Unsupervised Conditional Generation

Unsupervised Conditional Generation

Domain X  G  Domain Y

Transform an object from one domain to another
without paired data (e.g. style transfer)

Domain X		Domain Y
 male	⟷	 female
It is good. It’s a good day. I love you.	⟷	It is bad. It’s a bad day. I don’t love you.

Unsupervised Conditional Generation

- Approach 1: Direct Transformation

Domain X

$G_{X \to Y}$

Domain Y

For texture or color change

- Approach 2: Projection to Common Space

Domain X

EN_X

Encoder of domain X

Face Attribute

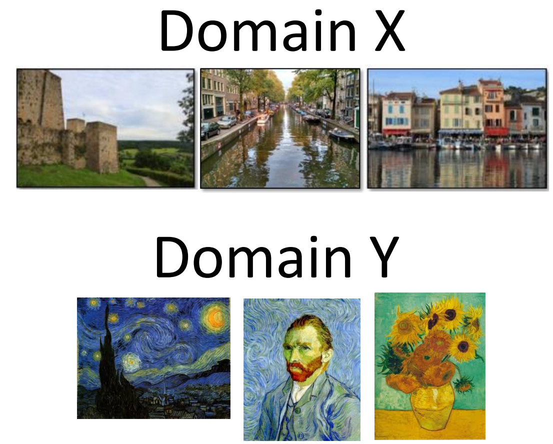
DE_Y

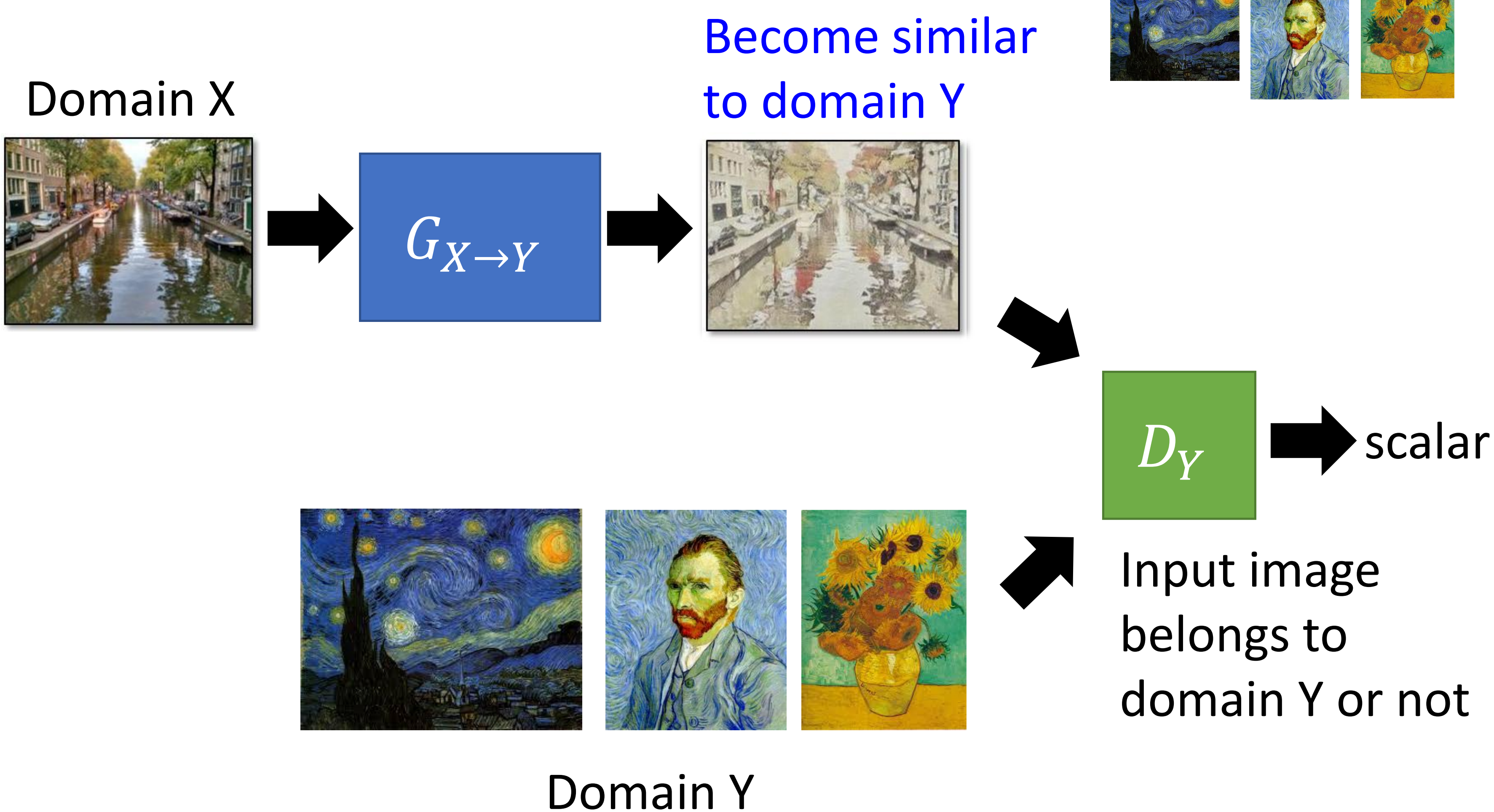
Decoder of domain Y

Domain Y

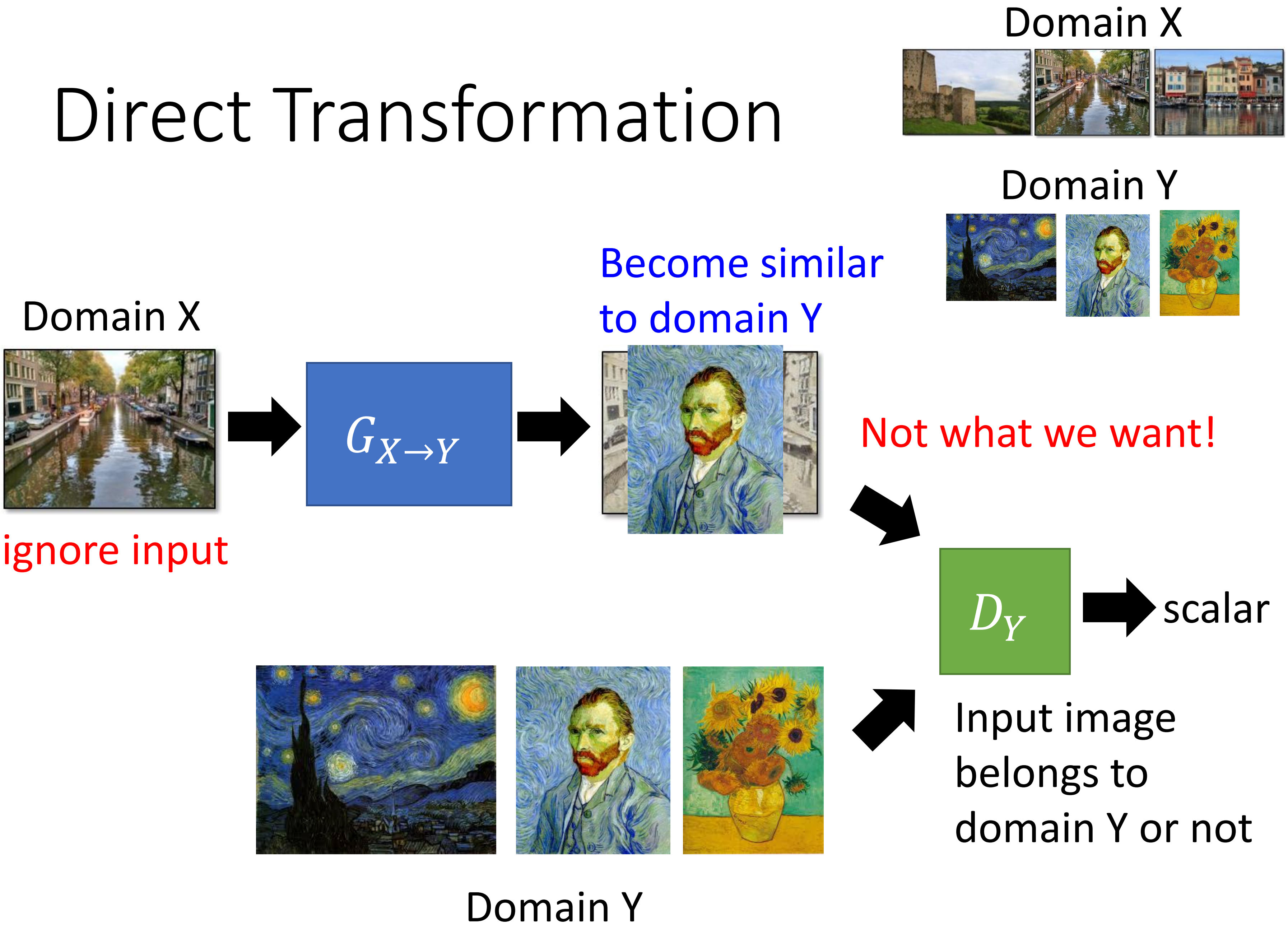
Larger change, only keep the semantics

Direct Transformation





Direct Transformation



Direct Transformation

Domain X

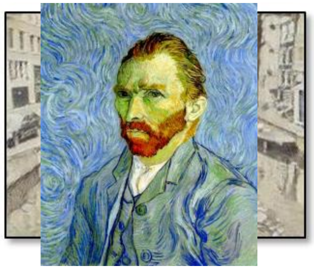
Domain Y

Domain X

ignore input

$G_{X \to Y}$

Become similar to domain Y



Not what we want!

