Wireless AI Laboratory

DiscoGAN

Presented by Trong-Binh Nguyen-202388548
Pusan National University

December 5th, 2023

Table of contents

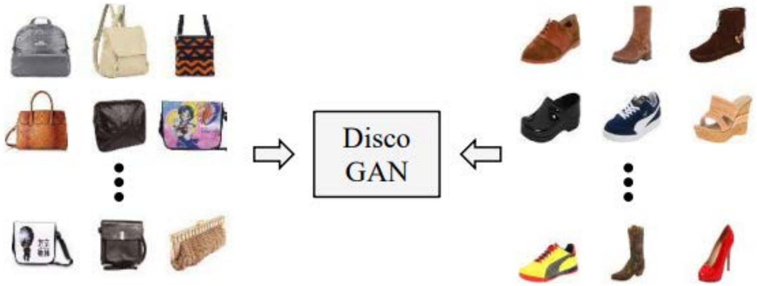
1. Introduction

Paper title: "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks" Kim Taeksoo, et al. International conference on machine learning. PMLR, 2017.

This work proposed method based on GANs that learn to discover relations between different domains

Using the discovered relations, this work successfully transfers style from one domain to another while preserving key attributes such as orientation and face identity.

1. Introduction



(a) Learning cross-domain relations **without any extra label**

(b) Handbag images (input) & **Generated** shoe images (output)

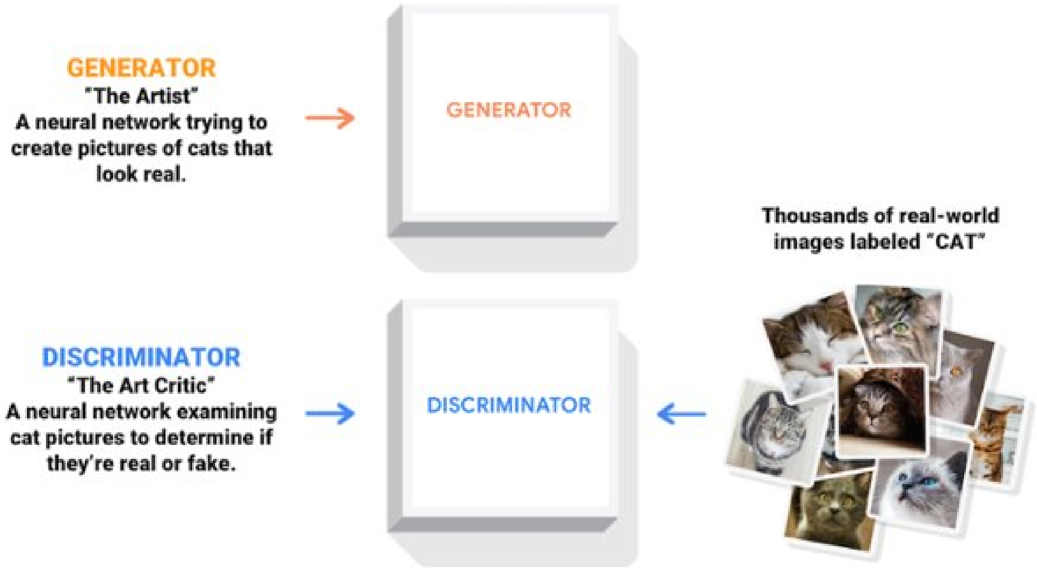
(c) Shoe images (input) & **Generated** handbag images (output)

2. GANs Revision

Adversarial Networks (GANs) are a class of artificial intelligence algorithms introduced by Ian Goodfellow and his colleagues in 2014.

GANs are designed for generative tasks, aiming to create new data instances that resemble a given dataset.

