Feature Extraction

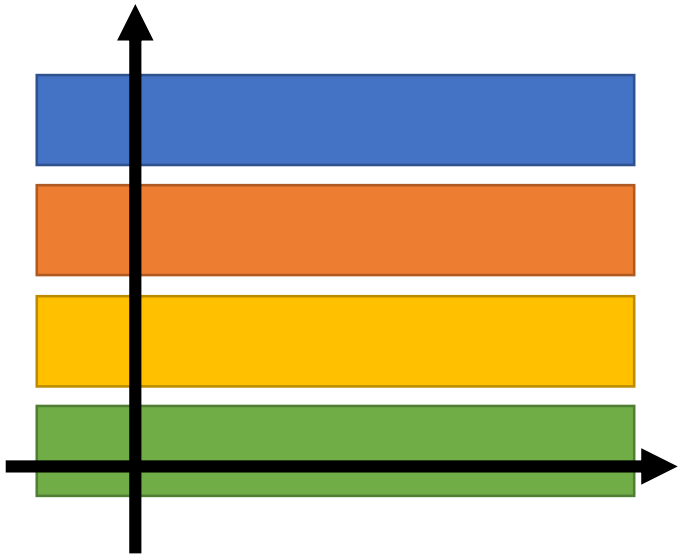
InfoGAN

(The colors represents the characteristics.)

