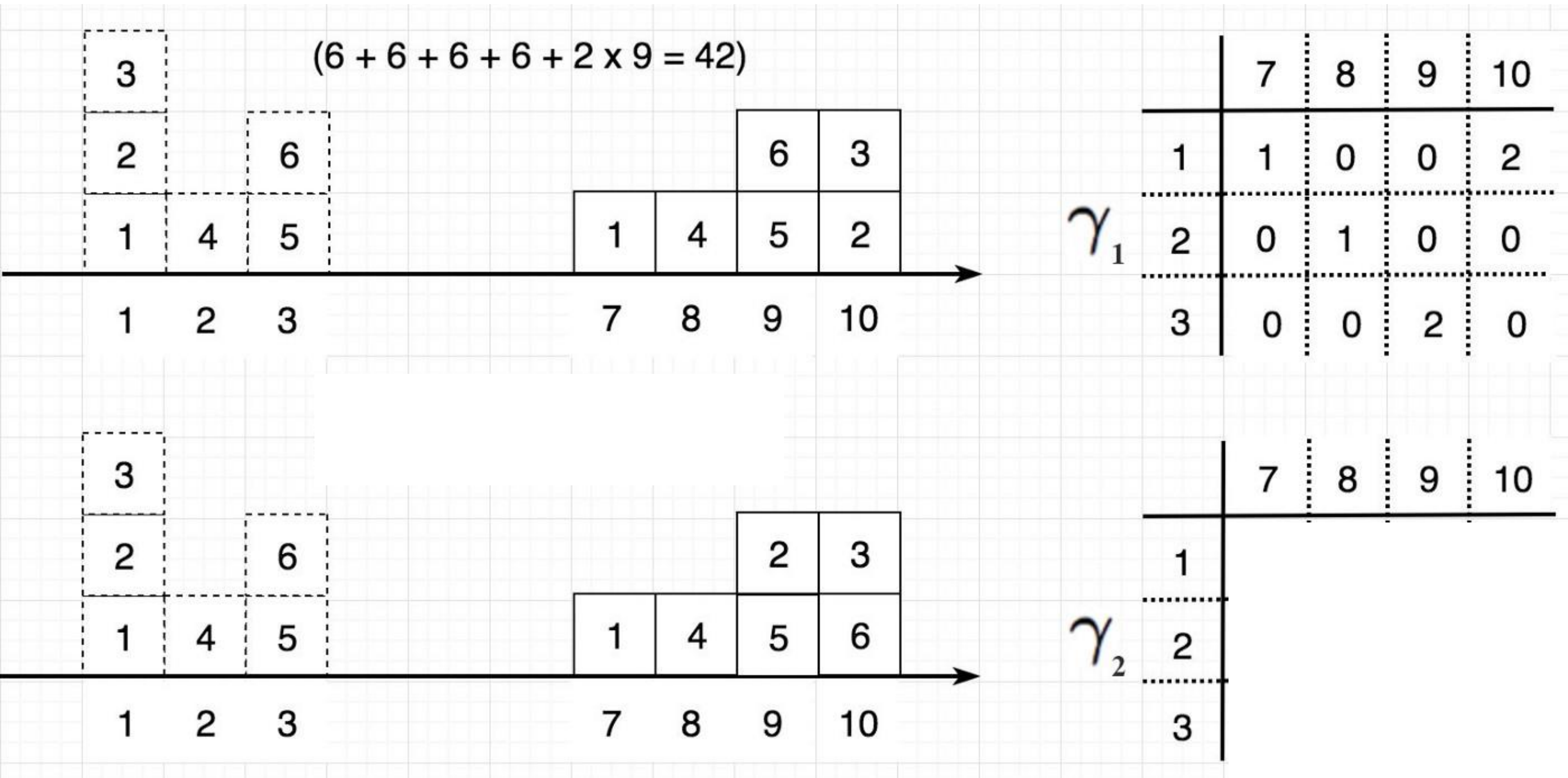
EM distance: cheapest transport plan to move the distribution from one to the other

