On Unifying Deep Generative Models

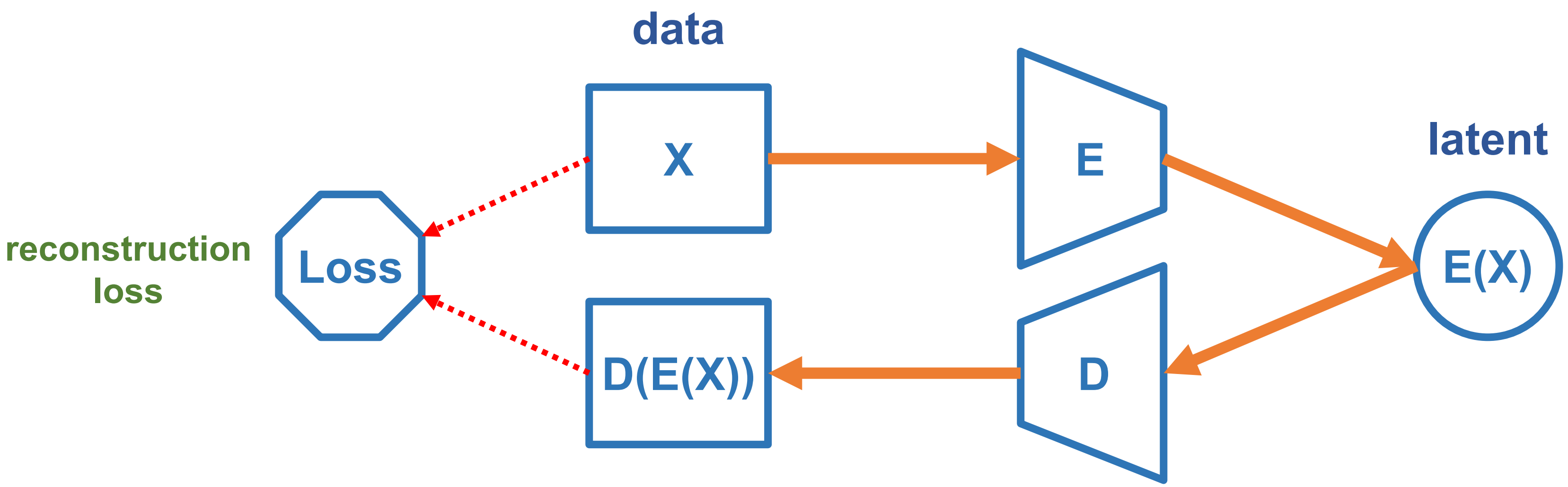
Herman Dong, 2017.8.15

Outline

- **Brief introduction to deep generative models**
  - AE (Autoencoder)
  - VAE (Variational Autoencoder)
  - GAN (Generative Adversarial Networks)
  - AAE (Adversarial Autoencoder)
  - VAE/GAN
  - ADA (Adversarial Domain Adaption)
- **Reformulation**
  - Graphical model representation
  - Connection to Wake-sleep Algorithm

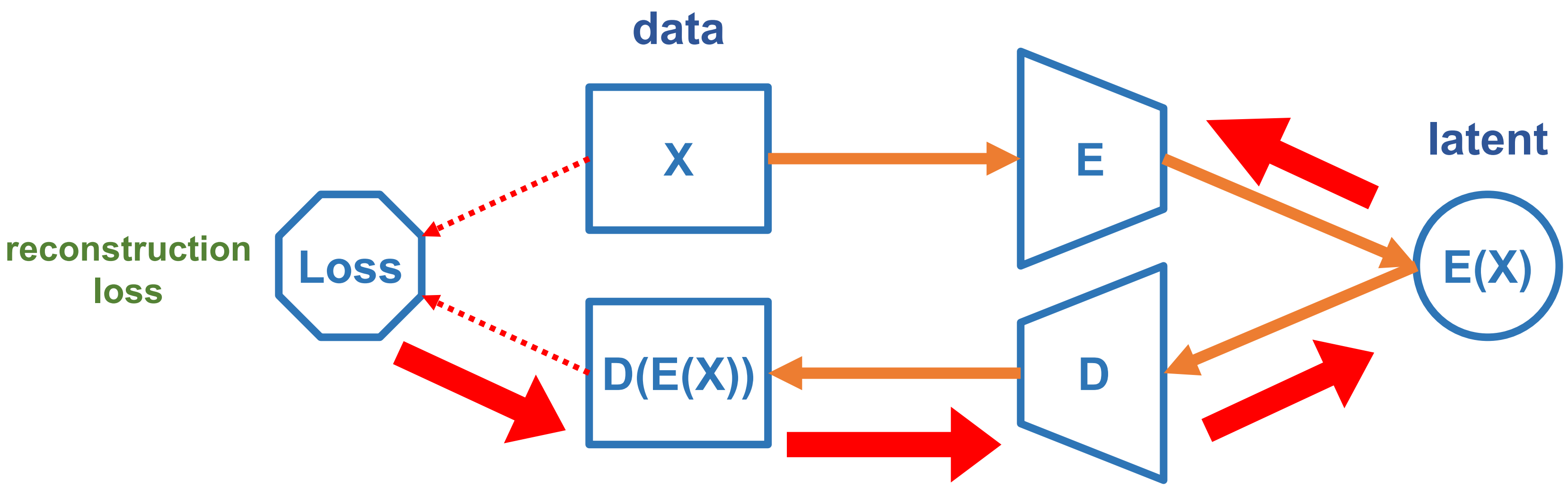
AE (Autoencoder)

- minimize **reconstruction loss**



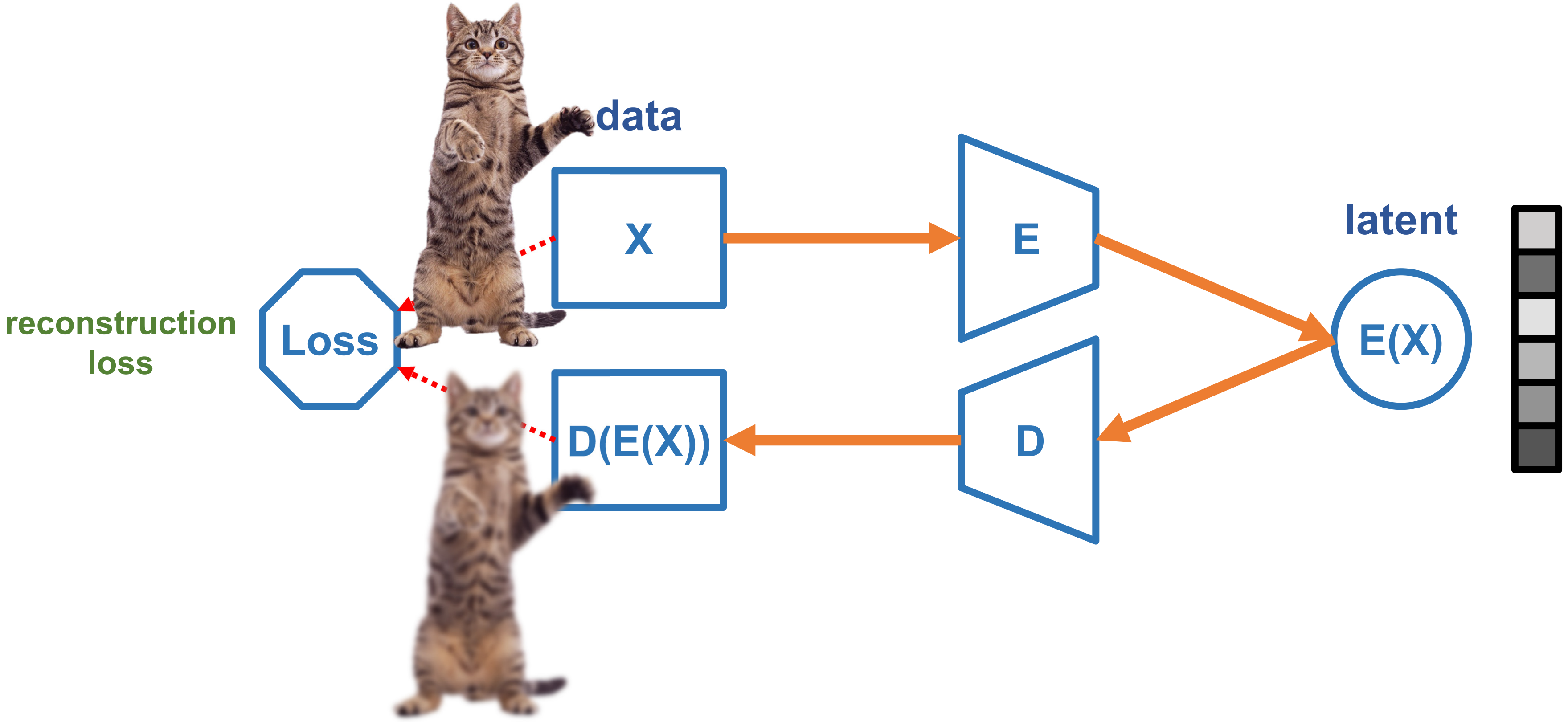
AE (Autoencoder)

- minimize **reconstruction loss**



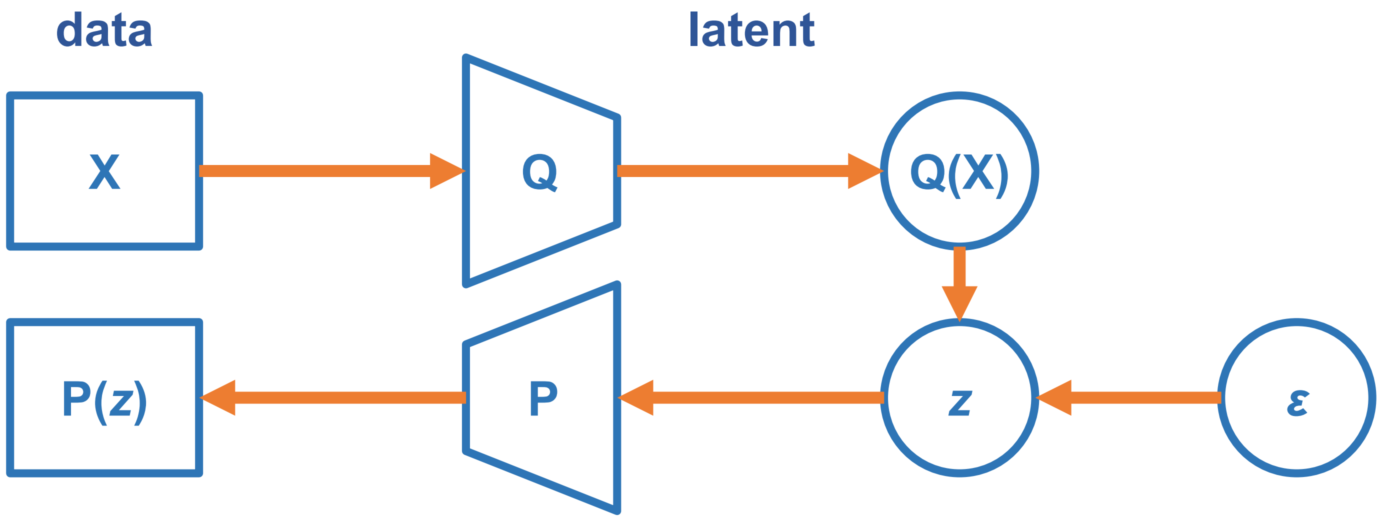
AE (Autoencoder)

- minimize **reconstruction loss**



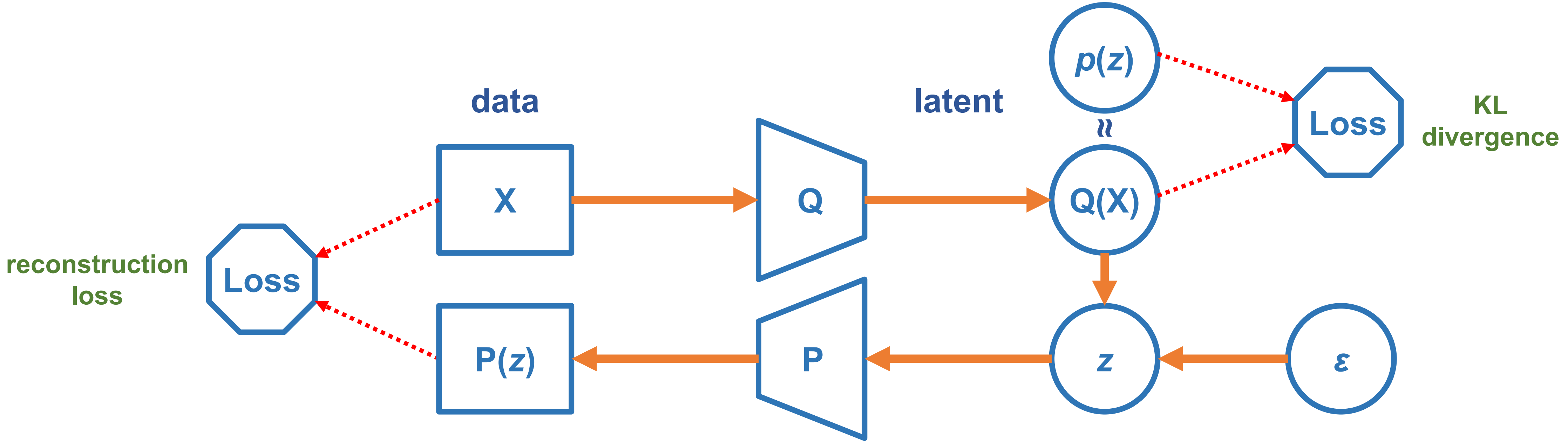
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



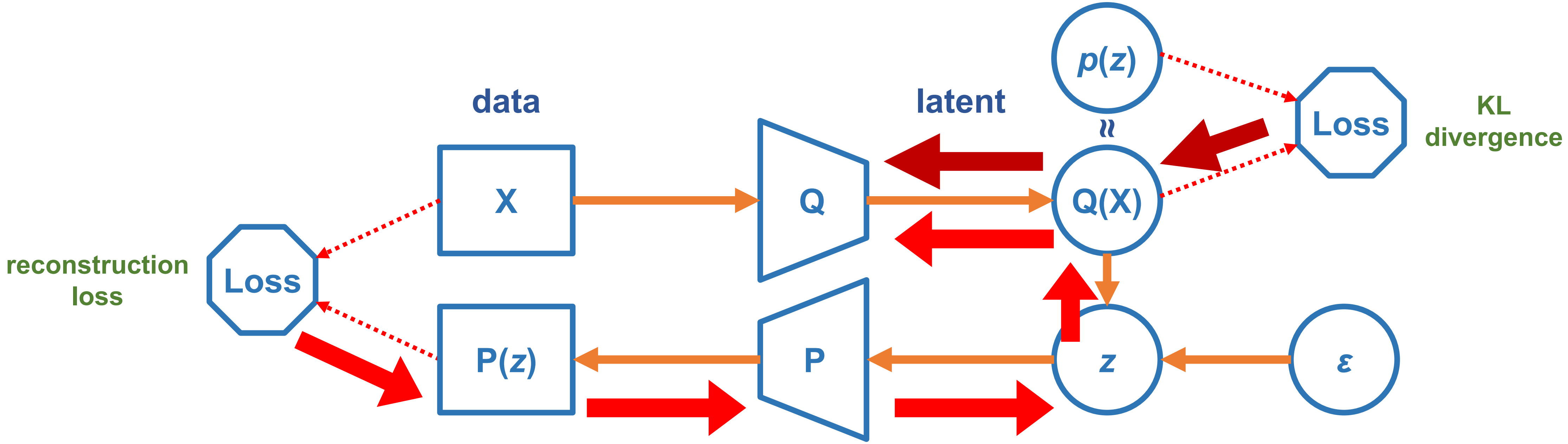
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



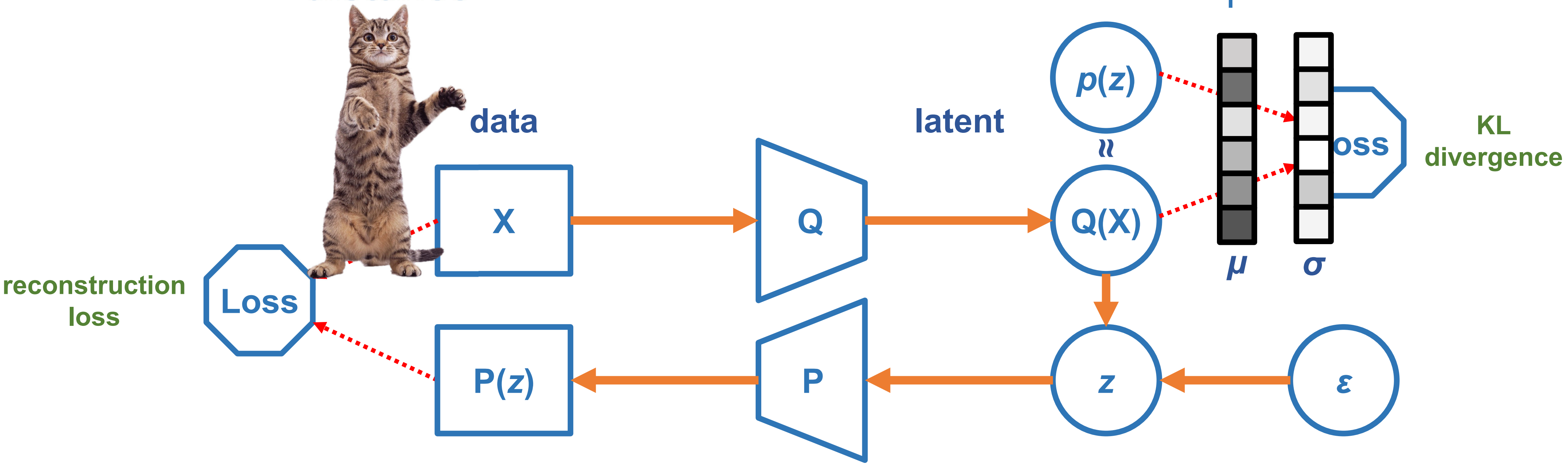
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



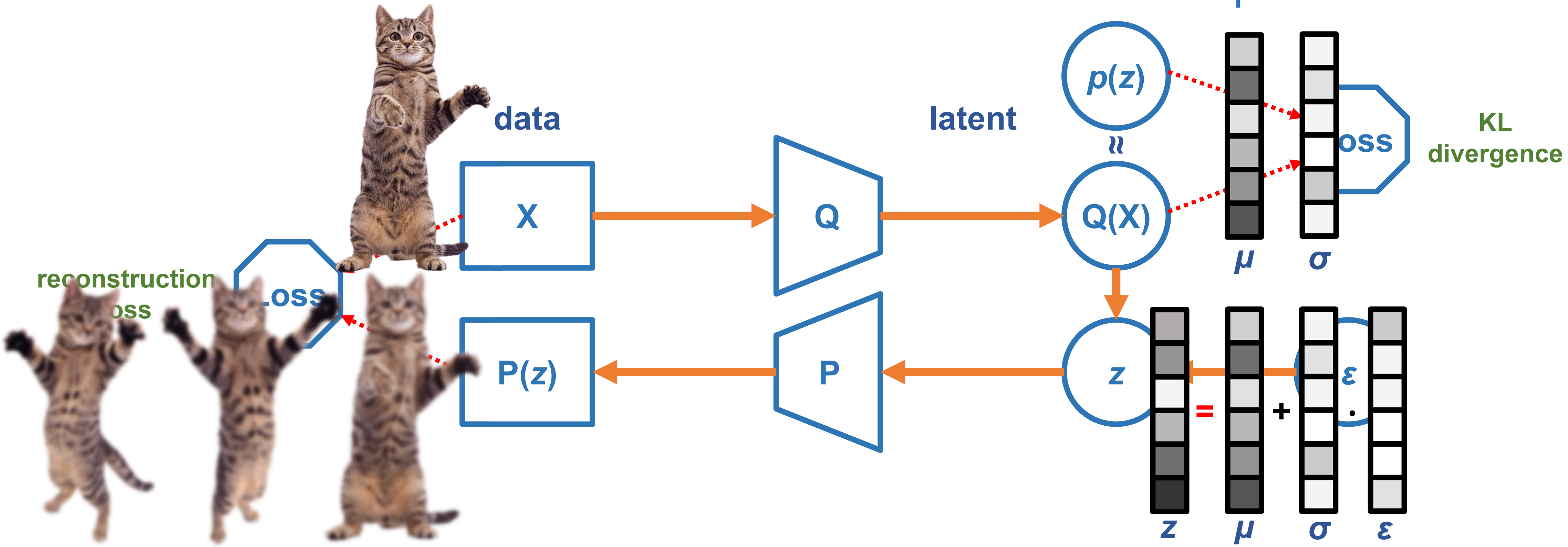
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



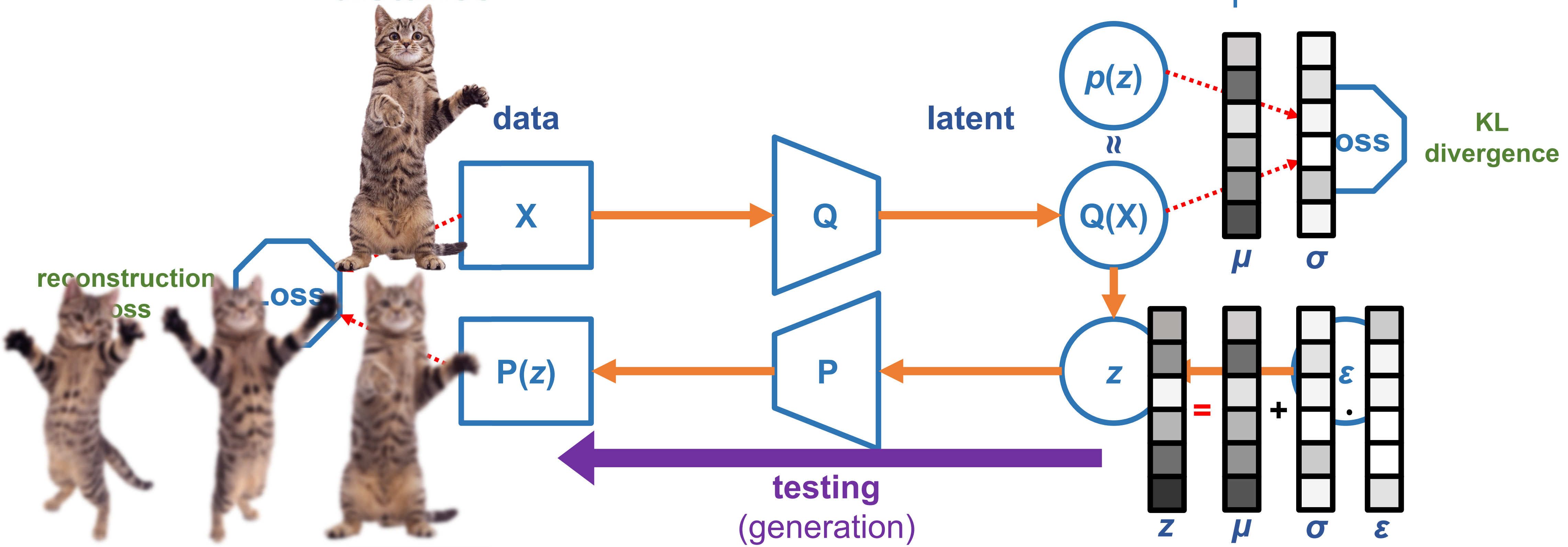
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



