



Application of Generative Models: Image-to-Image Translation

Hao Dong

Peking University



Application of Generative Models: Image-to-Image Translation

Why we learn im2im?

- The most classical generative model application ..
- The state-of-the-art methods are all based on GAN ...
- Understand GAN and the history better ...

Application of Generative Models: Image-to-Image Translation



- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

- Problem Definition
 - Image Inpainting / Reconstruction / Super Resolution
 - Pix2Pix: paired data
 - Discussion: ideal im2im
 - UNIT and CycleGAN: unpaired data
 - BiCycleGAN: multi-modality
 - MUNIT and Augmented CycleGAN: unpaired data + multi-modality
 - DRIT: disentangle domain-specific features
 - Attention CycleGAN: maintain background
 - StarGAN: label condition
 - Breaking the Cycle
 - GAN-CLS and SisGAN: text condition



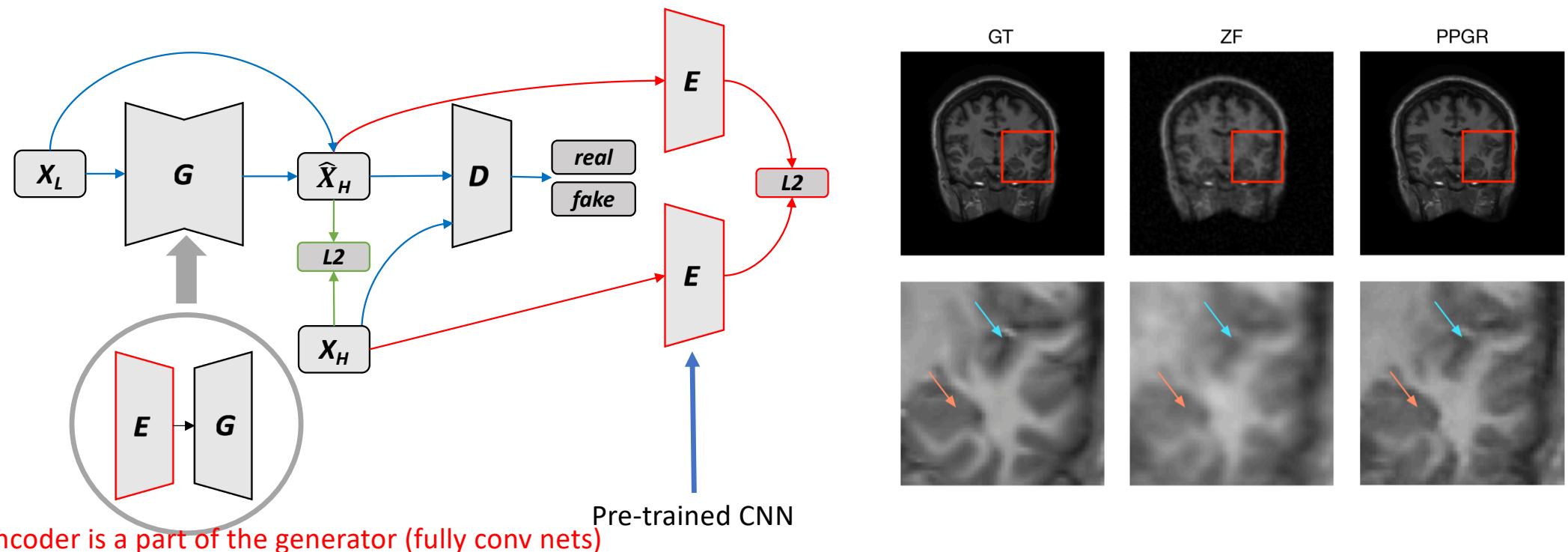
Problem Definition

- Supervised/Paired image-to-image translation
- Unsupervised/Unpaired image-to-image translation

- Problem Definition
- **Image Inpainting / Reconstruction / Super Resolution**
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

Image Inpainting / Reconstruction / Super Resolution

- Utilising Feature Information for Medical Image Reconstruction



Deep De-Aliasing for Fast Compressive Sensing MRI. S. Yu, H. Dong, G. Yang et al. arXiv:1705.07137 2017.

DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction.