GANs consist of two models: generator and discriminator



2. GANs Revision

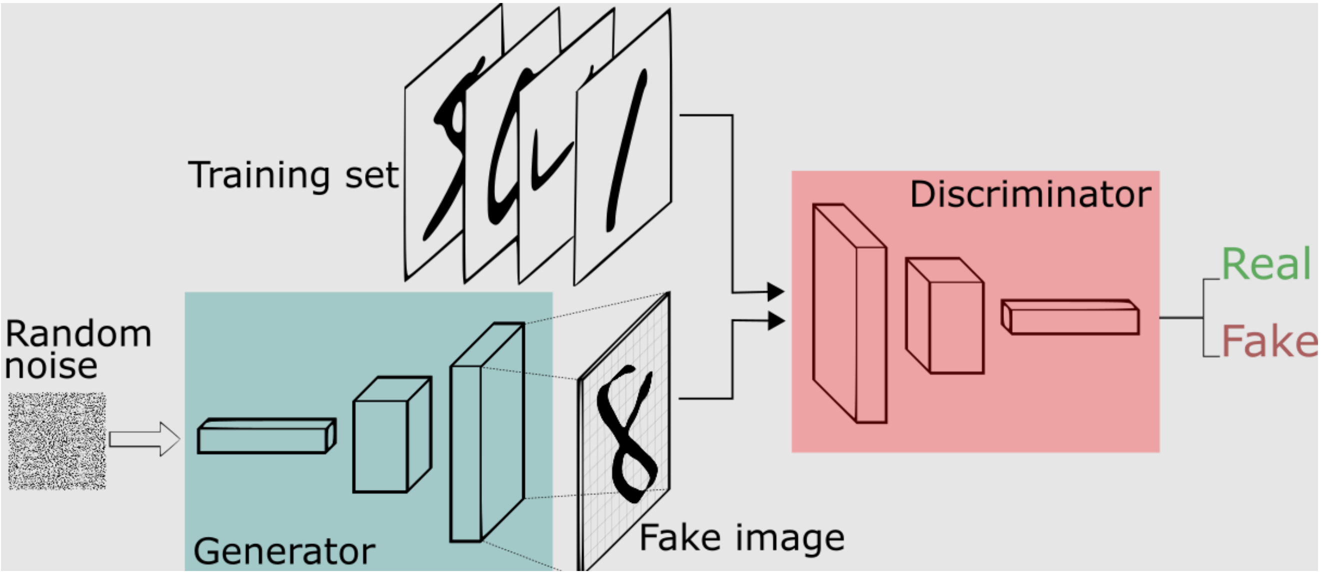
Generator (G):

- The generator is a neural network responsible for creating synthetic data.
- It takes random noise as input and transforms it into data that should resemble the real data.

Discriminator (D):

- The discriminator is another neural network that acts as a binary classifier.
- It distinguishes between real data from the dataset and fake data generated by the generator

Minimax two-player game between Generator vs Discriminator



2. GANs Revision

Minimax problem of GANs

Input random noise: $p_{\boldsymbol{z}}(\boldsymbol{z})$
Generator data: $G(\boldsymbol{z};\theta_g)$
Discriminator: $D(\boldsymbol{x};\theta_d)$

$D \in [0, 1]$

data is fake/real

$$\min_G \max_D V(G, D) = \underbrace{\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})]}_{\text{Discriminant for real data}} + \underbrace{\mathbb{E}_{z\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]}_{\text{Discriminant for generated data}}$$

Training of Generator	–	$\min_G[1 - D(G(z))] = 0$
Training of Discriminator	$\max_D D(x) = 1$	$\max_D[1 - D(G(z))] = 1$