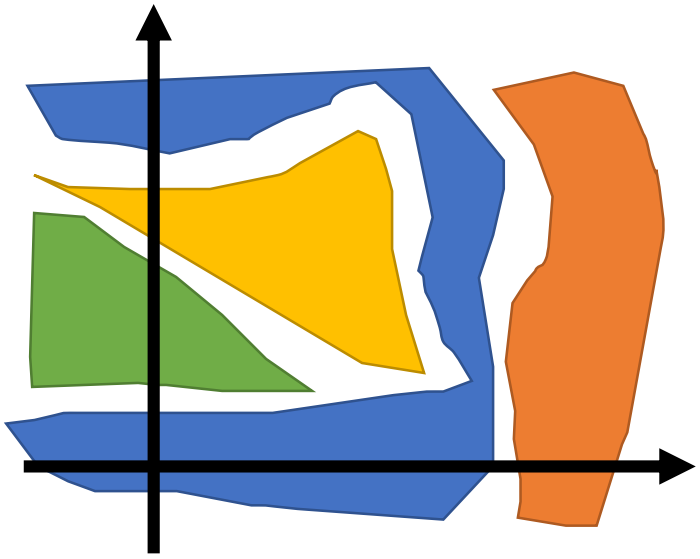
Regular GAN

Modifying a specific dimension, no clear meaning

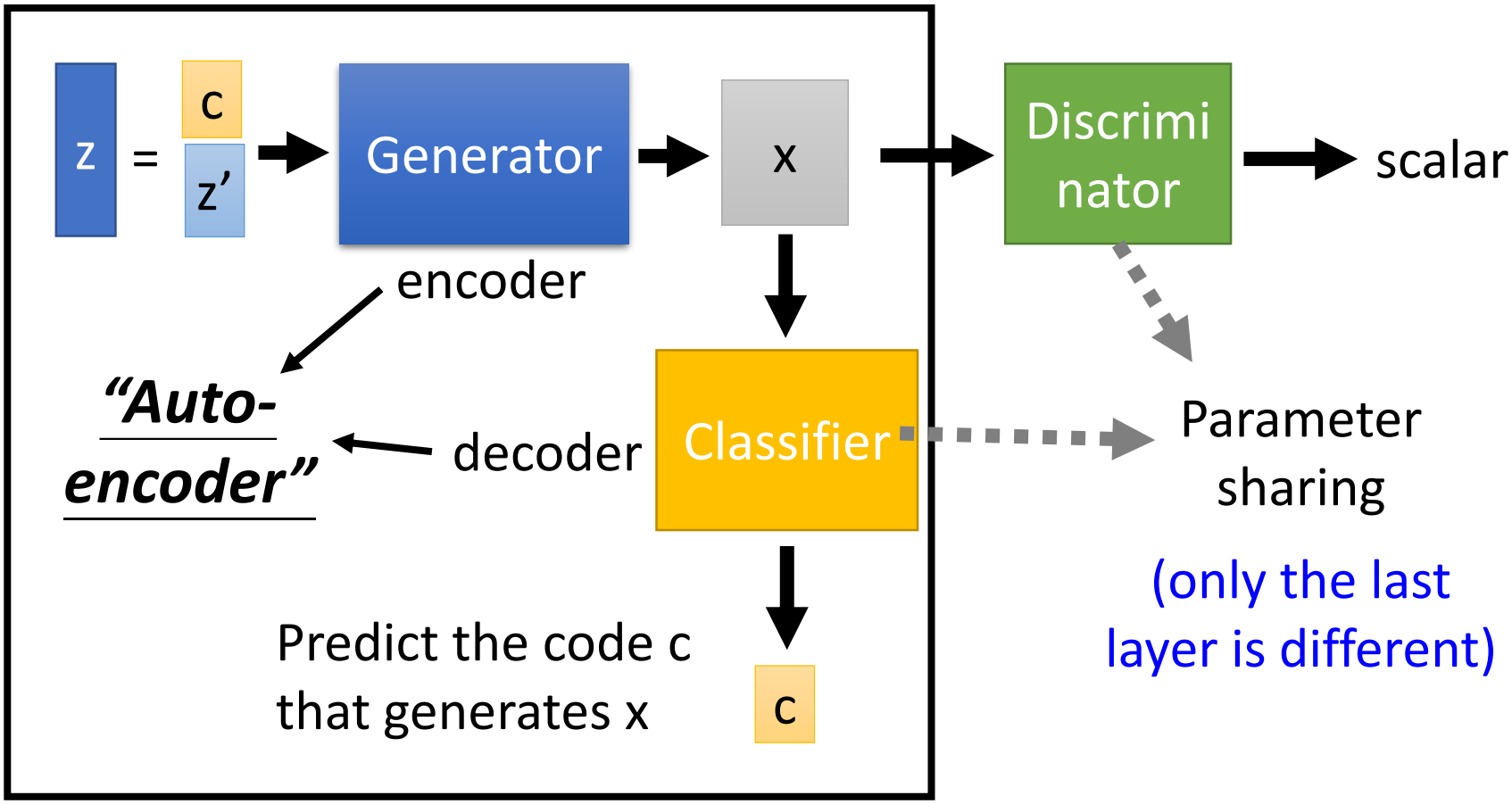
What we expect



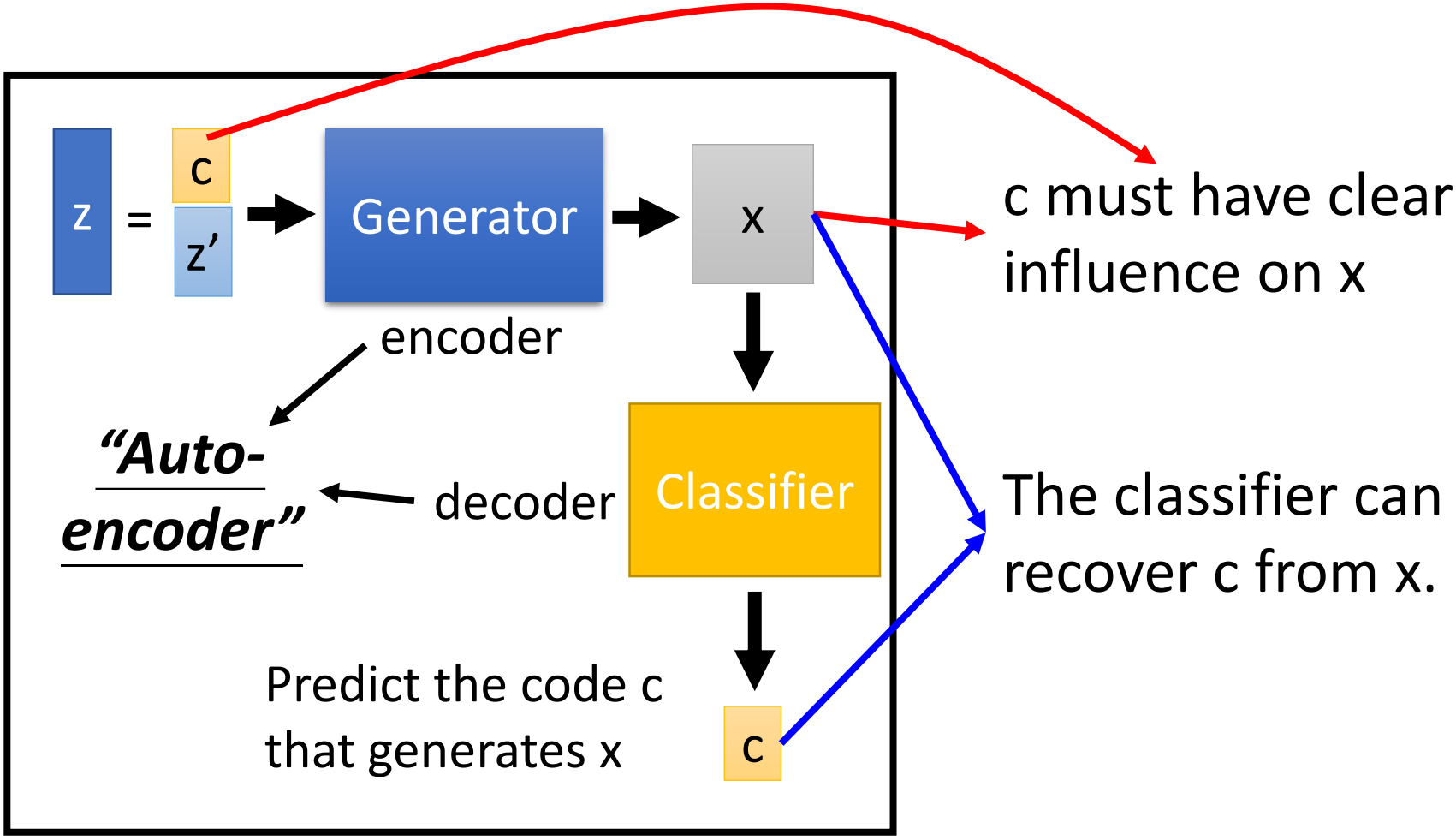
Actually …

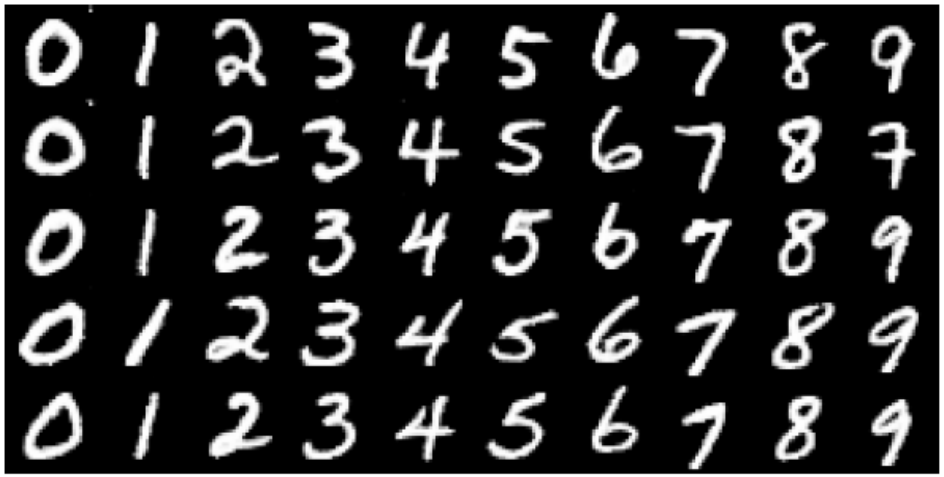


What is InfoGAN?

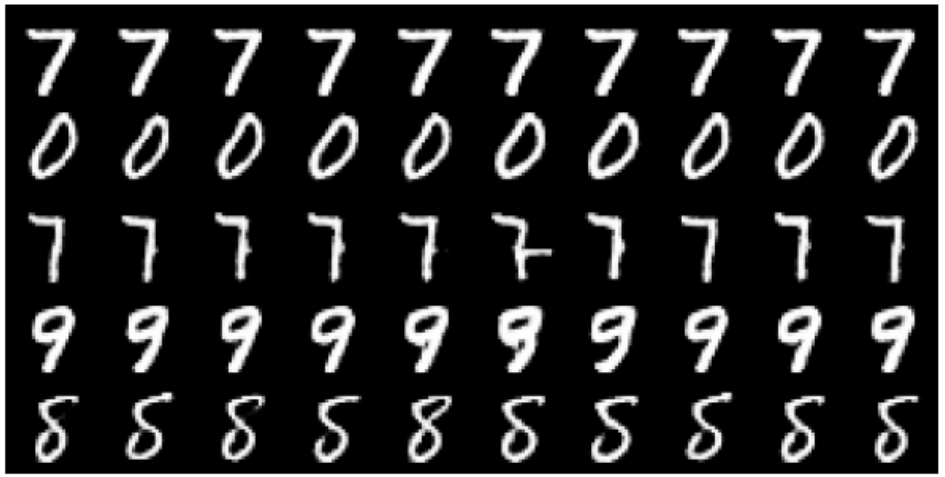


What is InfoGAN?

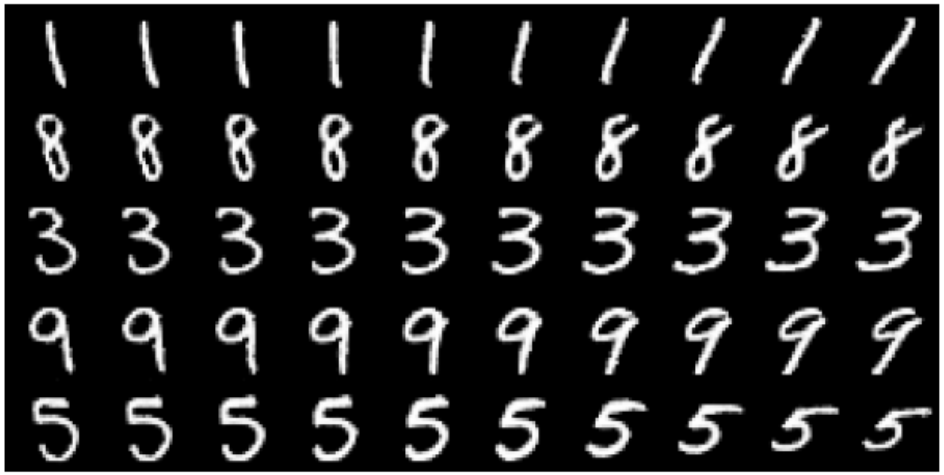




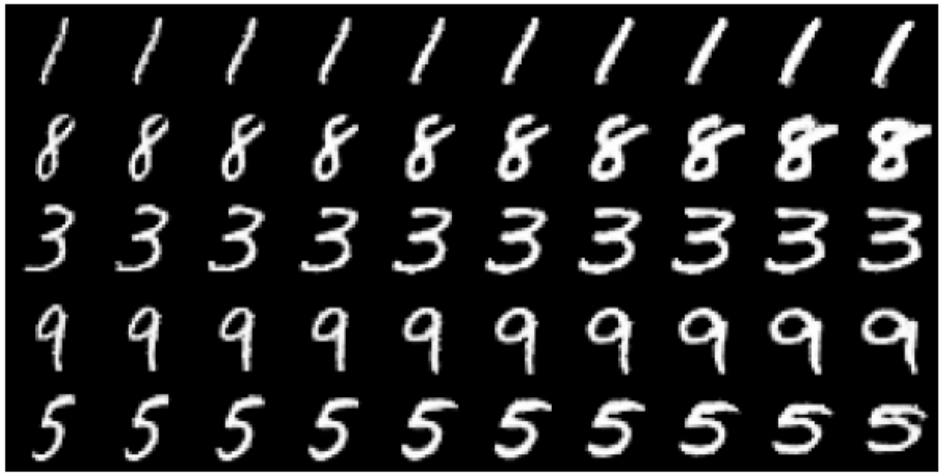
(a) Varying c_1 on InfoGAN (Digit type)