G. Yang, S. Yu, H. Dong et al. TMI 2017.

Image Inpainting / Reconstruction / Super Resolution

- Supervised image super resolution

Better feature reconstruction

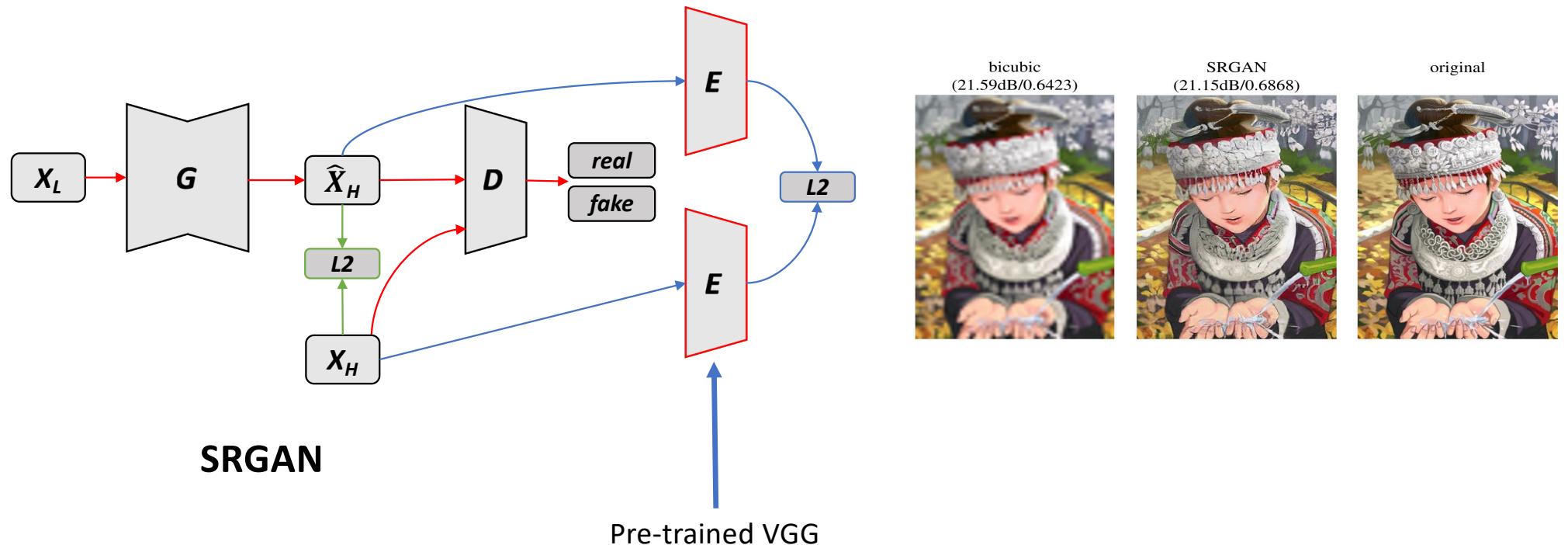


Image Inpainting / Reconstruction / Super Resolution

- Supervised image super resolution

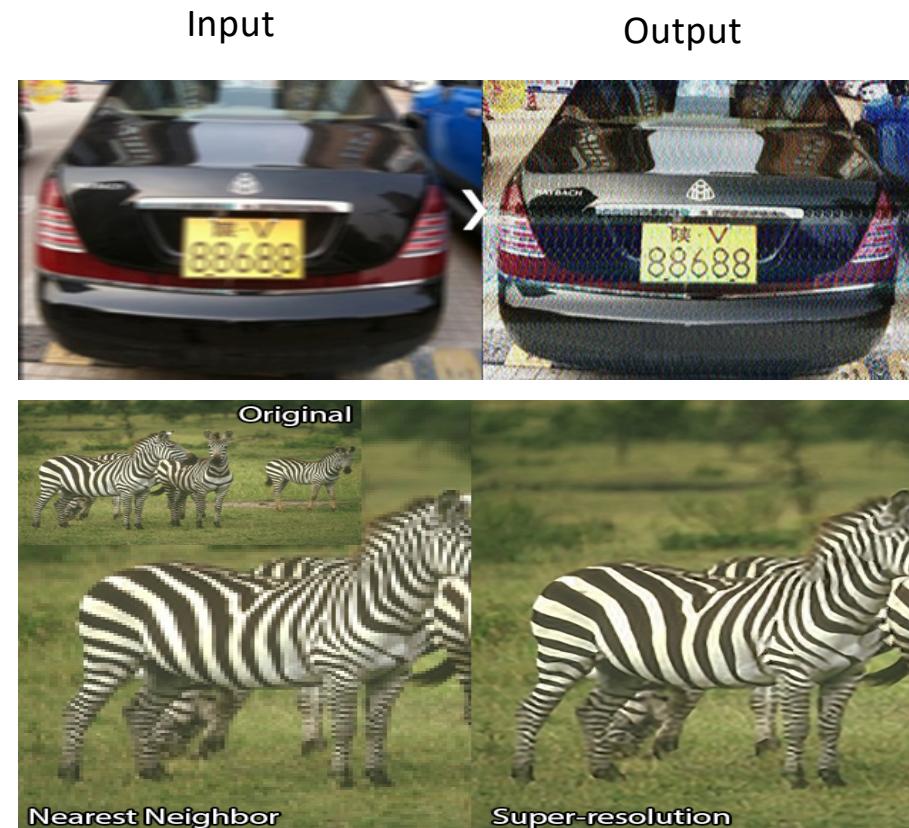
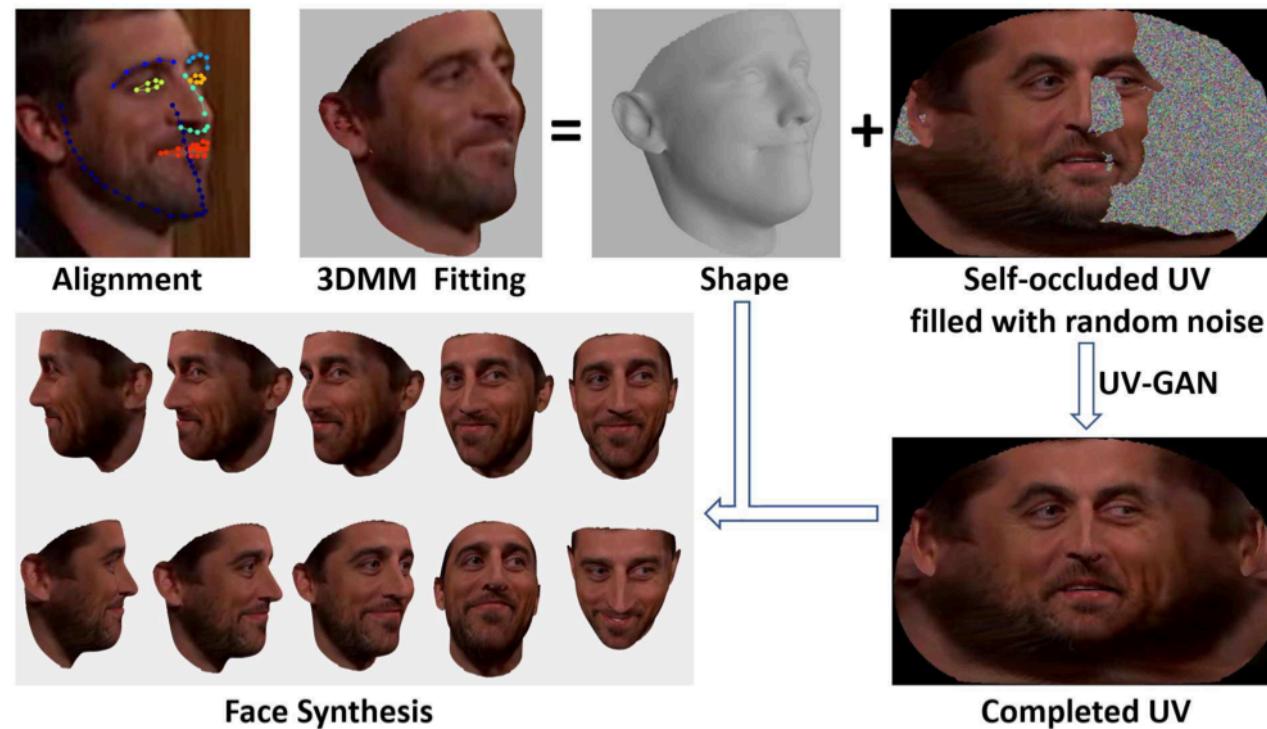