$V(D, G)$ has a saddle point at $D(G(\boldsymbol{z})) = \frac{1}{2}$

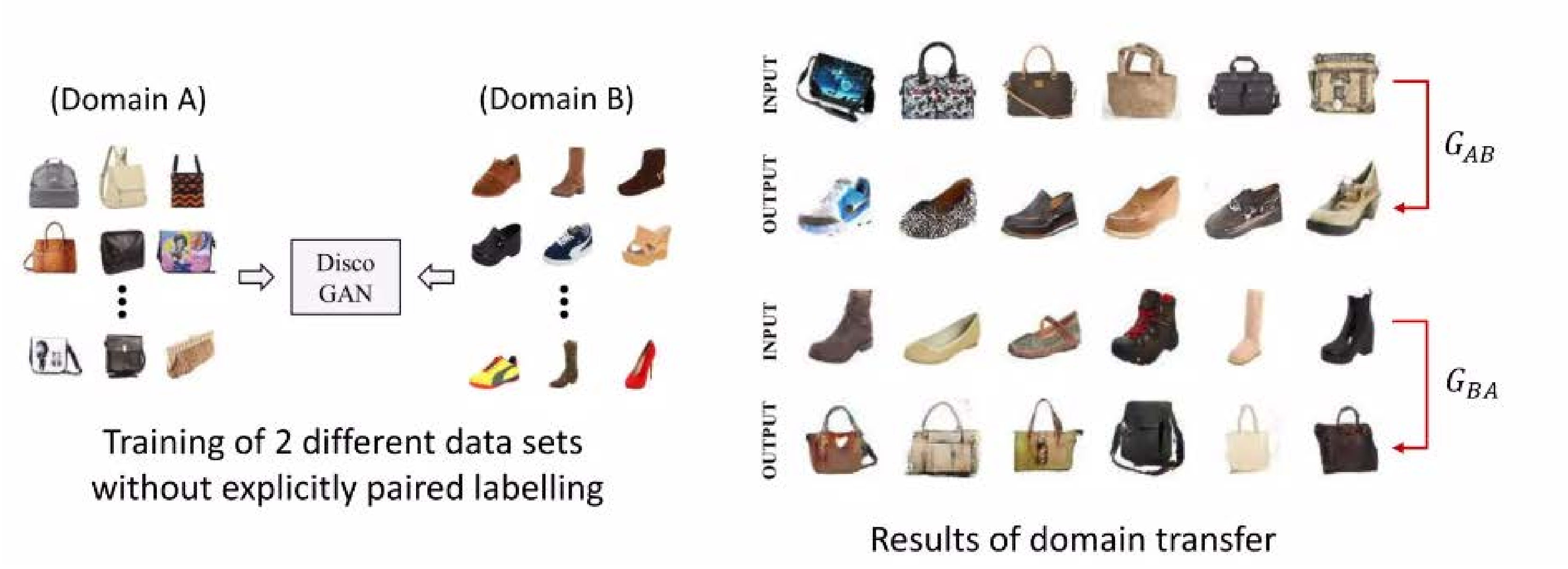
3. DiscoGAN

Discover Cross-Domain Relations with GANs.

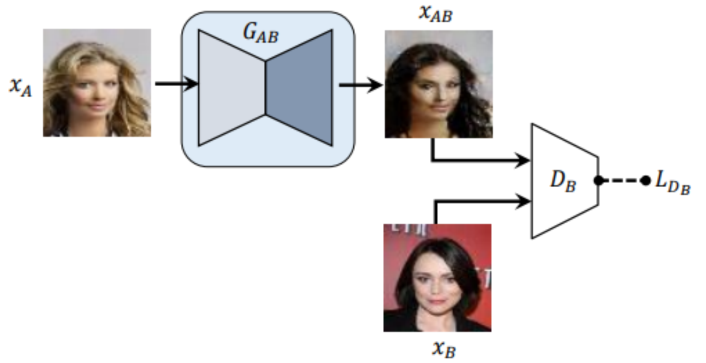
Previous AI (~2017) could also transfer data from one domain to another, preserving key attributes such as: Pix2Pix, …

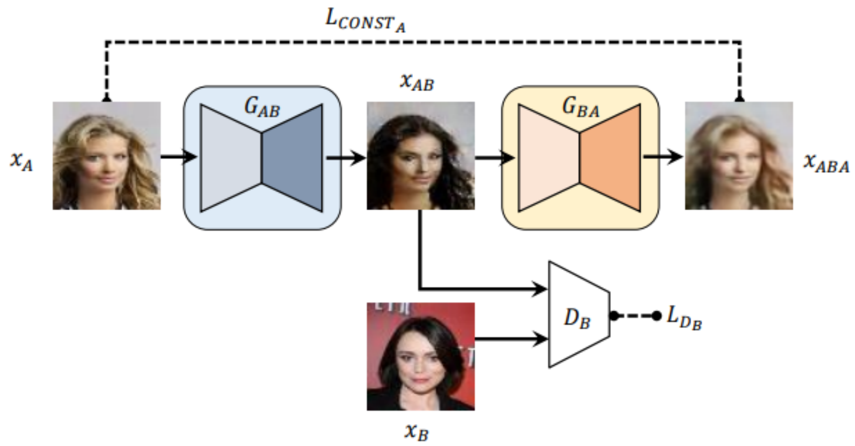
These training methods required paired data, that is costly and hard to collect