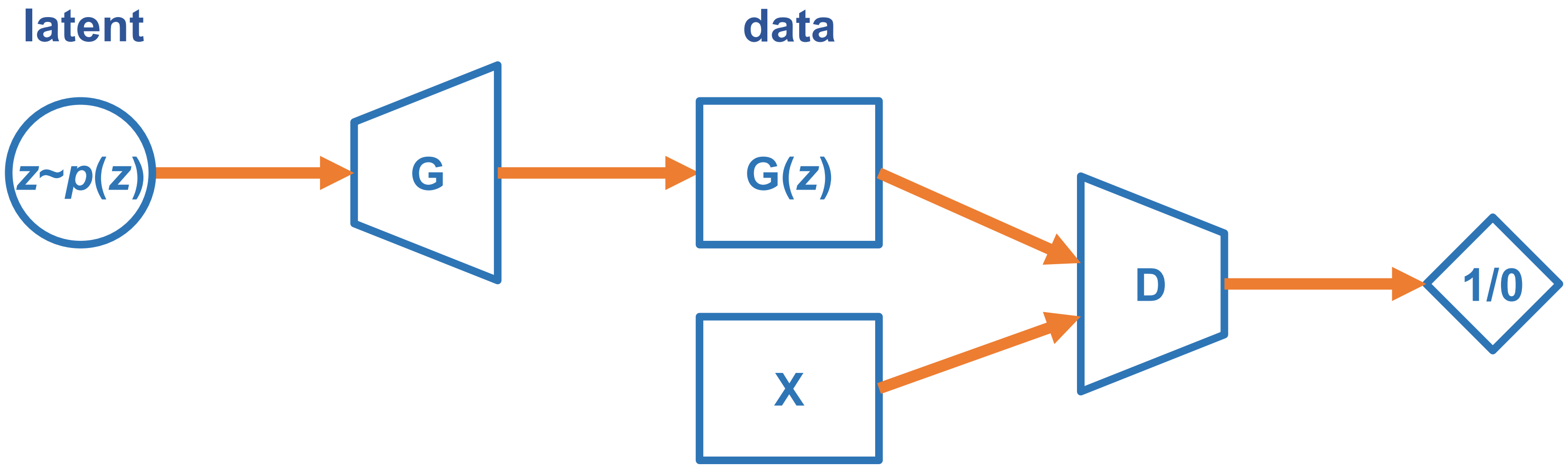
VAE (Variational Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



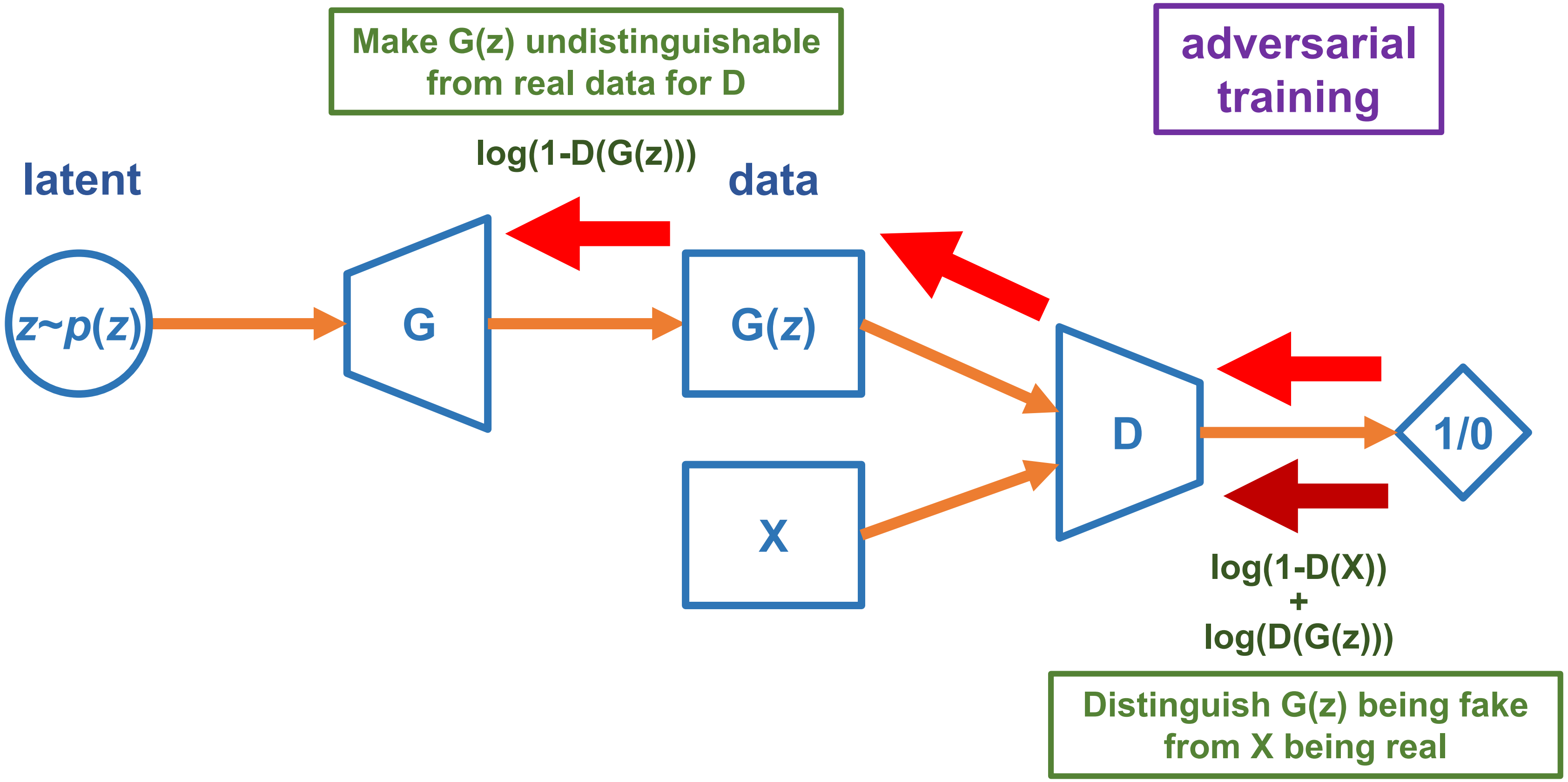
GAN (Generative Adversarial Network)

- minimize **distance** between the distribution of real data and generated samples



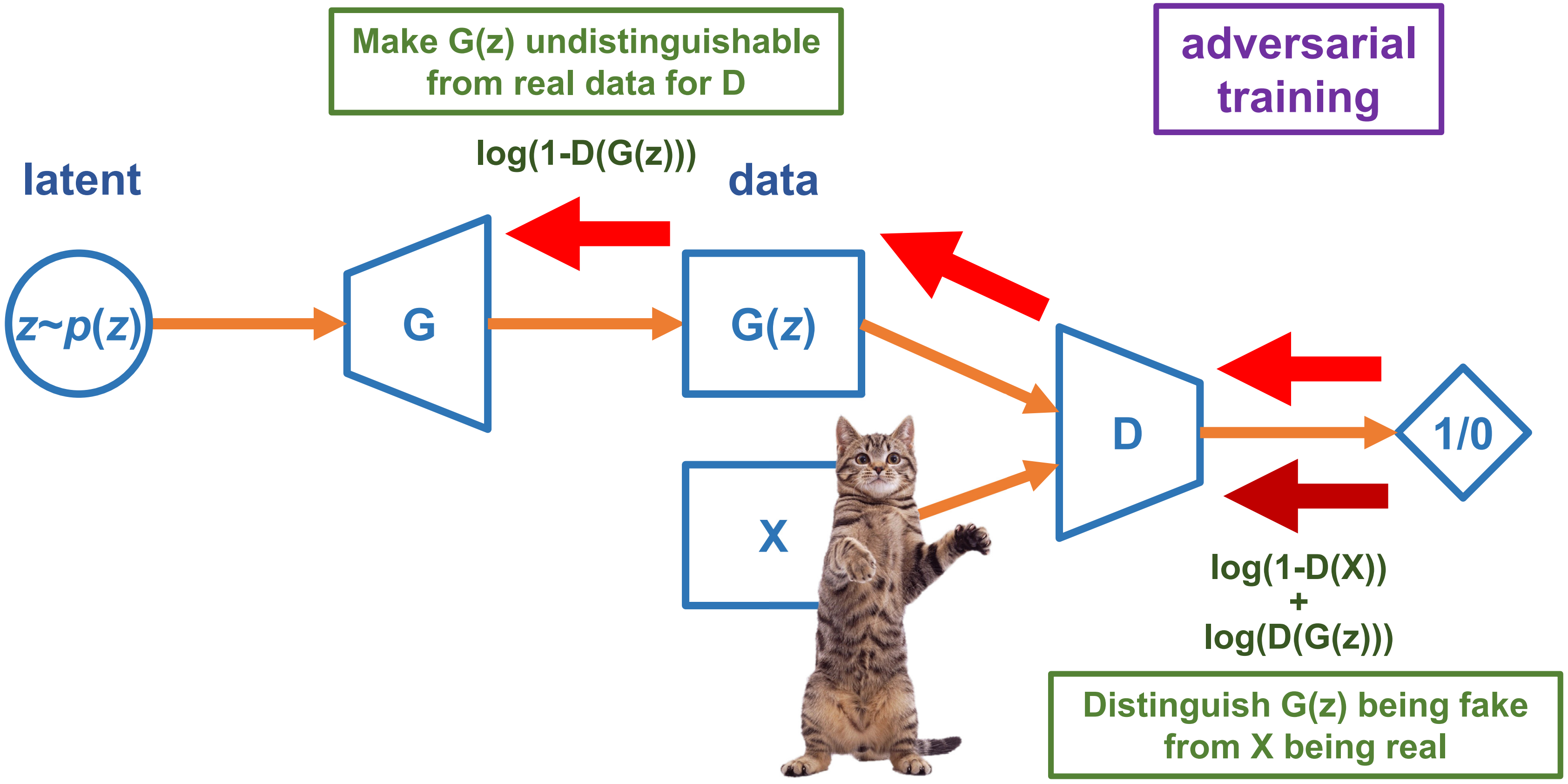
GAN (Generative Adversarial Network)

- minimize **distance** between the distribution of real data and generated samples



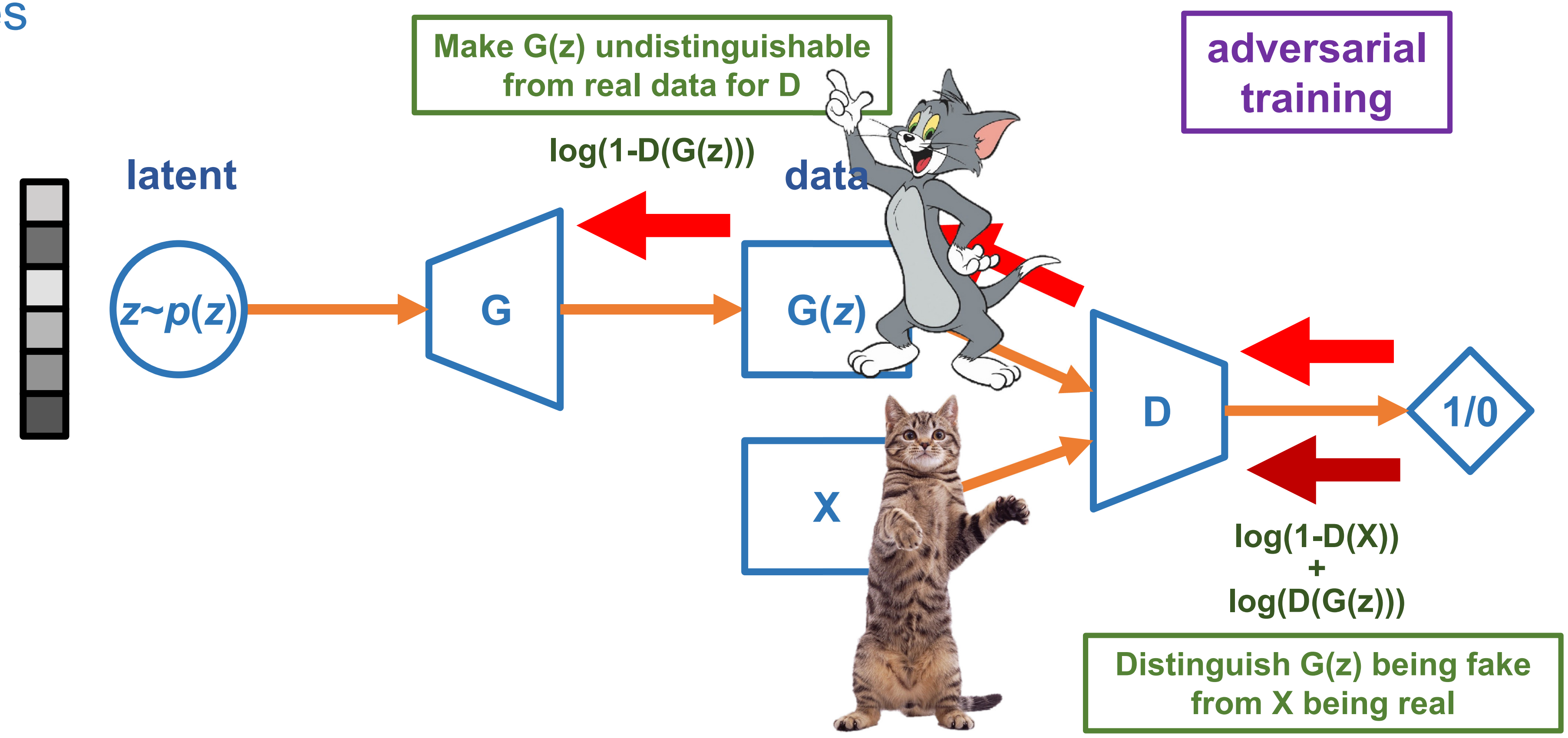
GAN (Generative Adversarial Network)

- minimize **distance** between the distribution of real data and generated samples



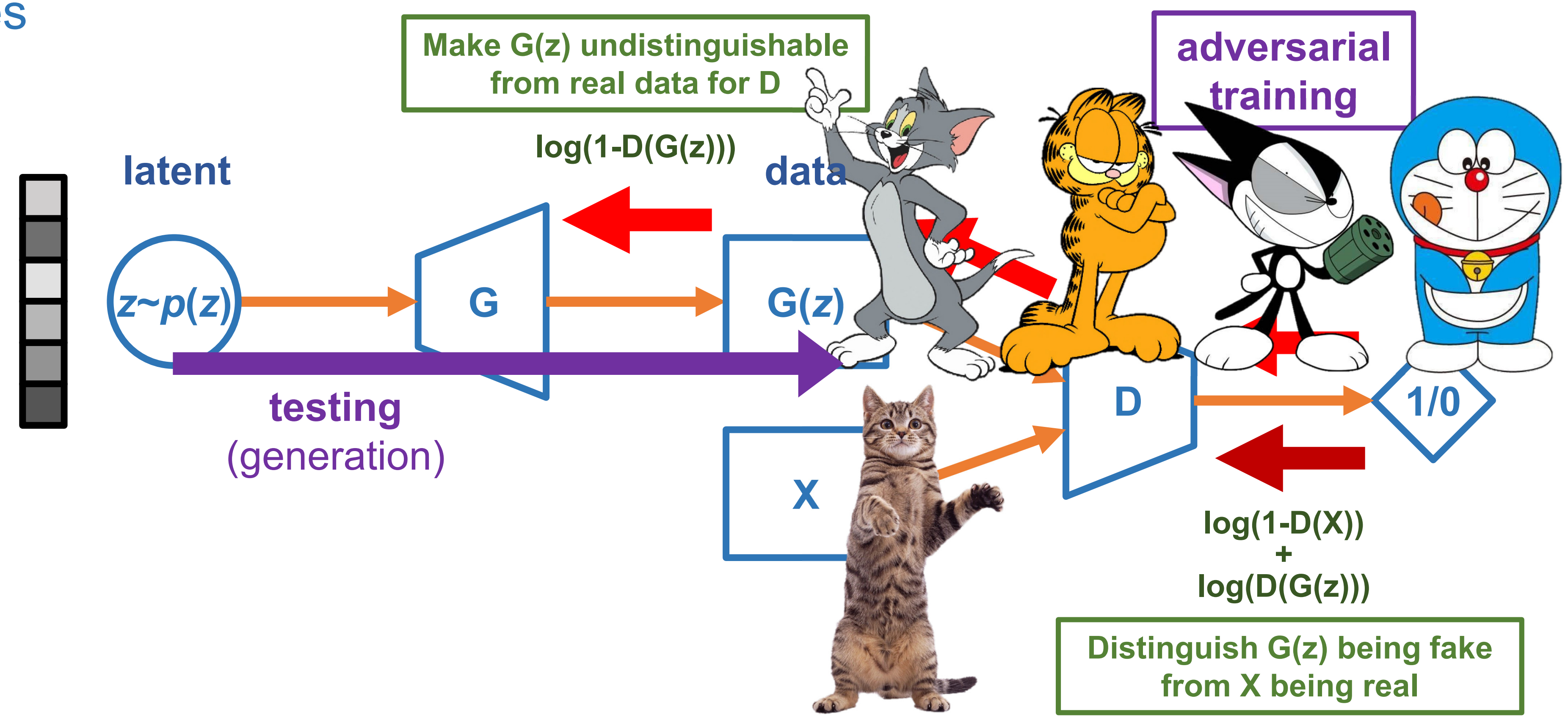
GAN (Generative Adversarial Network)

- minimize **distance** between the distribution of real data and generated samples



GAN (Generative Adversarial Network)

- minimize **distance** between the distribution of real data and generated samples

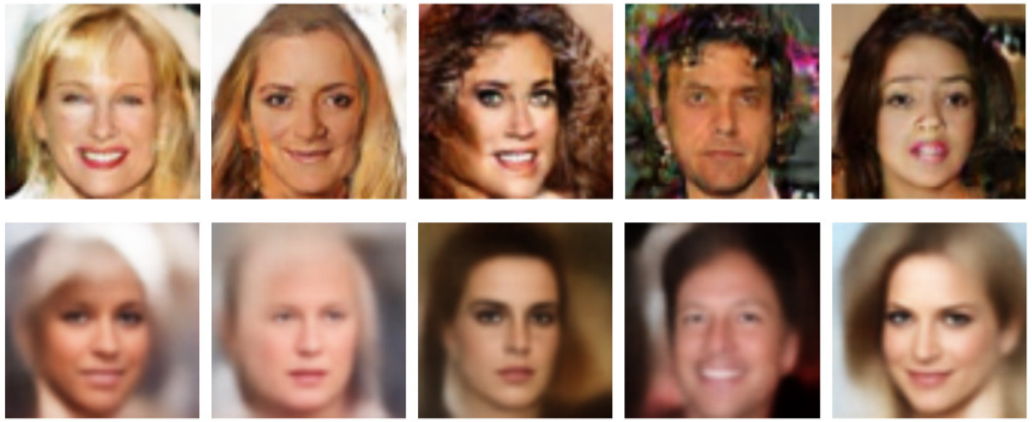


GAN vs VAE

- **GAN**
  - Generator aim to **fool the discriminator**
  - Discriminator aim to **distinguish generated data from real data**
  - output images are **sharper**
  - **higher diversity, lower stability**
- **VAE**
  - Objective: **reconstruct real data**
  - using pixel-to-pixel loss
  - output images are **more blurred**
  - **lower diversity, higher stability**

GAN

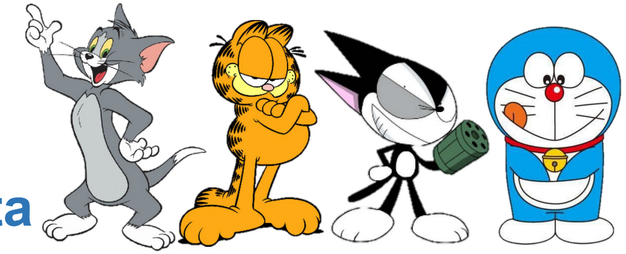
VAE



A. B. L. Larsen, S. K. Sønderby, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300, 2015.

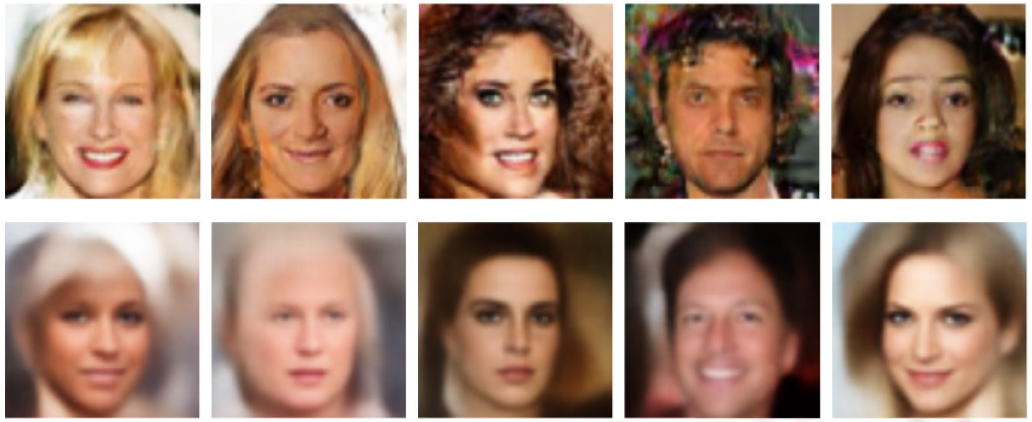
GAN vs VAE

- **GAN**
  - Generator aim to **fool the discriminator**
  - Discriminator aim to **distinguish generated data from real data**
  - output images are **<u>sharper</u>**
  - **higher diversity, lower stability**
- **VAE**
  - Objective: **reconstruct real data**
  - using pixel-to-pixel loss
  - output images are **<u>more blurred</u>**
  - **lower diversity, higher stability**