Earth-Mover distance / Wasserstein distance

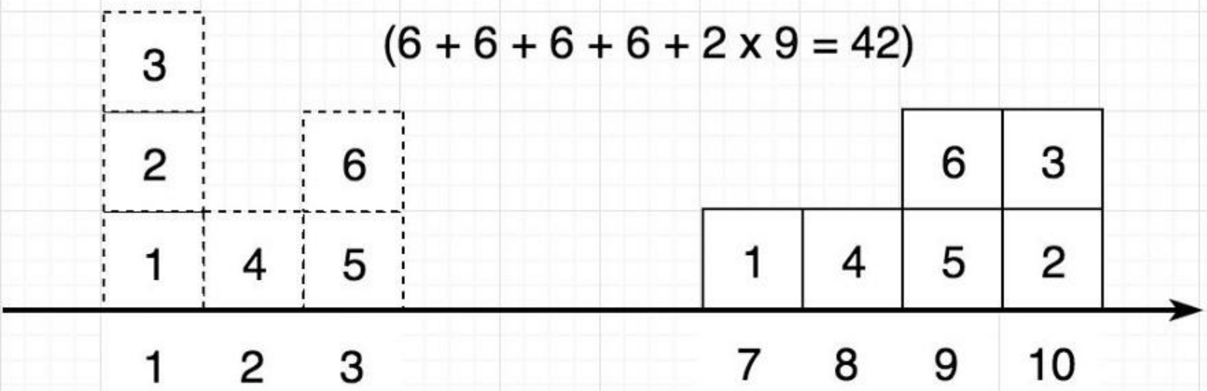


Many different ways to move
EM dist: the way with the minimum cost

Earth-Mover distance / Wasserstein distance

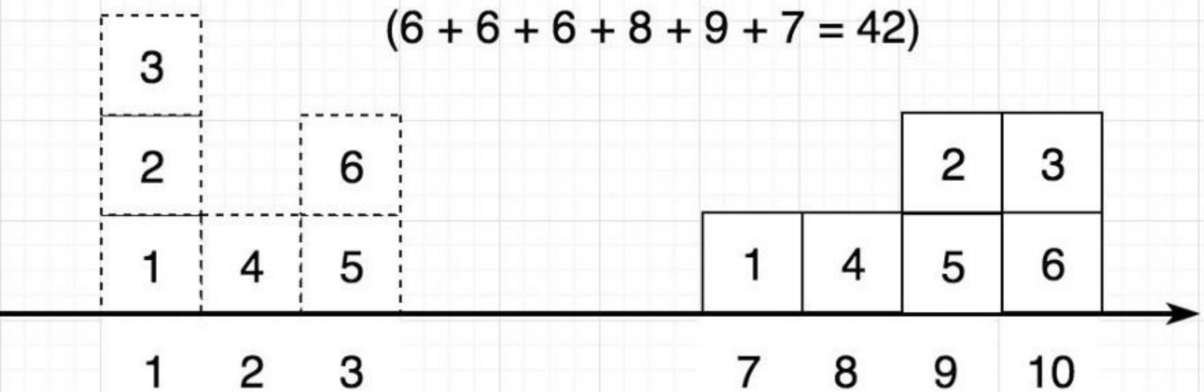


Earth-Mover distance / Wasserstein distance



γ_1

	7	8	9	10
1	1	0	0	2
2	0	1	0	0
3	0	0	2	0



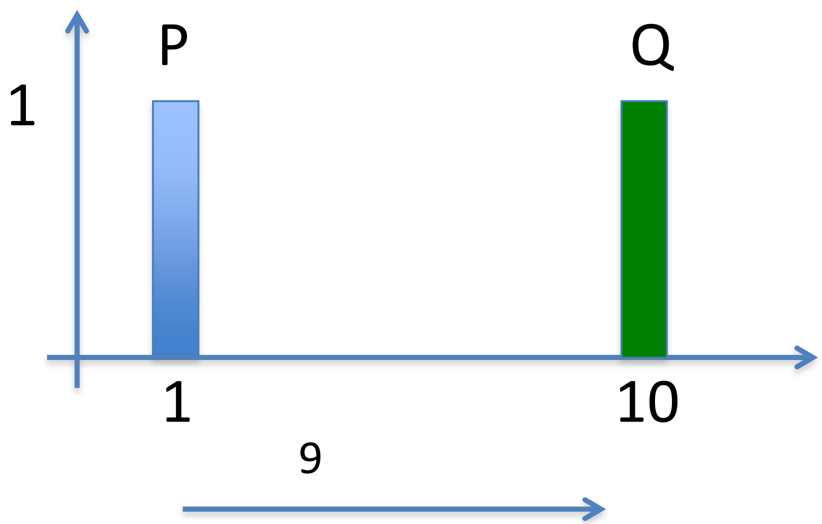
γ_2

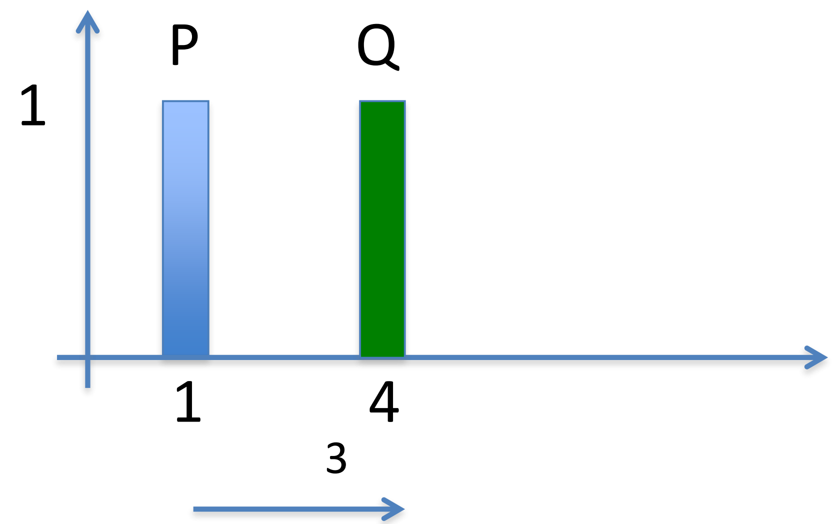
	7	8	9	10
1	1	0	1	1
2	0	1	0	0
3	0	0	1	1