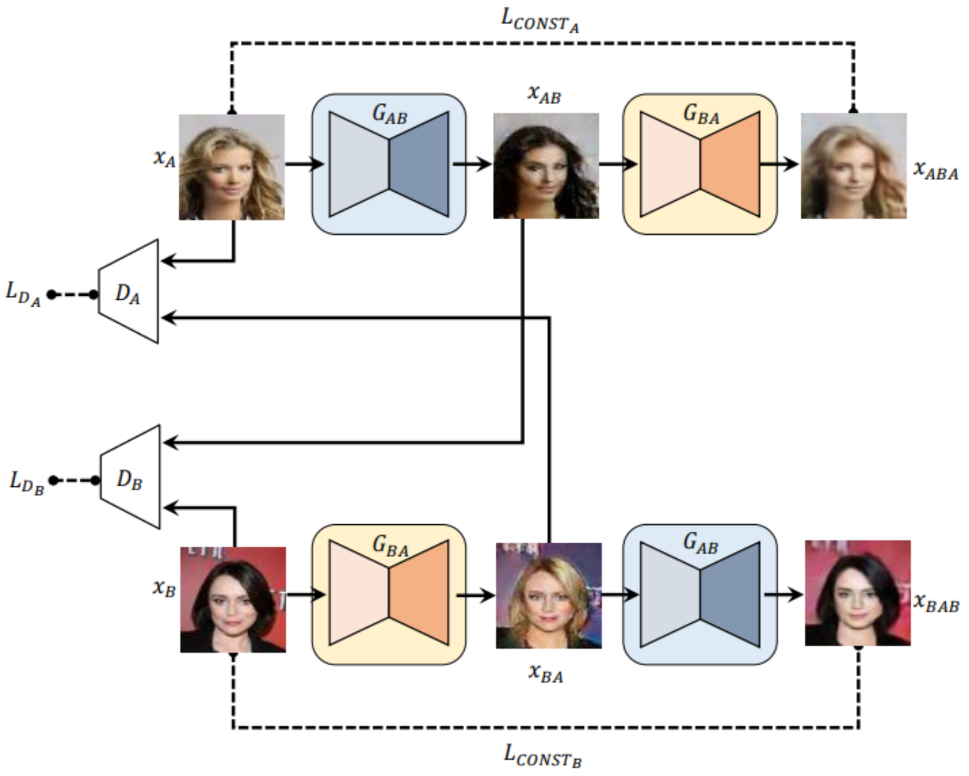
DiscoGAN required 2 different data sets **without any paired data**, and the results show better performance with robustness to model collapse



3. DiscoGAN: *Network models*







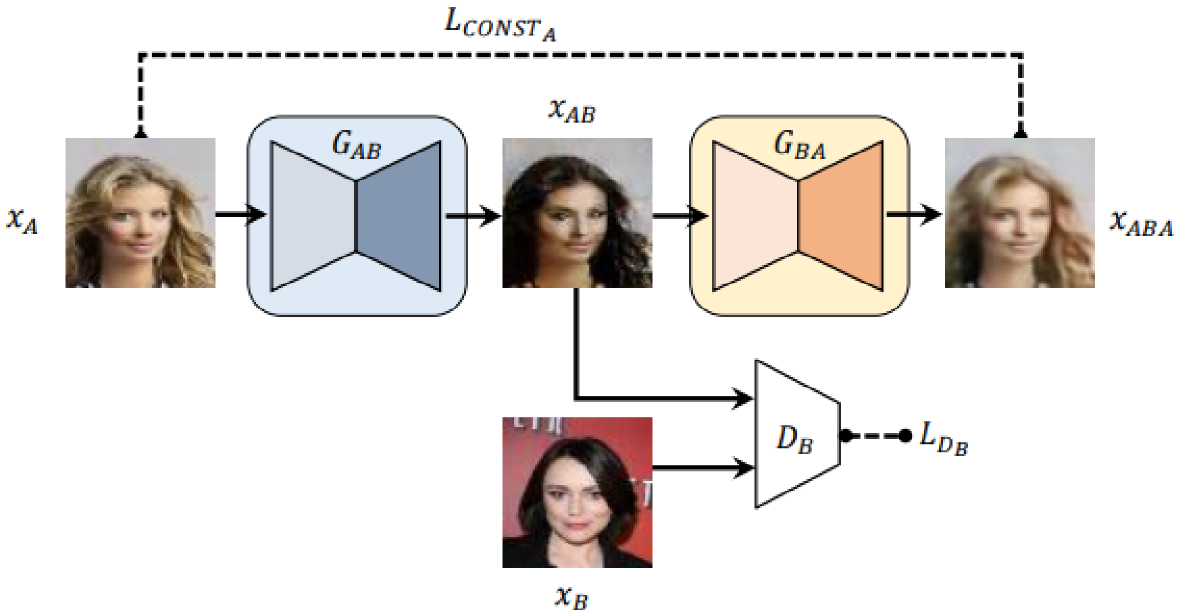
Each generator consists of encoder-decoder pair (input and output are images)
GAN loss (and the reconstruction) is to be minimized on training processes
In DiscoGAN, 2 coupled GANs map each domain to its counterpart domain (bijective)

3. DiscoGAN: *Idea*

Intuition of cross domain: constraint all images in one domain to be representable by images in the other domain