Image Inpainting / Reconstruction / Super Resolution

- UV-GAN



UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition.
J. Deng, S. Cheng et al. CVPR. 2018.



- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- **Pix2Pix: paired data**
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

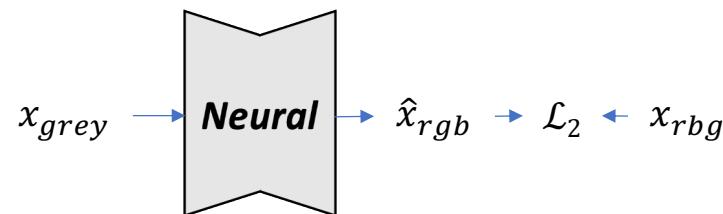
Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

- Beyond MLE: Adversarial Learning



- Question 1: What color are they?
Red? Blue? Yellow? ... obviously there are more than one solution
- Question 2: What if I train a neural net: input x_{grey} output x_{rgb} with MLE?



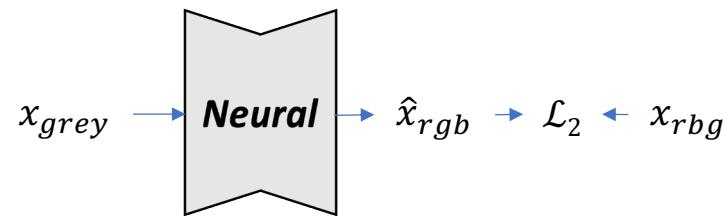
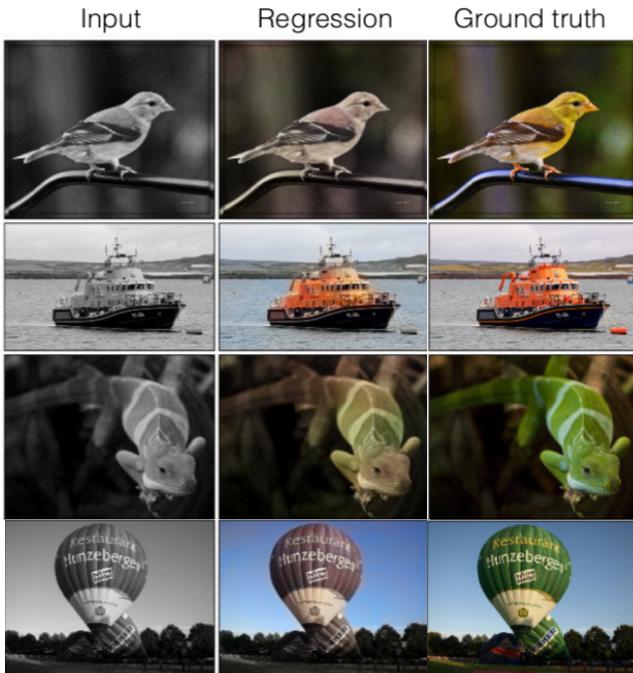
What is the problem??

Colorful Image Colorization. R. Zhang, P. Isola, A.A. Efros. ECCV. 2016.

Image-to-Image Translation with Conditional Adversarial Networks. P. Isola, J. Zhu et al. CVPR 2017.

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation
 - Beyond MLE: Adversarial Learning



Different colors will have conflicts,
(some want red, some want blue, ...)
resulting “grey” outputs