D_Y → scalar

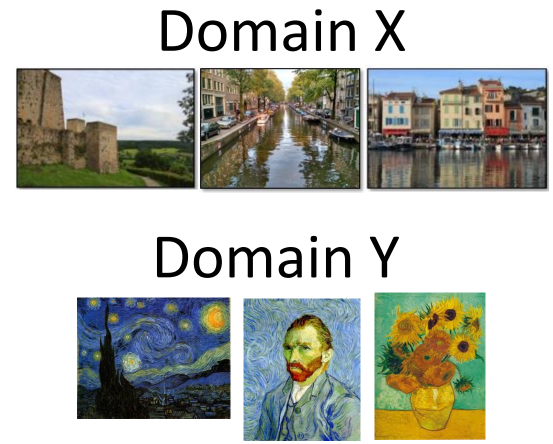
Input image belongs to domain Y or not

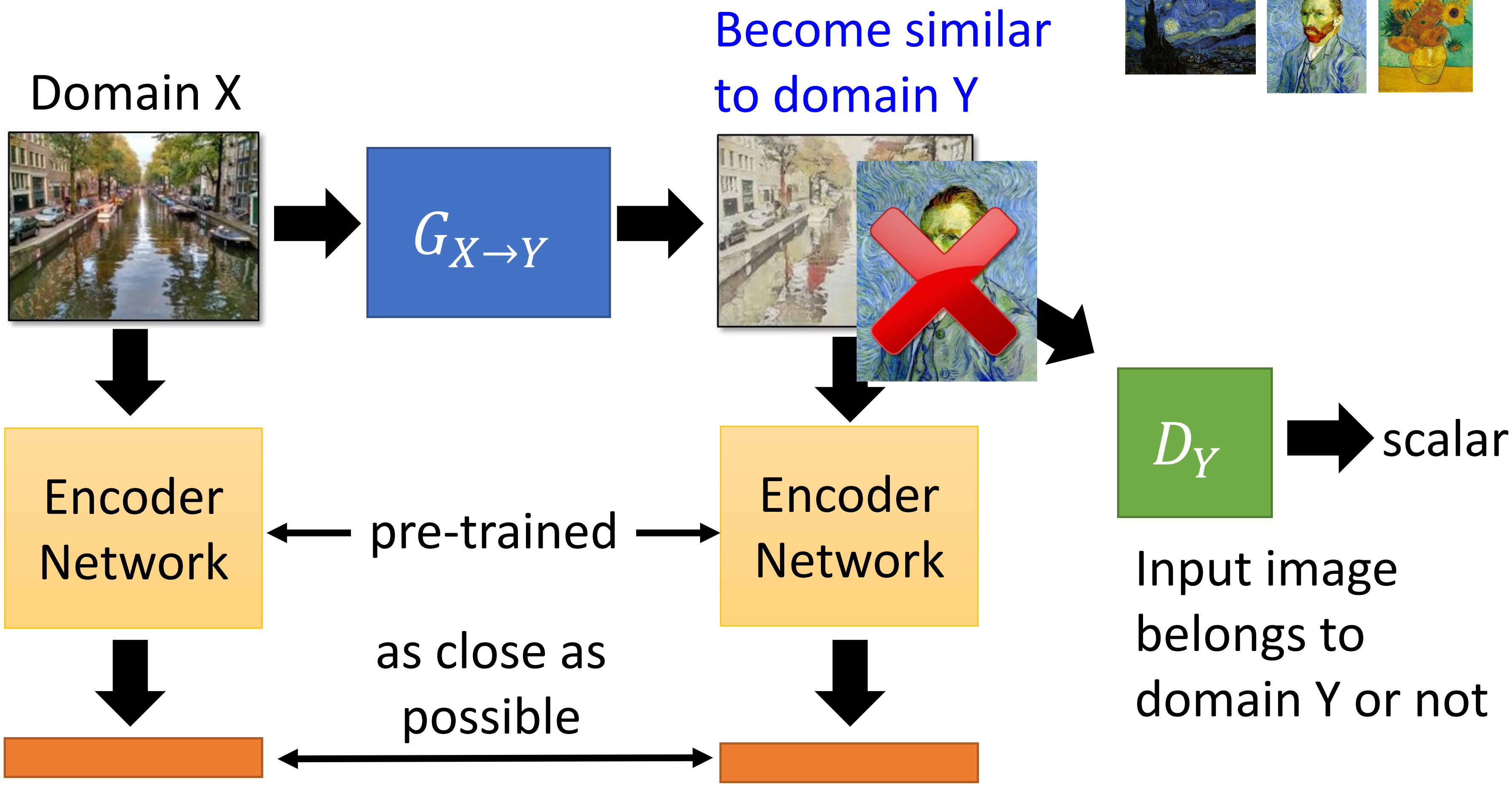
The issue can be avoided by network design.

Simpler generator makes the input and output more closely related.

[Tomer Galanti, et al. ICLR, 2018]

Direct Transformation

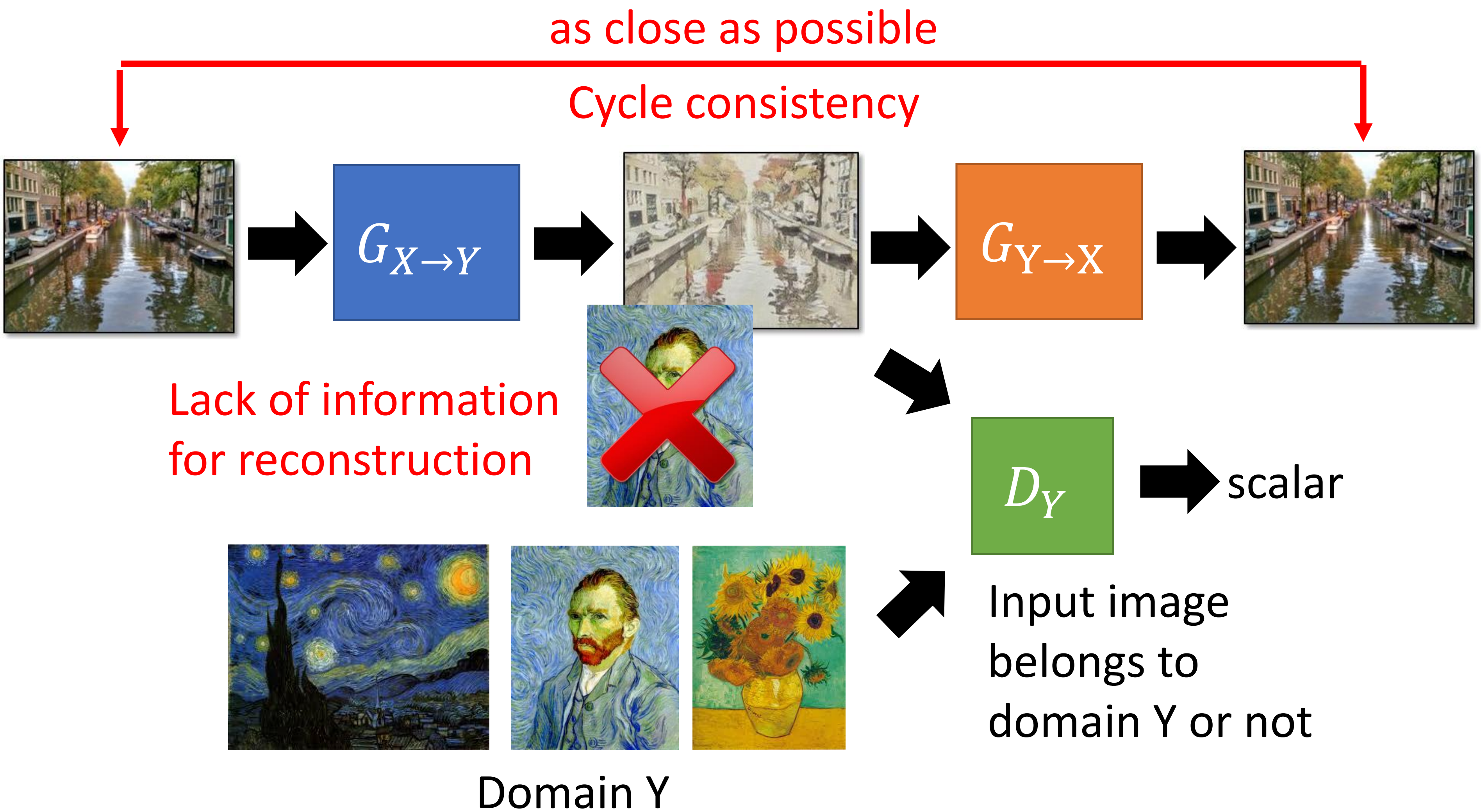




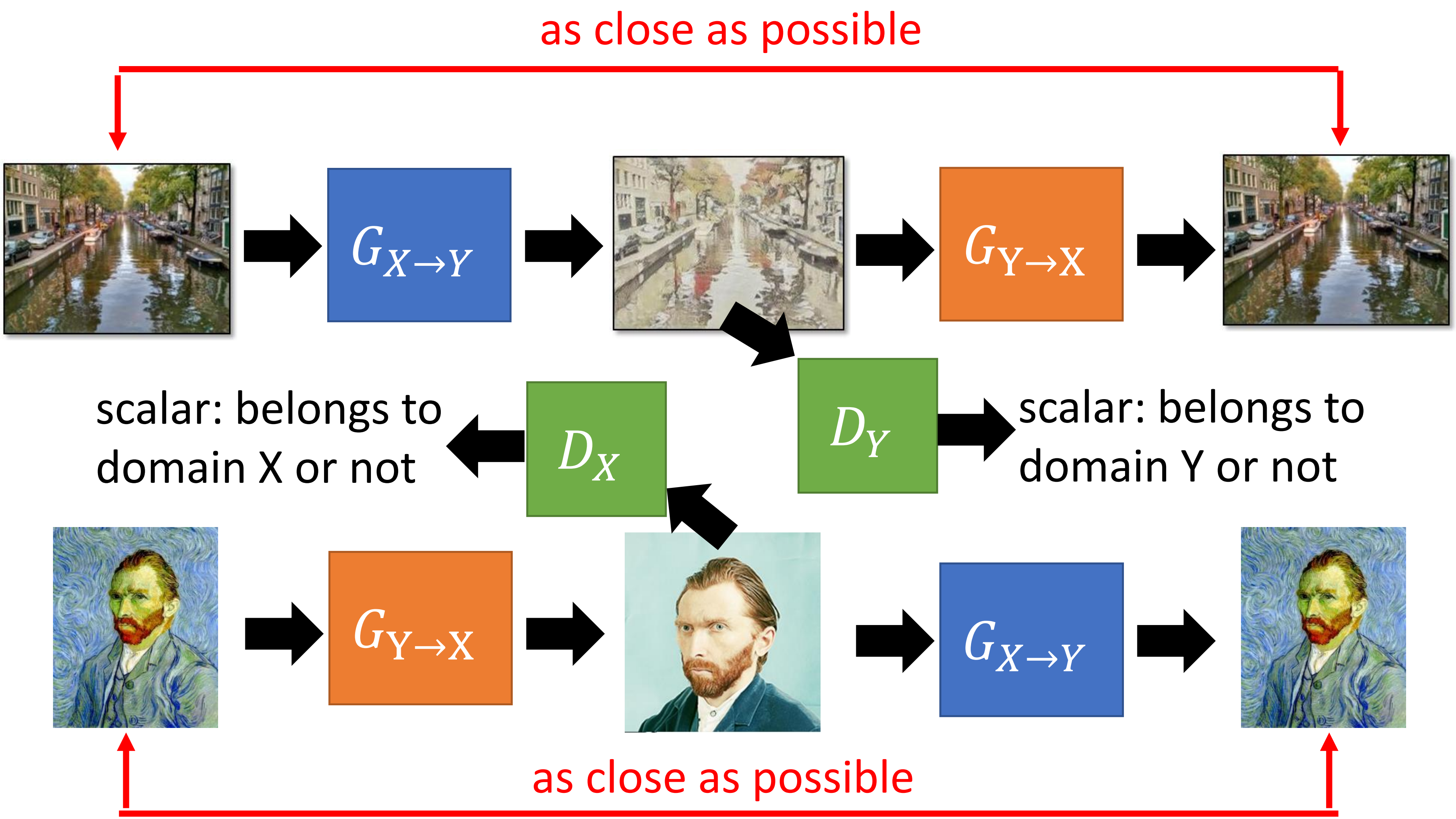
Baseline of DTN [Yaniv Taigman, et al., ICLR, 2017]