GAN

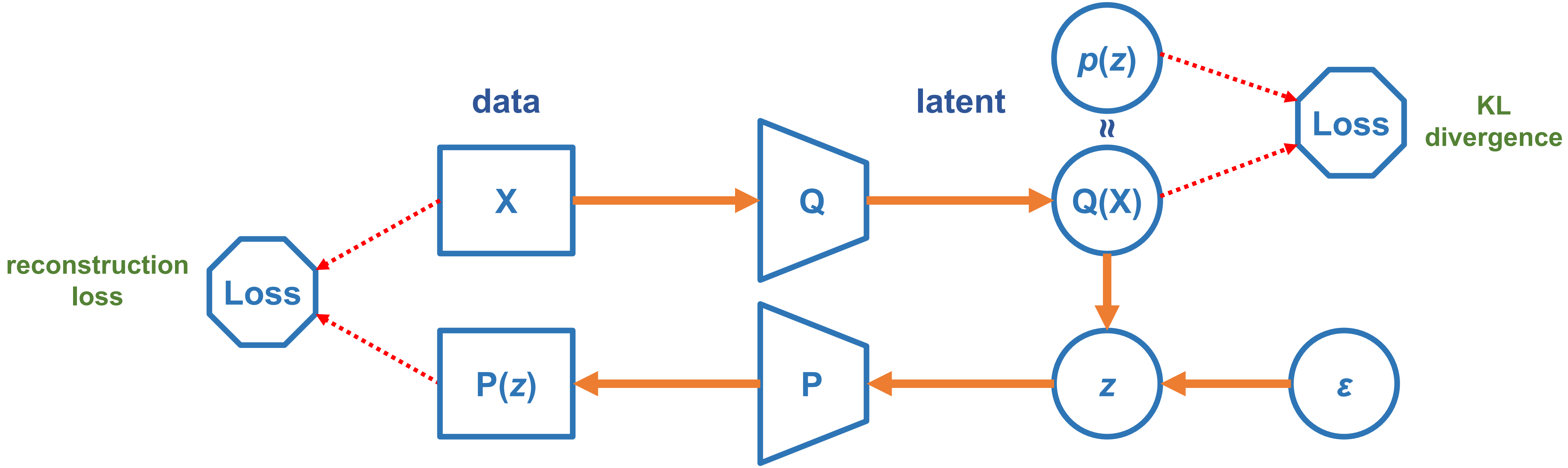
VAE



A. B. L. Larsen, S. K. Sønderby, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300, 2015.

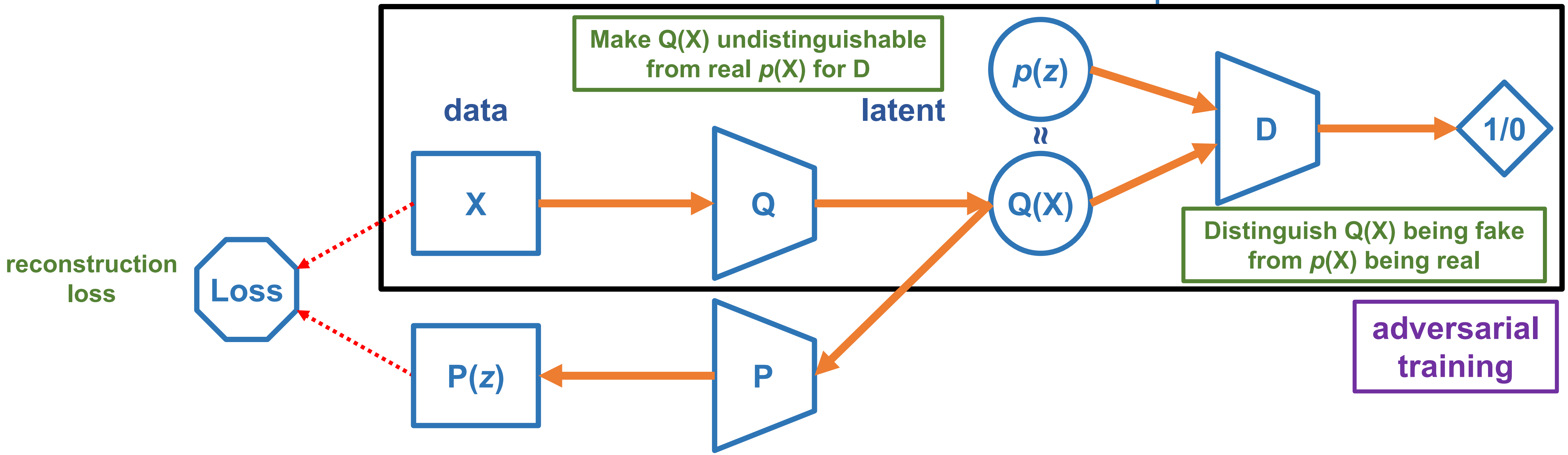
AAE (Adversarial Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



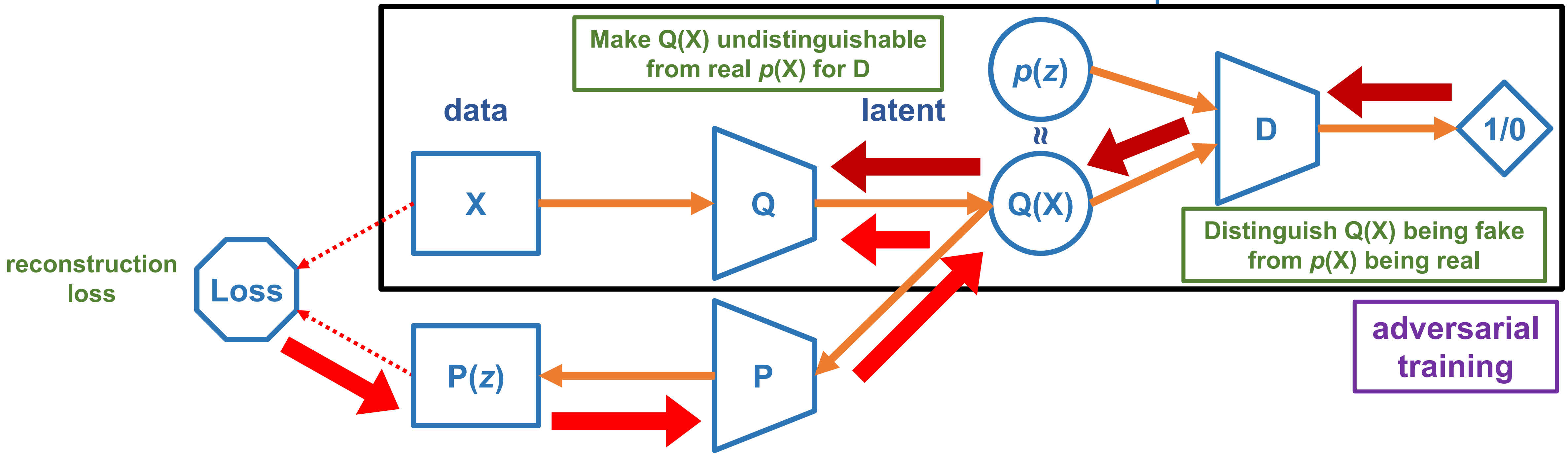
AAE (Adversarial Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



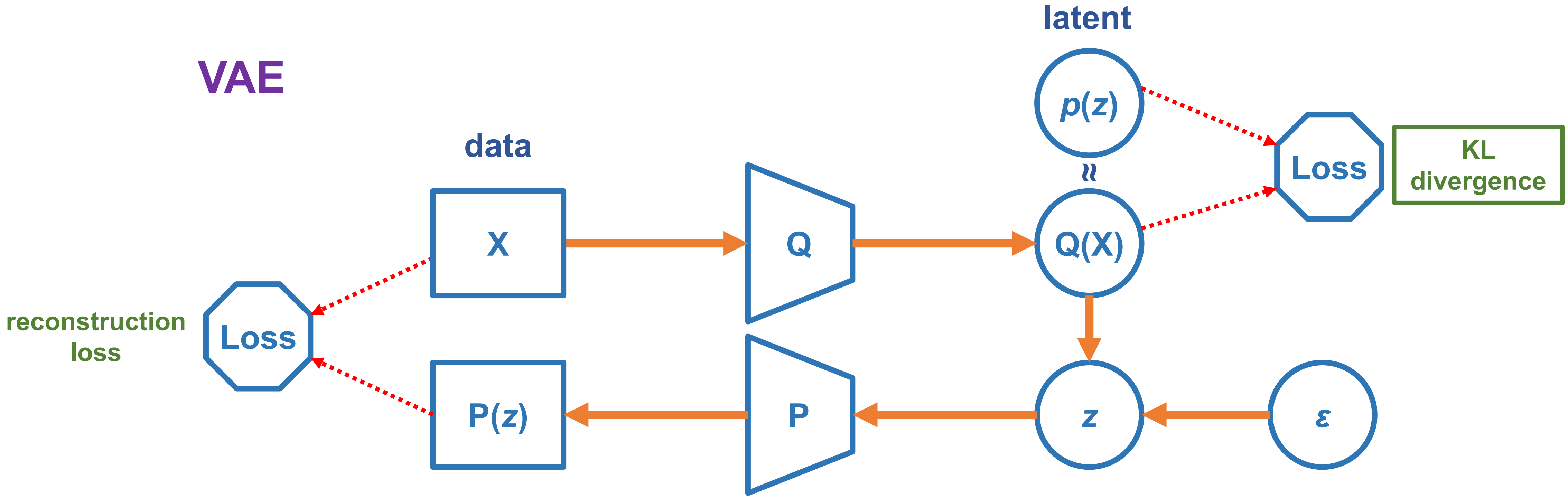
AAE (Adversarial Autoencoder)

- minimize **reconstruction loss**
- minimize **distance** between encoded latent distribution and prior distribution



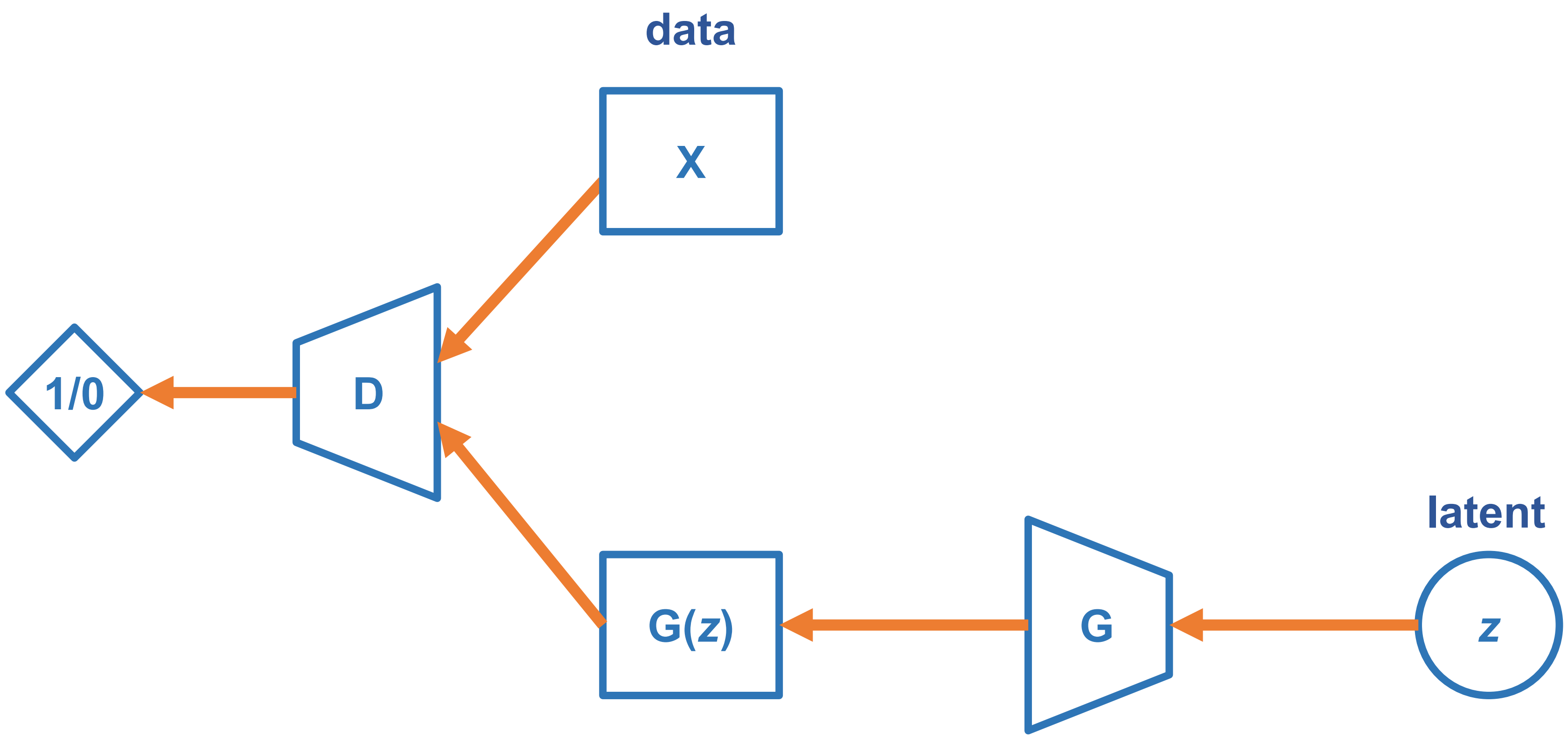
VAE/GAN

VAE

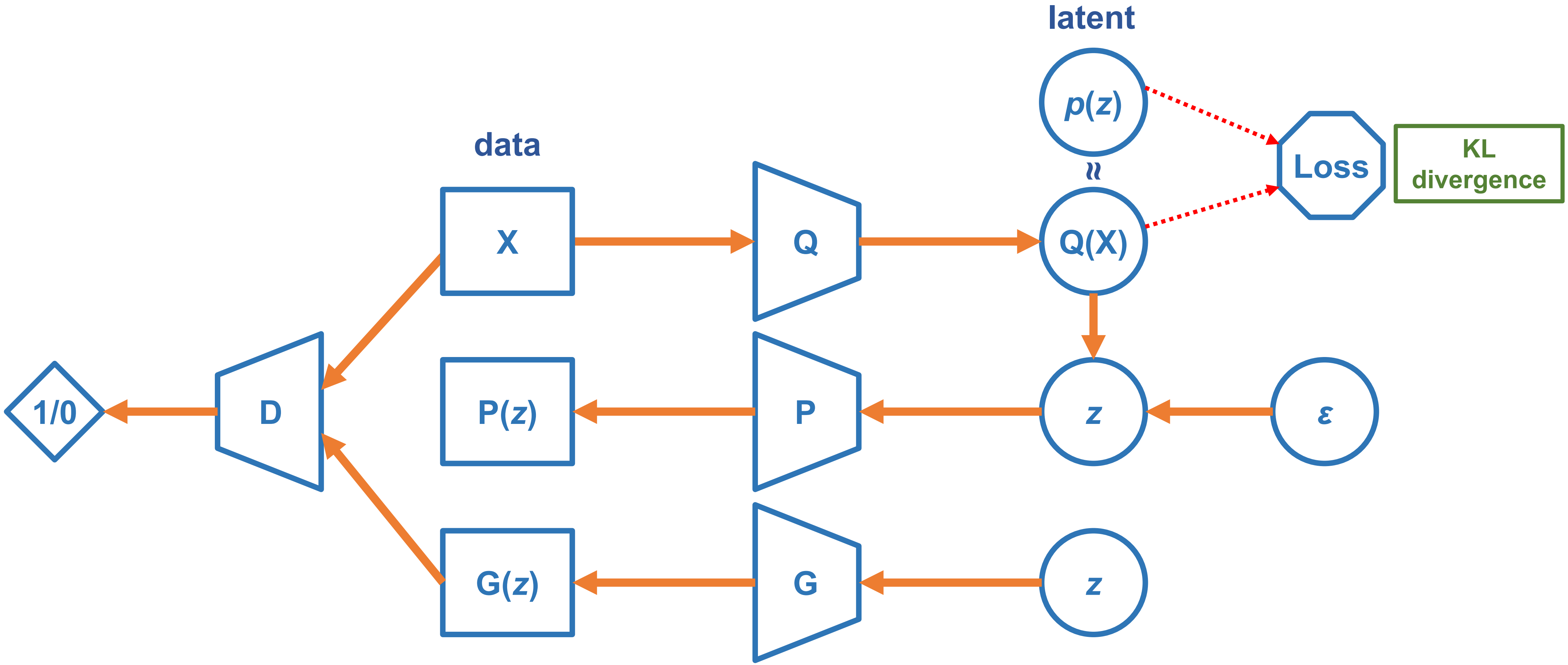

latent
data
p(z)
≈
Loss
KL divergence
X
Q
Q(X)
reconstruction loss
Loss
z
P(z)
P
ε