(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)

https://arxiv.org/abs/1606.03657

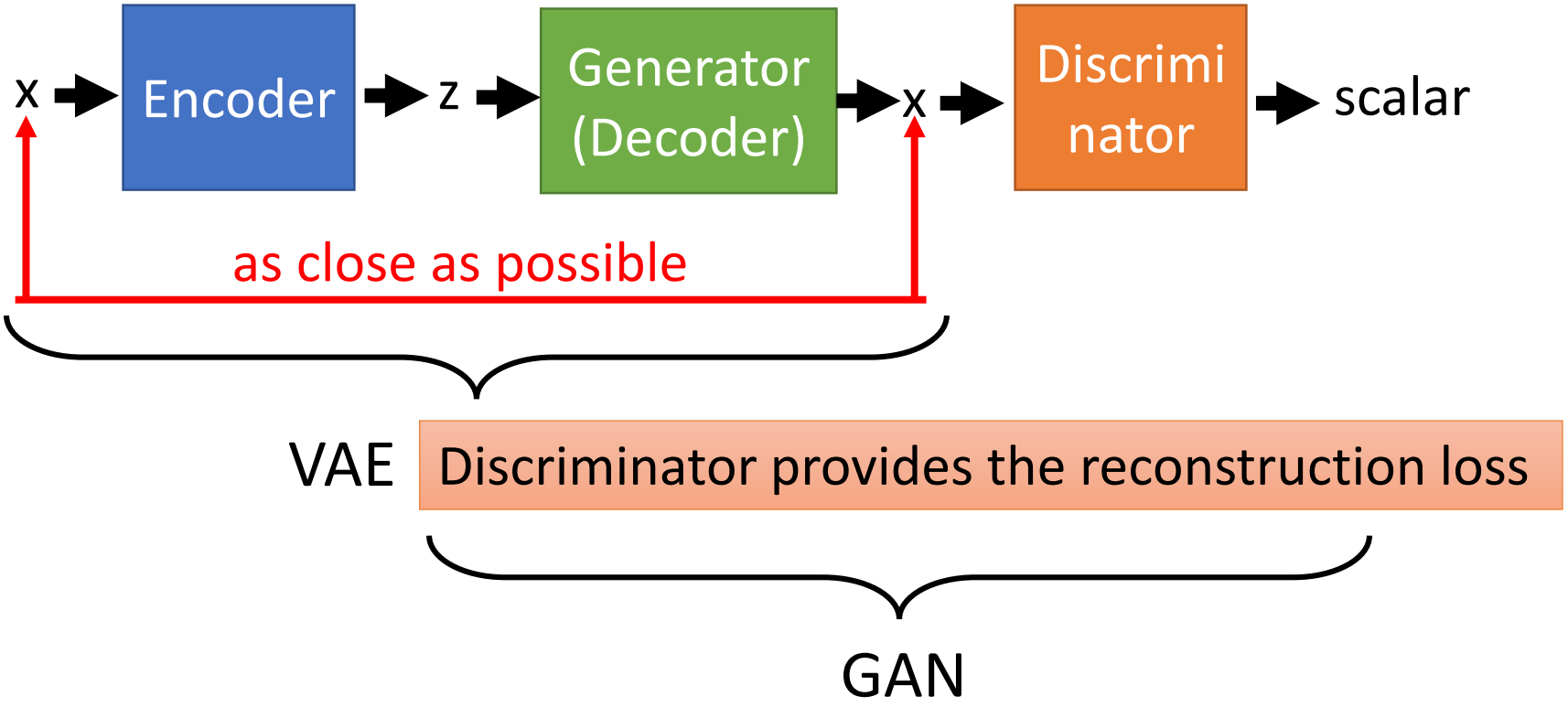
VAE-GAN

Anders Boesen, Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, Ole Winther, "Autoencoding beyond pixels using a learned similarity metric", ICML. 2016

- Minimize reconstruction error
- z close to normal

- Minimize reconstruction error
- Cheat discriminator

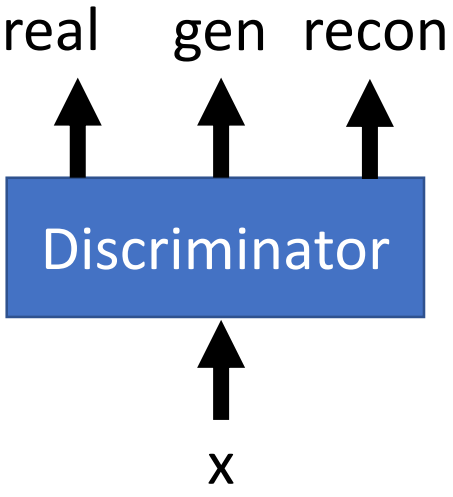
- Discriminate real, generated and reconstructed images