[Jun-Yan Zhu, et al., ICCV, 2017]

Direct Transformation

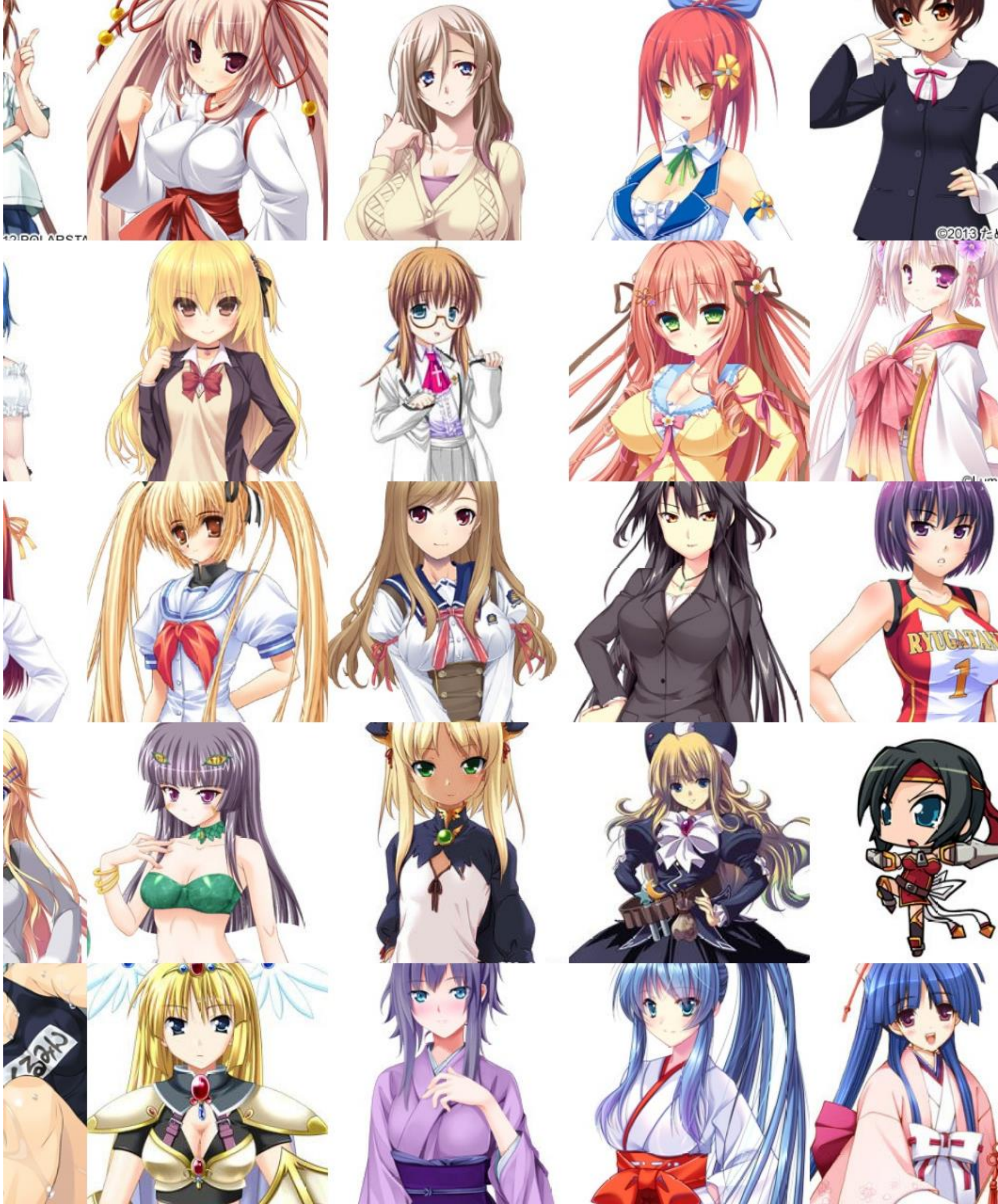


Direct Transformation



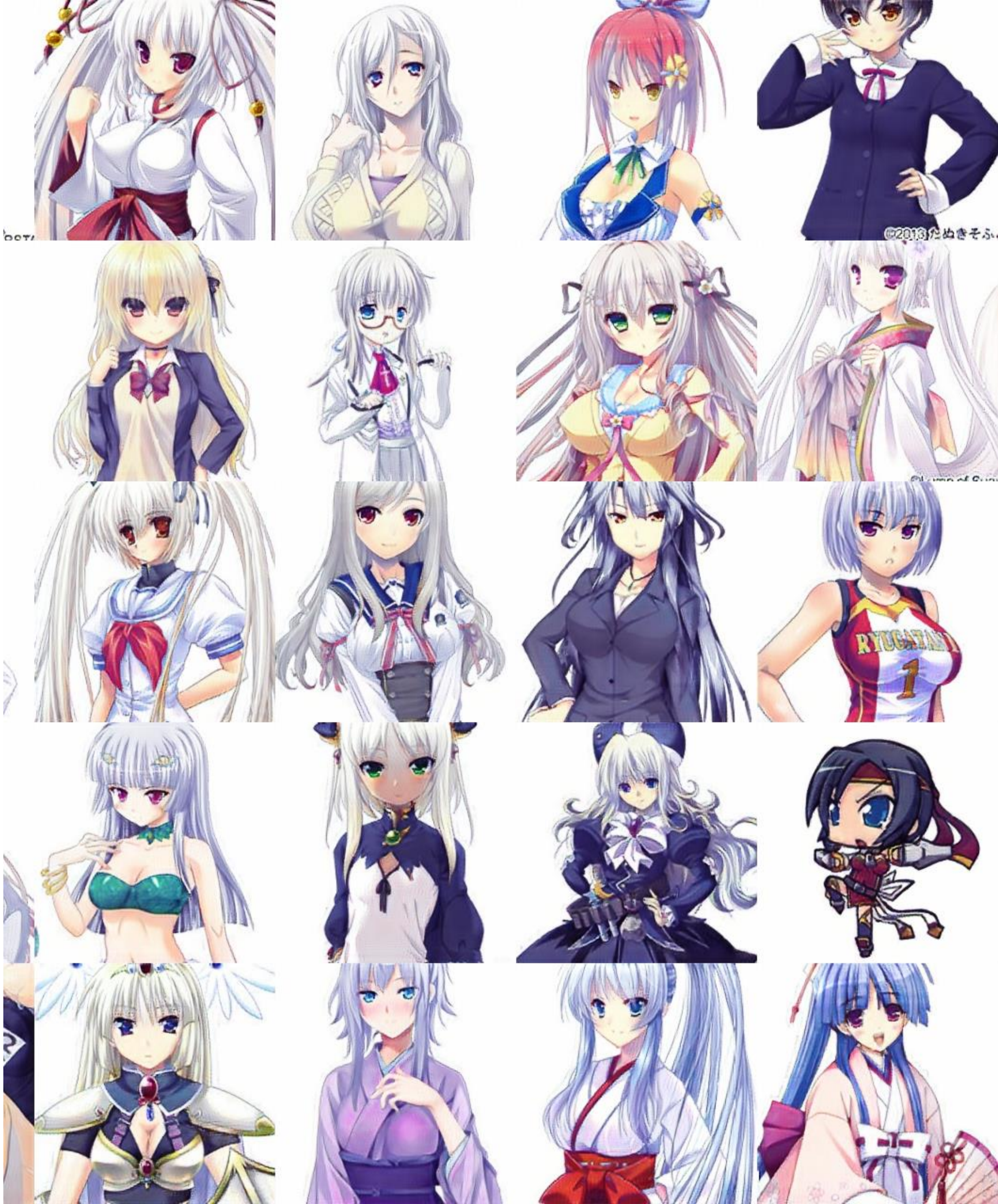
Cycle GAN – Silver Hair

- https://github.com/Aixile/chainer-cyclegan



Cycle GAN – Silver Hair

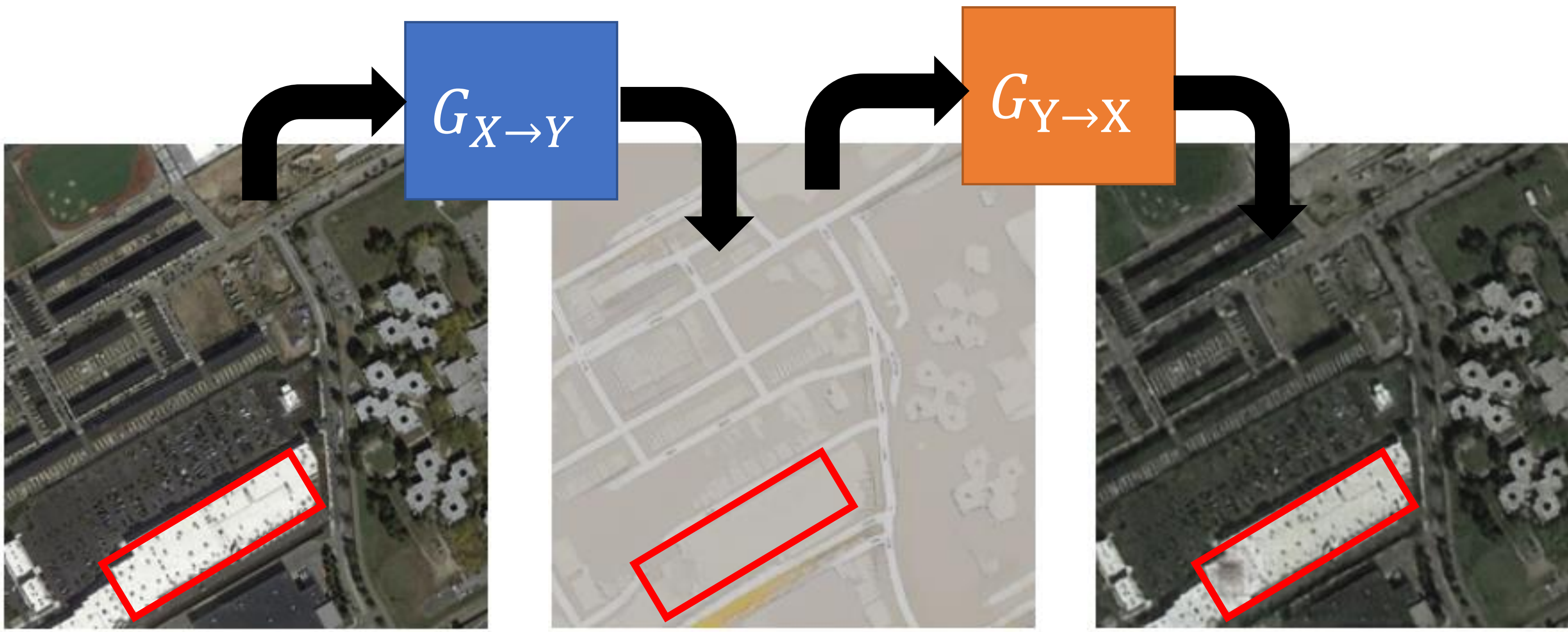
- https://github.com/Aixile/chainer-cyclegan