[Martin Arjovsky et al. 2017]

WGAN: use Wasserstein distance

Evaluate Wasserstein distance between p_{data} and p_G:



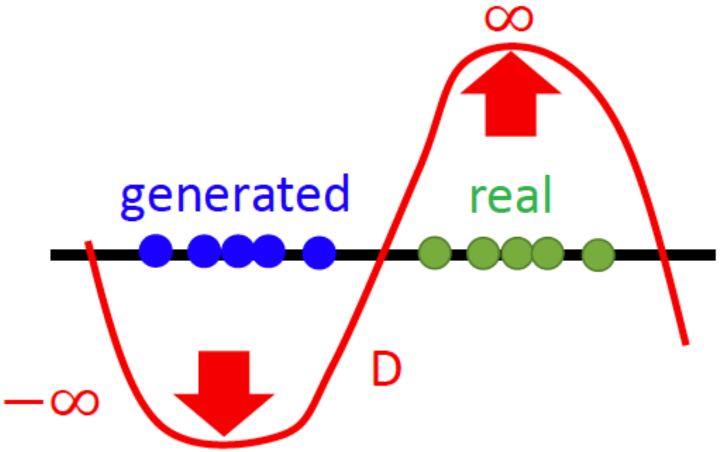


WGAN: use Wasserstein distance

Evaluate Wasserstein distance between p_{data} and p_G:

$$V(G,D) = \max_{D \in 1-Lipschitz} \{E_{x \sim P_{data}}[D(x)] - E_{x \sim P_G}[D(x)]\}$$

D has to be sufficiently smooth

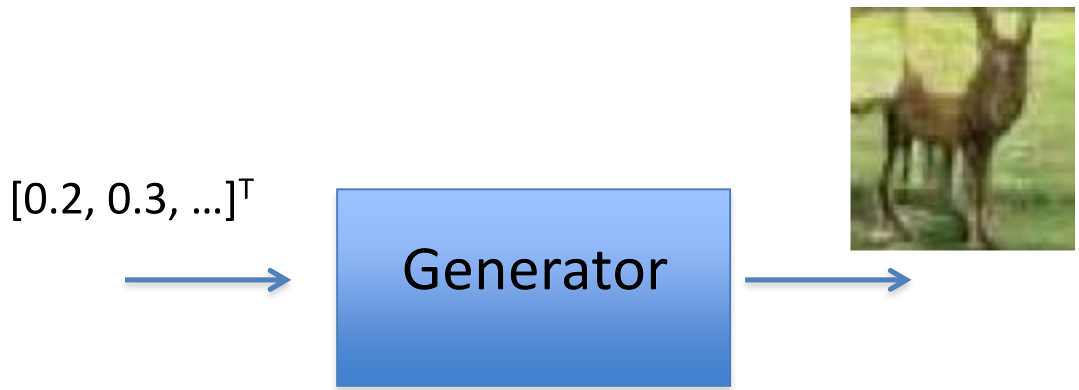


Different ways to handle this constraint: WGAN-GP, etc.