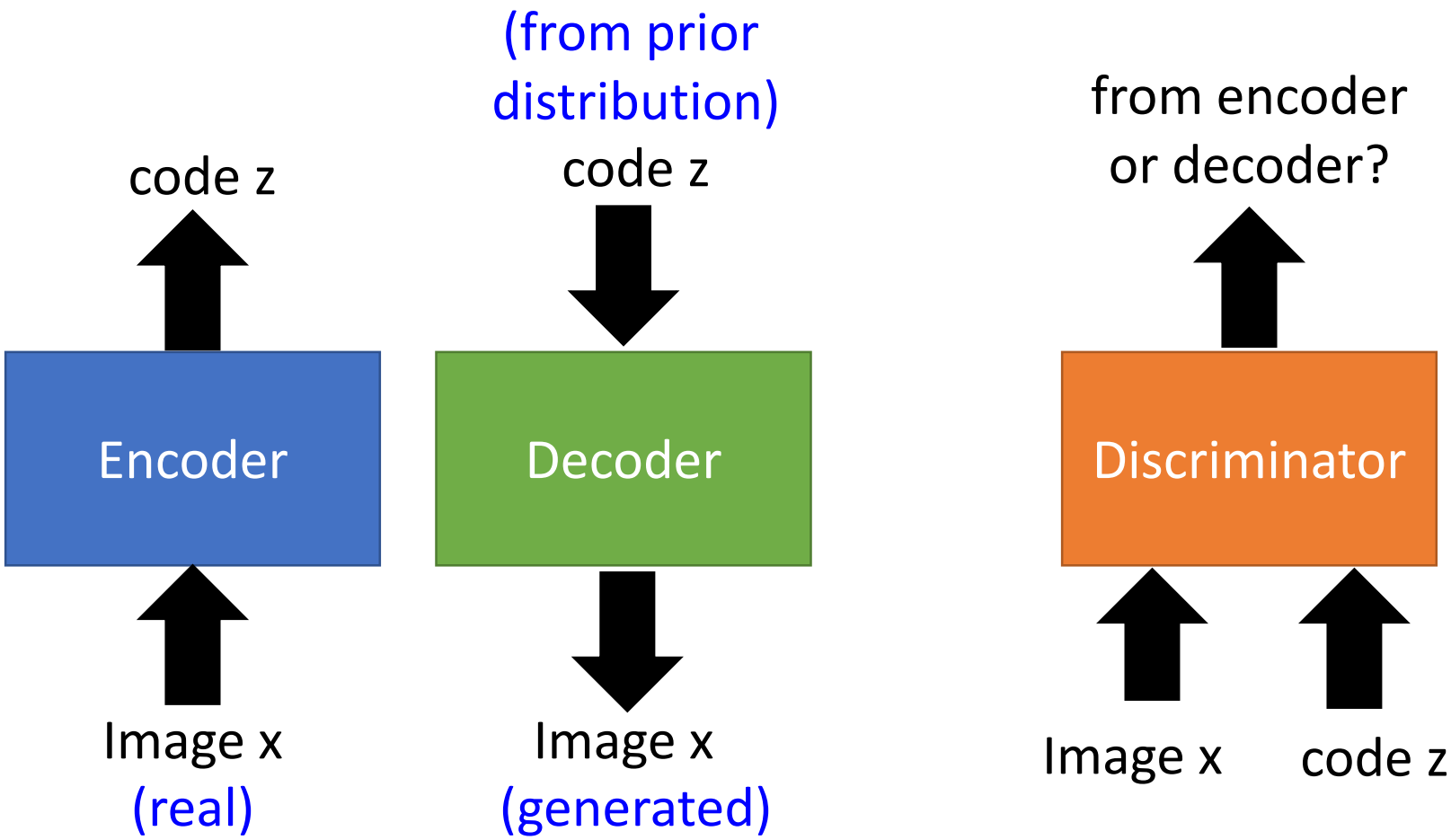
Algorithm

- Initialize En, De, Dis
- In each iteration:
  - Sample M images $x^1, x^2, \cdots, x^M$ from database
  - Generate M codes $\tilde{z}^1, \tilde{z}^2, \cdots, \tilde{z}^M$ from encoder
    - $\tilde{z}^i = En(x^i)$
  - Generate M images $\tilde{x}^1, \tilde{x}^2, \cdots, \tilde{x}^M$ from decoder
    - $\tilde{x}^i = De(\tilde{z}^i)$
  - Sample M codes $z^1, z^2, \cdots, z^M$ from prior P(z)
  - Generate M images $\hat{x}^1, \hat{x}^2, \cdots, \hat{x}^M$ from decoder
    - $\hat{x}^i = De(z^i)$
  - Update En to decrease $\|\tilde{x}^i - x^i\|$, decrease KL(P($\tilde{z}^i$|x^i)||P(z))
  - Update De to decrease $\|\tilde{x}^i - x^i\|$, increase $Dis(\tilde{x}^i)$ and $Dis(\hat{x}^i)$
  - Update Dis to increase $Dis(x^i)$, decrease $Dis(\tilde{x}^i)$ and $Dis(\hat{x}^i)$

Another kind of discriminator:

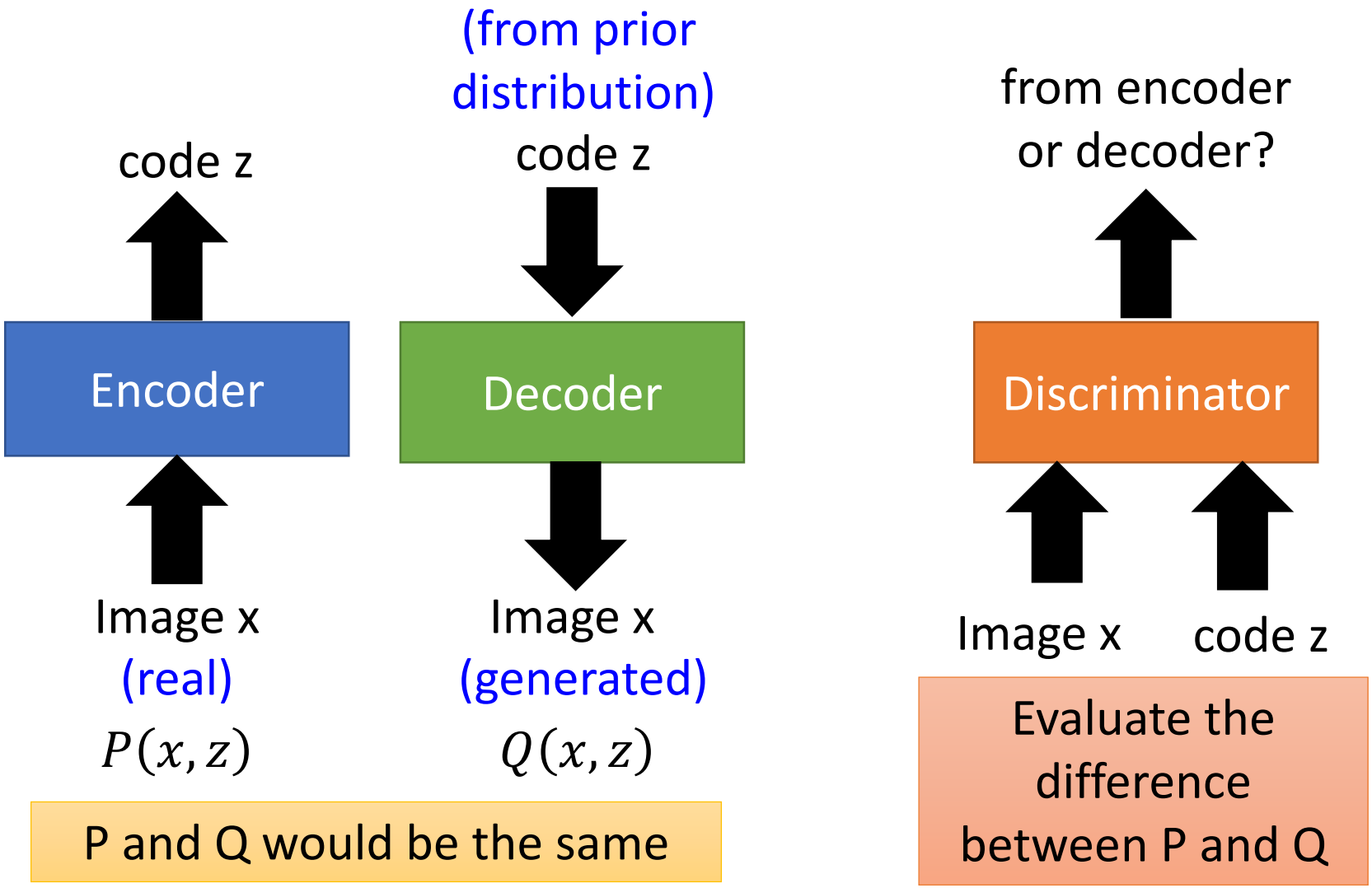


BiGAN


code z
Encoder
Image x
(real)
(from prior
distribution)
code z
Decoder
Image x
(generated)
from encoder
or decoder?
Discriminator
Image x
code z

Algorithm

- Initialize encoder En, decoder De, discriminator Dis
- In each iteration:
  - Sample M images $x^1, x^2, \cdots, x^M$ from database
  - Generate M codes $\tilde{z}^1, \tilde{z}^2, \cdots, \tilde{z}^M$ from encoder
    - $\tilde{z}^i = En(x^i)$
  - Sample M codes $z^1, z^2, \cdots, z^M$ from prior P(z)
  - Generate M codes $\tilde{x}^1, \tilde{x}^2, \cdots, \tilde{x}^M$ from decoder
    - $\tilde{x}^i = De(z^i)$
  - Update Dis to increase $Dis(x^i, \tilde{z}^i)$, decrease $Dis(\tilde{x}^i, z^i)$
  - Update En and De to decrease $Dis(x^i, \tilde{z}^i)$, increase $Dis(\tilde{x}^i, z^i)$



Optimal encoder and decoder:

En(x') = z' ⇒ De(z') = x' For all x'

De(z'') = x'' ⇒ En(x'') = z'' For all z''