Issue of Cycle Consistency

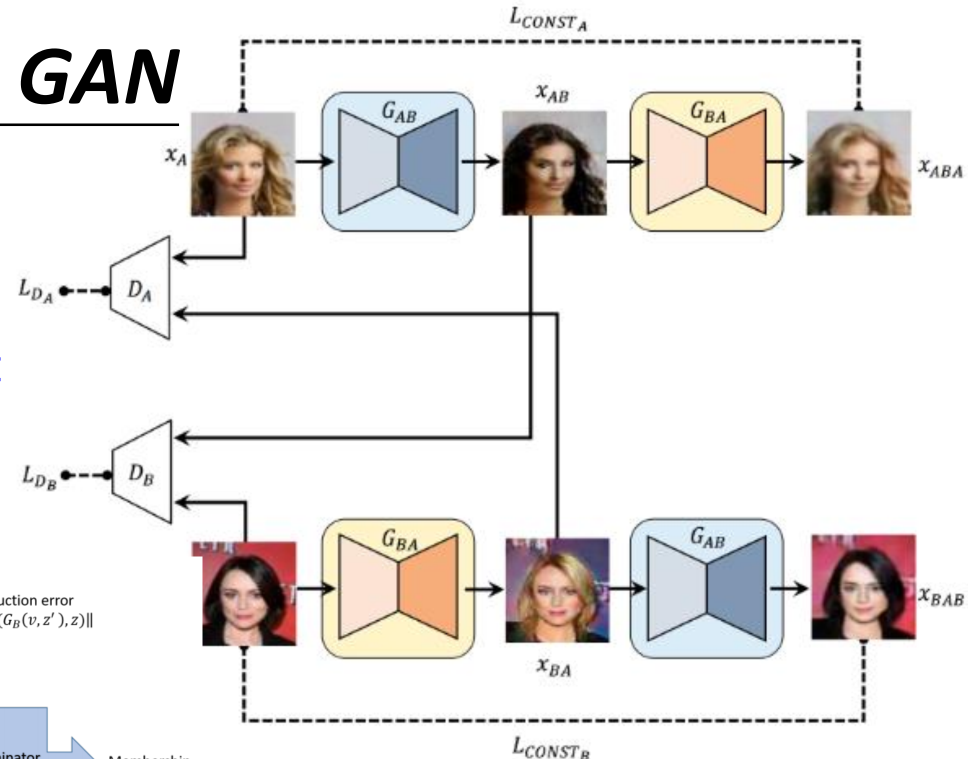
- **CycleGAN: a Master of Steganography (隱寫術)**

[Casey Chu, et al., NIPS workshop, 2017]



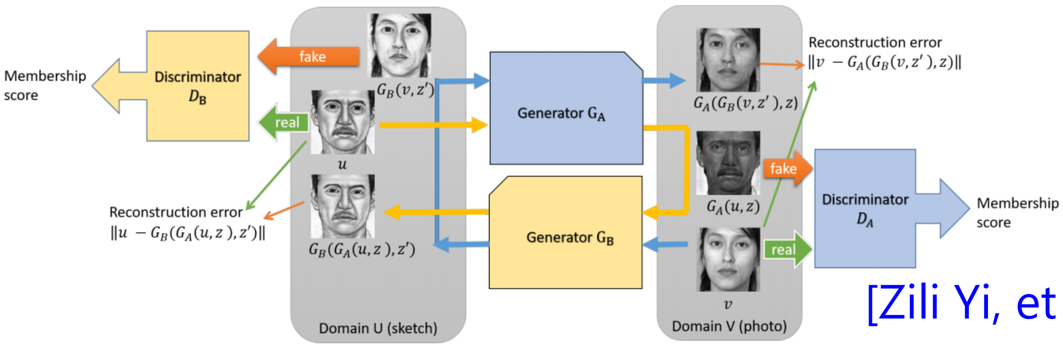
The information is hidden.

Disco GAN



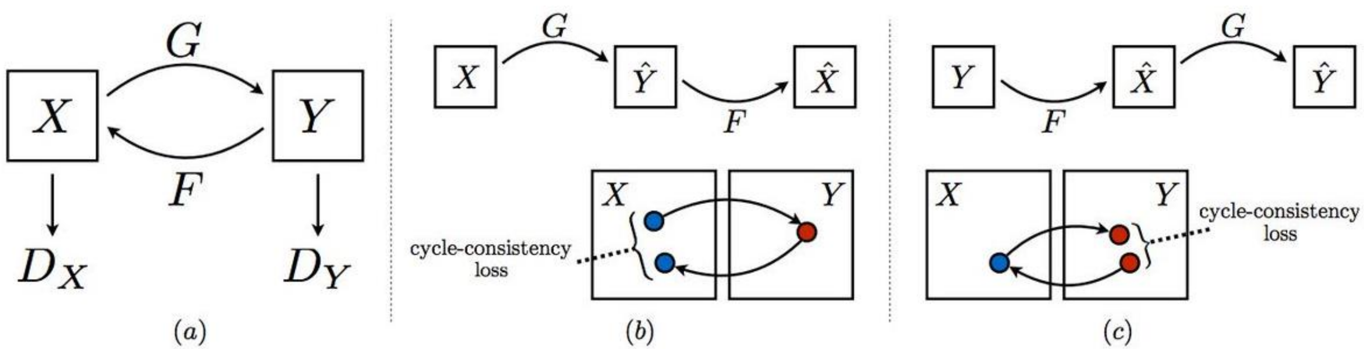
[Taeksoo Kim, et al., ICML, 2017]

Dual GAN



[Zili Yi, et al., ICCV, 2017]

Cycle GAN

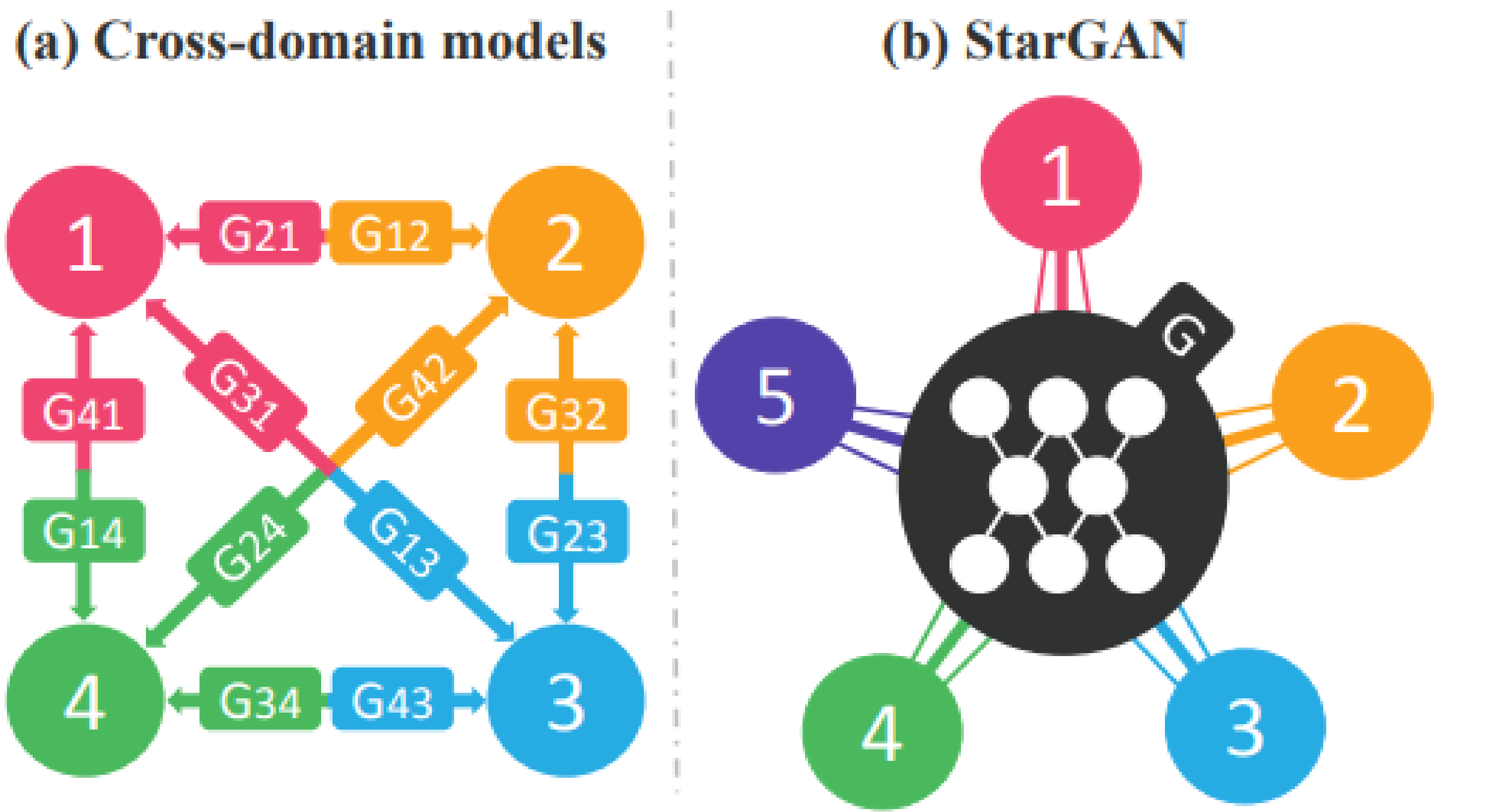


[Jun-Yan Zhu, et al., ICCV, 2017]