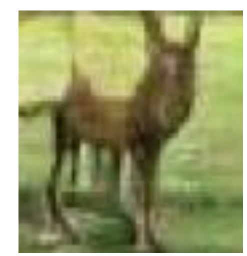
[Scott Reed et al. 2016]

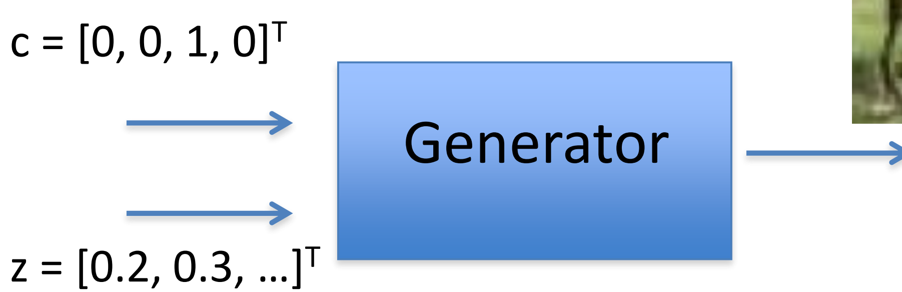
Conditional GAN

- Original GAN: no control over modes of the generated data
- CGAN: Add the condition to generate corresponding images
- CGAN requires labeled data in training



1-hot vector to encode class / condition





Learn this through the discriminator

Conditional GAN

1-hot vector to encode class / condition

$c = [0, 0, 1, 0]^T$

$z = [0.2, 0.3, \ldots]^T$

Generator

Discriminator

Scalar D: Image is real/fake

Conditional GAN

1-hot vector to encode class / condition

$c = [0, 0, 1, 0]^T$

$z = [0.2, 0.3, \ldots]^T$

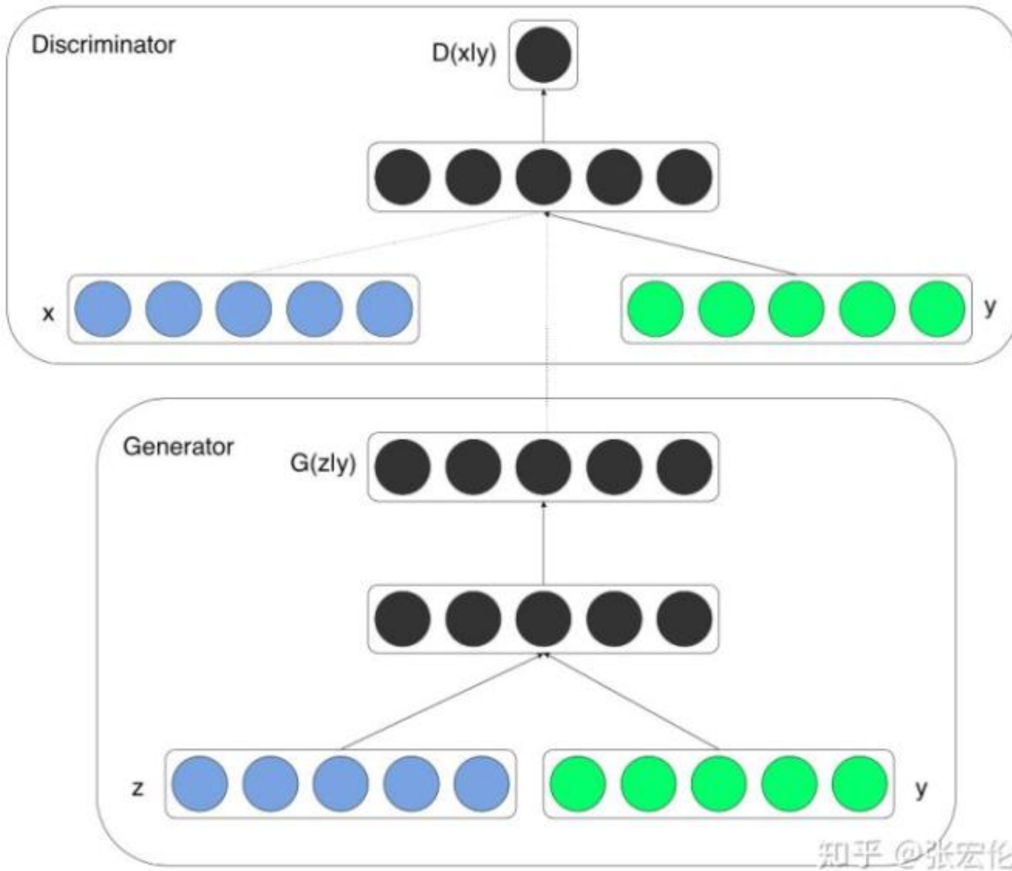
Generator

$c = [0, 0, 1, 0]^T$

Discriminator

Scalar D: Image is real/fake

Conditional GAN



$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x|y)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z|y)))]$$

[Chen et al. 2016]