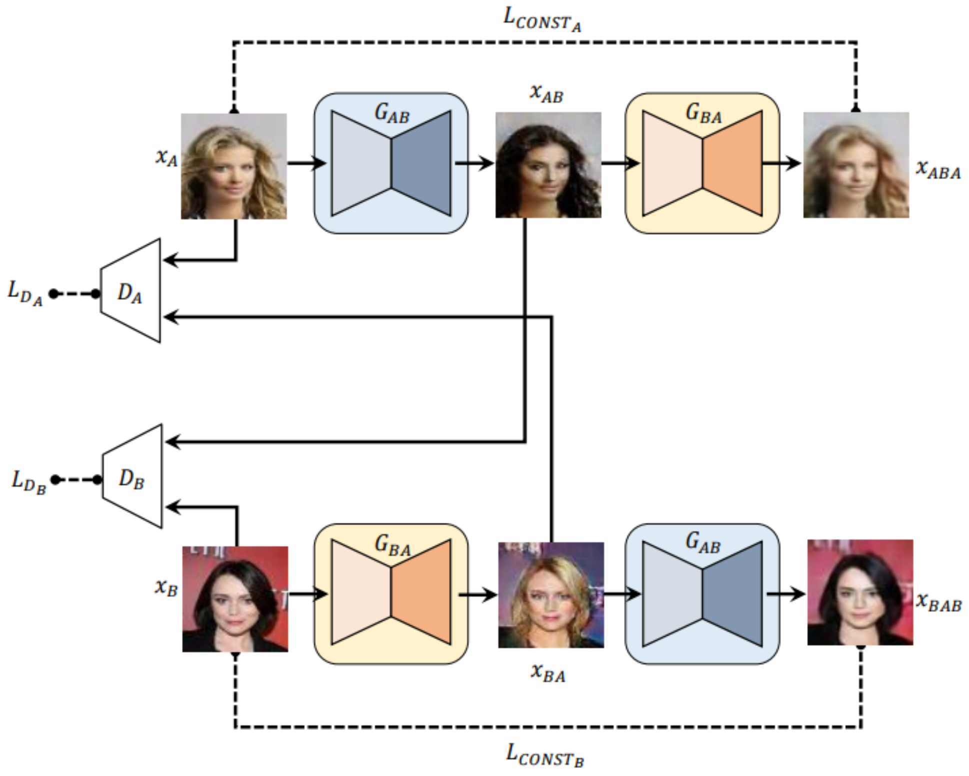
Reconstruct loss: measures how well original input is reconstructed after a sequence of two generation: $L_{CONST_A} = d(\mathbf{G}_{BA} \circ \mathbf{G}_{AB}(x_A), x_A)$

Gans Loss: measures how good image is in domain B



This is only injection mapping
One side mapping

3. DiscoGAN: *Idea*



- Bijection: ideally $G_{AB}^{-1} = G_{BA}$
  $\rightarrow \min_{G_{AB}}(L_{CONST_A}), \min_{G_{BA}}(L_{CONST_B})$
- Domain transition: ideally $\boldsymbol{x}_{AB} \in \mathbb{R}^B, \boldsymbol{x}_{BA} \in \mathbb{R}^A$
  $\rightarrow \min_{D_B}(L_{D_B}), \min_{D_A}(L_{D_A})$

2 models coupled to guarantee bijection and domain transition

3. DiscoGAN: *Idea*

GAN with a reconstruction loss
and Gan loss

$$L_{G_{AB}} = L_{GAN_B} + L_{CONST_A}$$

$$\begin{aligned} L_{D_B} \quad = \quad & -\mathbb{E}_{x_B \sim P_B}[\log \mathbf{D}_B(x_B)] \\ & -\mathbb{E}_{x_A \sim P_A}[\log(1 - \mathbf{D}_B(\mathbf{G}_{AB}(x_A)))] \end{aligned}$$

DiscoGAN

$$\begin{aligned} L_G &= L_{G_{AB}} + L_{G_{BA}} \\ &= L_{GAN_B} + L_{CONST_A} + L_{GAN_A} + L_{CONST_B} \end{aligned}$$