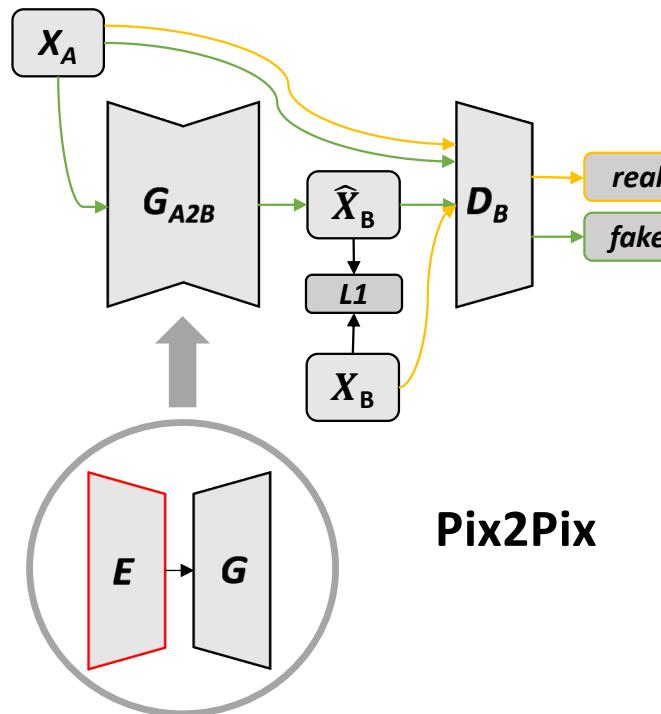
Colorful Image Colorization. *R. Zhang, P. Isola, A.A. Efros. ECCV. 2016.*

Image-to-Image Translation with Conditional Adversarial Networks. *P. Isola, J. Zhu et al. CVPR 2017.*

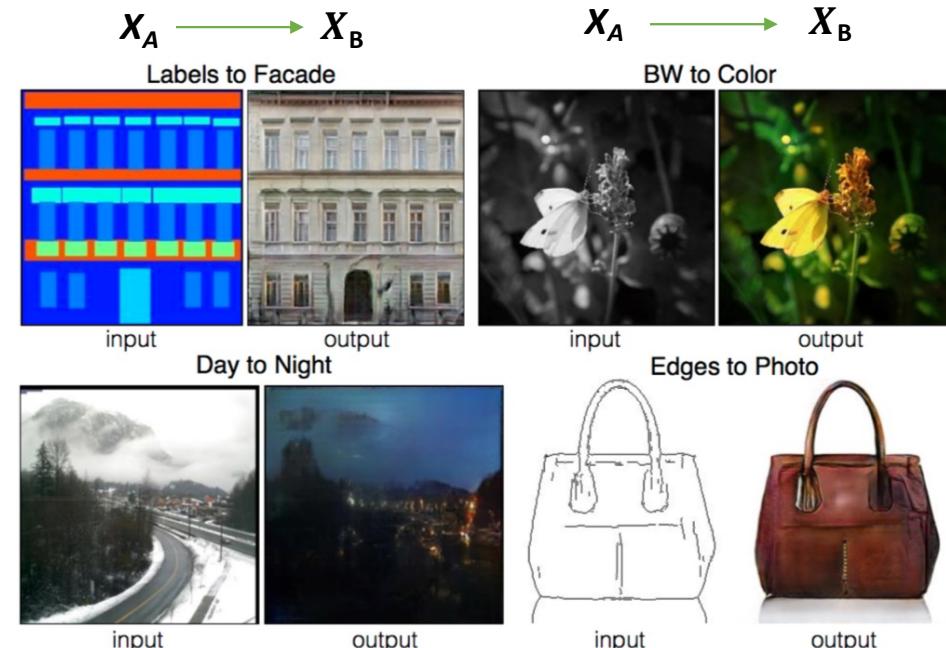
Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

- Beyond MLE: Adversarial Learning



Encoder is a part of the generator (fully conv nets)



$$\mathcal{L}_D = \mathbb{E}_{x \sim p_{data}} [\log D(x_A, x_B)] + \mathbb{E}_{x \sim p_{data}} [\log(1 - D(x_A, G(x_A)))]$$