VAE/GAN

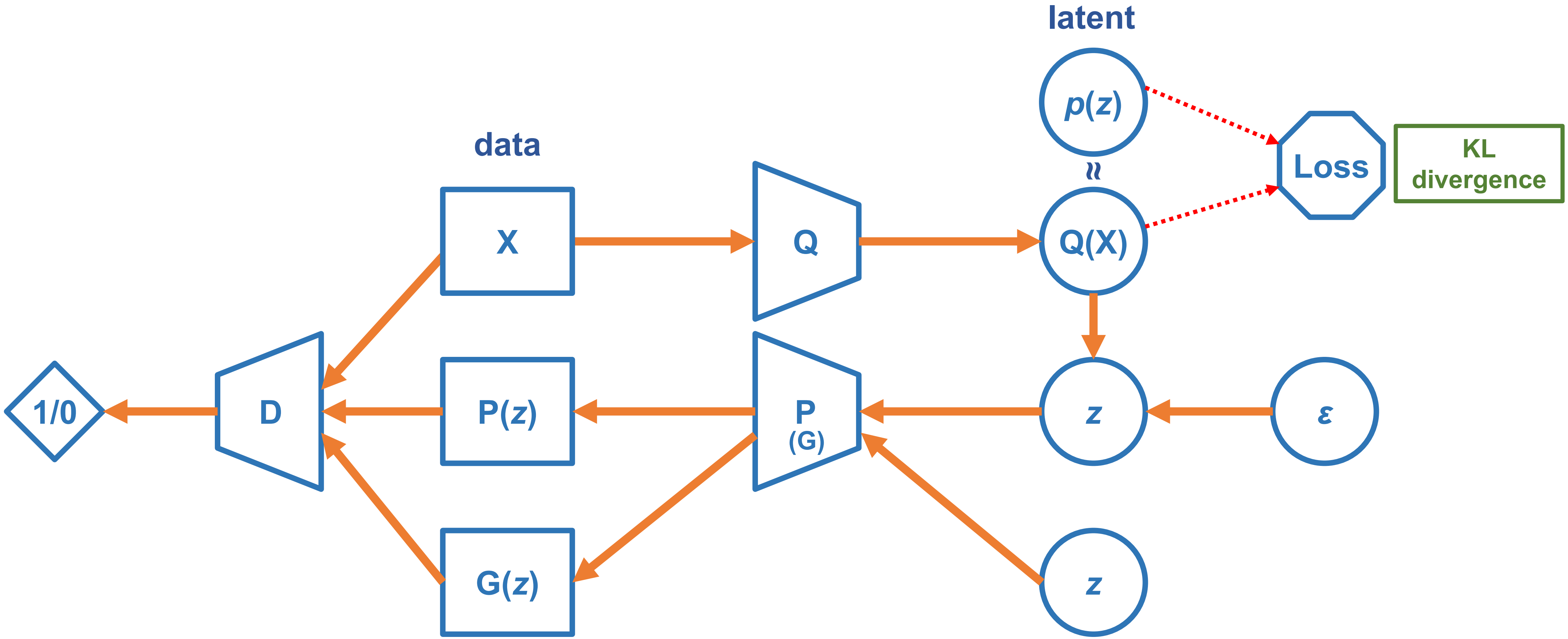
GAN



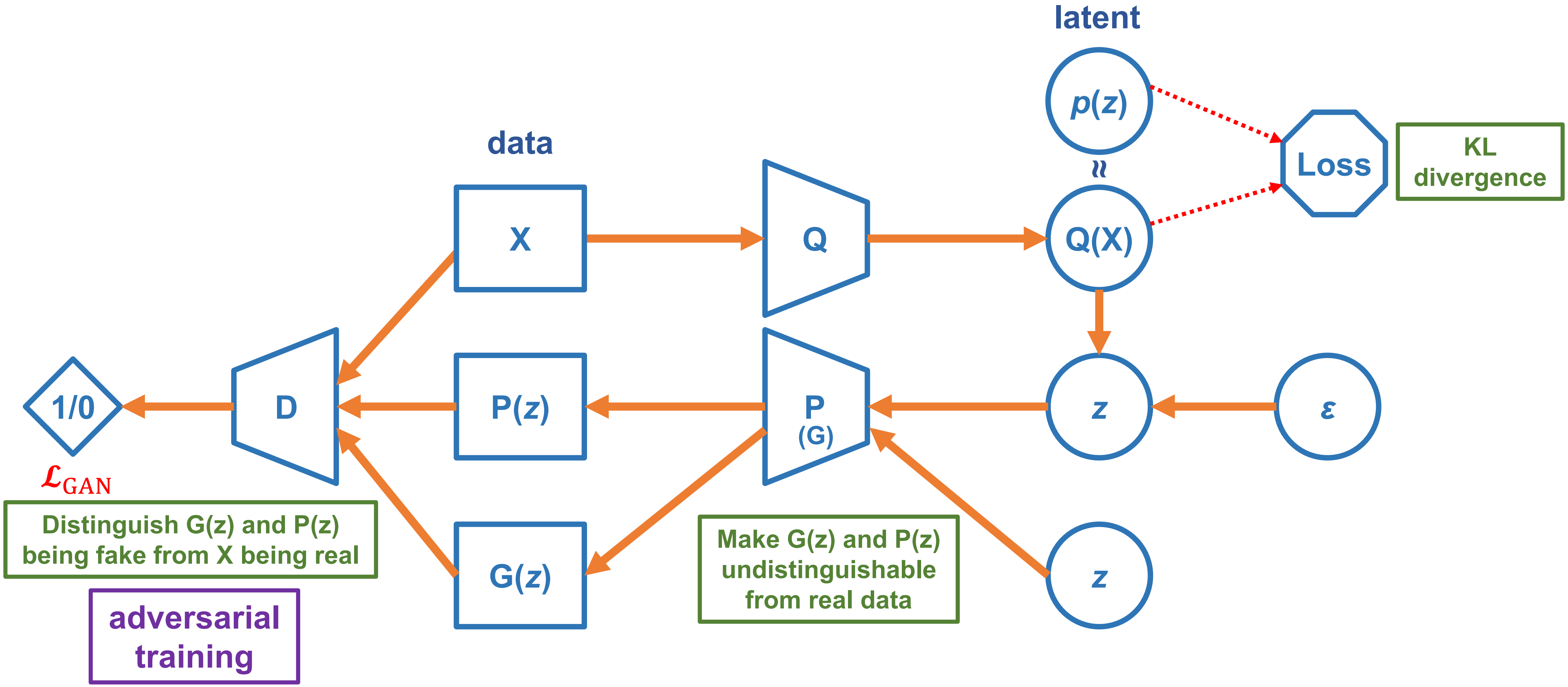
VAE/GAN



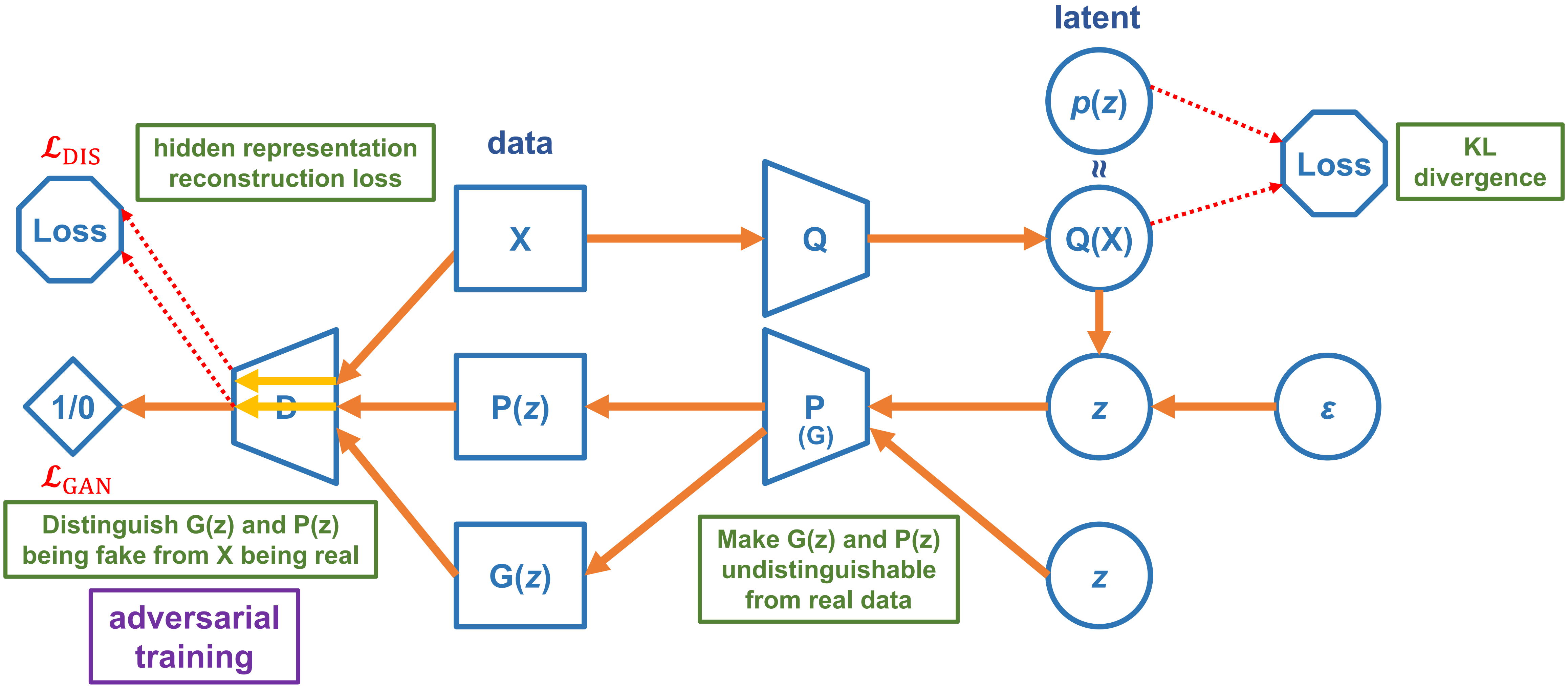
VAE/GAN


latent
data
$p(z)$
≀
Loss
KL divergence
X
Q
Q(X)
1/0
D
P(z)
P
(G)
z
ε
G(z)
z

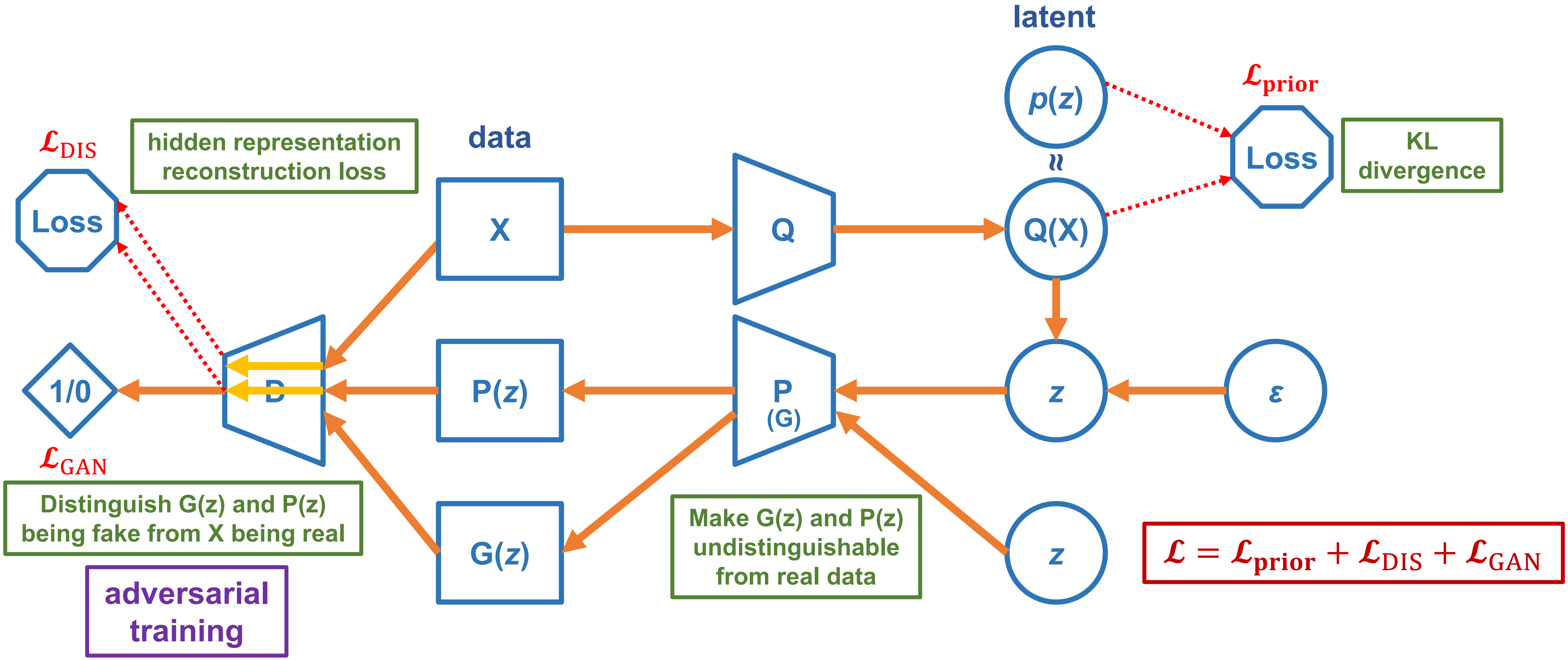
VAE/GAN


latent
data
p(z)
≈
Loss
KL divergence
X
Q
Q(X)
1/0
D
P(z)
P
(G)
z
ε
$\mathcal{L}_{\text{GAN}}$
Distinguish G(z) and P(z) being fake from X being real
G(z)
Make G(z) and P(z) undistinguishable from real data
z
adversarial training

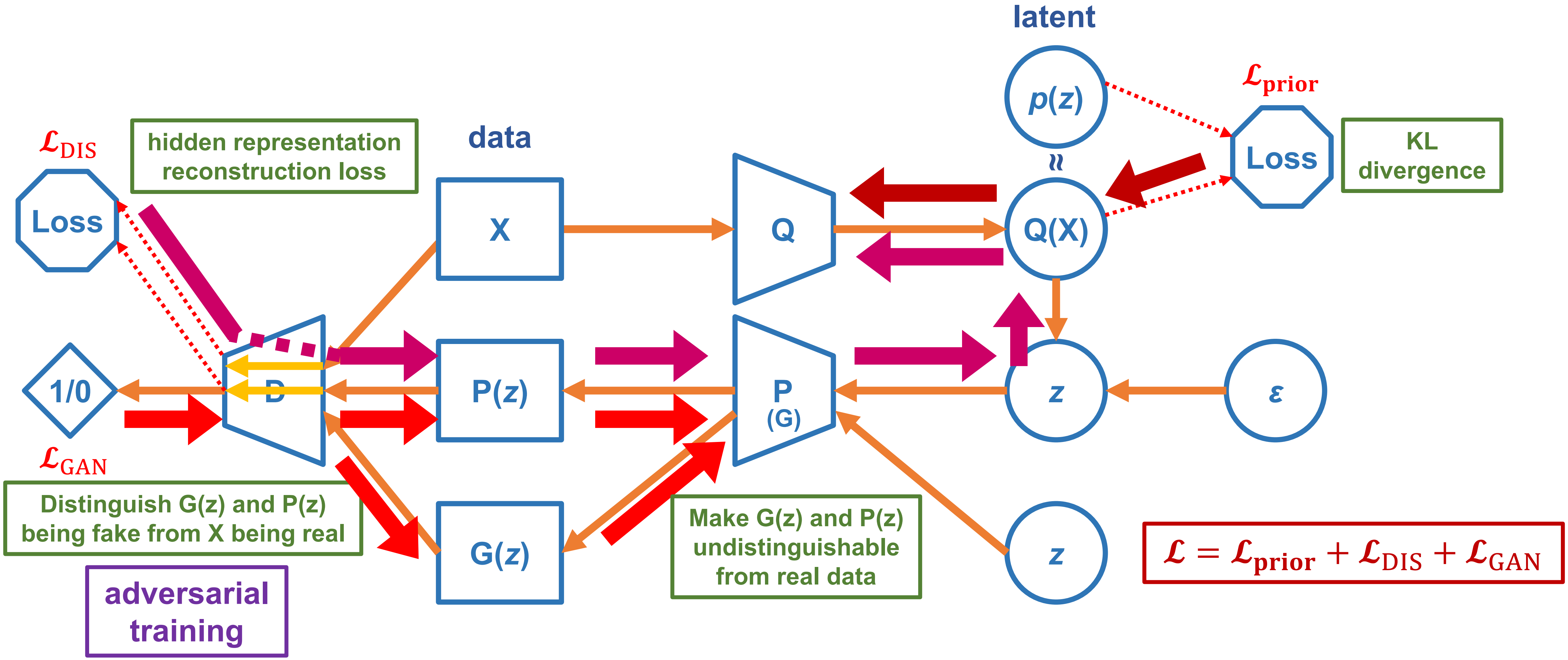
VAE/GAN


latent
p(z)
≈
Q(X)
Loss
KL divergence
$\mathcal{L}_{\text{DIS}}$
hidden representation reconstruction loss
data
X
Q
Loss
1/0
D
P(z)
P
(G)
z
ε
$\mathcal{L}_{\text{GAN}}$
Distinguish G(z) and P(z) being fake from X being real
adversarial training
G(z)
Make G(z) and P(z) undistinguishable from real data
z

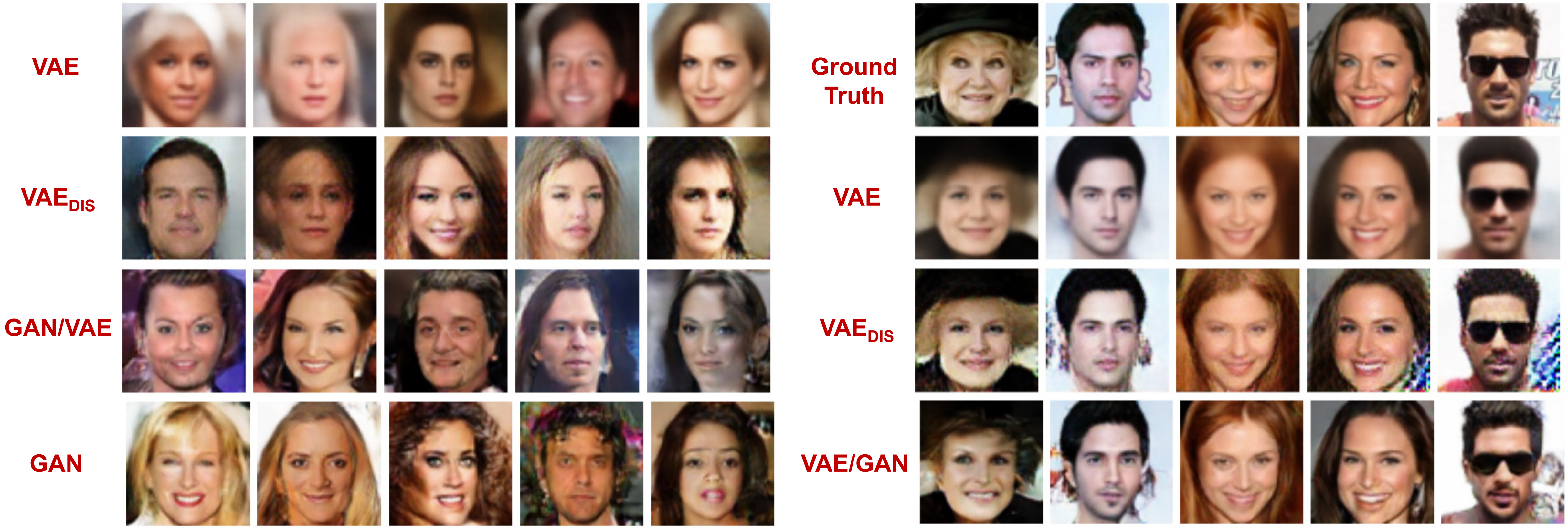
VAE/GAN


latent
$\mathcal{L}_{\text{prior}}$
p(z)
≈
Loss
KL divergence
$\mathcal{L}_{\text{DIS}}$
hidden representation reconstruction loss
data
Loss
X
Q
Q(X)
D
P(z)
P
(G)
z
ε
1/0
$\mathcal{L}_{\text{GAN}}$
Distinguish G(z) and P(z) being fake from X being real
adversarial training
G(z)
Make G(z) and P(z) undistinguishable from real data
z
$\mathcal{L} = \mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{DIS}} + \mathcal{L}_{\text{GAN}}$

VAE/GAN


latent
data
$\mathcal{L}_{\text{prior}}$
$p(z)$
≈
Loss
KL divergence
$\mathcal{L}_{\text{DIS}}$
hidden representation reconstruction loss
Loss
X
Q
Q(X)
1/0
D
P(z)
P (G)
z
ε
$\mathcal{L}_{\text{GAN}}$
Distinguish G(z) and P(z) being fake from X being real
adversarial training
G(z)
Make G(z) and P(z) undistinguishable from real data
z
$\mathcal{L} = \mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{DIS}} + \mathcal{L}_{\text{GAN}}$