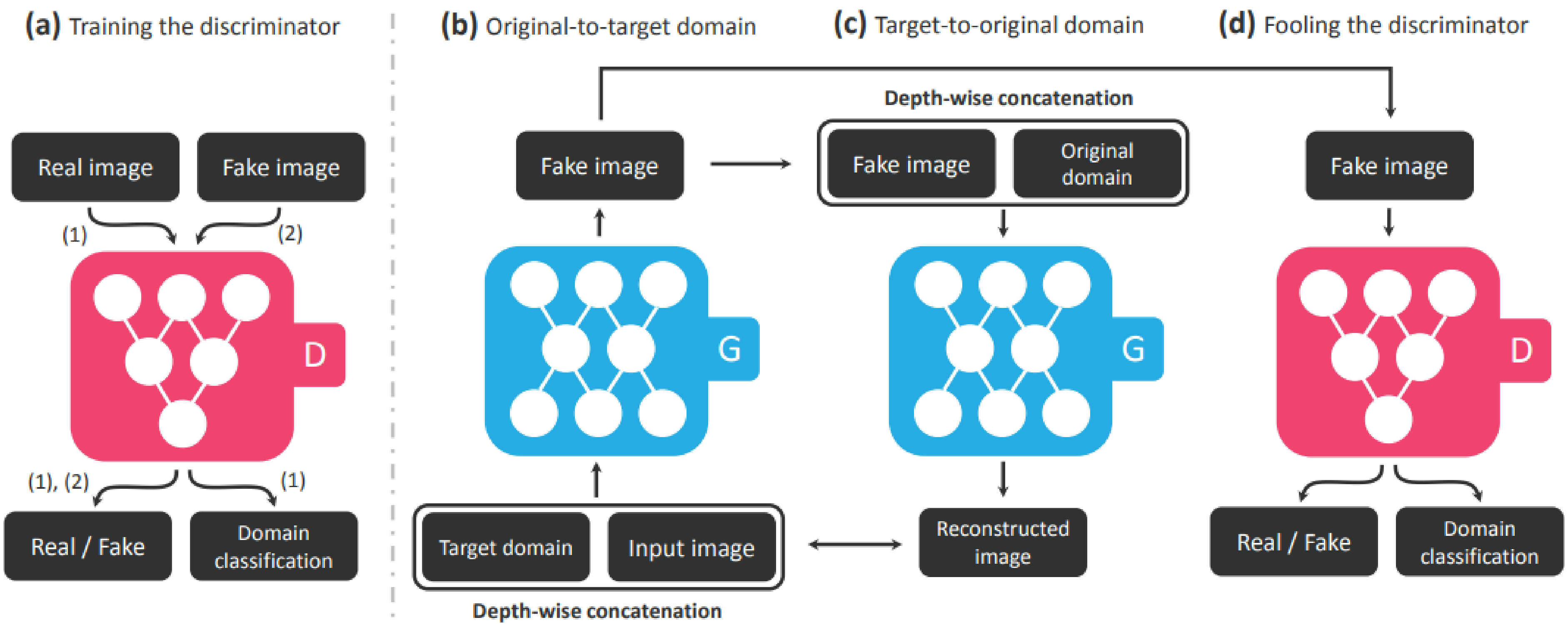
StarGAN

For multiple domains, considering starGAN

[Yunjey Choi, arXiv, 2017]

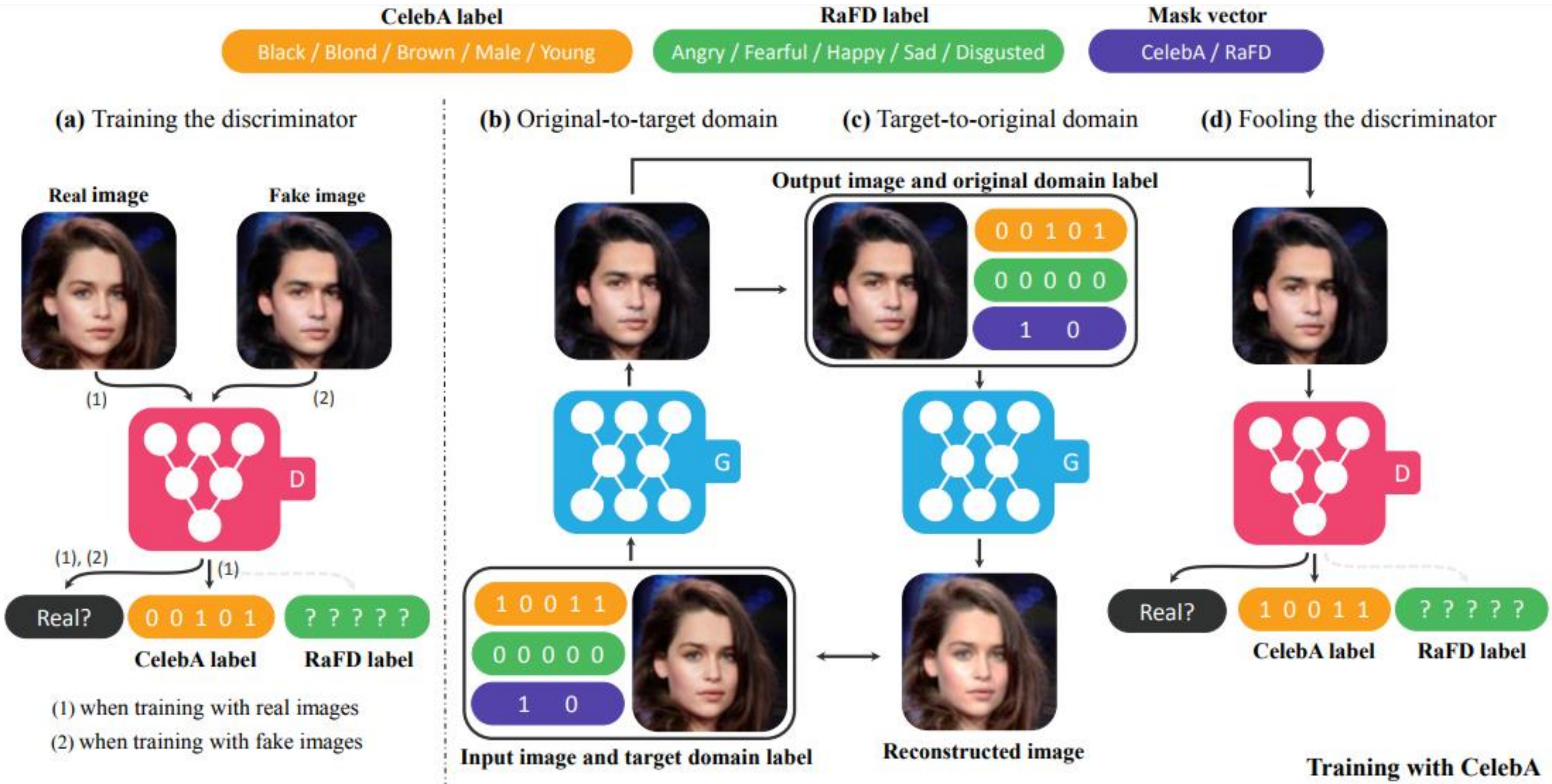


StarGAN



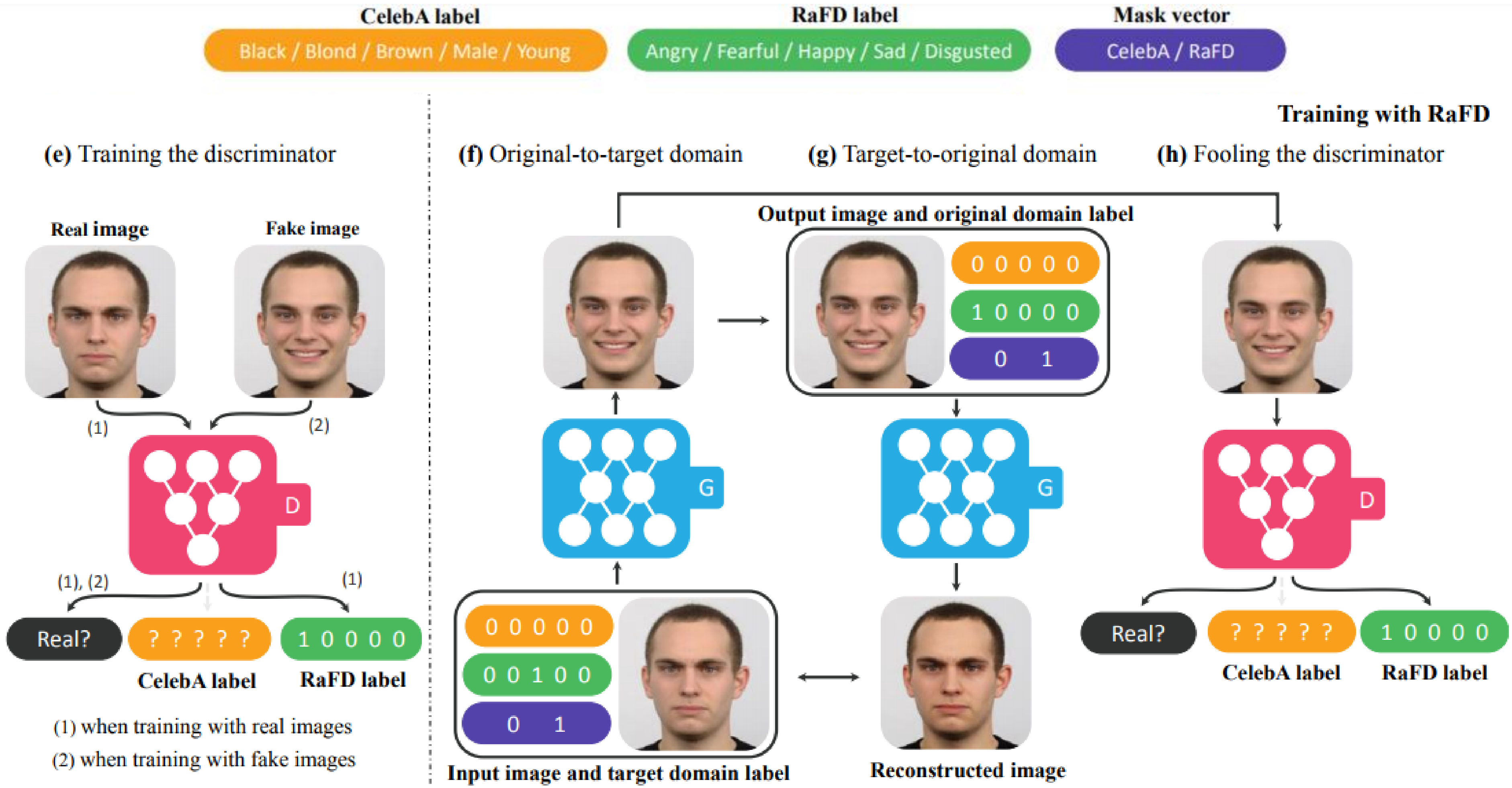
(a) Training the discriminator
(b) Original-to-target domain
(c) Target-to-original domain
(d) Fooling the discriminator
Real image
Fake image
(1)
(2)
D
(1), (2)
(1)
Real / Fake
Domain classification
Fake image
G
Target domain
Input image
Depth-wise concatenation
Depth-wise concatenation
Fake image
Original domain
G
Reconstructed image
Fake image
D
Real / Fake
Domain classification

StarGAN



CelebA label
Black / Blond / Brown / Male / Young
RaFD label
Angry / Fearful / Happy / Sad / Disgusted
Mask vector
CelebA / RaFD
(a) Training the discriminator
(b) Original-to-target domain
(c) Target-to-original domain
(d) Fooling the discriminator
Real image
Fake image
(1)
(2)
D
(1), (2)
(1)
Real?
0 0 1 0 1
? ? ? ? ?
CelebA label
RaFD label
(1) when training with real images
(2) when training with fake images
Output image and original domain label
0 0 1 0 1
0 0 0 0 0
1 0
G
G
1 0 0 1 1
0 0 0 0 0
1 0
Input image and target domain label
Reconstructed image
D
Real?
1 0 0 1 1
? ? ? ? ?
CelebA label
RaFD label
Training with CelebA