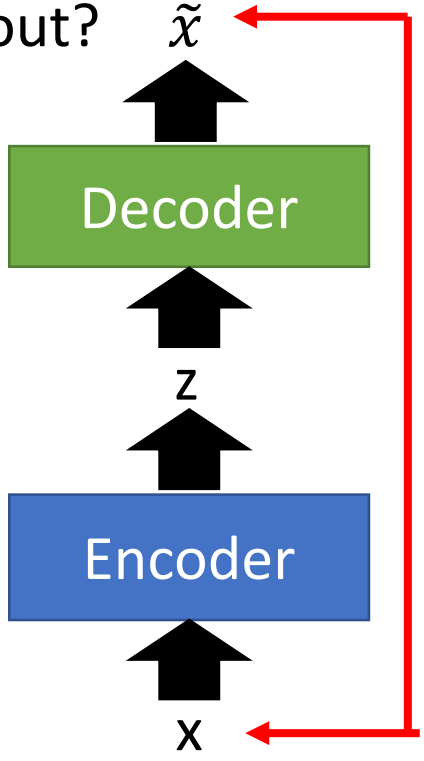
BiGAN

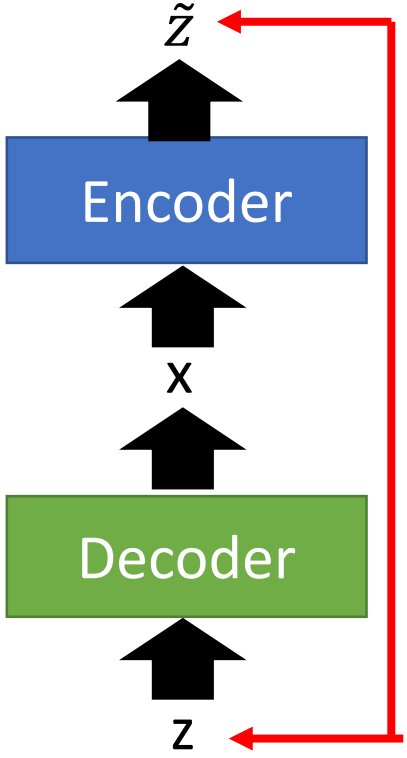
Optimal encoder and decoder:

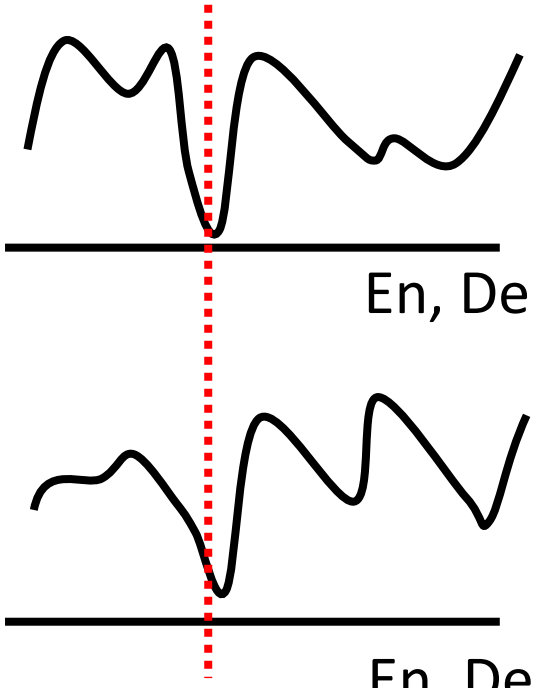
En(x') = z' ⇒ De(z') = x' For all x'

De(z'') = x'' ⇒ En(x'') = z'' For all z''

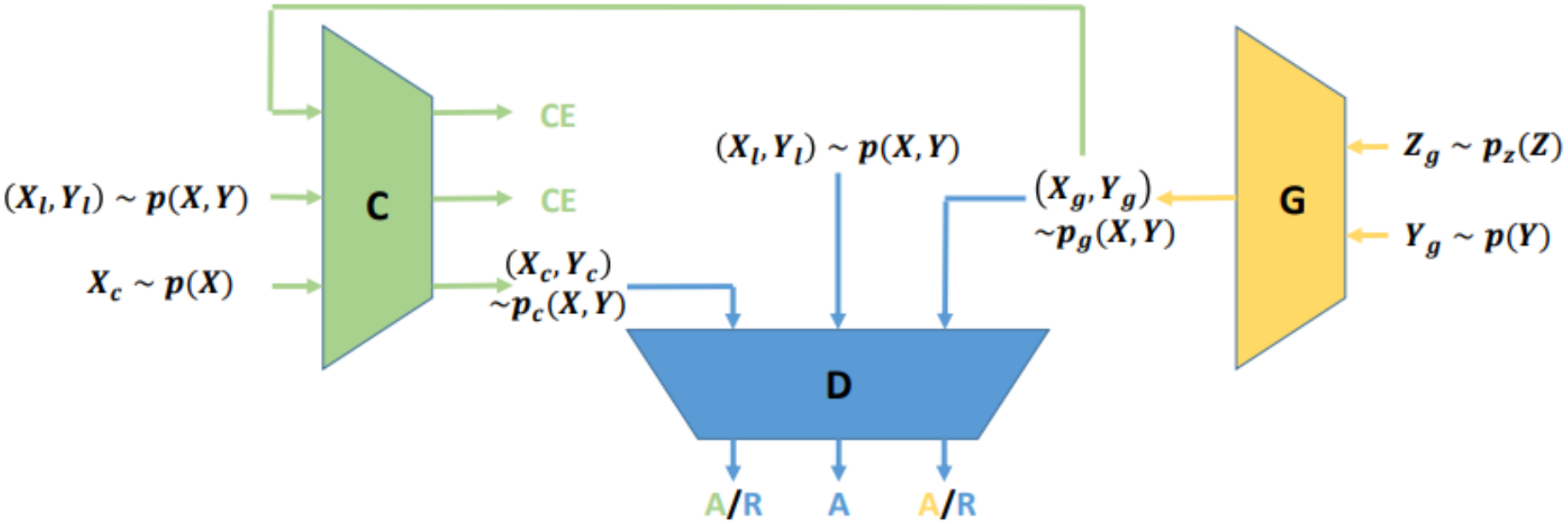
How about?