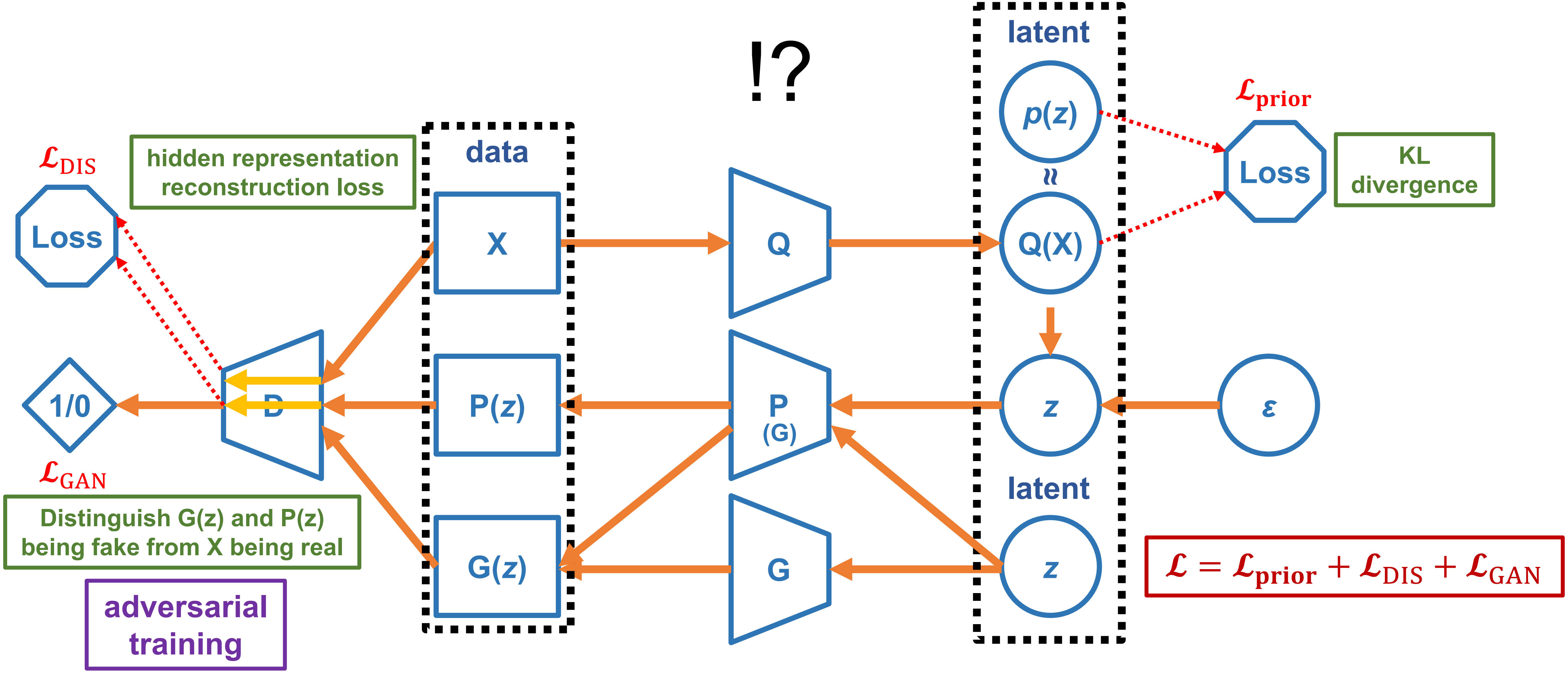
VAE/GAN



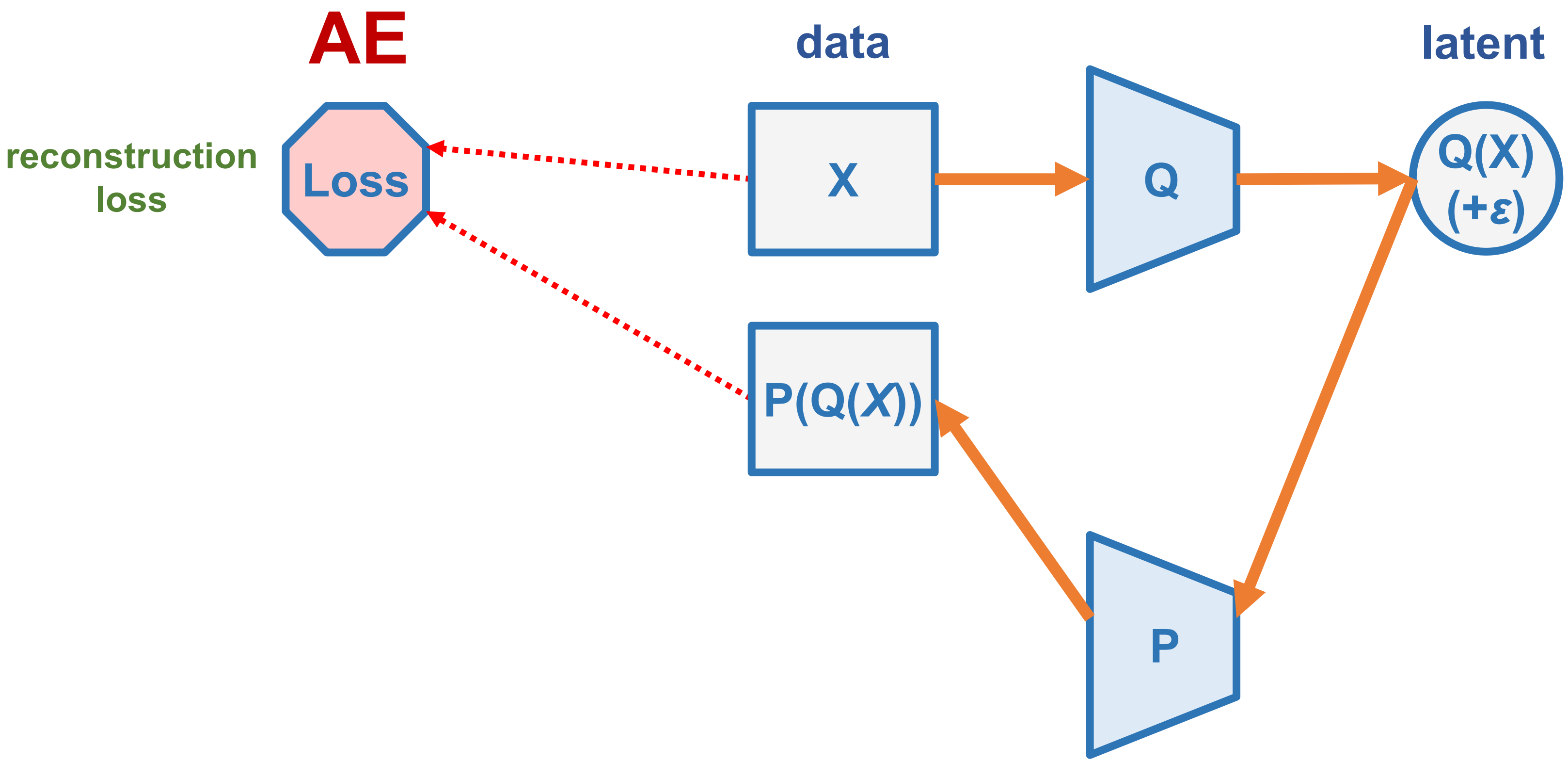
Generation test **Reconstruction test**

A. B. L. Larsen, S. K. Sønderby, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300, 2015.

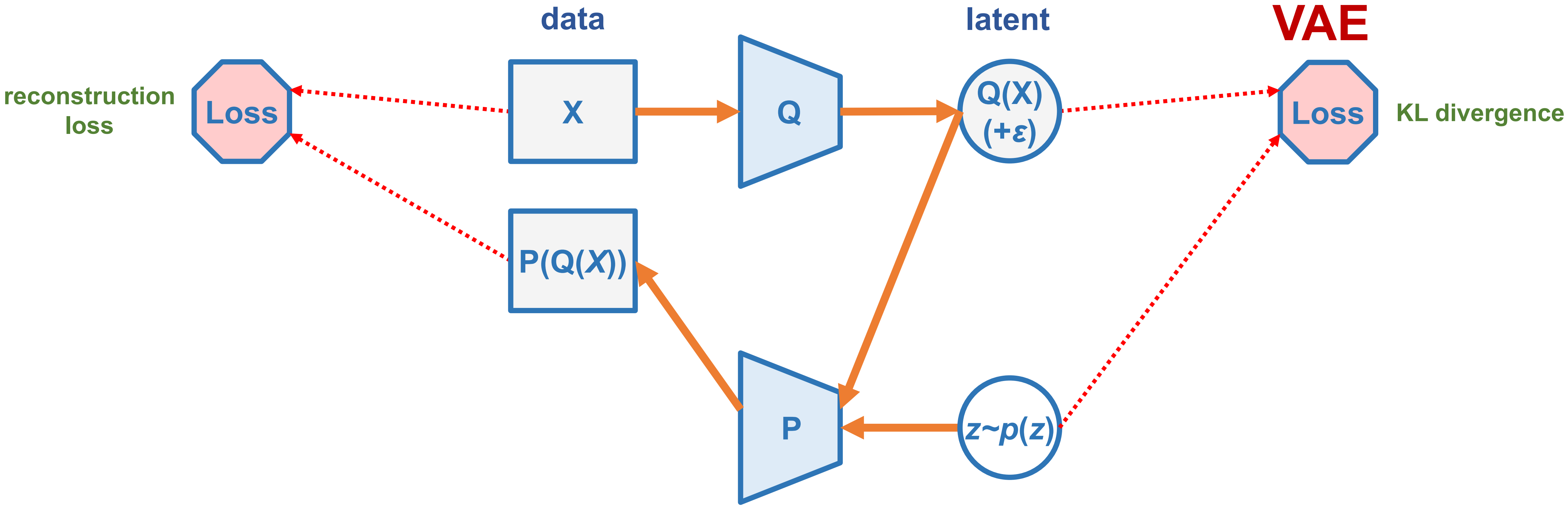
VAE/GAN


!?
latent
$p(z)$
$\wr$
Q(X)
z
latent
z
$\mathcal{L}_{\text{prior}}$
Loss
KL divergence
ε
data
X
P(z)
G(z)
Q
P (G)
G
$\mathcal{L}_{\text{DIS}}$
hidden representation reconstruction loss
Loss
D
1/0
$\mathcal{L}_{\text{GAN}}$
Distinguish G(z) and P(z) being fake from X being real
adversarial training
$\mathcal{L} = \mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{DIS}} + \mathcal{L}_{\text{GAN}}$

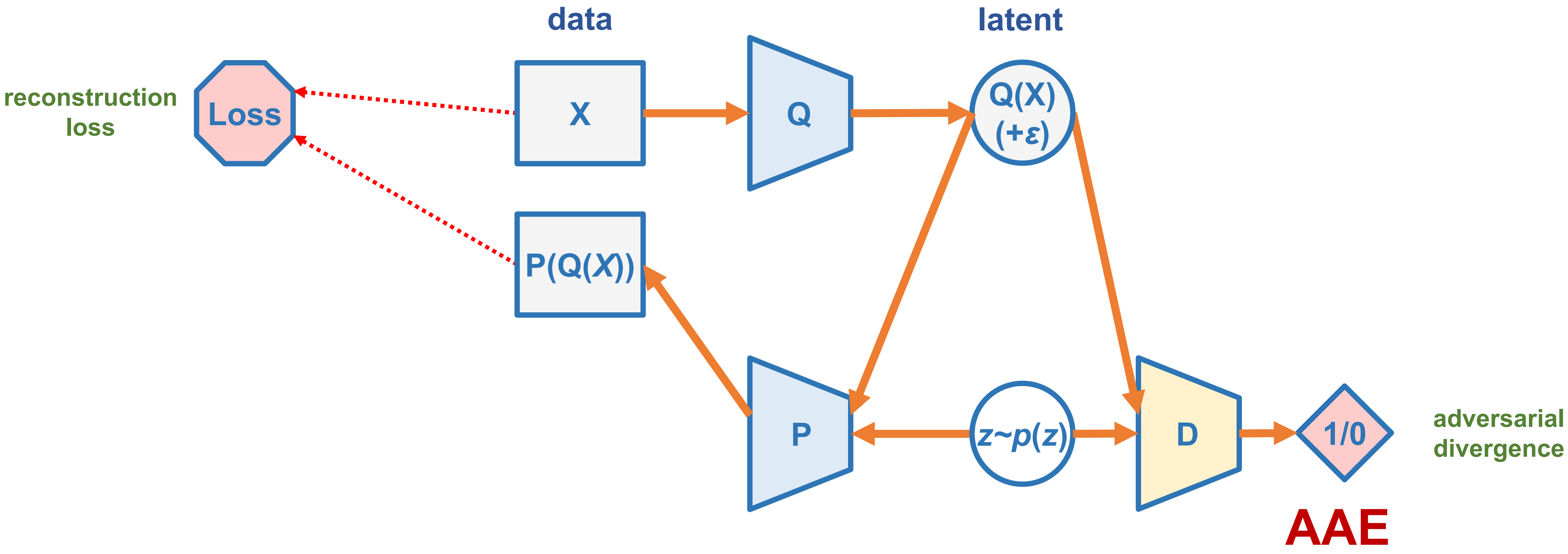
What's going on?



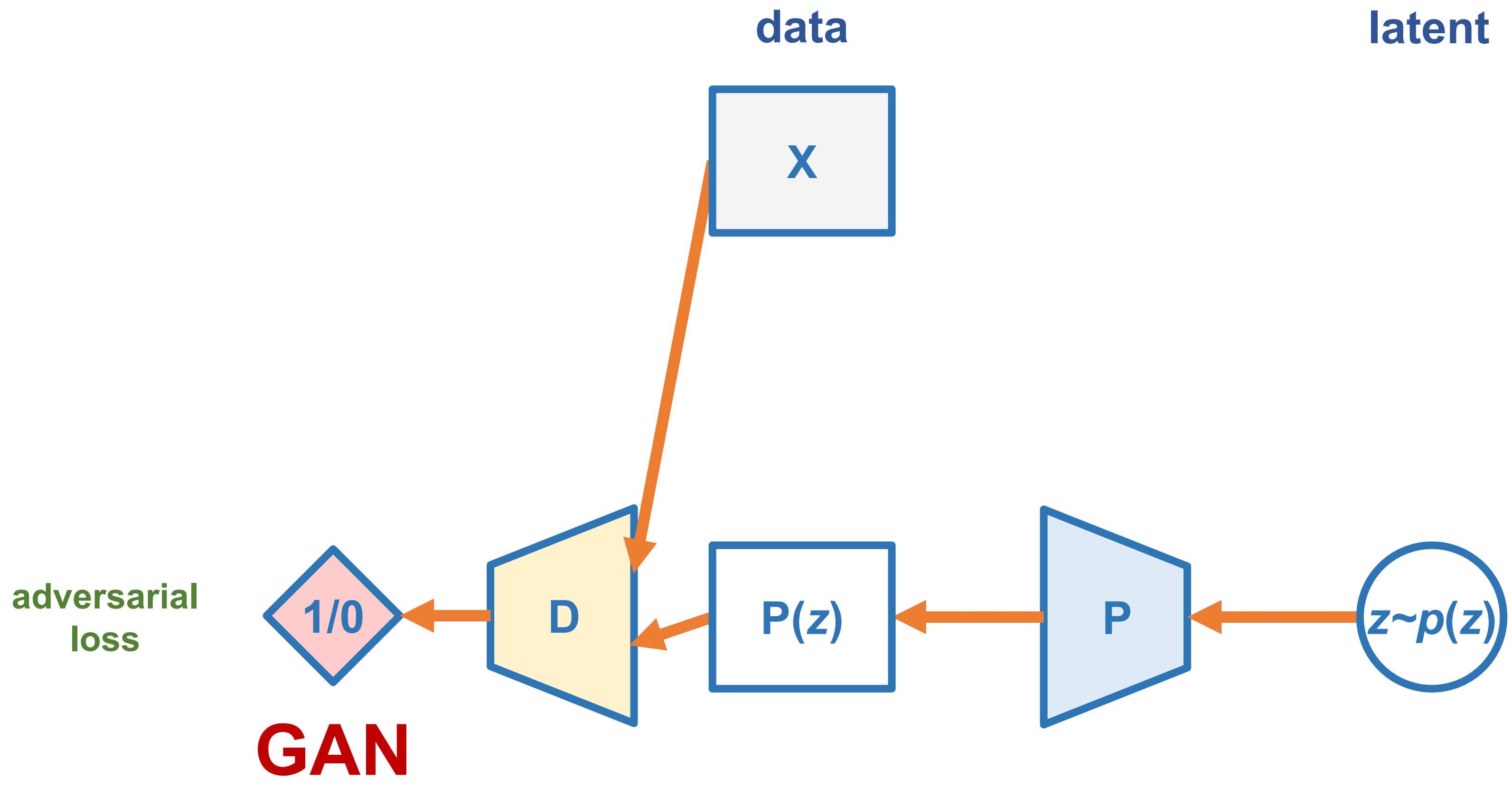
What's going on?



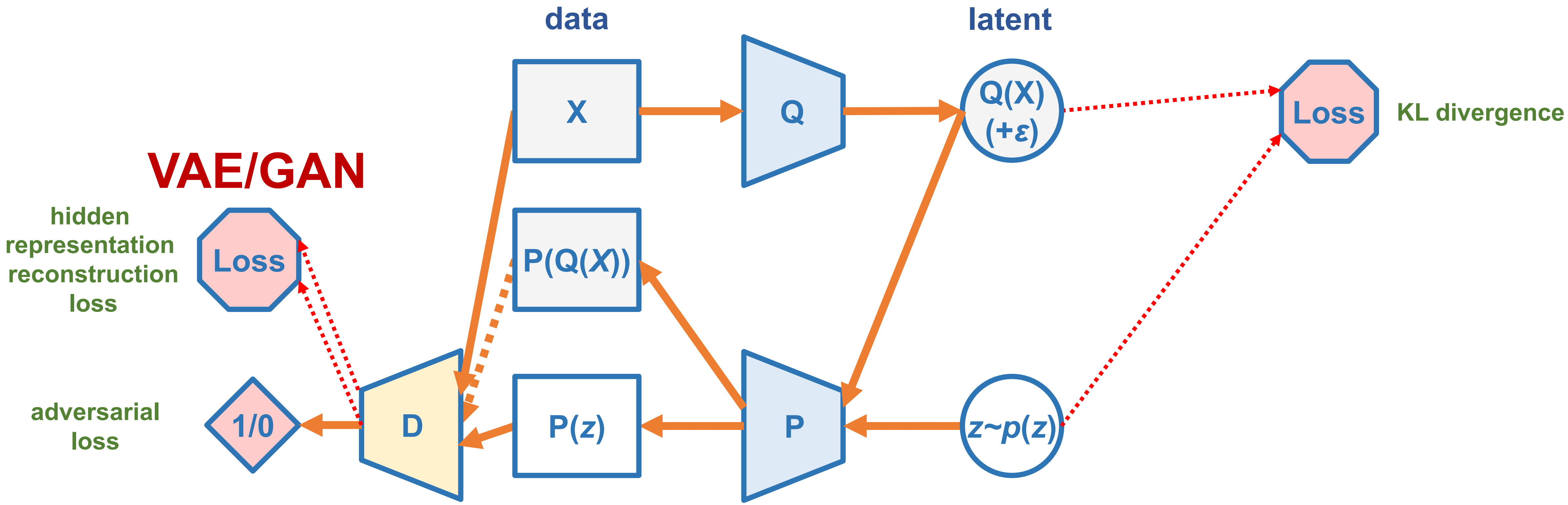
What's going on?



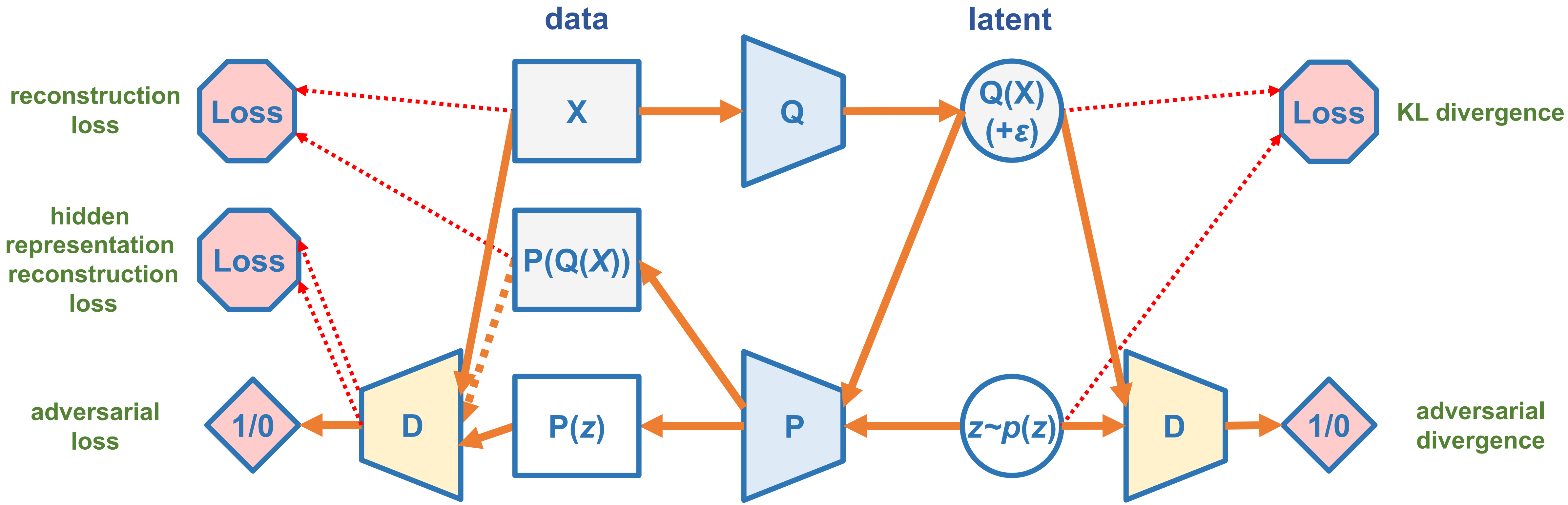
What’s going on?



What's going on?

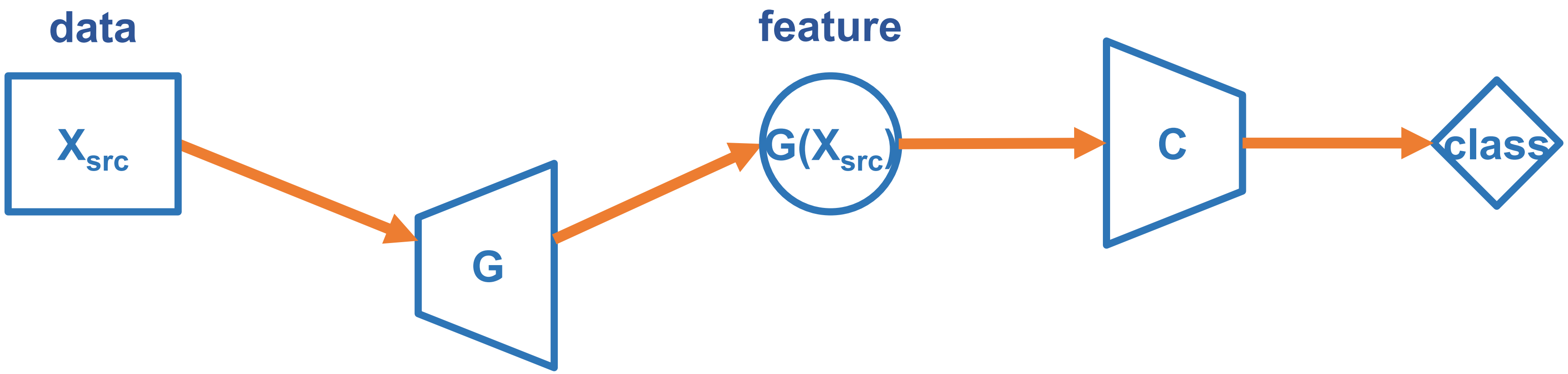


What's going on?



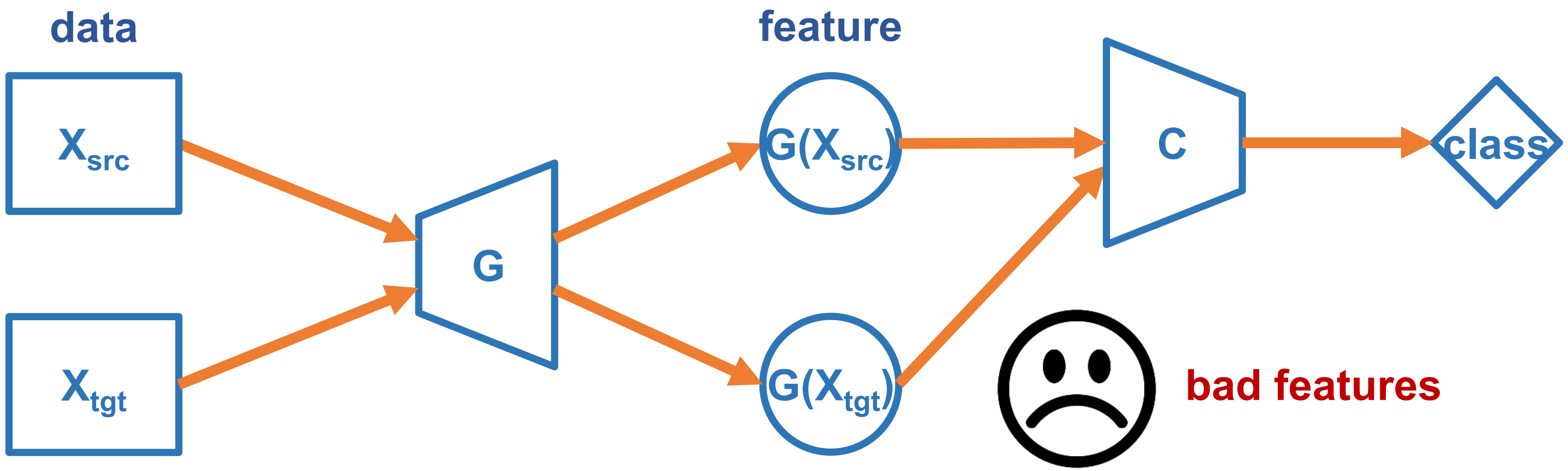
ADA (Adversarial Domain Adaption)

- Goal: given labeled data in source domain, aim to classify unlabeled data in target domain.