$$\mathcal{L}_G = \mathbb{E}_{x \sim p_{data}} [\log D(x_A, G(x_A))]$$

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

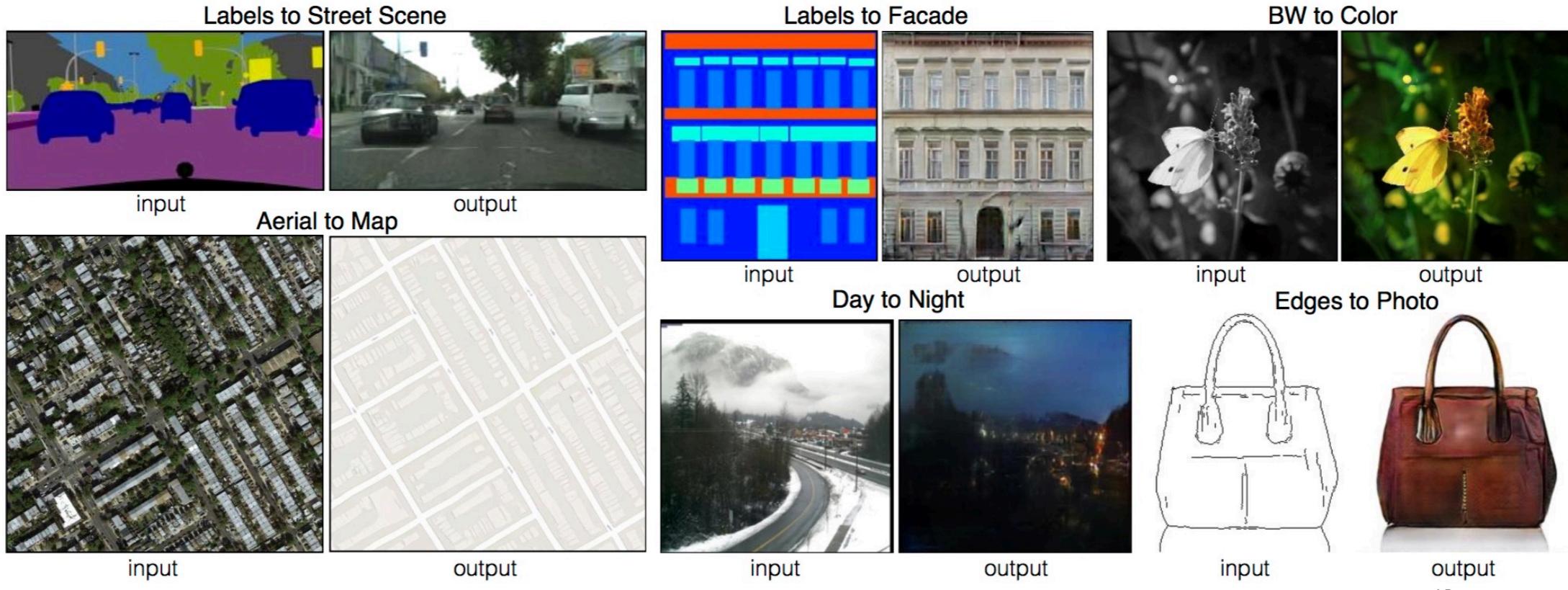


Image-to-Image Translation with Conditional Adversarial Networks. P. Isola, J. Zhu et al. CVPR 2017.

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation



Image-to-Image Translation with Conditional Adversarial Networks. P. Isola, J. Zhu et al. CVPR 2017.

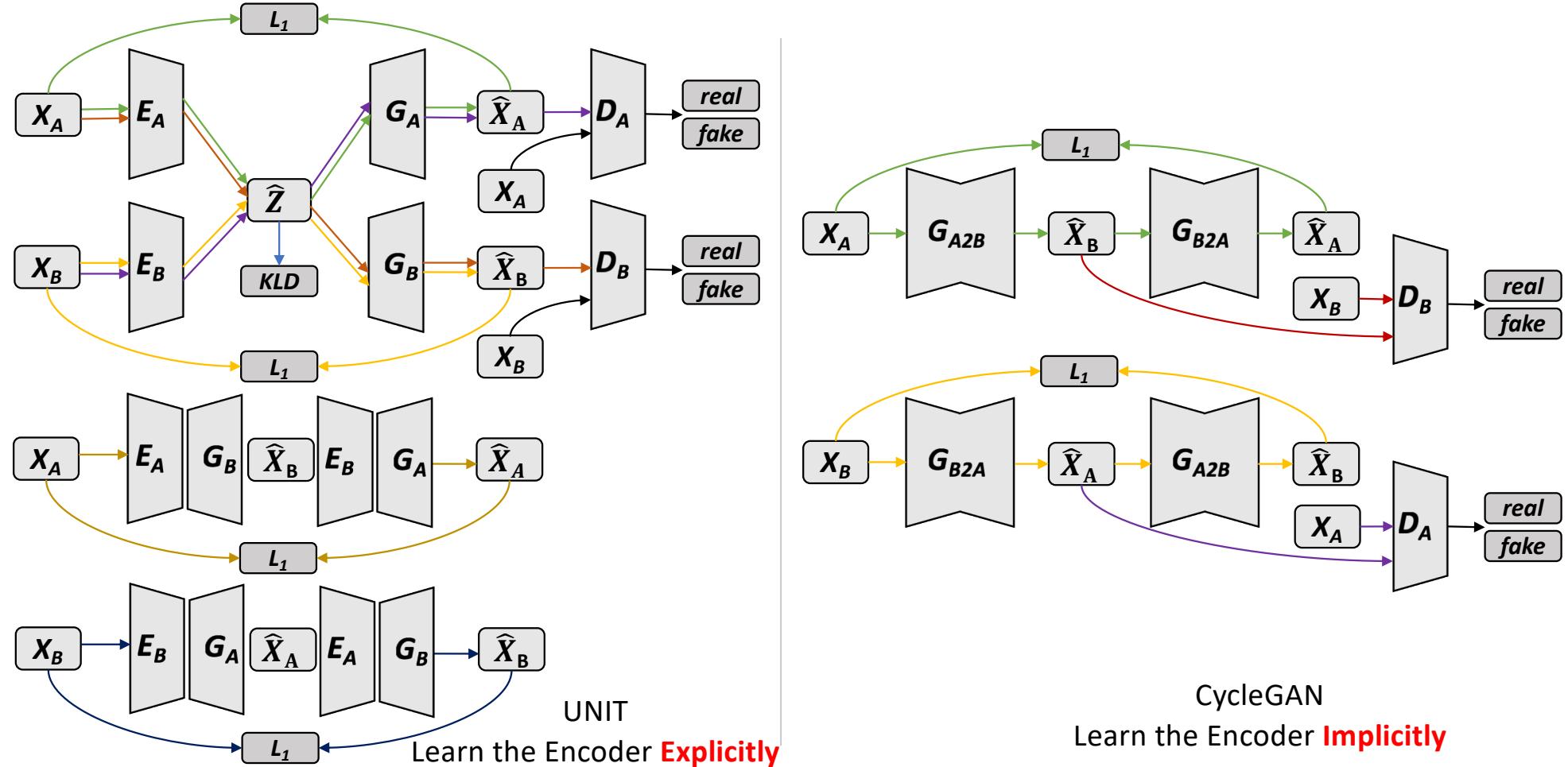
- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- **Discussion: ideal im2im**
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

Discussion: ideal im2im

- What should the ideal image-to-image translation to be?
 - Unpaired data
 - Maintain background
 - Multi-modality
 - Disentanglement
 - Multi-domain
 - Conditional translation

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- **UNIT and CycleGAN: unpaired data**
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