$$L_D = L_{D_A} + L_{D_B}$$

3. DiscoGAN: *Mode Collapse*

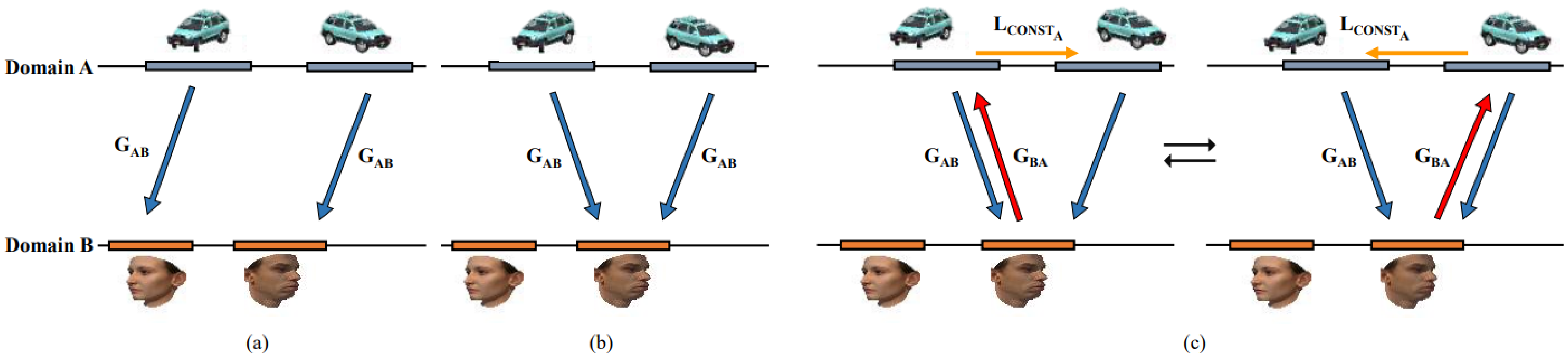
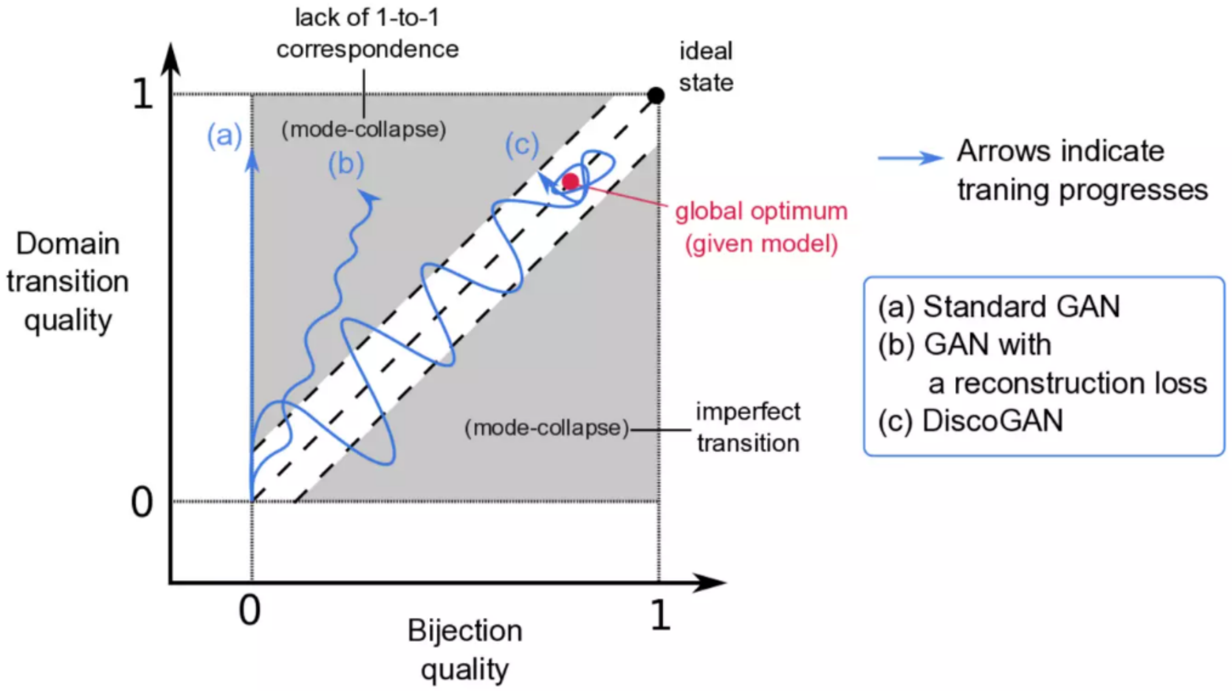


Figure 3. Illustration of our models on simplified one dimensional domains. (a) ideal mapping from domain A to domain B in which the two domain A modes map to two different domain B modes, (b) GAN model failure case, (c) GAN with reconstruction model failure case.