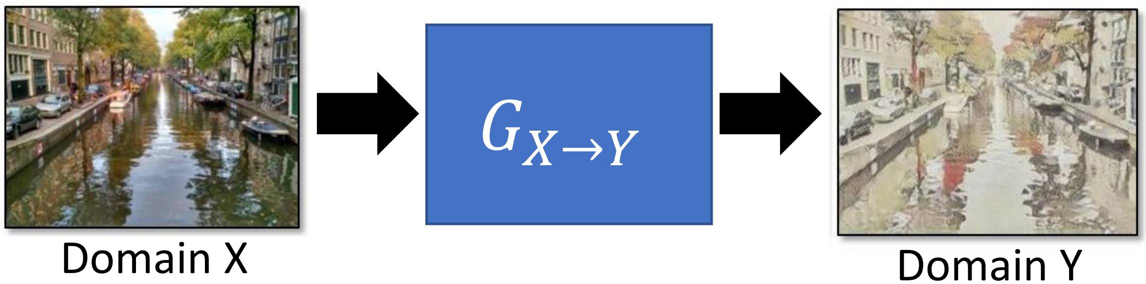
StarGAN



CelebA label
Black / Blond / Brown / Male / Young
RaFD label
Angry / Fearful / Happy / Sad / Disgusted
Mask vector
CelebA / RaFD
Training with RaFD
(e) Training the discriminator
(f) Original-to-target domain
(g) Target-to-original domain
(h) Fooling the discriminator
Real image
Fake image
(1)
(2)
D
(1), (2)
(1)
Real?
? ? ? ? ?
1 0 0 0 0
CelebA label
RaFD label
(1) when training with real images
(2) when training with fake images
Output image and original domain label
0 0 0 0 0
1 0 0 0 0
0 1
G
G
0 0 0 0 0
0 0 1 0 0
0 1
Input image and target domain label
Reconstructed image
D
Real?
? ? ? ? ?
1 0 0 0 0
CelebA label
RaFD label

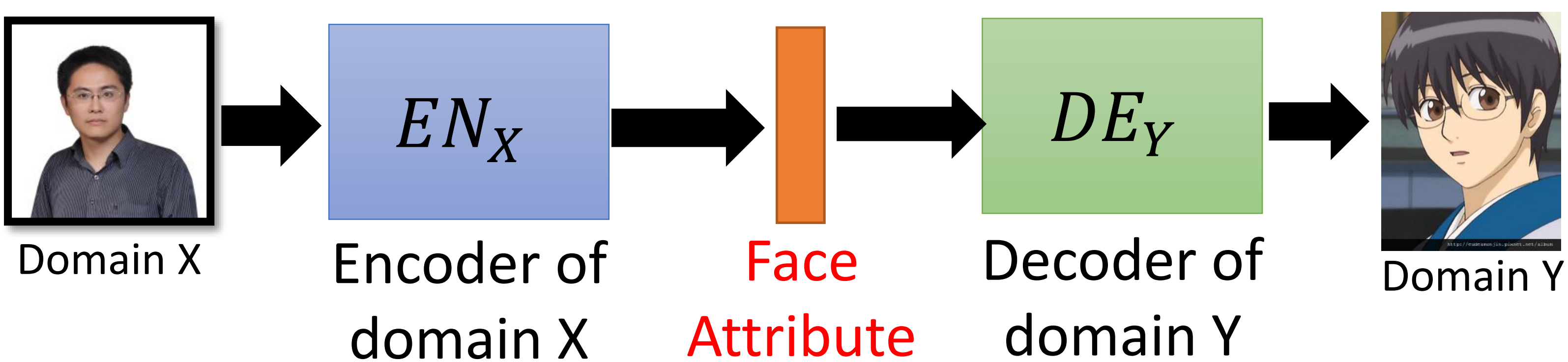
Unsupervised Conditional Generation

- Approach 1: Direct Transformation



For texture or color change

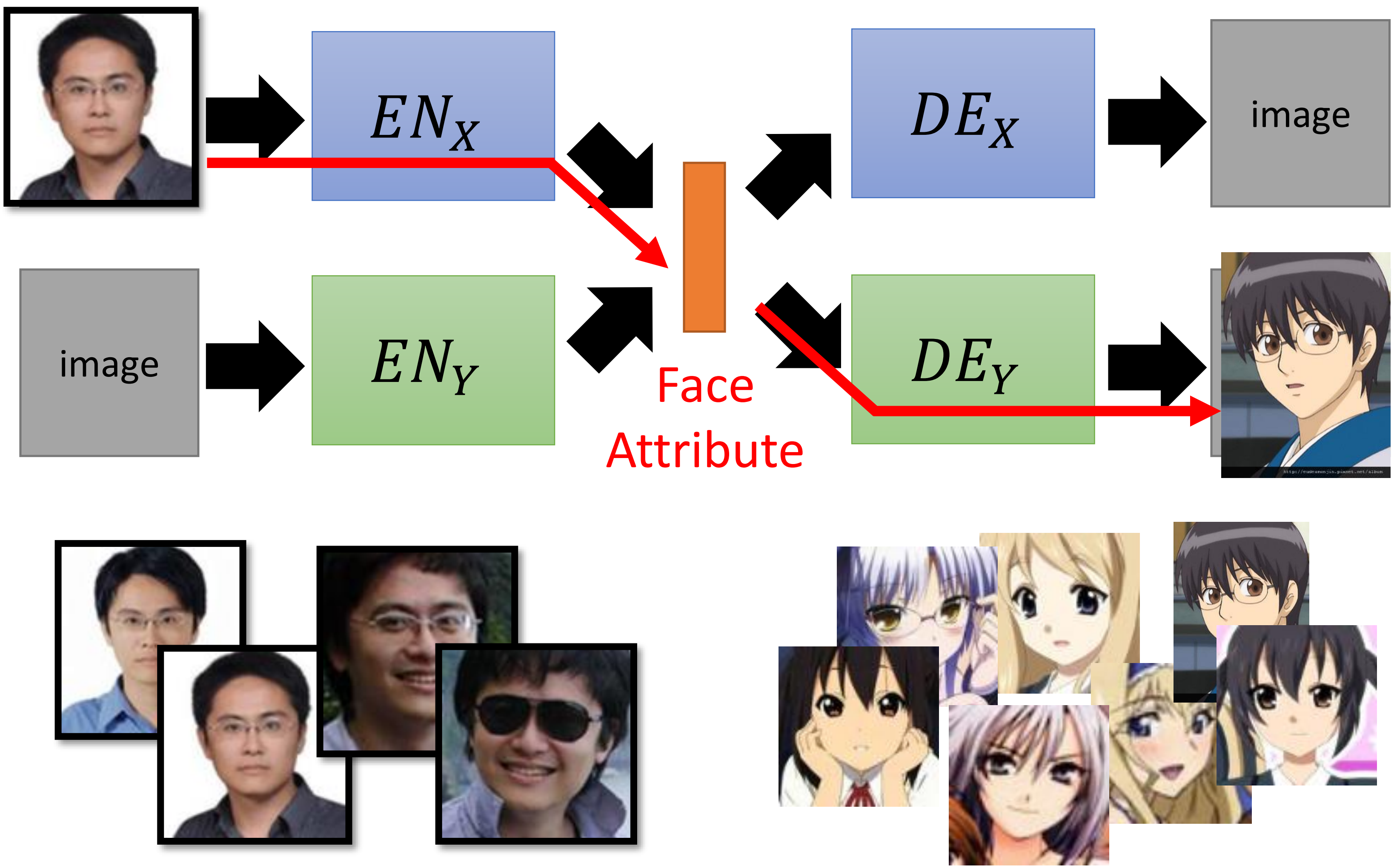
- Approach 2: Projection to Common Space