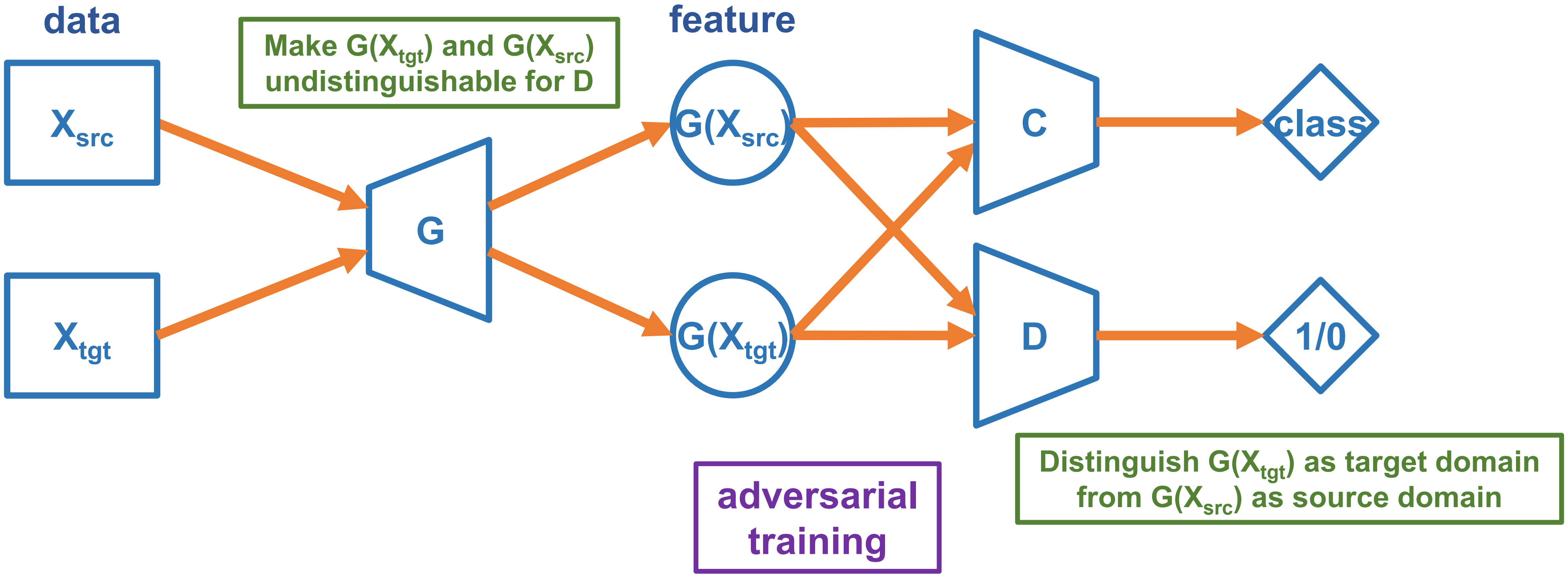
ADA (Adversarial Domain Adaption)

- Goal: given labeled data in source domain, aim to classify unlabeled data in target domain.



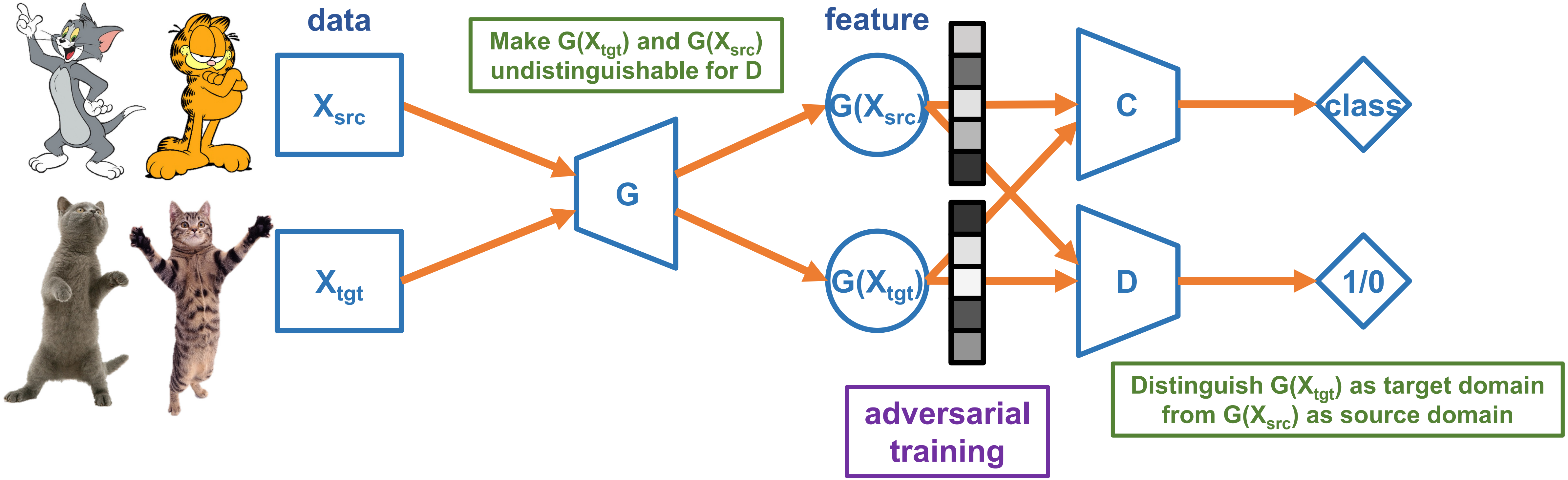
ADA (Adversarial Domain Adaption)

- Goal: given labeled data in source domain, aim to classify unlabeled data in target domain.



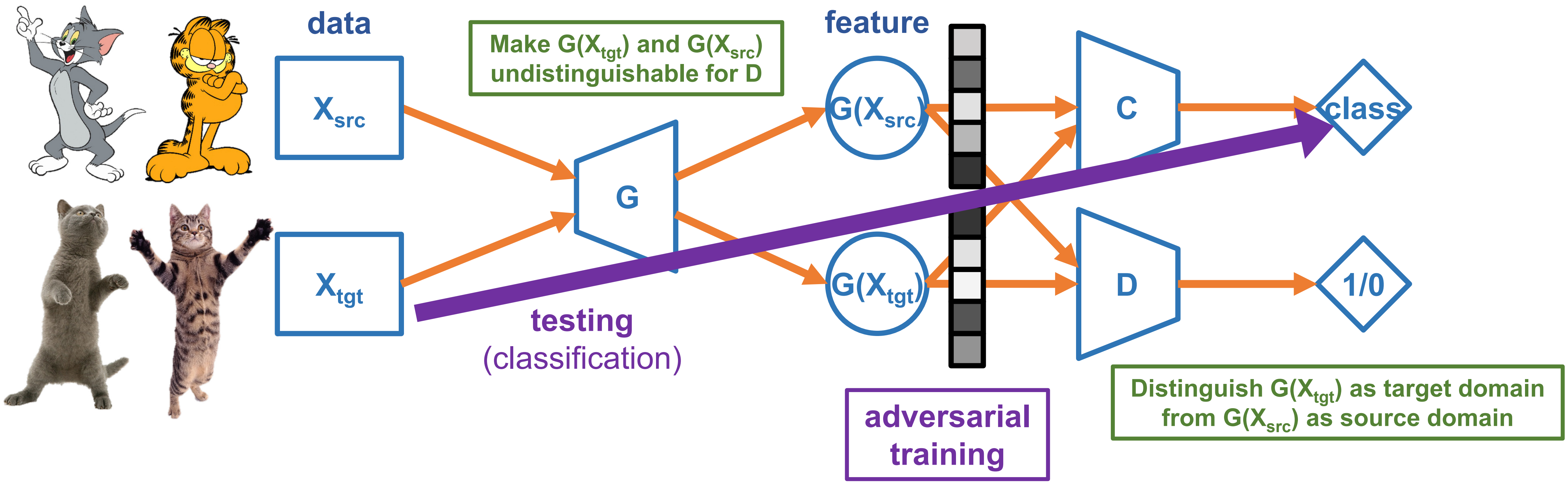
ADA (Adversarial Domain Adaption)

- Goal: given labeled data in source domain, aim to classify unlabeled data in target domain.

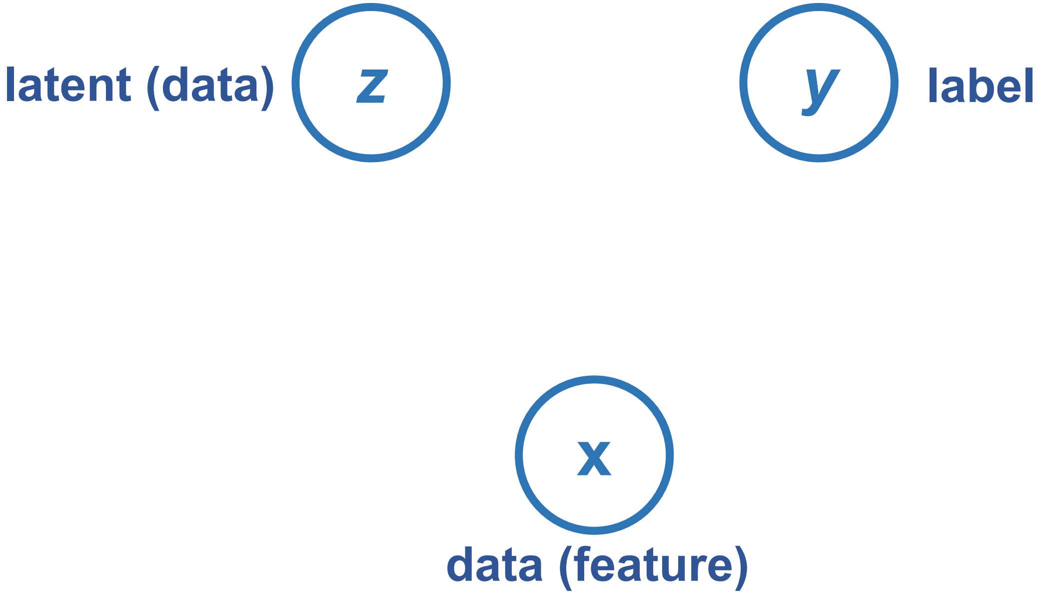


ADA (Adversarial Domain Adaption)

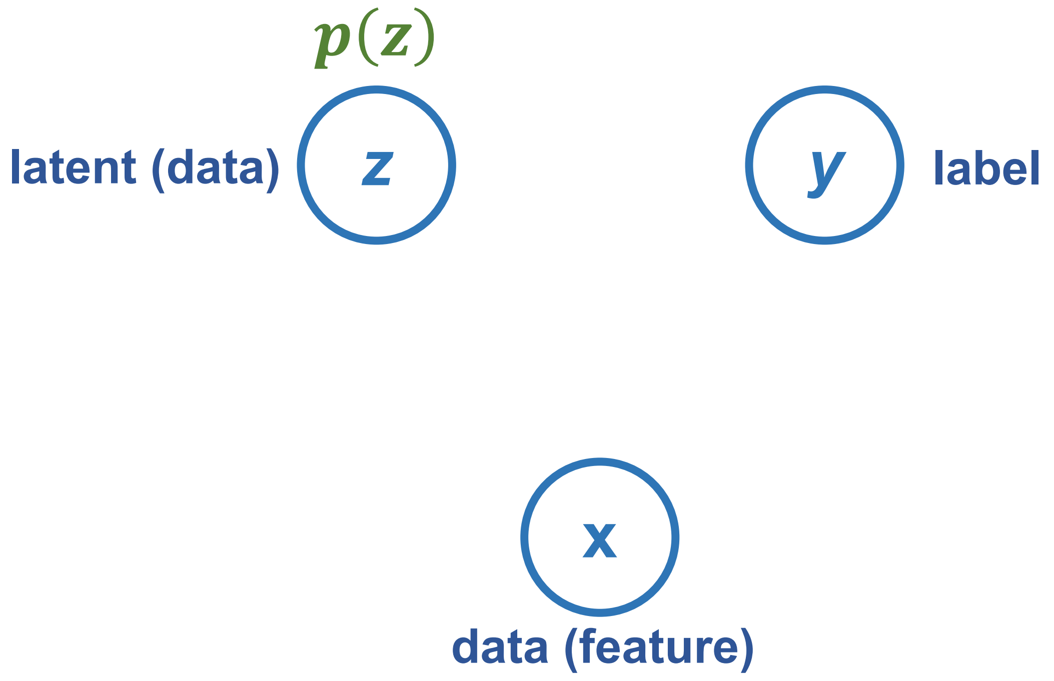
- Goal: given labeled data in source domain, aim to classify unlabeled data in target domain.



On Unifying Deep Generative Models

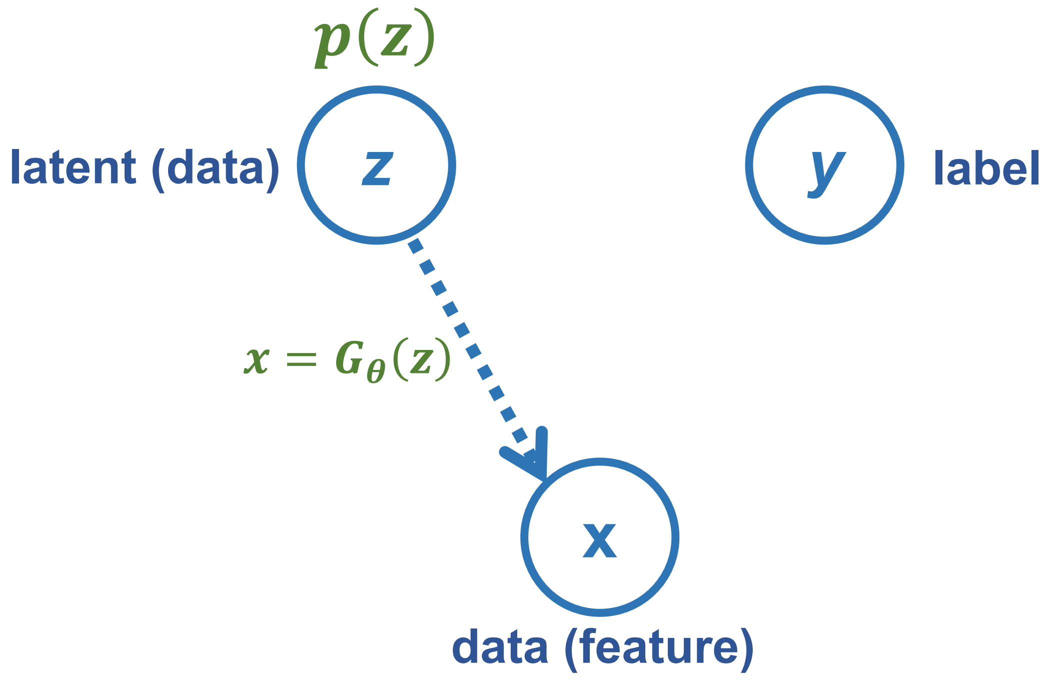


On Unifying Deep Generative Models



On Unifying Deep Generative Models

- $\boldsymbol{G_\theta} - \theta$ are parameters in generator
- $\boldsymbol{D_\phi} - \phi$ are parameters in generator

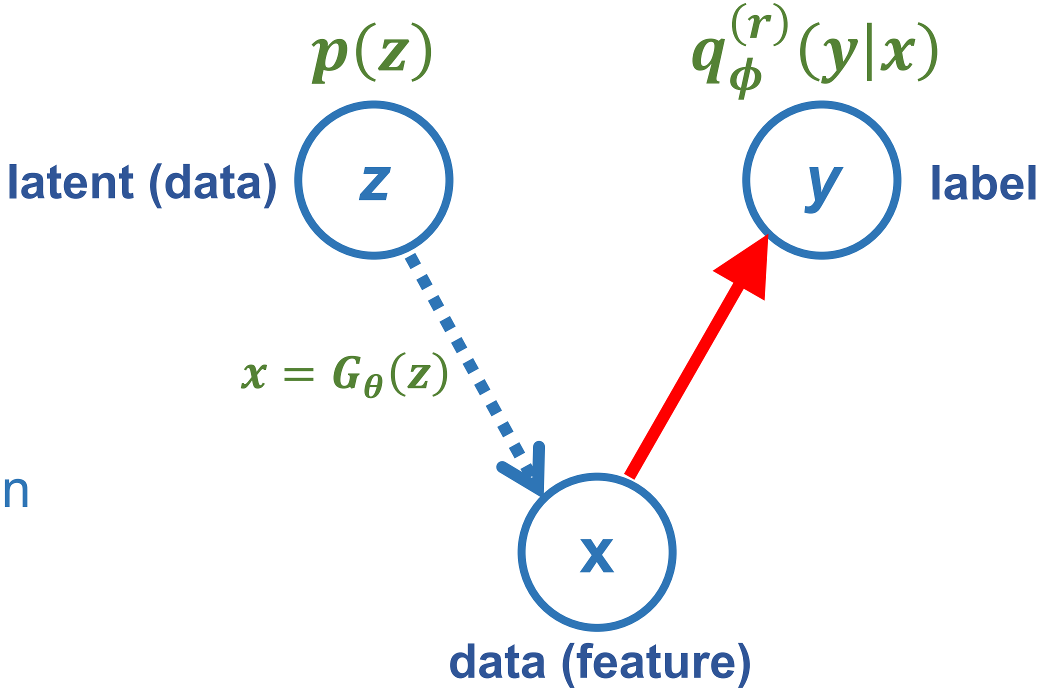


On Unifying Deep Generative Models

- G_θ – θ are parameters in generator
- D_ϕ – ϕ are parameters in generator

- **Solid line** – generative process
- **Dashed line** – inference process
- **Hollow arrow** – deterministic transformation
- **Red arrow** – adversarial mechanism

- $q_\phi^{(r)}(y|x)$ denotes $q_\phi(y|x)$ and $q_\phi(1-y|x)$



GAN $y = \begin{cases} 1, & if\ x\ is\ real \\ 0, & if\ x\ is\ fake \end{cases}$

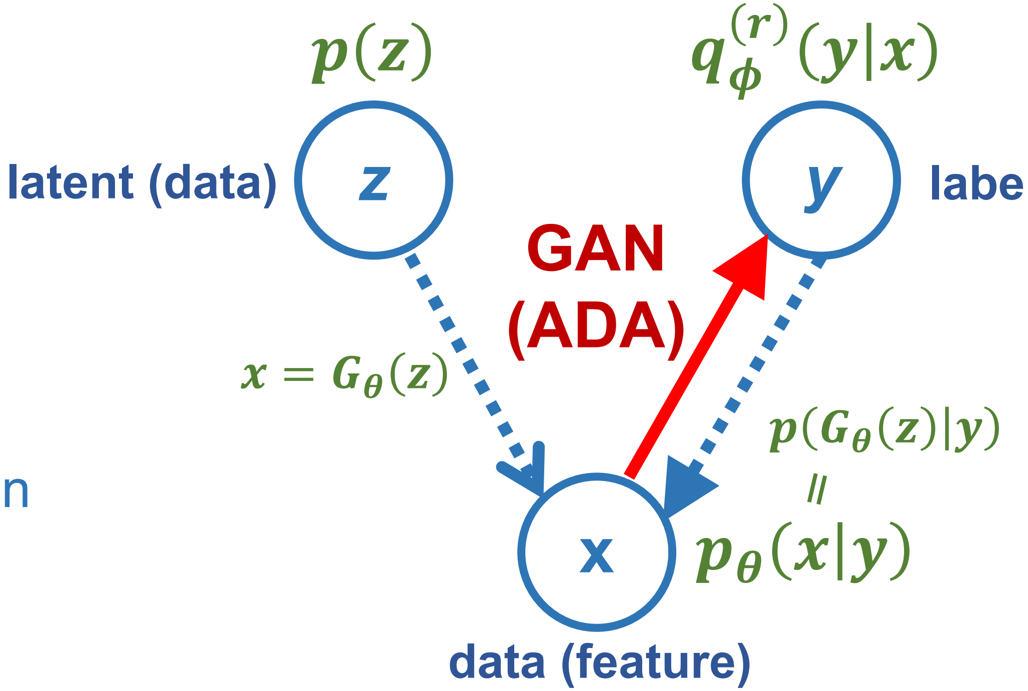
ADA $y = \begin{cases} 1, & if\ x\ is\ in\ source\ domain \\ 0, & if\ x\ is\ in\ target\ domain \end{cases}$

On Unifying Deep Generative Models

- G_θ – θ are parameters in generator
- D_ϕ – ϕ are parameters in generator

- **Solid line** – generative process
- **Dashed line** – inference process
- **Hollow arrow** – deterministic transformation
- **Red arrow** – adversarial mechanism

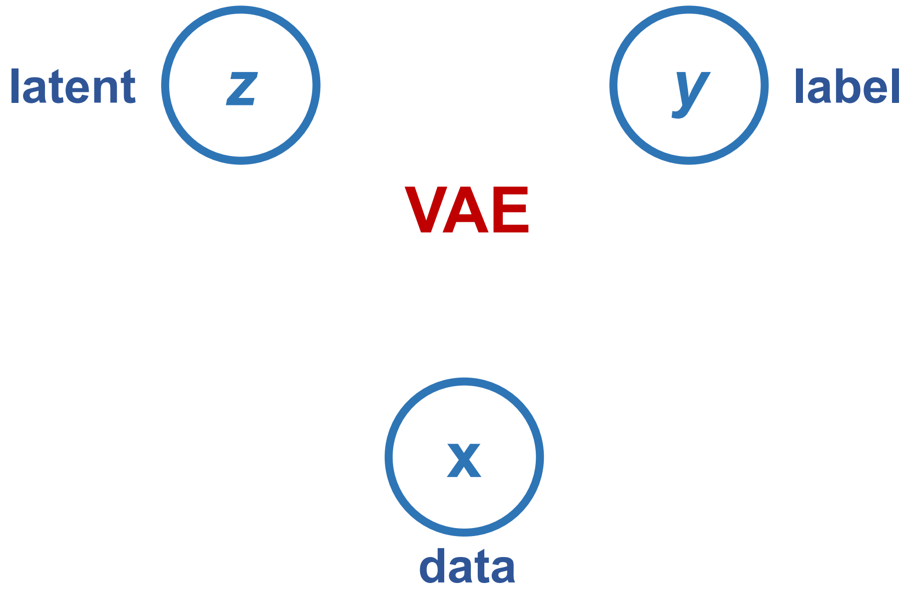
- $q_\phi^{(r)}(y|x)$ denotes $q_\phi(y|x)$ and $q_\phi(1-y|x)$