3. DiscoGAN

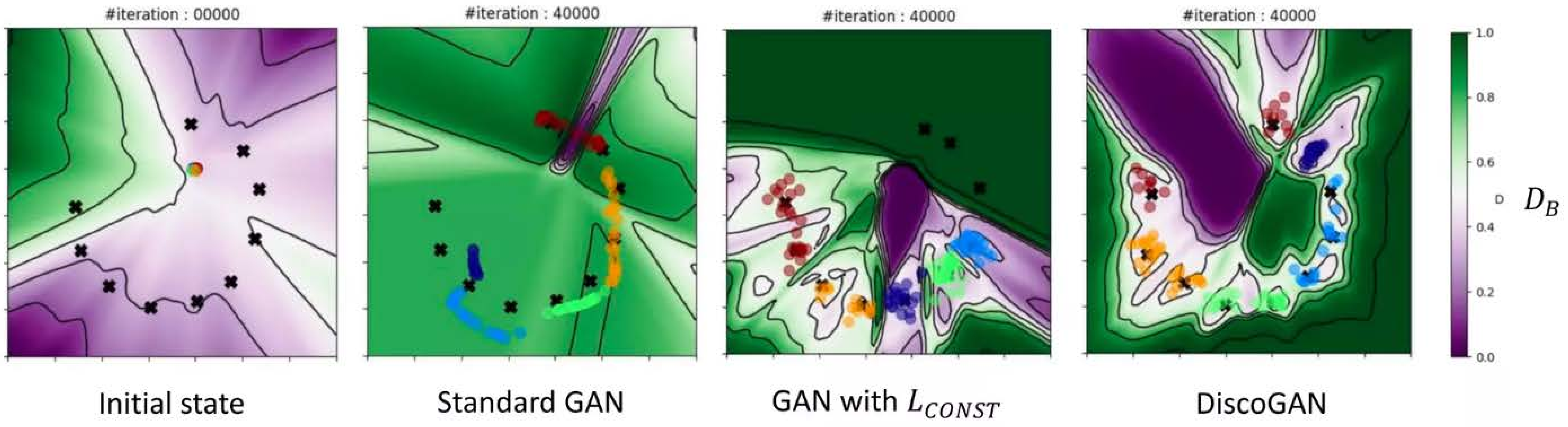
Why DiscoGAN is robust to mode-collapse?



- In DiscoGAN, two coupled models are trained together simultaneously. G_{AB}'s and G_{BA}'s share parameters
- Constraints of coupled reconstruction losses lead to the strict bijection

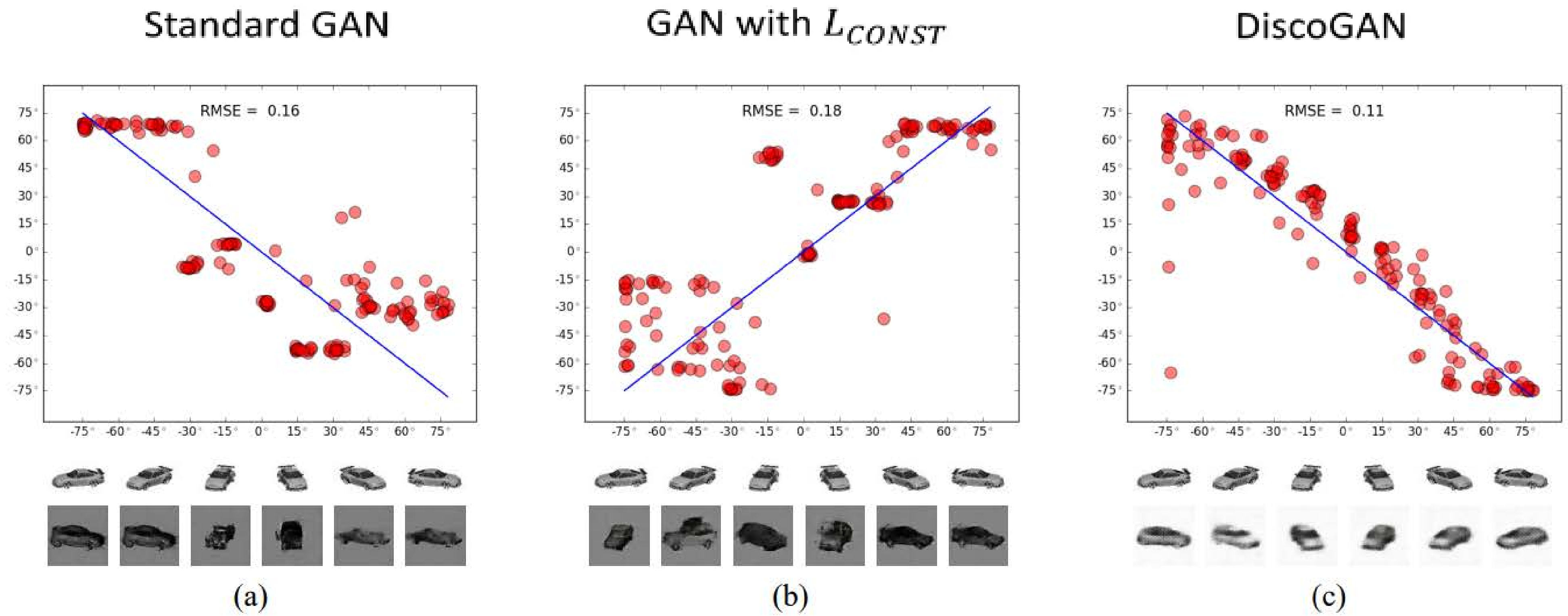
3. DiscoGAN: *Performance*

Colored points: samples in domain A
Black x's: target modes in domain B



- In DiscoGAN, discriminator B is perfectly fooled by translated sampled from domain A
- DiscoGAN prevents mode-collapse by translating into distinct well-bounded regions that do not overlap