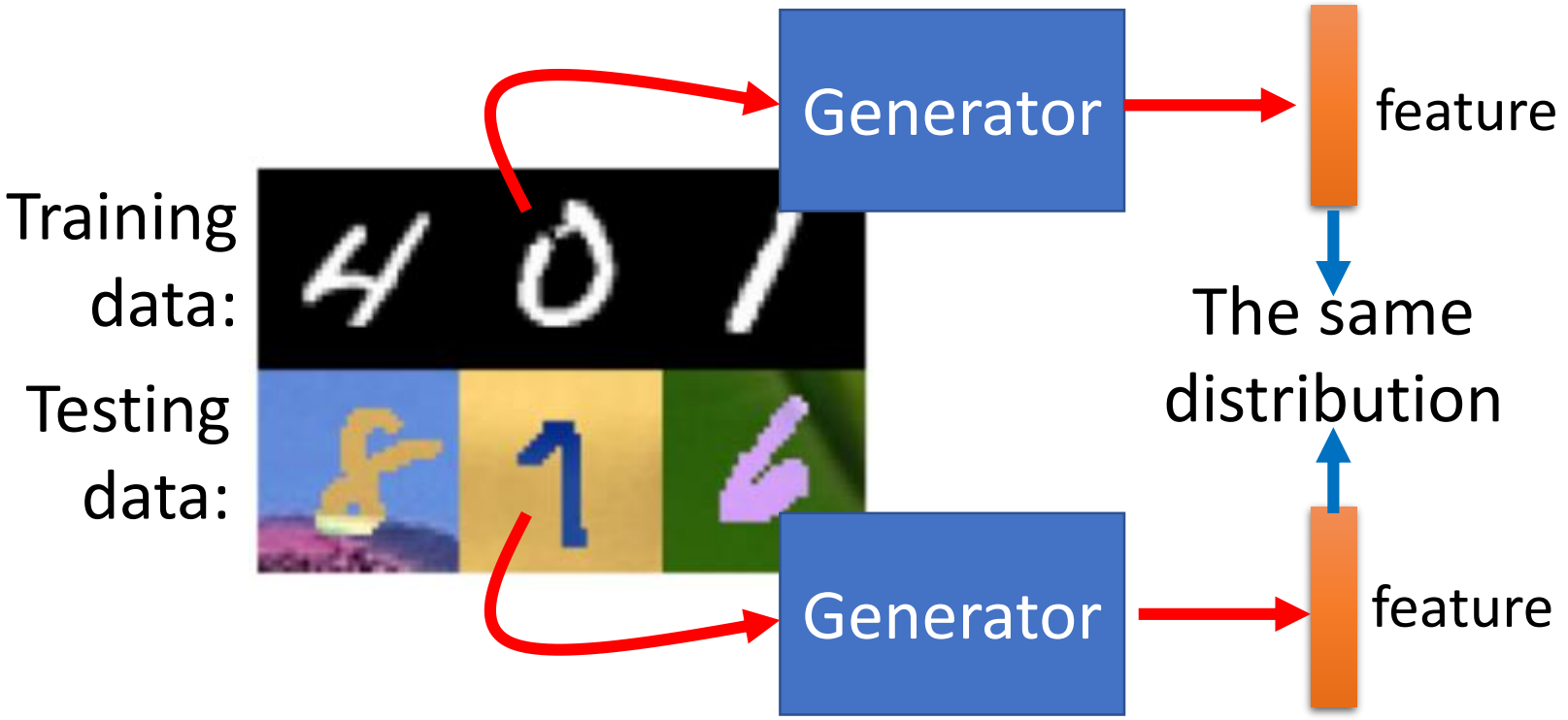
Triple GAN



Chongxuan Li, Kun Xu, Jun Zhu, Bo Zhang, “Triple Generative Adversarial Nets”, arXiv 2017

Domain-adversarial training

- Training and testing data are in different domains

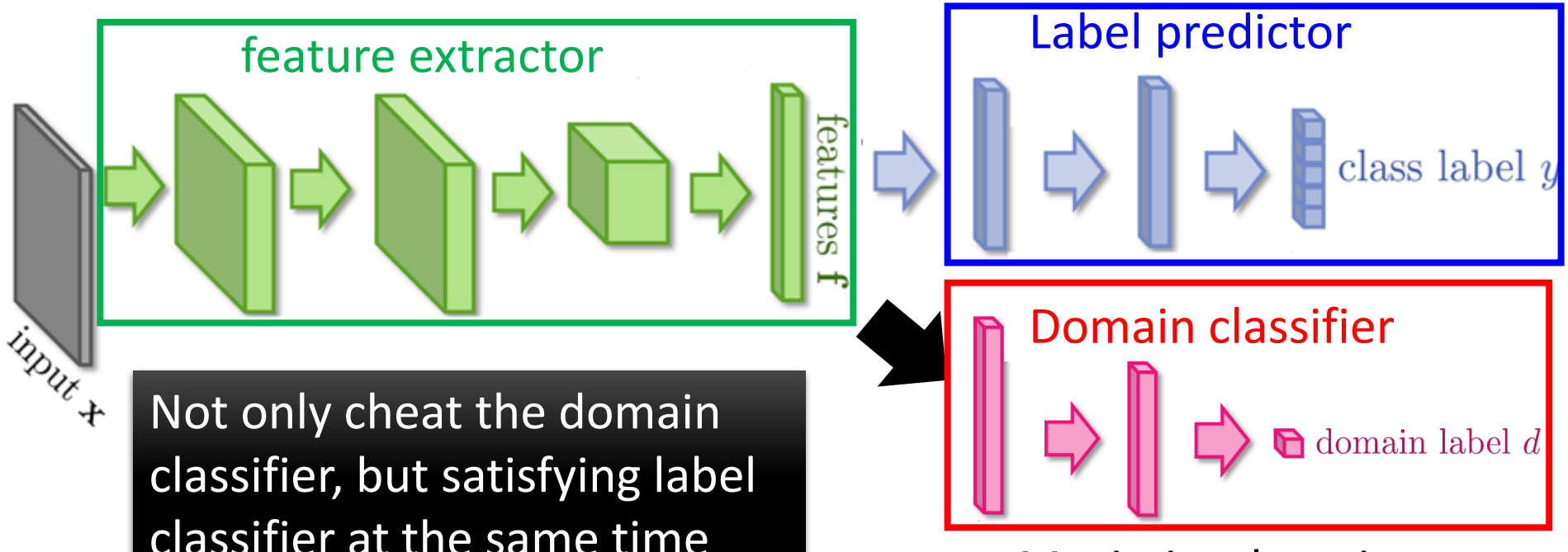


Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, Domain-Adversarial Training of Neural Networks, JMLR, 2016

Domain-adversarial training

Maximize label classification accuracy + minimize domain classification accuracy

Maximize label classification accuracy

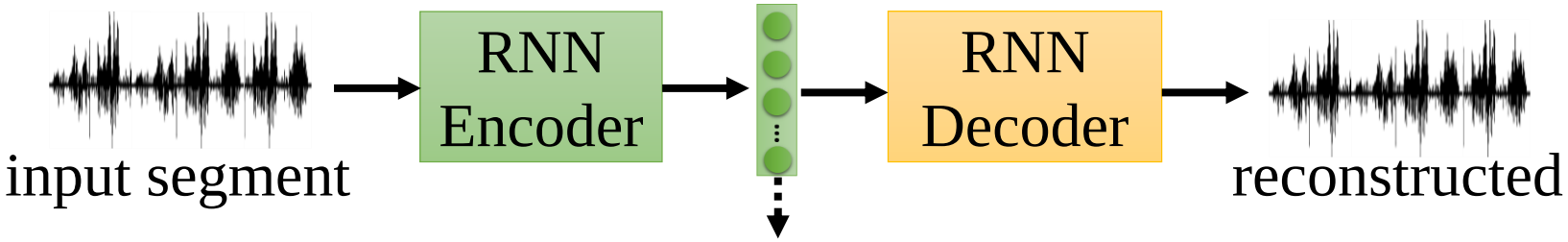


Not only cheat the domain classifier, but satisfying label classifier at the same time

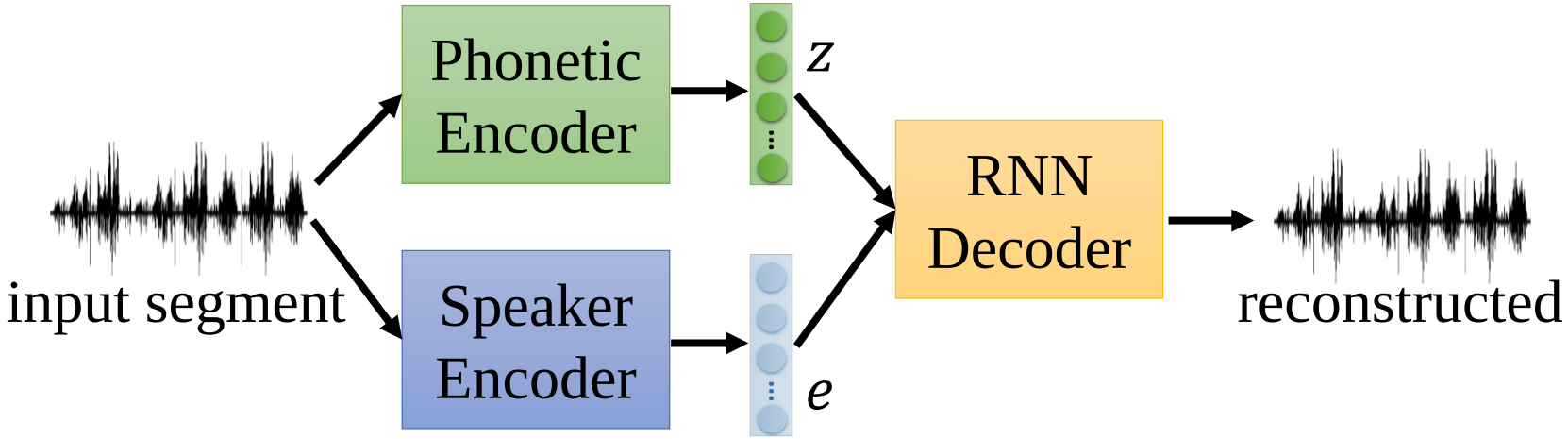
Maximize domain classification accuracy

This is a big network, but different parts have different goals.