Larger change, only keep the semantics

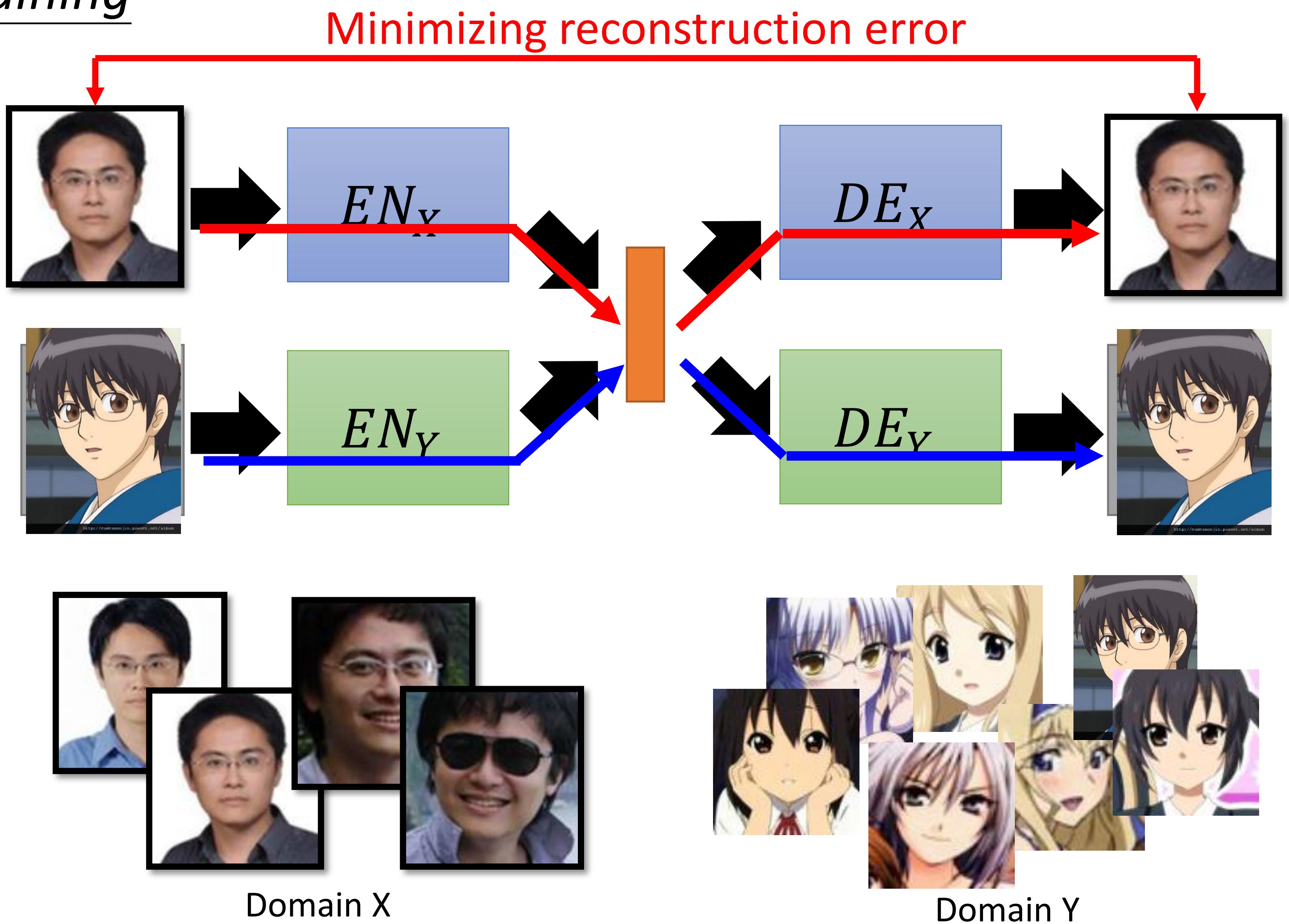
Projection to Common Space

Target



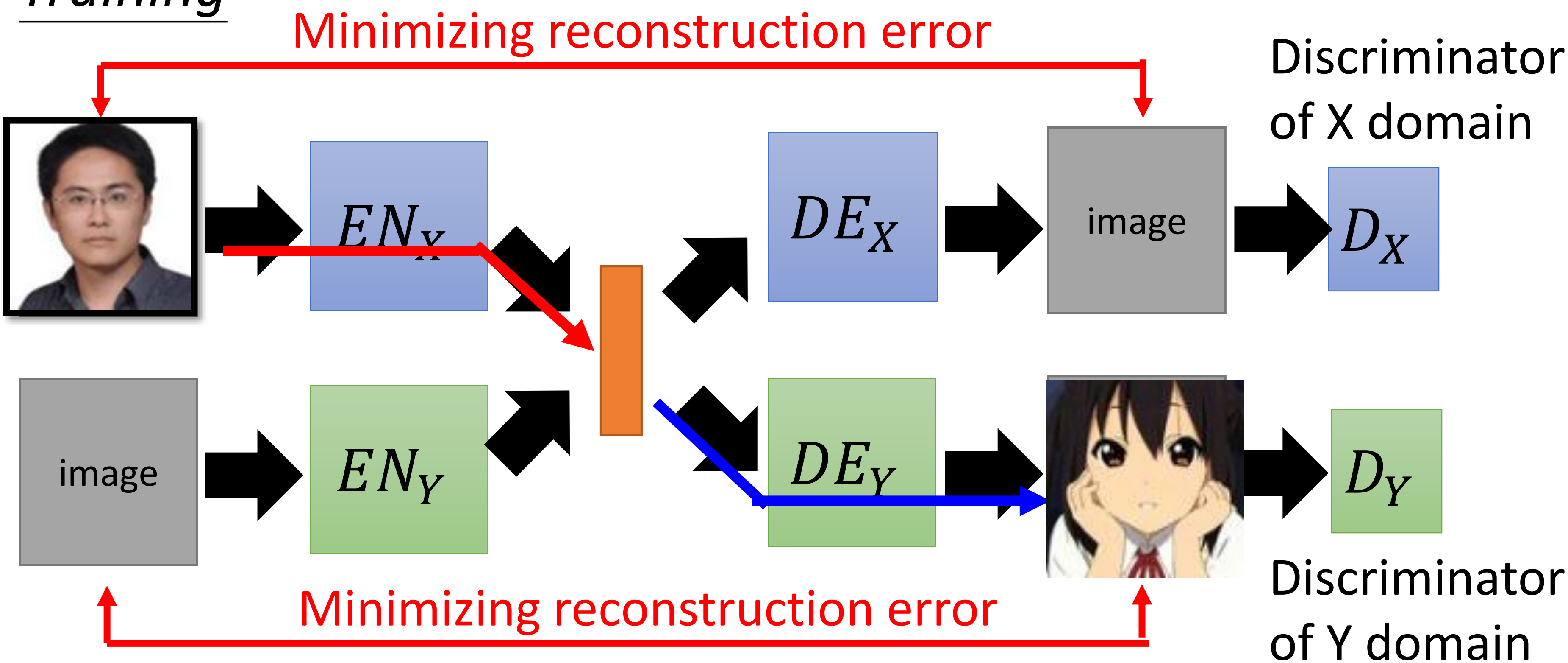
Projection to Common Space

Training



Projection to Common Space

Training

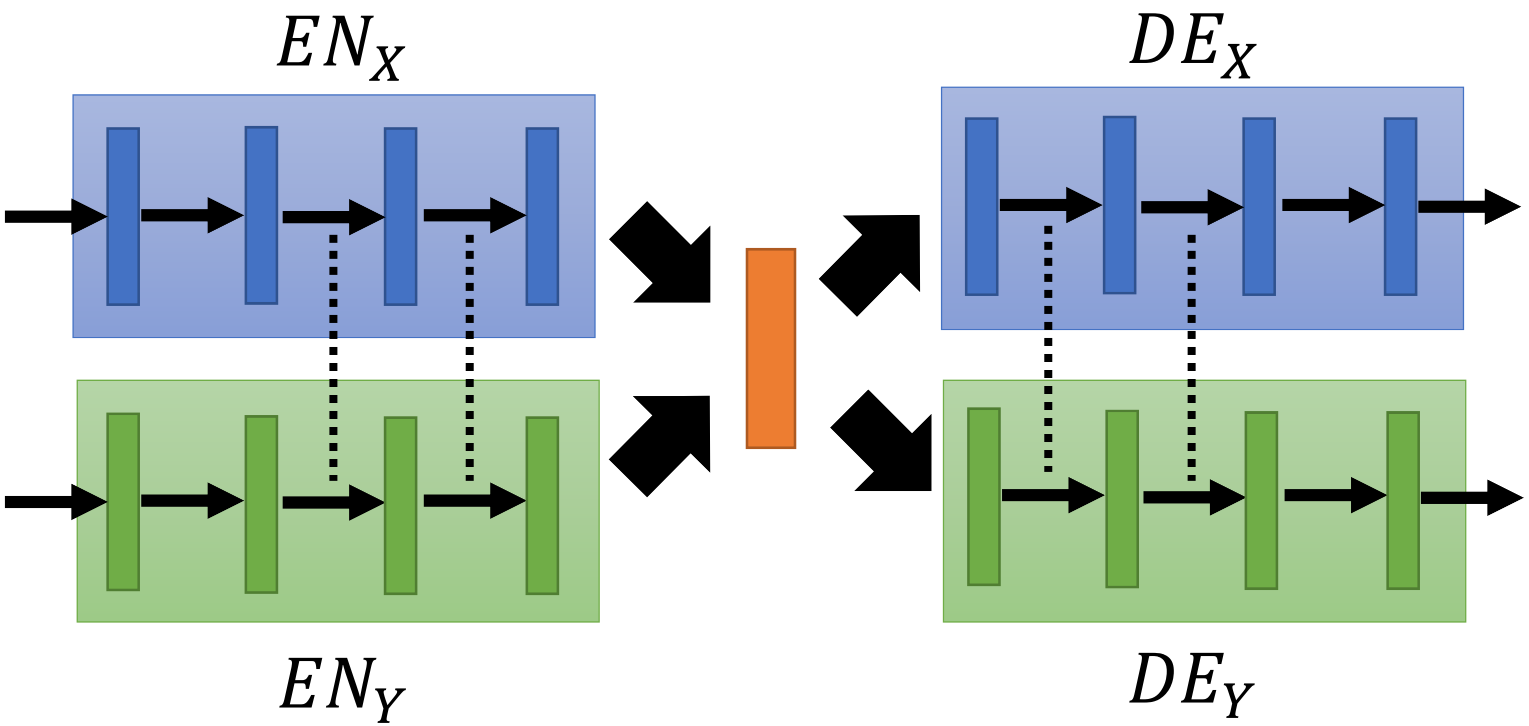


Because we train two auto-encoders separately …

The images with the same attribute may not project to the same position in the latent space.

Projection to Common Space

Training



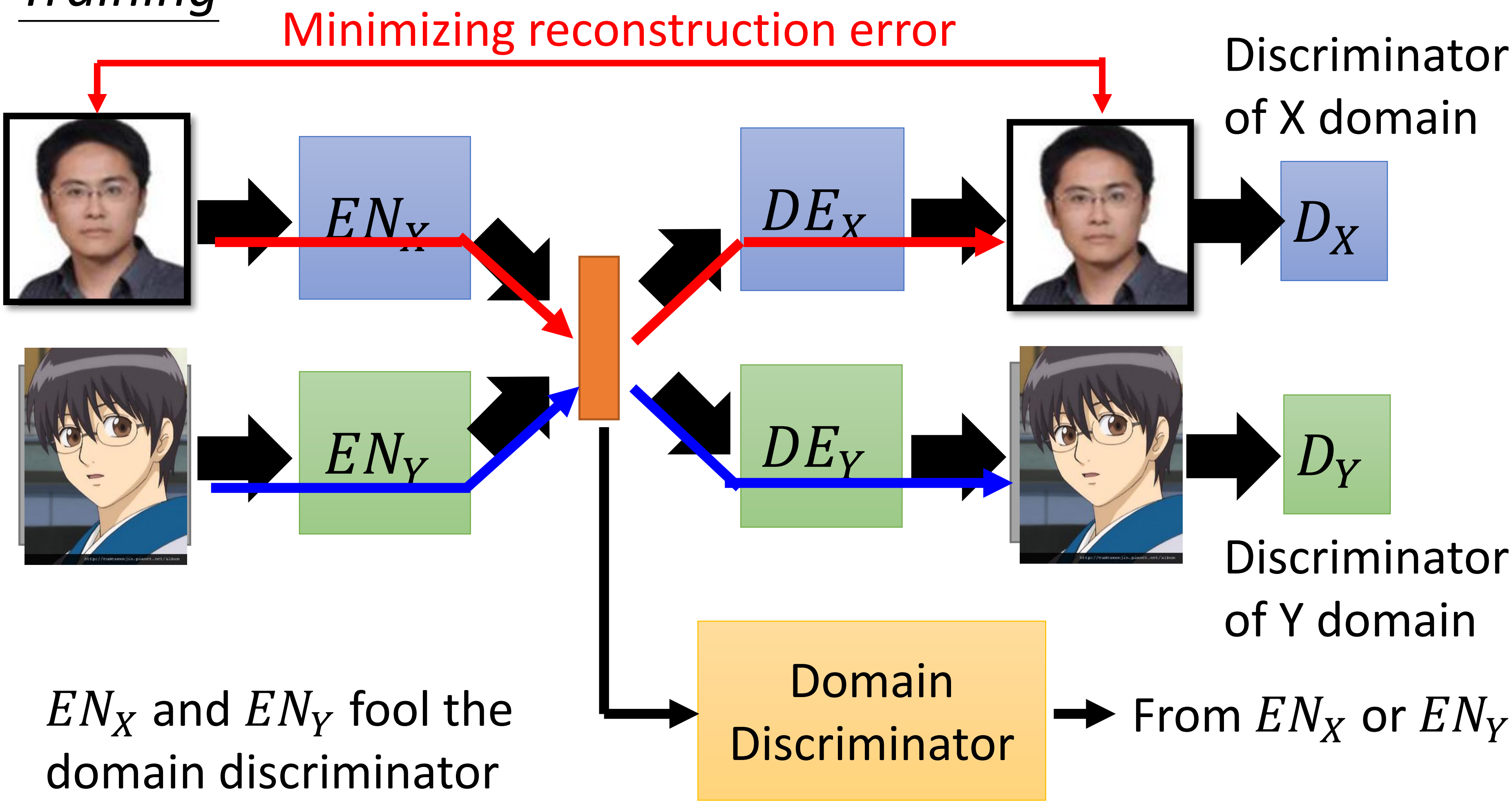
Sharing the parameters of encoders and decoders

Couple GAN[Ming-Yu Liu, et al., NIPS, 2016]

UNIT[Ming-Yu Liu, et al., NIPS, 2017]