GAN with Encoder -- Unsupervised Image-to-Image Translation

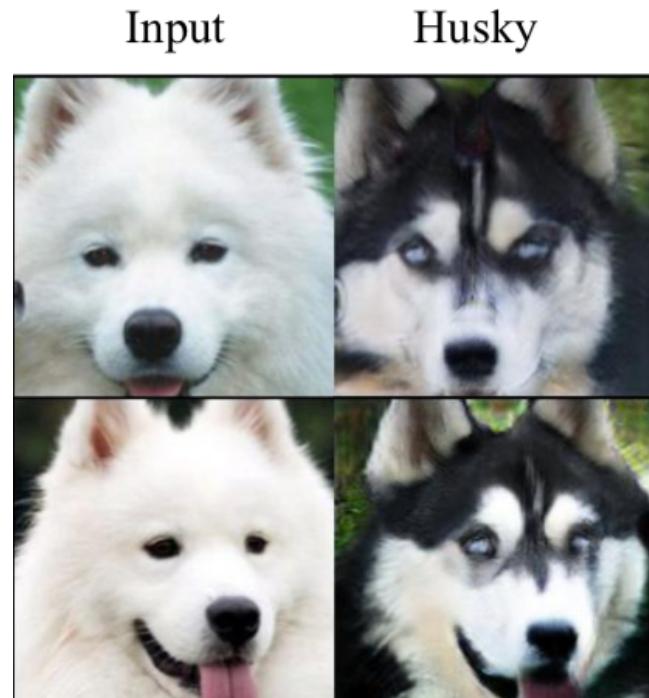


Unsupervised image-to-image translation networks. M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. J. Zhu, T. Park et al. ICCV 2017.

UNIT and CycleGAN: unpaired data

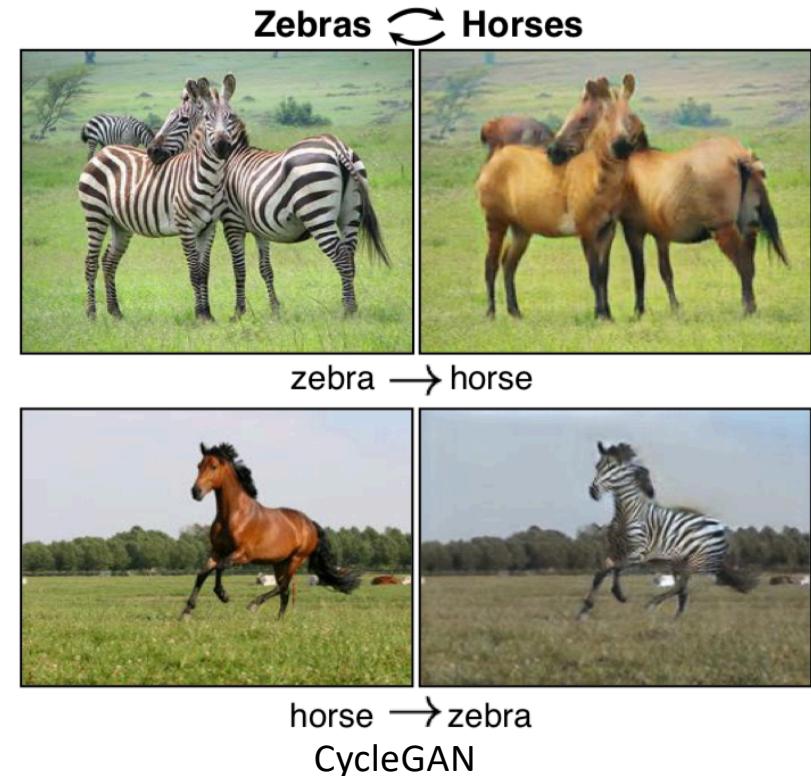
- CycleGAN: Unpaired Image-to-Image Translation



Liu et al.

Learn the Encoder **Explicitly**

Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*
 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*



Learn the Encoder **Implicitly**

UNIT and CycleGAN: unpaired data

- CycleGAN: Unpaired Image-to-Image Translation



Input GTA5 CG

<https://blog.csdn.net/gdymind>



Output image with German street view style blog.csdn.net/gdymind

Unsupervised image-to-image translation networks. M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. J. Zhu, T. Park et al. ICCV 2017.

UNIT and CycleGAN: unpaired data

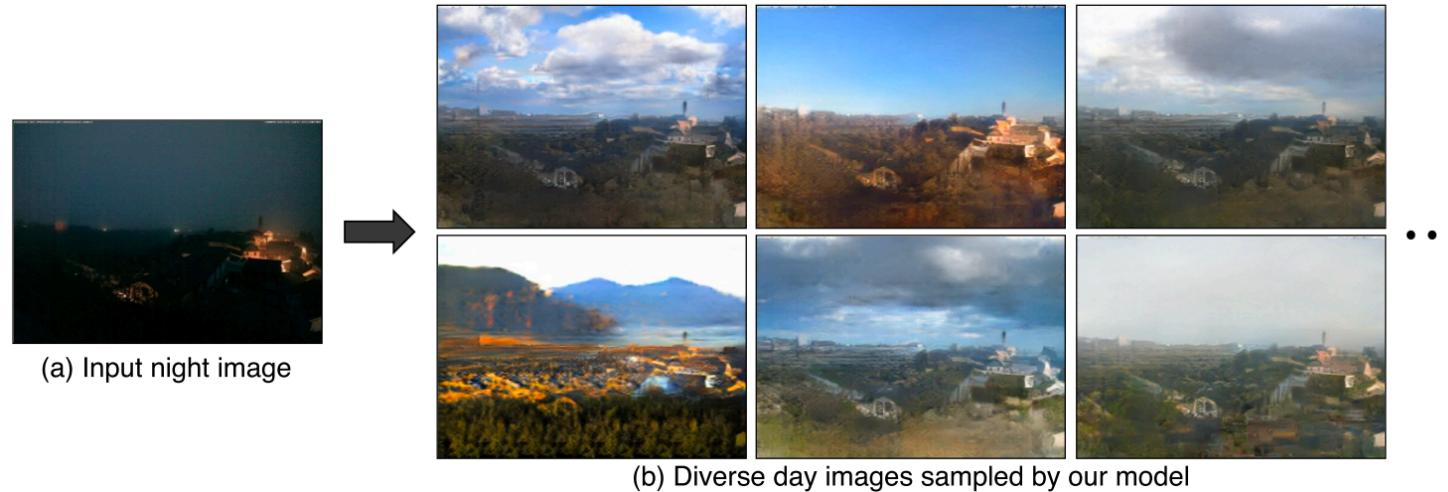
- Discussion: are they unsupervised learning?
 - NO, two image domains == binary labels.
- Why the background / shape can be maintained?
 - Fully convolutional networks → inductive bias
 - Cycle-consistency loss
- Questions?



- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- **BiCycleGAN: multi-modality**
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

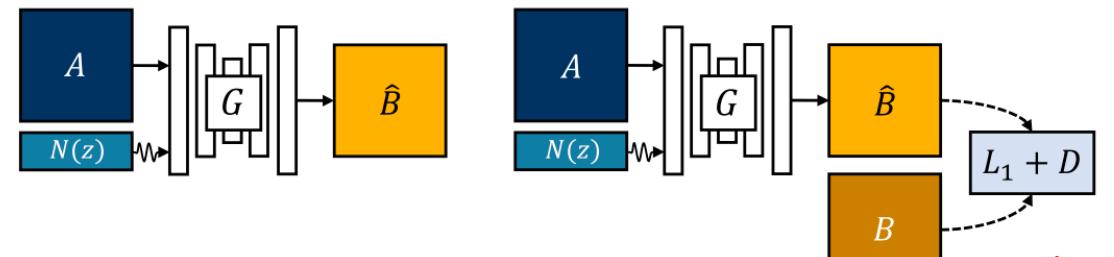
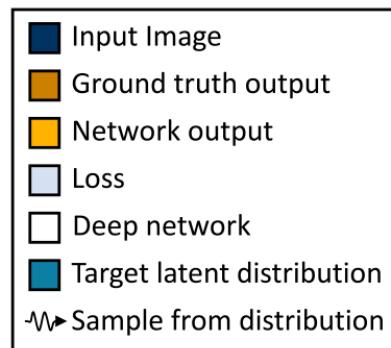
BiCycleGAN: multi-modality