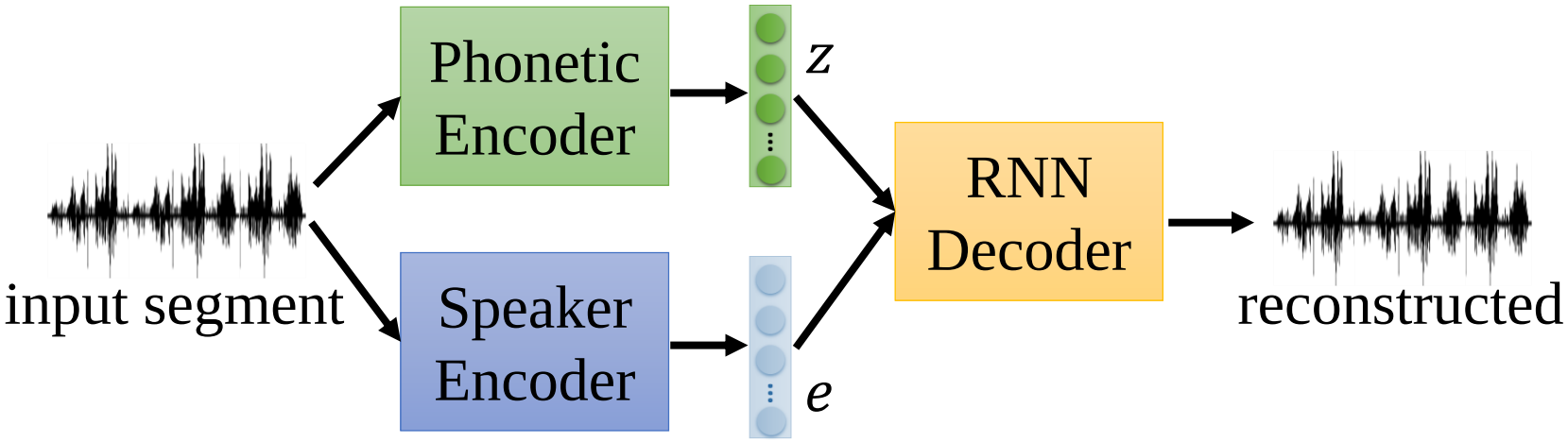
Original Seq2seq Auto-encoder

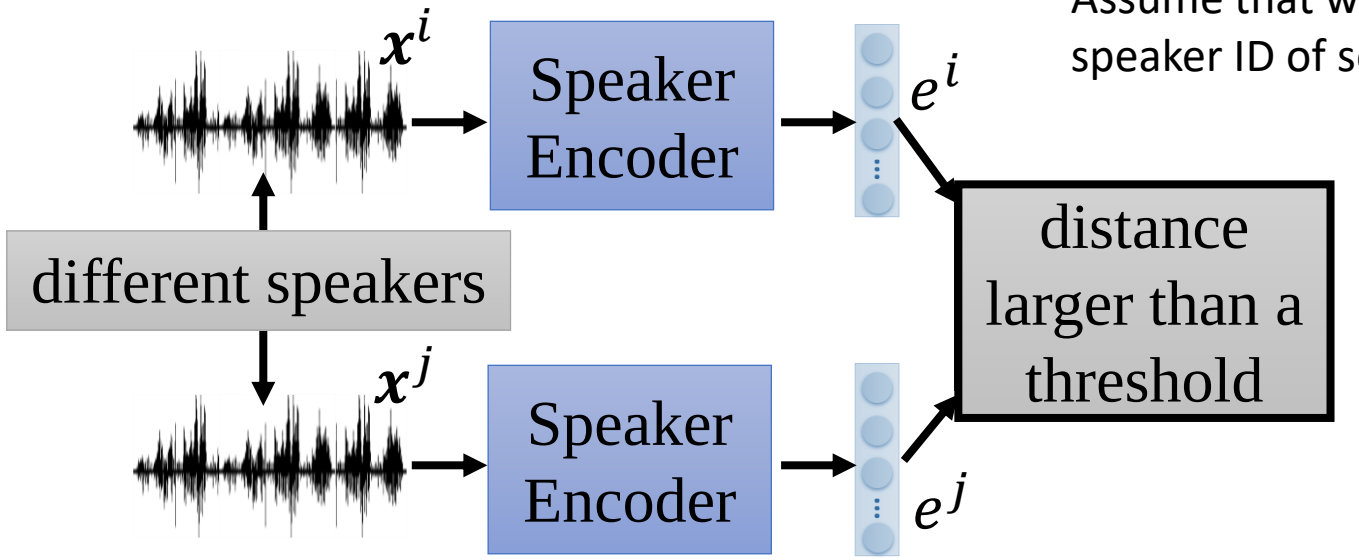


Feature Disentangle



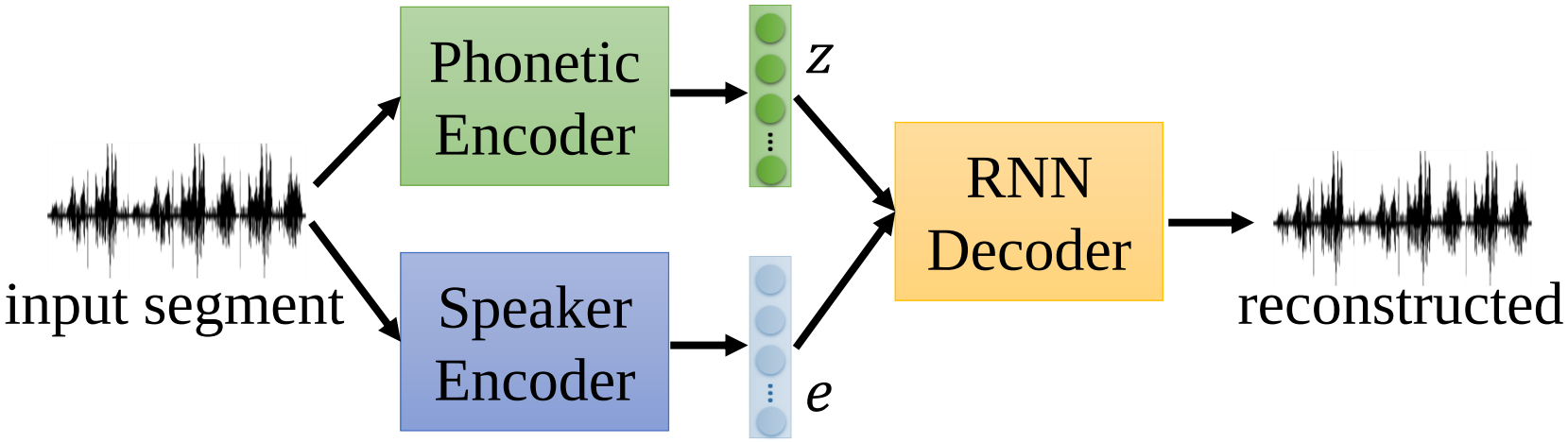
Feature Disentangle



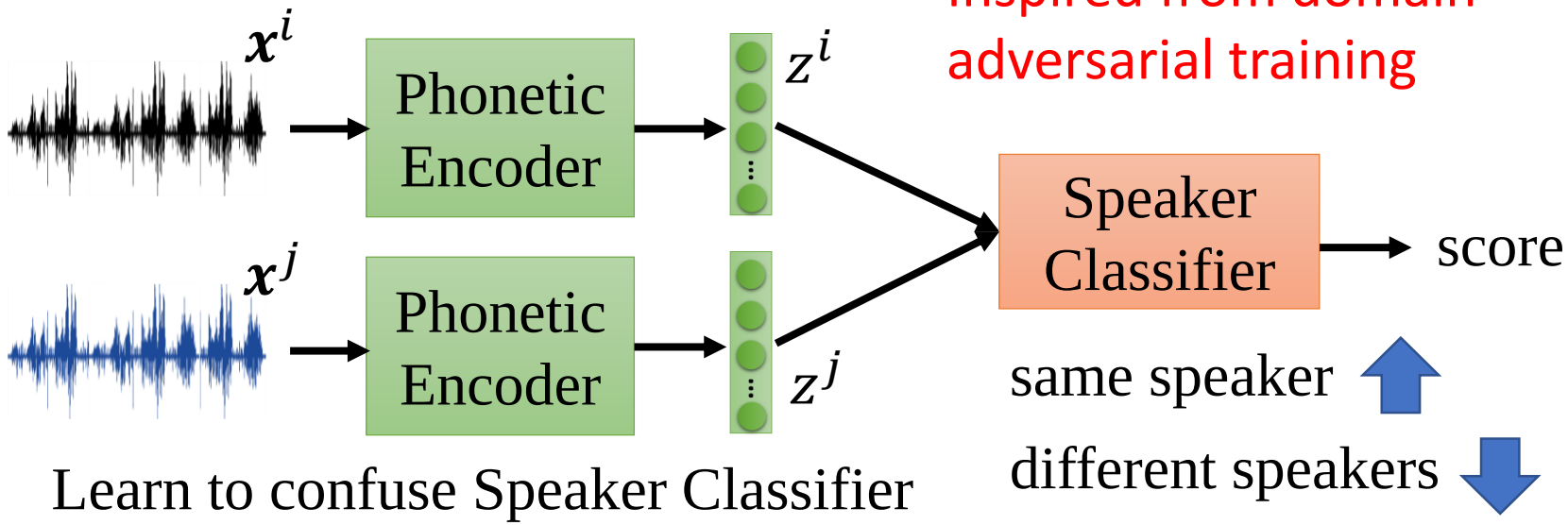


Assume that we know speaker ID of segments.

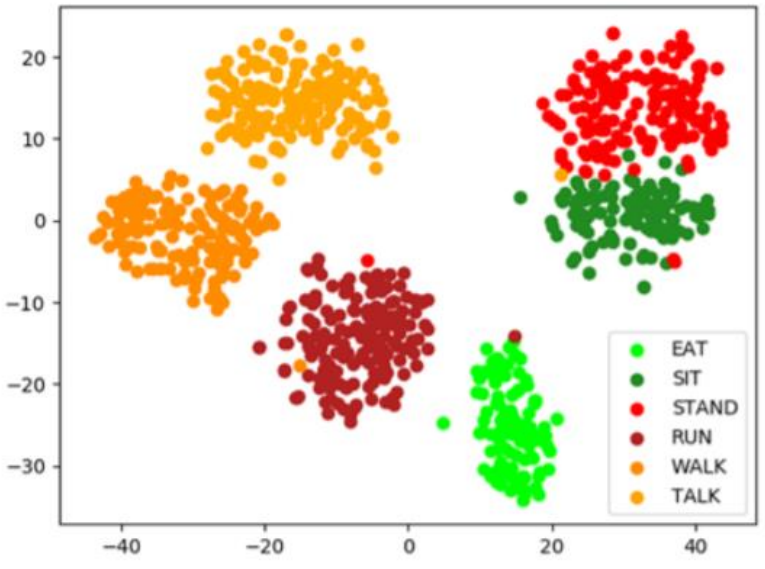
Feature Disentangle