Projection to Common Space

Training

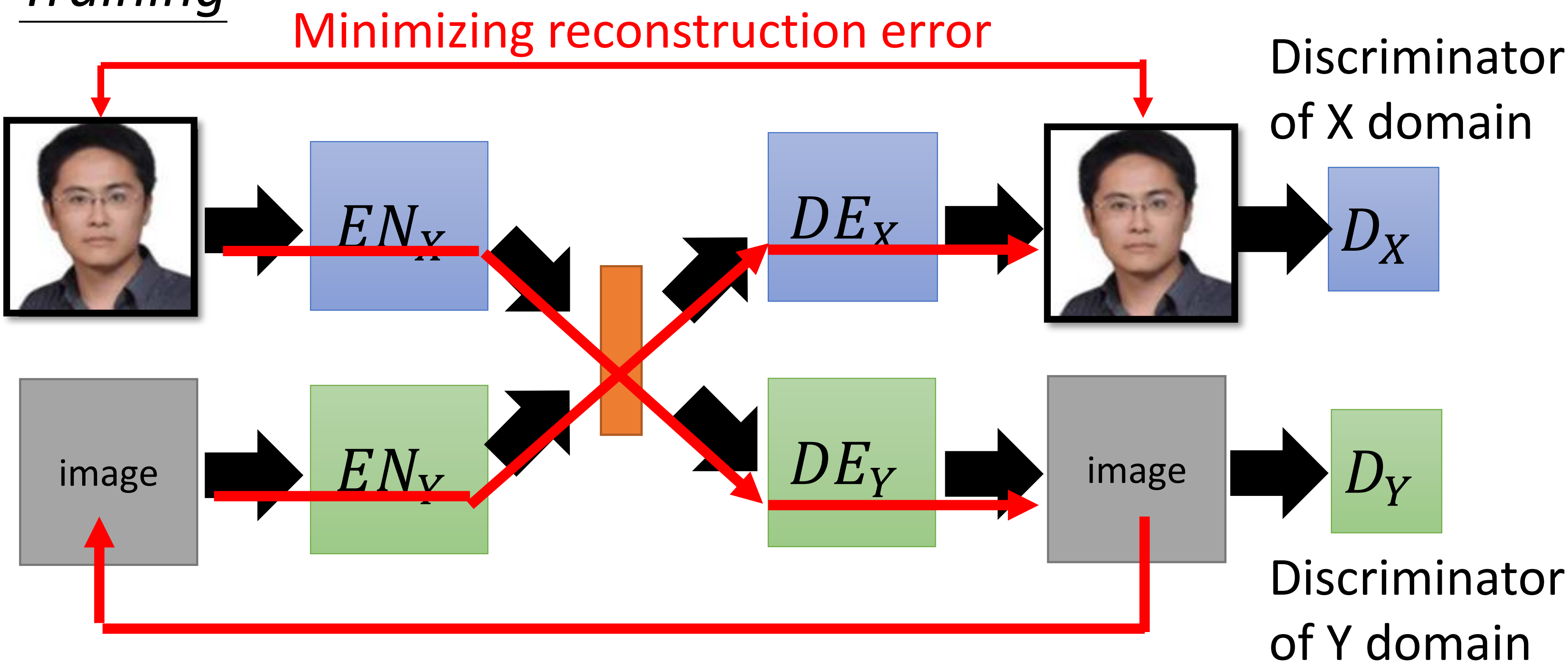


EN_X and EN_Y fool the domain discriminator

The domain discriminator forces the output of EN_X and EN_Y have the same distribution. [Guillaume Lample, et al., NIPS, 2017]

Projection to Common Space

Training

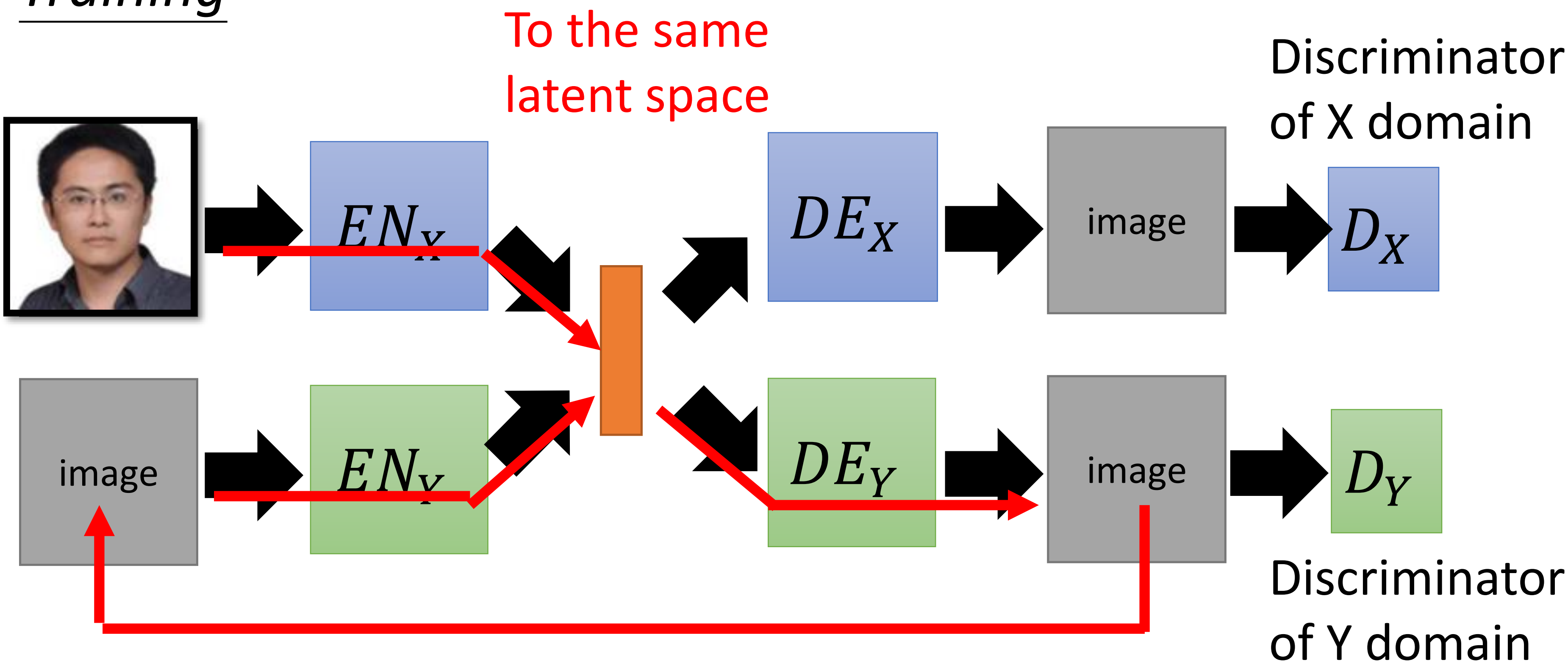


Cycle Consistency:

Used in ComboGAN [Asha Anoosheh, et al., arXiv, 017]

Projection to Common Space

Training



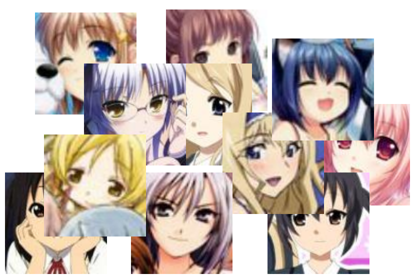
Semantic Consistency:

Used in DTN [Yaniv Taigman, et al., ICLR, 2017] and XGAN [Amélie Royer, et al., arXiv, 2017]

世界二次元化

- Using the code: https://github.com/Hi-king/kawaii_creator
- It is not cycle GAN, Disco GAN

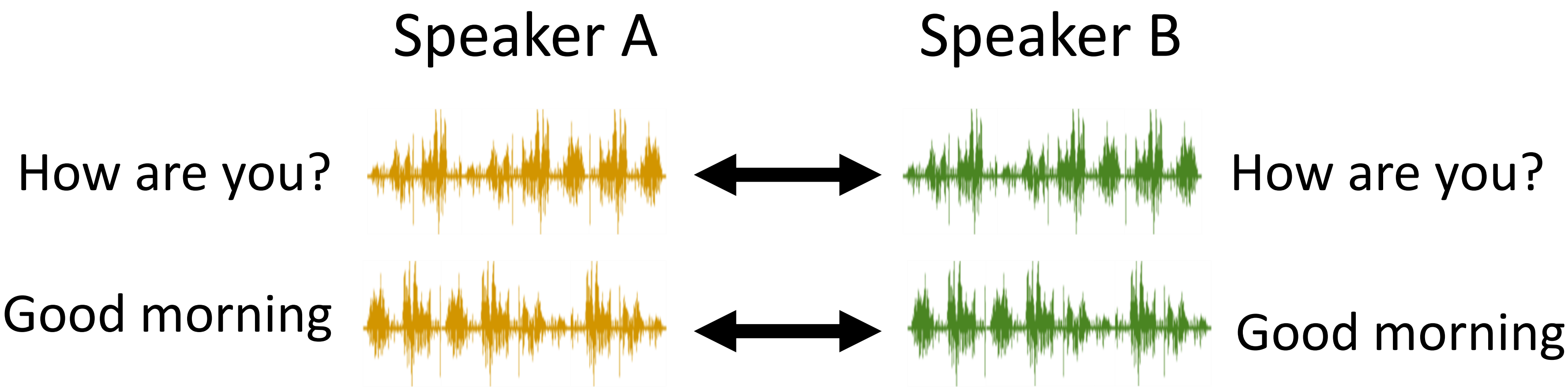
input



output ***domain***

Voice Conversion

In the past



Today

Speakers A and B are talking about completely different things.

Speaker A	Speaker B
我	

感謝周儒杰同學提供實驗結果

Reference

- Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV, 2017
- Zili Yi, Hao Zhang, Ping Tan, Minglun Gong, DualGAN: Unsupervised Dual Learning for Image-to-Image Translation, ICCV, 2017
- Tomer Galanti, Lior Wolf, Sagie Benaim, The Role of Minimal Complexity Functions in Unsupervised Learning of Semantic Mappings, ICLR, 2018
- Yaniv Taigman, Adam Polyak, Lior Wolf, Unsupervised Cross-Domain Image Generation, ICLR, 2017
- Asha Anoosheh, Eirikur Agustsson, Radu Timofte, Luc Van Gool, ComboGAN: Unrestrained Scalability for Image Domain Translation, arXiv, 2017
- Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, Kevin Murphy, XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings, arXiv, 2017

Reference

- Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, Marc'Aurelio Ranzato, Fader Networks: Manipulating Images by Sliding Attributes, NIPS, 2017
- Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, Jiwon Kim, Learning to Discover Cross-Domain Relations with Generative Adversarial Networks, ICML, 2017
- Ming-Yu Liu, Oncel Tuzel, “Coupled Generative Adversarial Networks”, NIPS, 2016
- Ming-Yu Liu, Thomas Breuel, Jan Kautz, Unsupervised Image-to-Image Translation Networks, NIPS, 2017
- Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo, StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, arXiv, 2017