3. DiscoGAN: *Performance*



3. DiscoGAN: *Performance*

Translation of gender

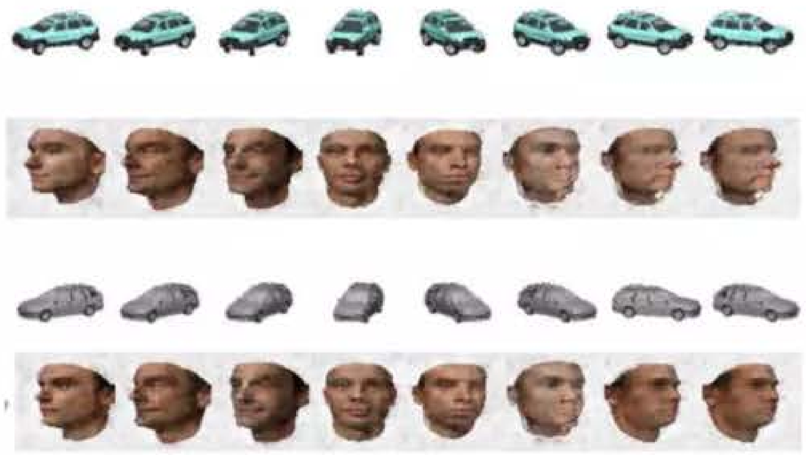
Blond to black,
Black to blond hair

Glasses to non-glasses,
non-glasses to glasses

- DiscoGAN translates specific feature, preserving other facial features

3. DiscoGAN: *Performance*

Chair to car



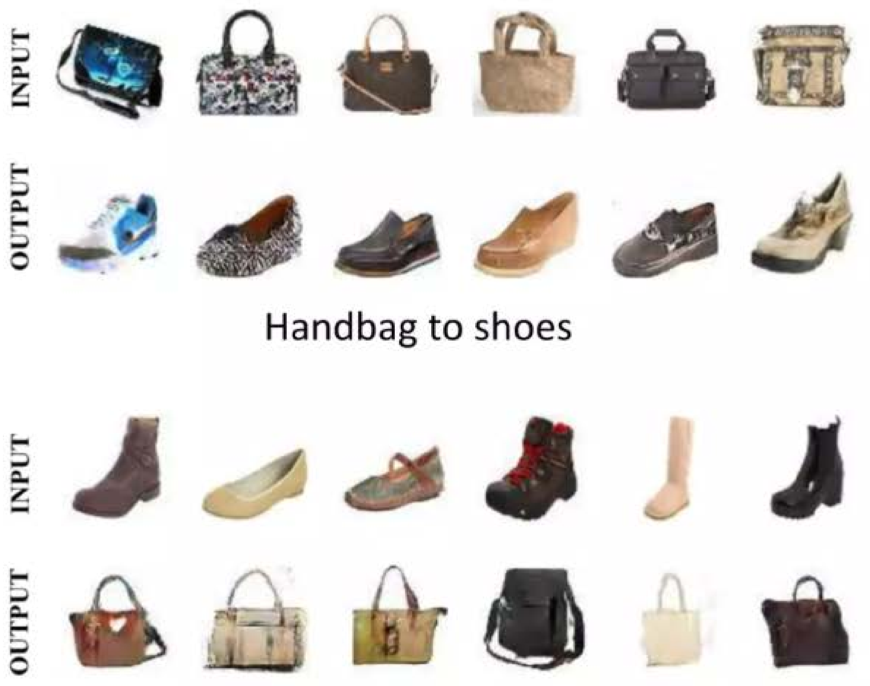
Car to face

- Note that training is implemented without any paired data
- The main attribute (azimuth) is preserved

3. DiscoGAN: *Performance*

- 1-to-N problem

3. DiscoGAN: *Performance*



Handbag to shoes

Shoes to handbag

- Same style is discovered

3. DiscoGAN: *Summary*

- DiscoGAN is proposed as a learning method to discover over cross-domain relations without any pair labels
- Results show better performance with robustness to mode collapse
- Symmetry design by coupling 2 GANs
- https://github.com/SKTBrain/DiscoGAN.git

QUESTION