InfoGAN

- InfoGAN: unsupervised

- Learn interpretable, **disentangled** representation
  - a separate set of dimensions (in the input vector) for each of the attributes
  - e.g. face dataset: eye color, hairstyle, presence or absence of eyeglasses, etc.
  - e.g. Digit dataset: 10-state discrete variable

- Discover visual concepts

InfoGAN

c: Structured latent variables

- associate with some attributes
- learn these association in an unsupervised way
- modify generator objective, discriminator same as before

$c = [c1, c2, \ldots]^T$

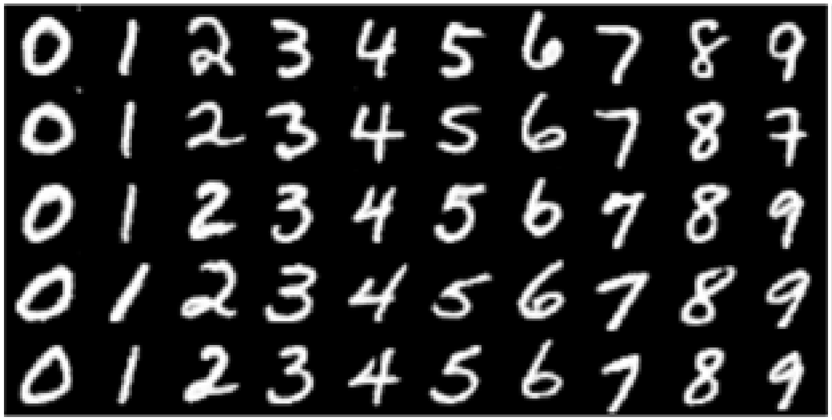
$z = [0.2, 0.3, \ldots]^T$

Generator

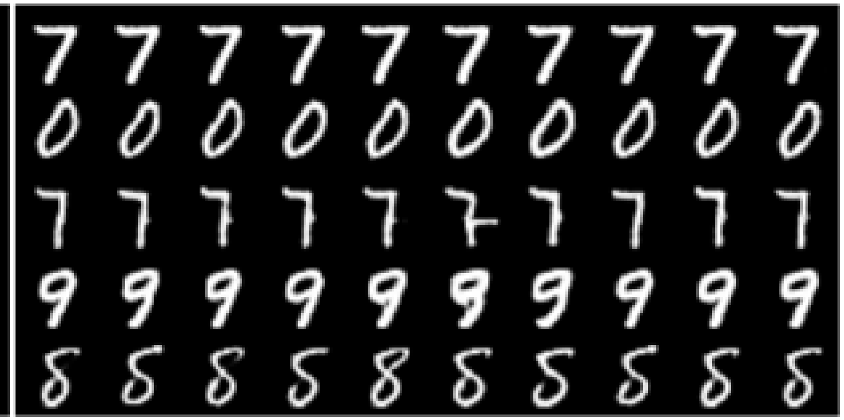
Discriminator

real/fake

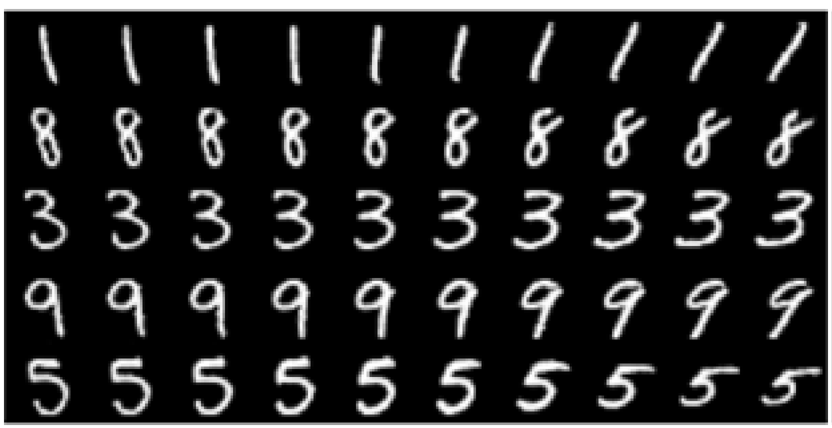
InfoGAN



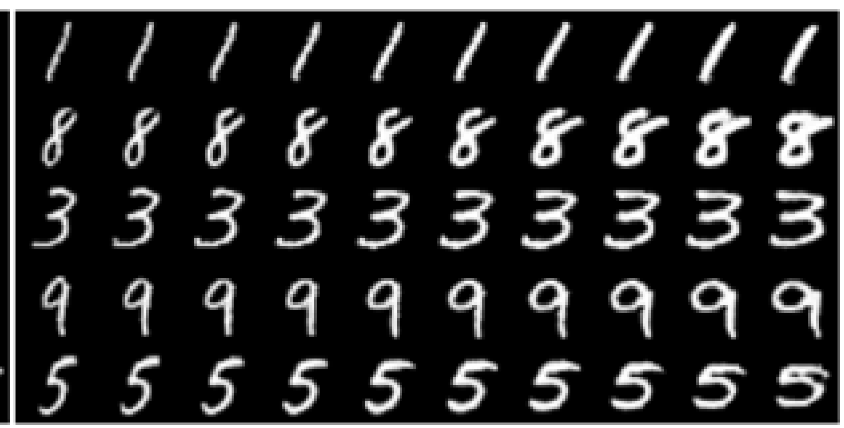
(a) Varying c_1 on InfoGAN (Digit type)



(b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)



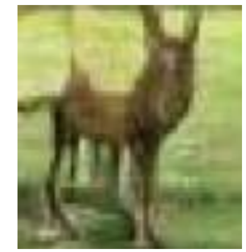
(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

InfoGAN

$c = [c1, c2, \ldots]^T$

Generator

$z = [0.2, 0.3, \ldots]^T$



Idea: promote association between c and G(z,c)

$$V(D,G) = \mathbb{E}_{x\sim P_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim\text{noise}}[\log(1 - D(G(z)))]$$

Same as before

Mutual information

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

Mutual Information and conditional entropy

- MI of two random variables: mutual dependence between two variables
- Measures the information that X and Y share
- If X and Y are independent, $I(X;Y) = 0$

$$\mathrm{I}(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

p(x,y): joint probability mass function
p(x), p(y): marginal pmf

$$\mathrm{I}(X;Y) = \mathrm{H}(Y) - \mathrm{H}(Y|X)$$

(Marginal) entropy

Conditional entropy:
information/uncertainty re. Y

$$\mathrm{H}(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}$$

If X and Y are independent, H(Y|X) = H(Y), I(X;Y) = 0

Mutual Information and conditional entropy

H(dog| ): ?

H(dog| ): ?

Conditional entropy: information/uncertainty re. Y given X

If X and Y are independent, H(Y|X) = H(Y), I(X;Y) = 0

Mutual Information and conditional entropy

H(dog| ): small

H(dog| ): large

Conditional entropy: information/uncertainty re. Y given X

If X and Y are independent, $H(Y|X) = H(Y)$, $I(X;Y) = 0$

InfoGAN

(a) Azimuth (pose)