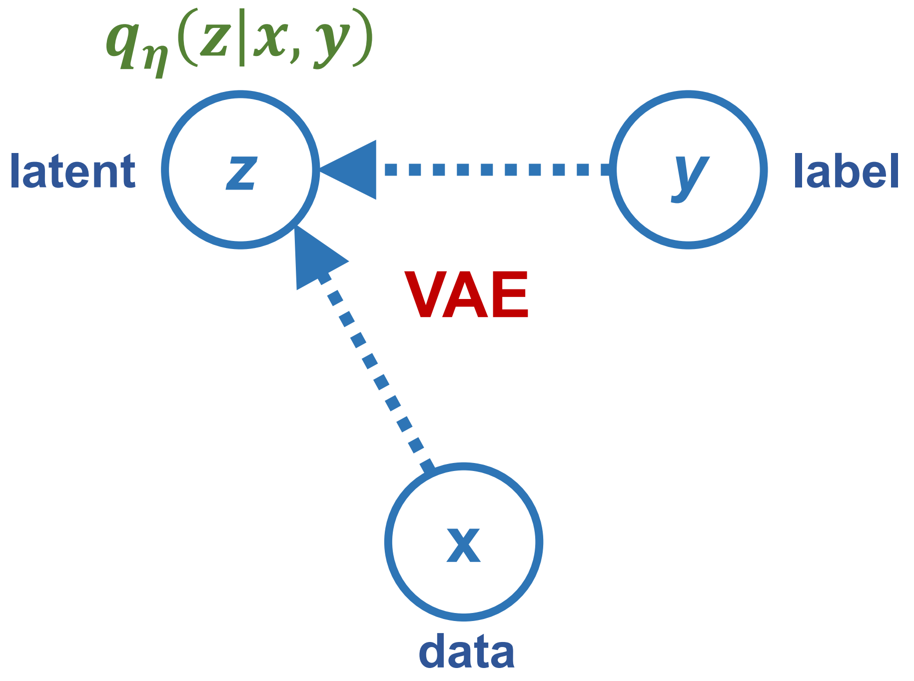
GAN $y = \begin{cases} 1, & if\ x\ is\ real \\ 0, & if\ x\ is\ fake \end{cases}$

ADA $y = \begin{cases} 1, & if\ x\ is\ in\ source\ domain \\ 0, & if\ x\ is\ in\ target\ domain \end{cases}$

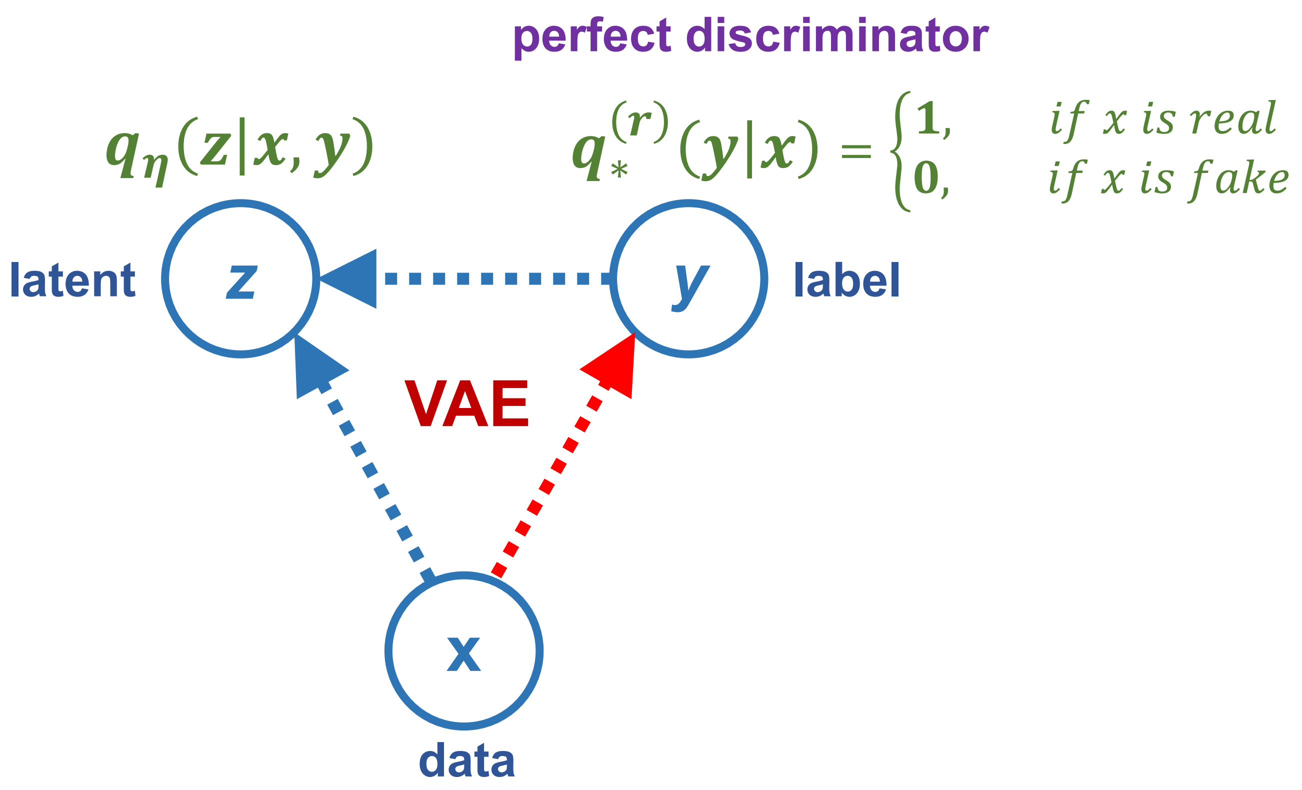
On Unifying Deep Generative Models



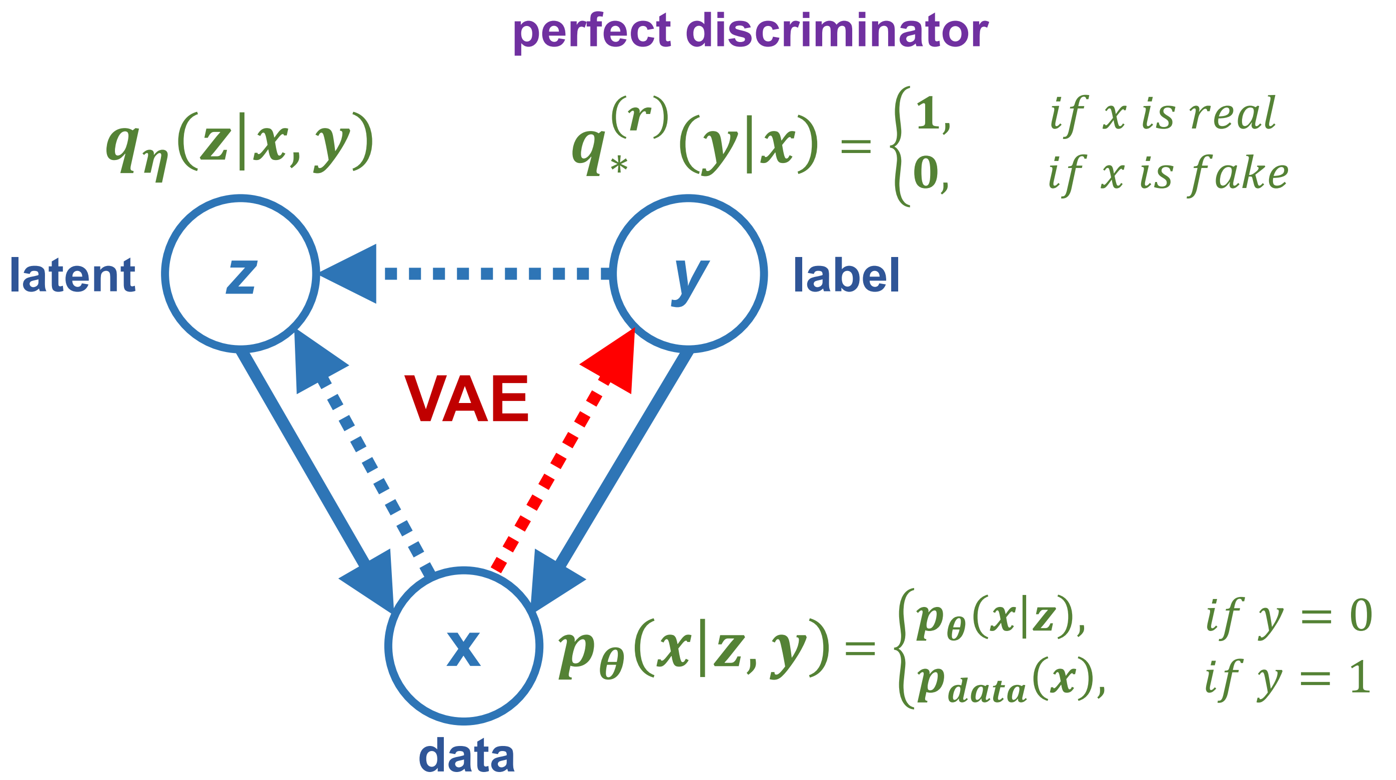
On Unifying Deep Generative Models



On Unifying Deep Generative Models



On Unifying Deep Generative Models

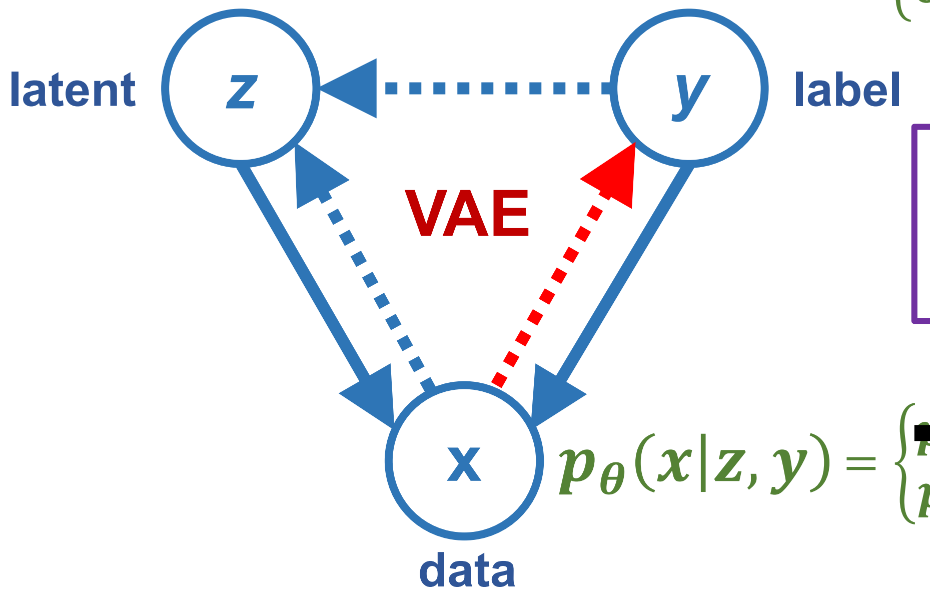


On Unifying Deep Generative Models

perfect discriminator

$$q_*^{(r)}(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} \mathbf{1}, & if\ x\ is\ real \\ \cancel{\mathbf{0}}, & \cancel{if\ x\ is\ fake} \end{cases}$$

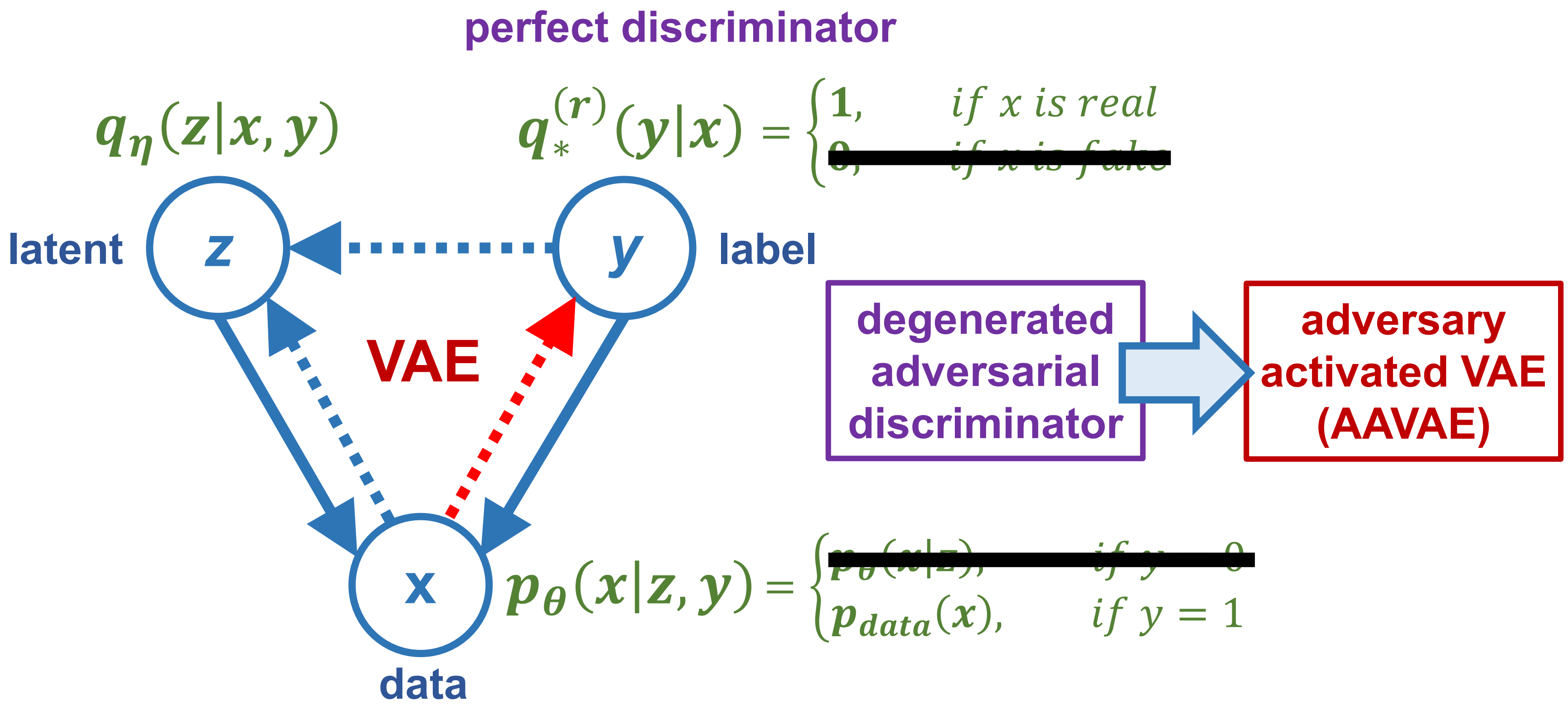
$$q_\eta(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{y})$$



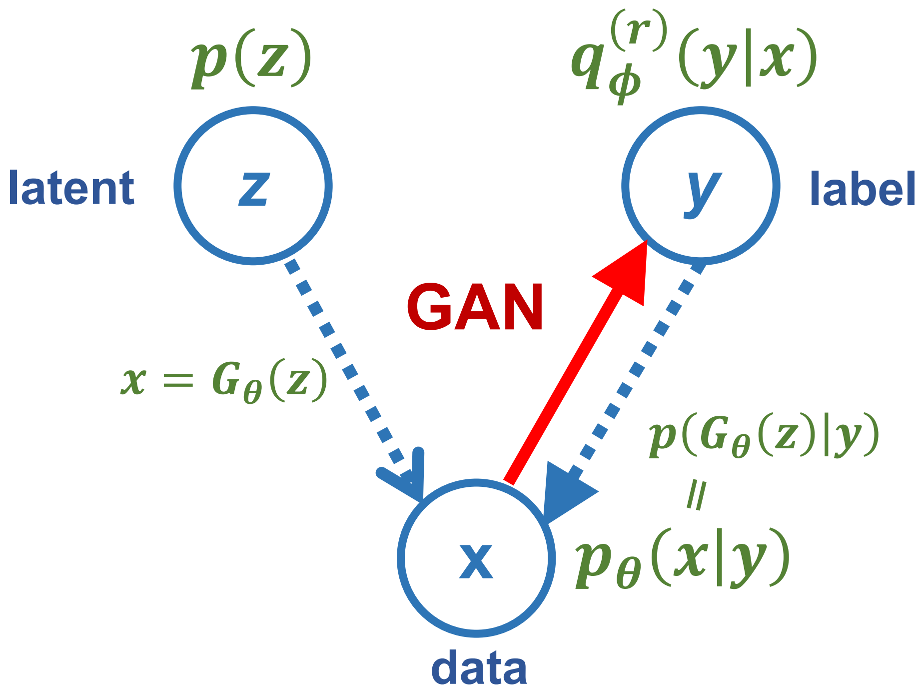
degenerated adversarial discriminator

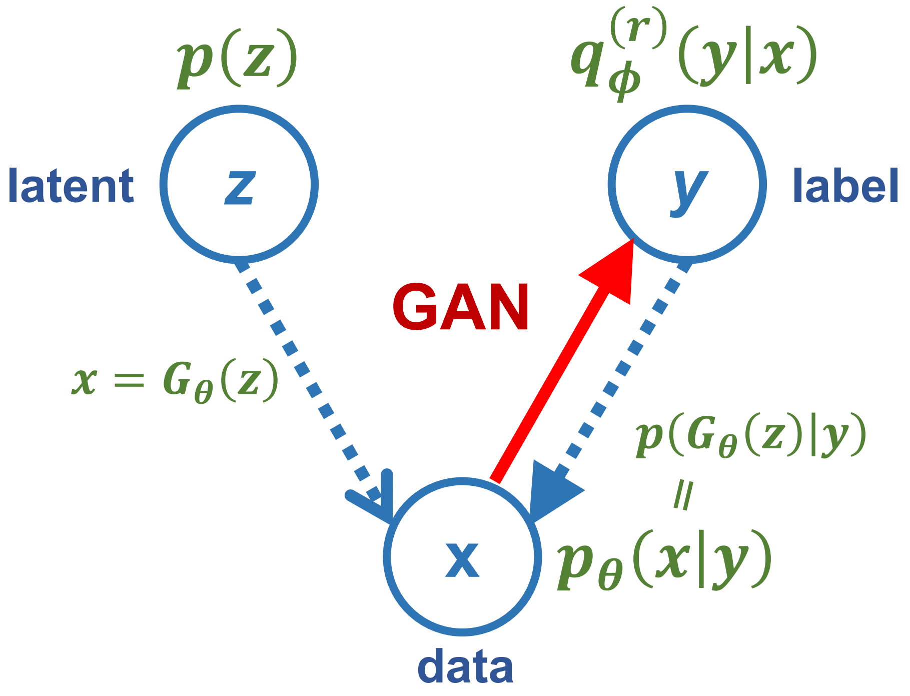
$$p_\theta(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{y}) = \begin{cases} \cancel{p_\theta(\boldsymbol{x}|\boldsymbol{z})}, & \cancel{if\ y = 0} \\ p_{data}(\boldsymbol{x}), & if\ y = 1 \end{cases}$$