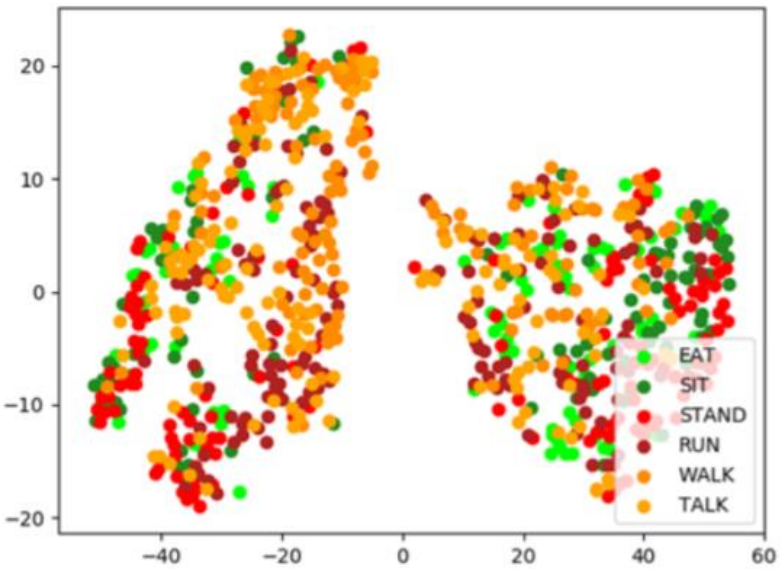
Inspired from domain adversarial training



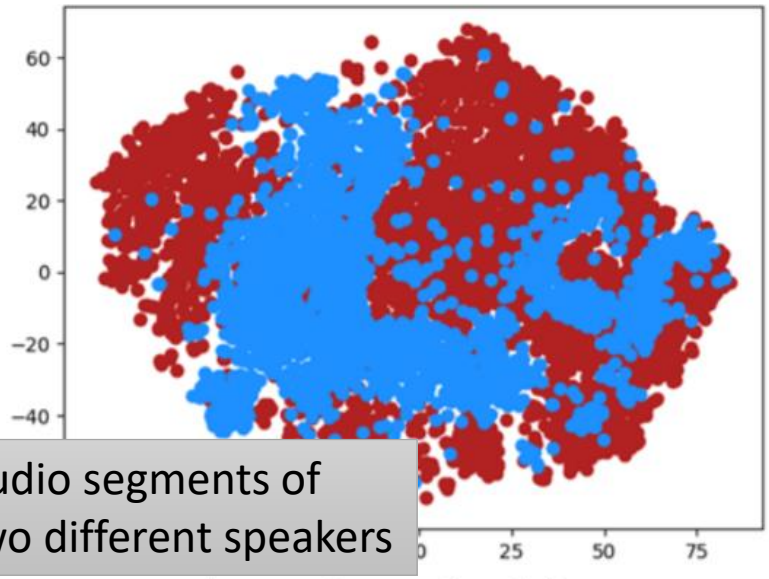
Learn to confuse Speaker Classifier



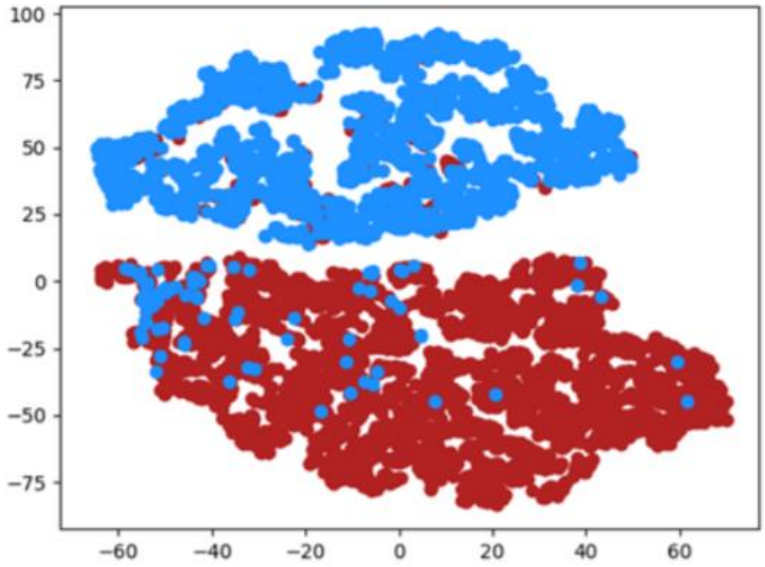
phonetic embedding z



speaker embedding e



phonetic embedding z



speaker embedding e

Audio segments of two different speakers

Acknowledgement

- 感謝 許傑盛 同學發現投影片 上的打字錯誤