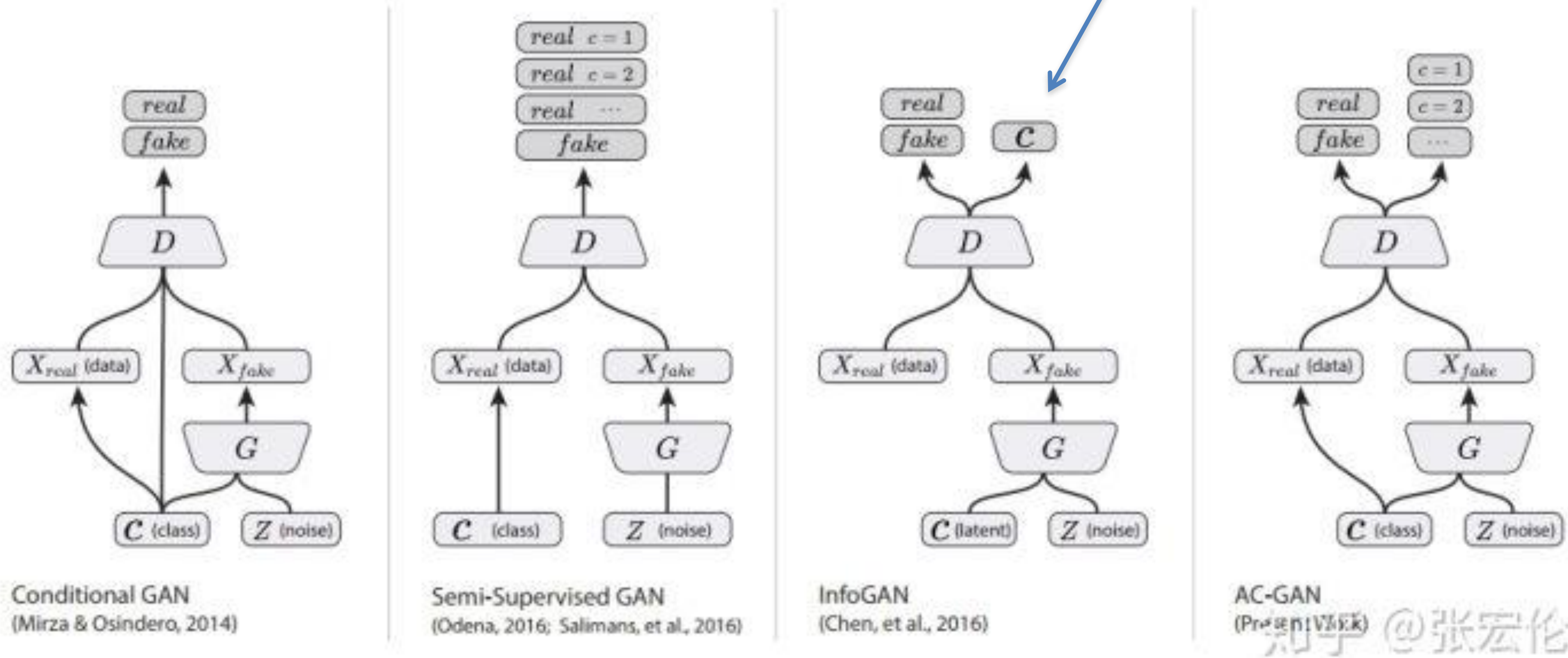
(b) Presence or absence of glasses

(c) Hair style

(d) Emotion

InfoGAN

Need Q(c|x) to maximize I(c; G(z,c))



$$V(D,G) = \mathbb{E}_{x\sim P_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim \text{noise}}[\log(1 - D(G(z)))]$$