On Unifying Deep Generative Models

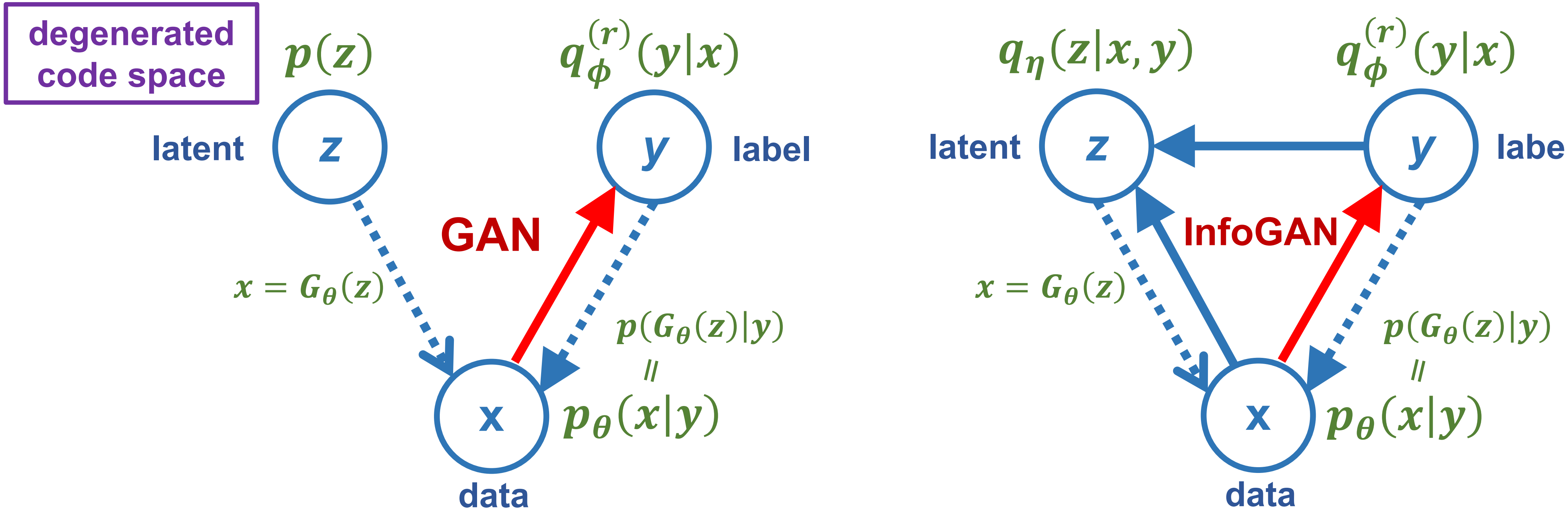


On Unifying Deep Generative Models

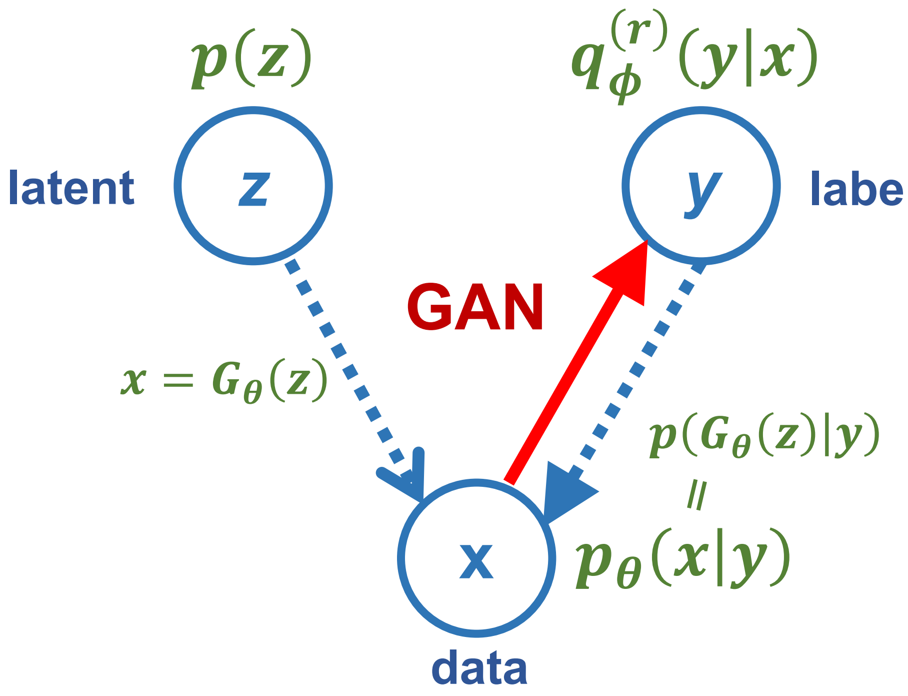


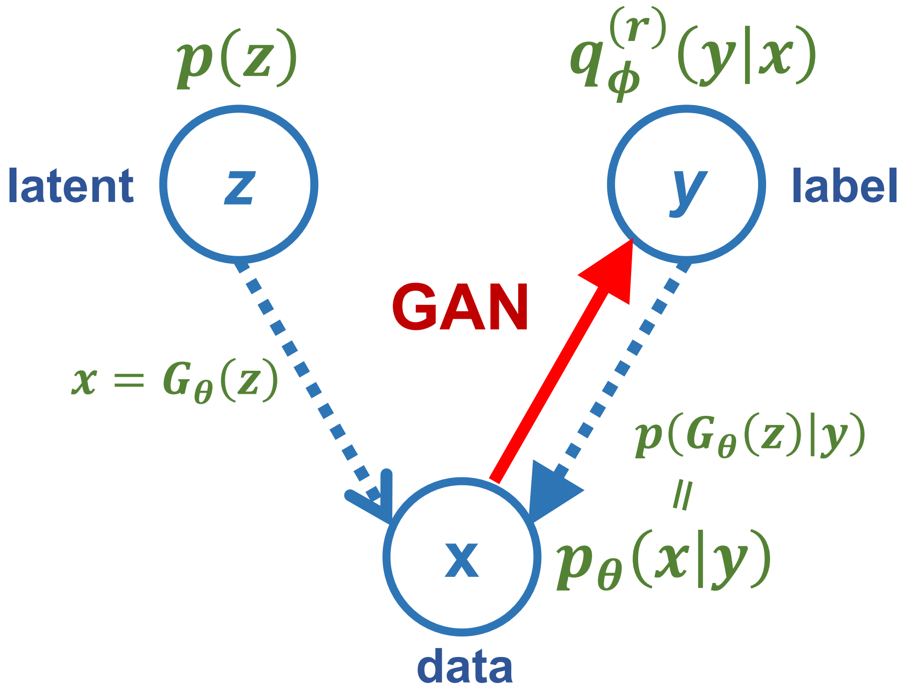


On Unifying Deep Generative Models

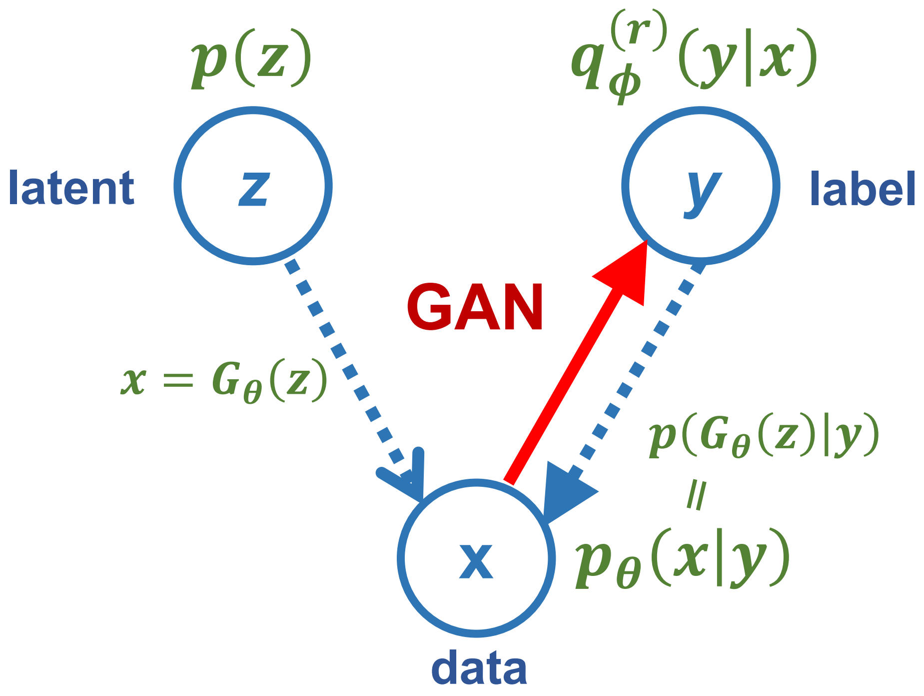


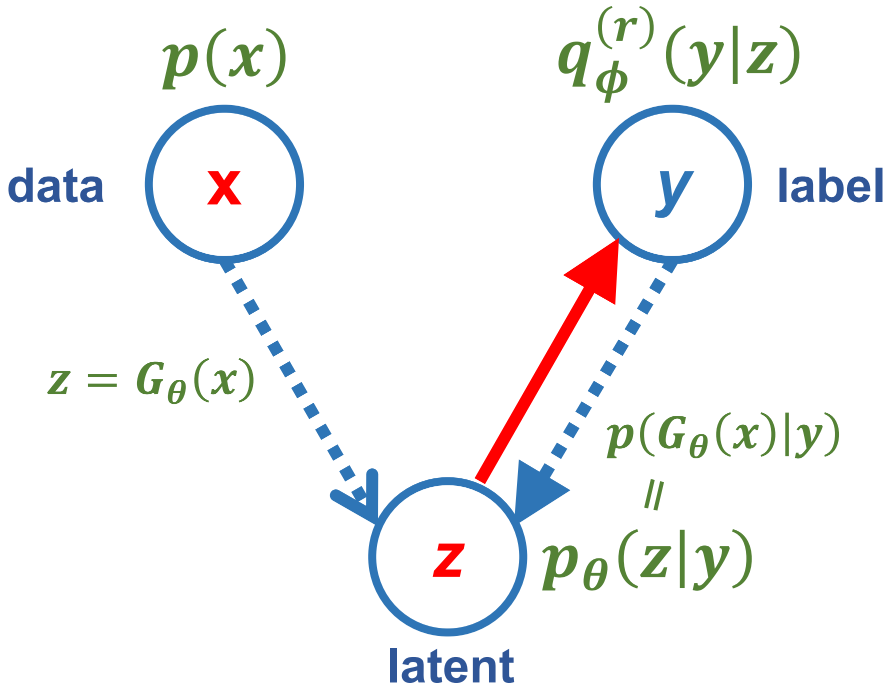
On Unifying Deep Generative Models



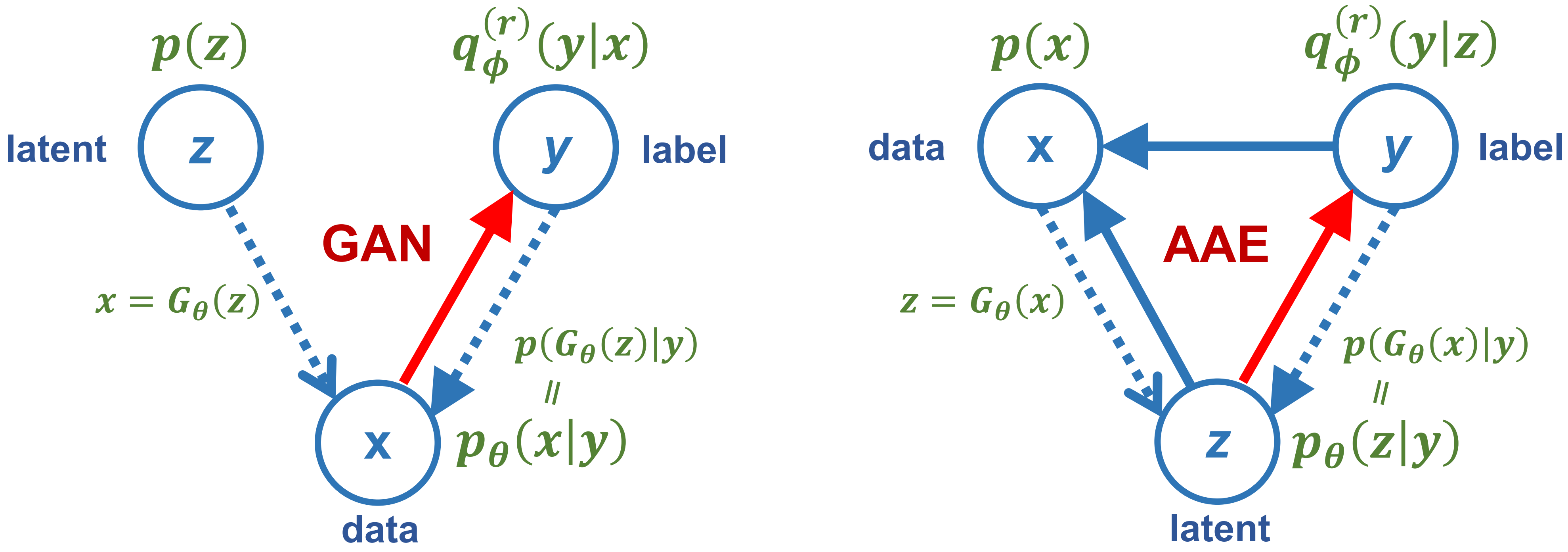


On Unifying Deep Generative Models

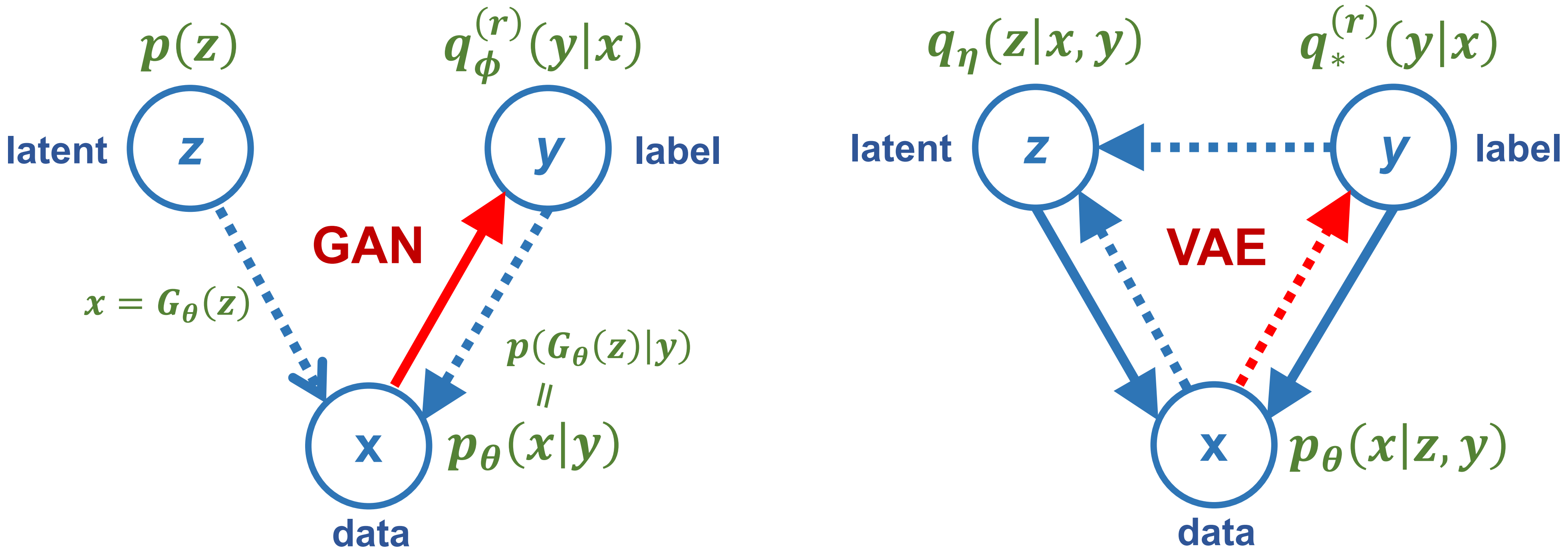




On Unifying Deep Generative Models

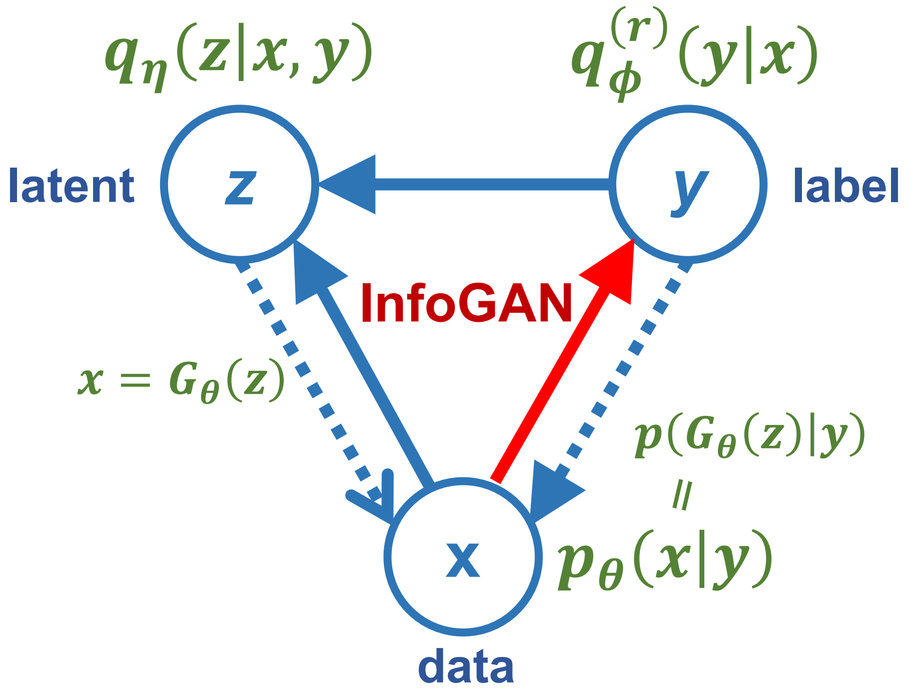


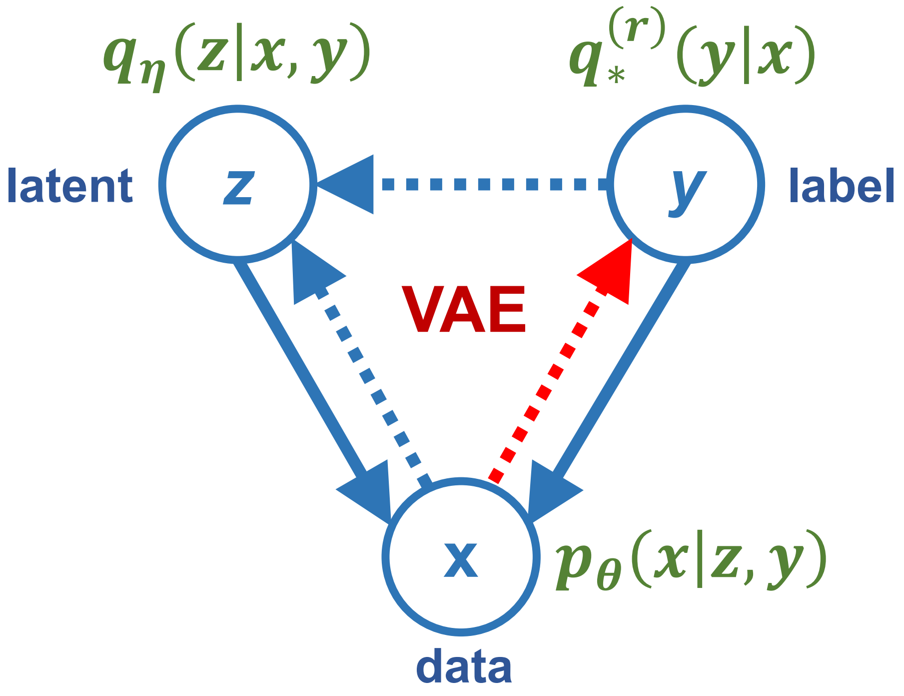
GAN vs VAE



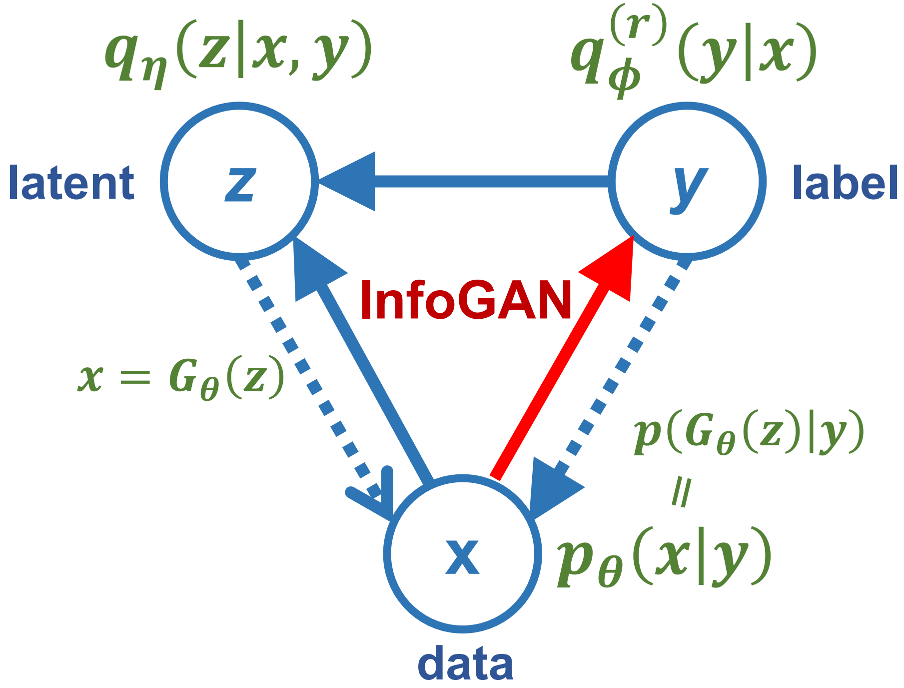
$p(z)$
latent
z
$x = G_\theta(z)$
GAN
$q_\phi^{(r)}(y|x)$
y
label
$p(G_\theta(z)|y)$
$||$
x
$p_\theta(x|y)$
data
$q_\eta(z|x,y)$
latent
z
VAE
$q_*^{(r)}(y|x)$
y
label
x
$p_\theta(x|z,y)$
data

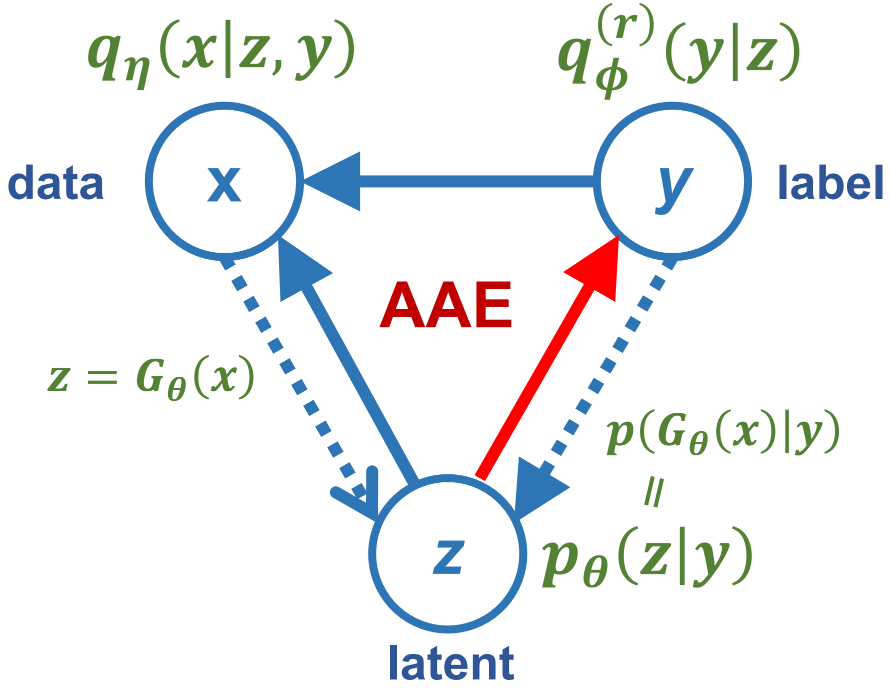
InfoGAN vs VAE





InfoGAN vs AAE





Wake-sleep Algorithm

- $\boldsymbol{h}$ - general latent variables
- $\boldsymbol{\lambda}$ - general parameters
- $\boldsymbol{\theta}$ - generator parameters

$$\text{Wake:} \quad \max_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{q_\lambda(h|x)p_{data}(x)}} [\log p_\theta(x|h)]$$
$$\text{Sleep:} \quad \max_{\lambda} \mathbb{E}_{\boldsymbol{p_\theta(x|h)p(h)}} [\log q_\lambda(h|x)]$$

- In **wake** phase, update $\boldsymbol{\theta}$ by fitting $\boldsymbol{p_\theta(x|h)}$ to $\boldsymbol{x}$ and $\boldsymbol{h}$ inferred by $\boldsymbol{q_\lambda(h|x)}$.
- In **sleep** phase, update $\boldsymbol{\lambda}$ based on generated samples.

- **VAE:** $\boldsymbol{h} \to \boldsymbol{z}, \boldsymbol{\lambda} \to \boldsymbol{\eta}$
- **GAN**: $\boldsymbol{h} \to \boldsymbol{y}, \boldsymbol{\lambda} \to \boldsymbol{\phi}$

References

- D. P. Kingma, M. Welling. **Auto-Encoding Variational Bayes**. arXiv preprint arXiv:1312.6114, 2014
- I. J. GoodFellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. **Genrative Adversrial Nets**. arXiv:preprint arXiv:1406.2661, 2014
- A. B. L. Larsen, S. K. Sønderby and O. Winther. **Autoencoding beyond pixels using a learned similarity metric**. arXiv preprint arXiv:1512.09300, 2015
- A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow and B. Frey. **Adversarial Autoencoders**. arXiv preprint arXiv:1511.05644, 2016
- Z. Hu, Z. Yang, R. Salakhutdinov and E. P. Xing. **On unifying deep generative models**. arXiv preprint arXiv:1706.00550, 2017