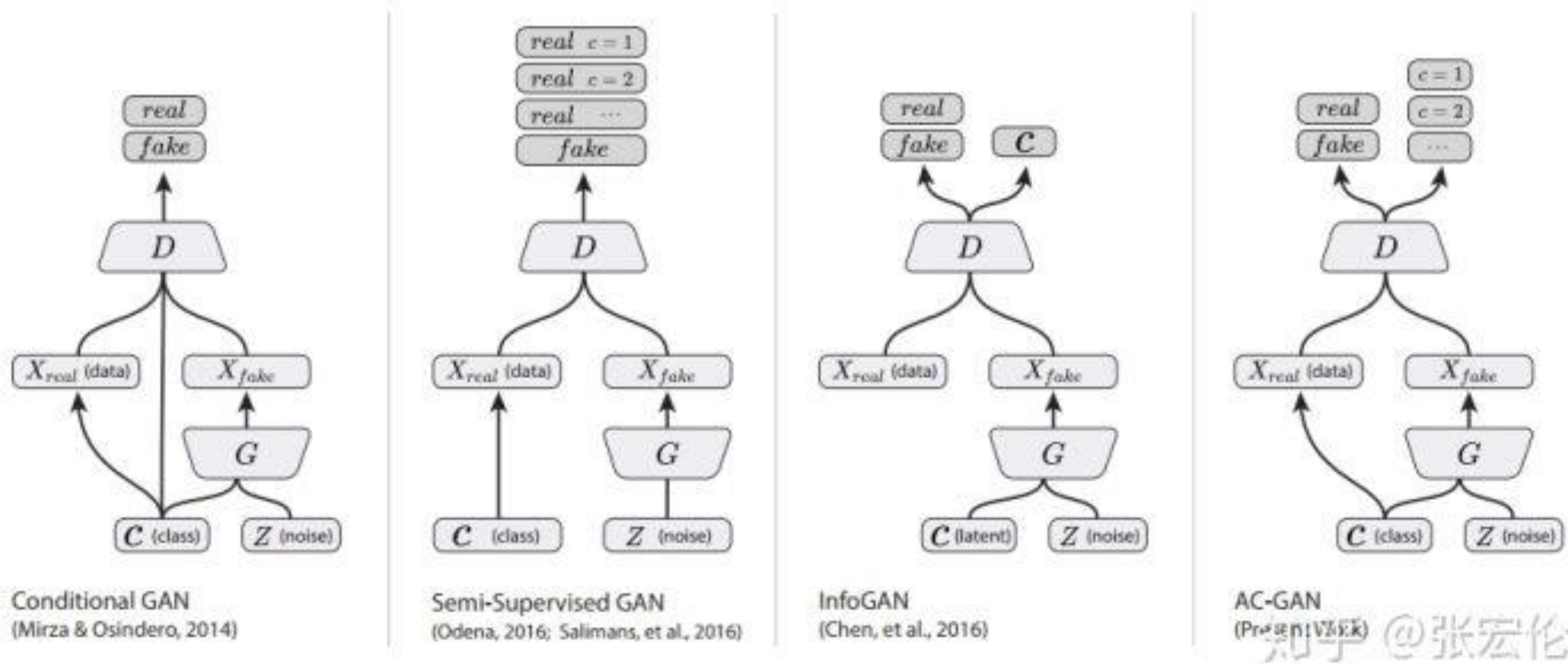
Same as before

Mutual information

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

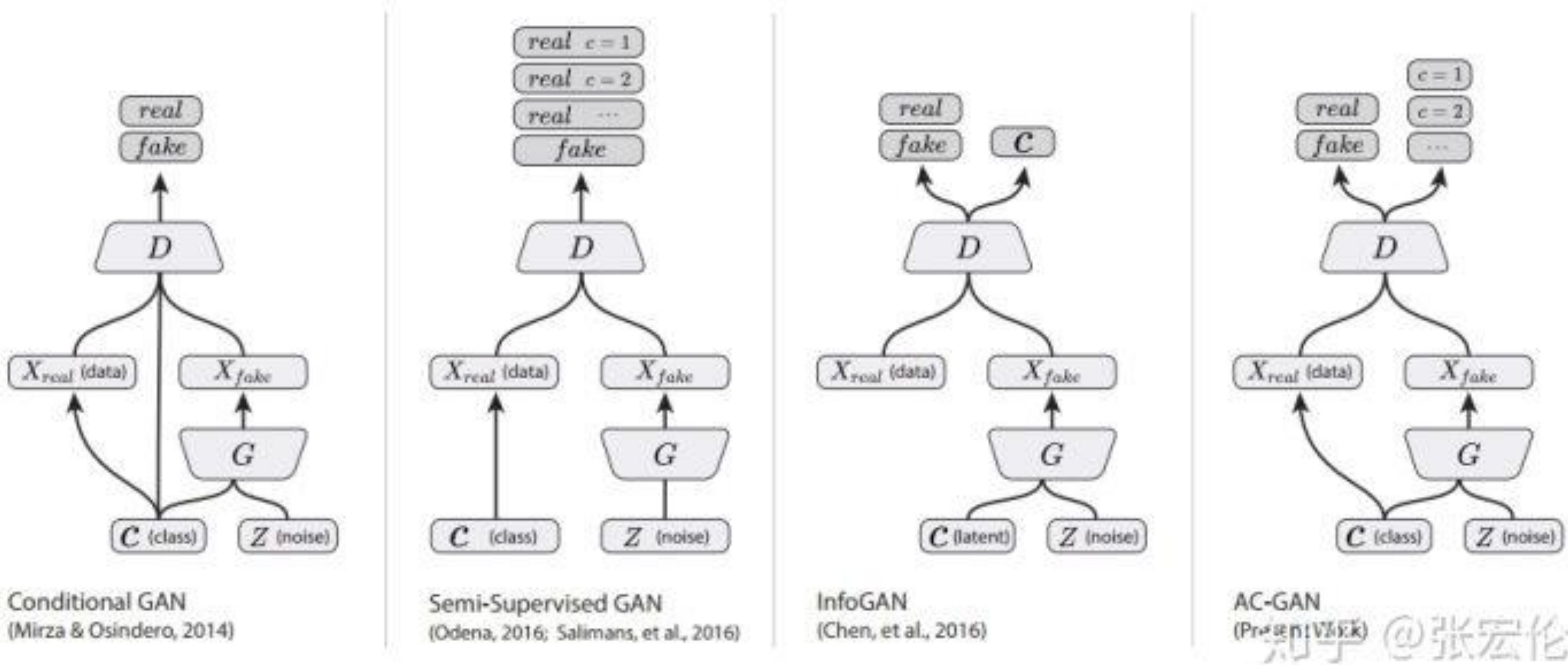
AC-GAN

[Odena, et al 2017]



Auxiliary Classifier GAN

AC-GAN



$$L_{adv}(D) = -\mathbb{E}_{x\sim p_{data}}[\log D(x)] - \mathbb{E}_{z\sim p_z, c\sim p_c}[\log(1 - D(G(z,c)))]$$

$$L_{cls}(D) = \mathbb{E}_{x\sim p_{data}}[L_D(c_x|x)]$$

$$L_{adv}(G) = \mathbb{E}_{z\sim p_z, c\sim p_c}[\log(1 - D(G(z,c)))]$$

$$L_{cls}(G) = \mathbb{E}_{z\sim p_z, c\sim p_c}[L_D(c|G(z,c))]$$

Separating large datasets into subsets by class