- Support diverse (multi-modal) outputs but still need paired data



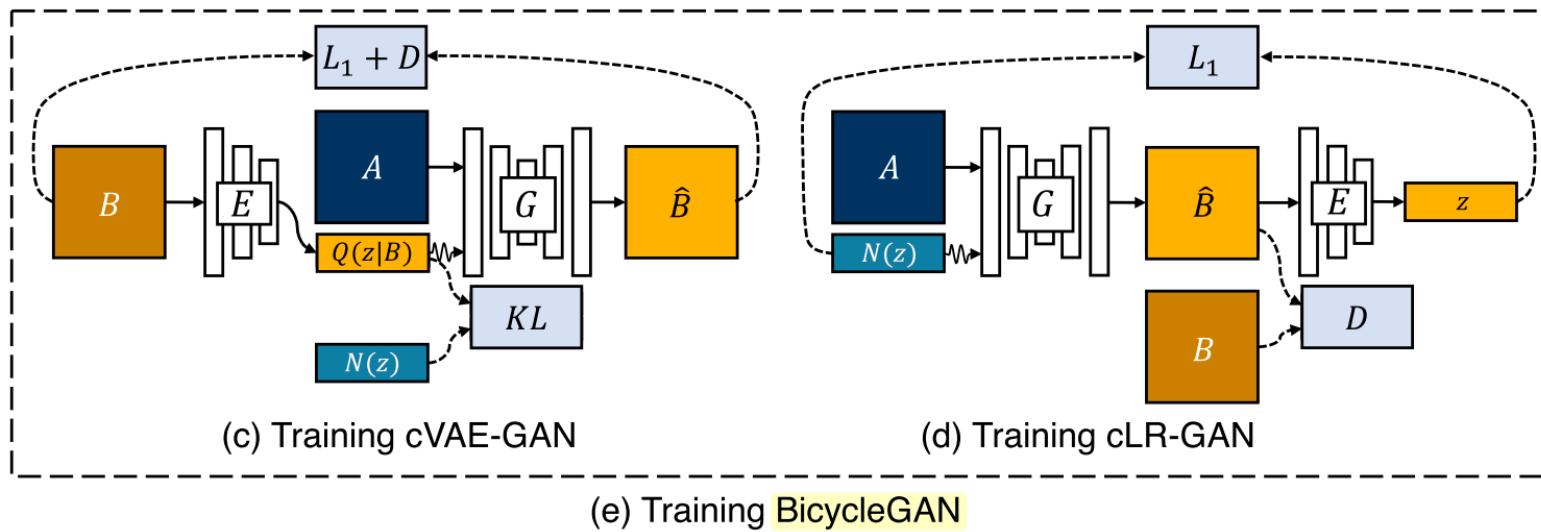
BiCycleGAN: multi-modality

- Cycle on latent noises + Cycle on translated images



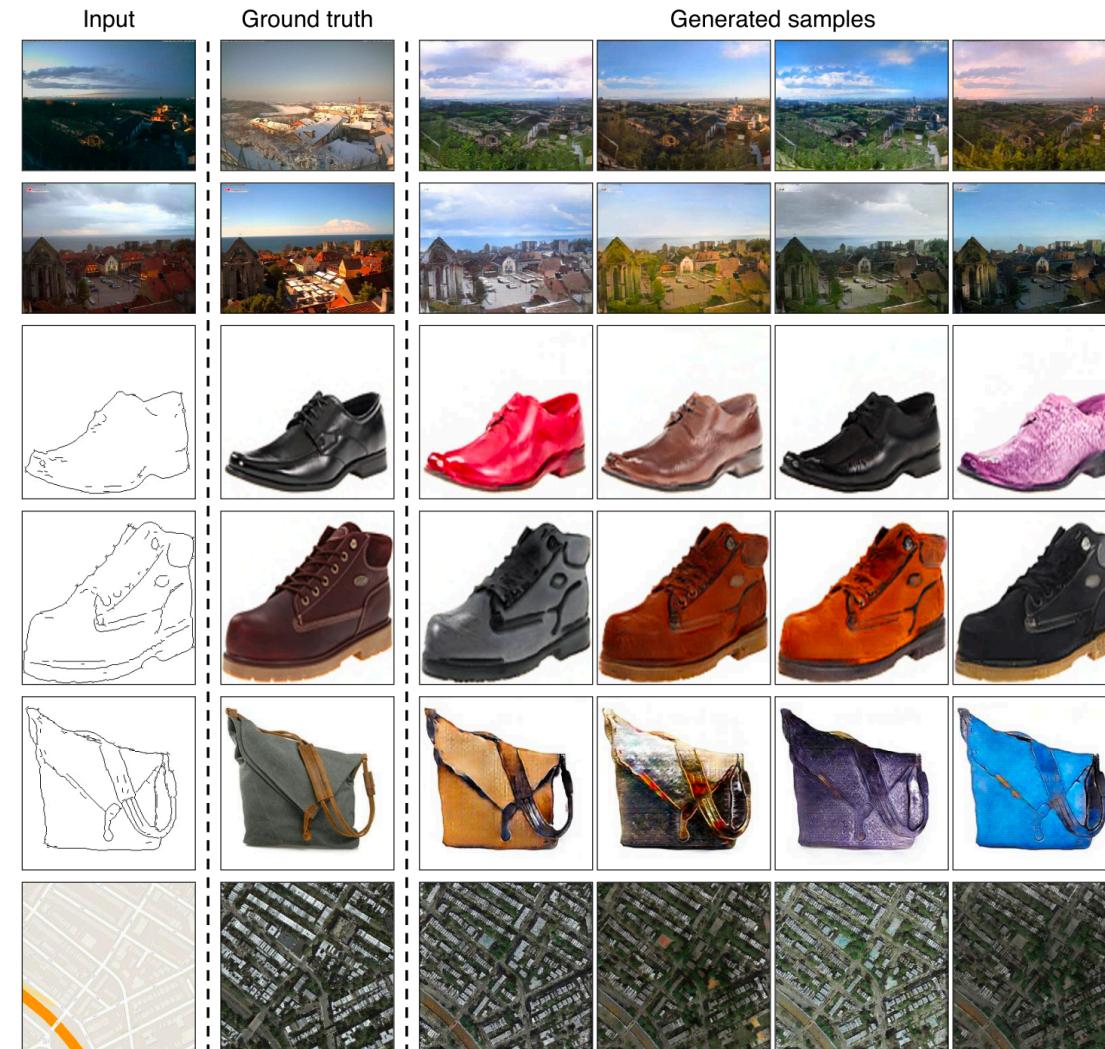
(a) Testing Usage for all models

(b) Training pix2pix+noise



BiCycleGAN: multi-modality

- Result





- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- **MUNIT and Augmented CycleGAN: unpaired data + multi-modality**
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

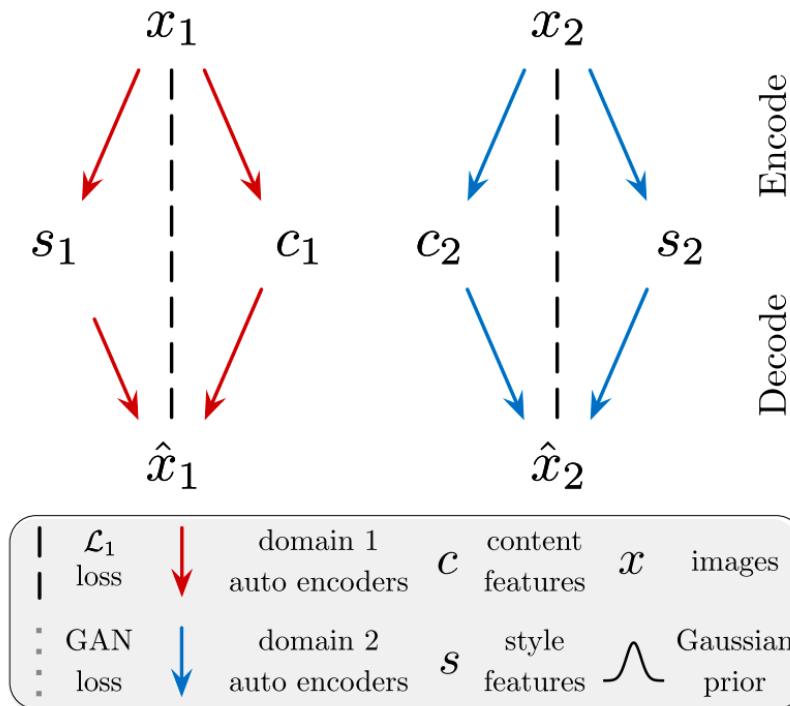
MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Goal: unpaired + multi-modal results

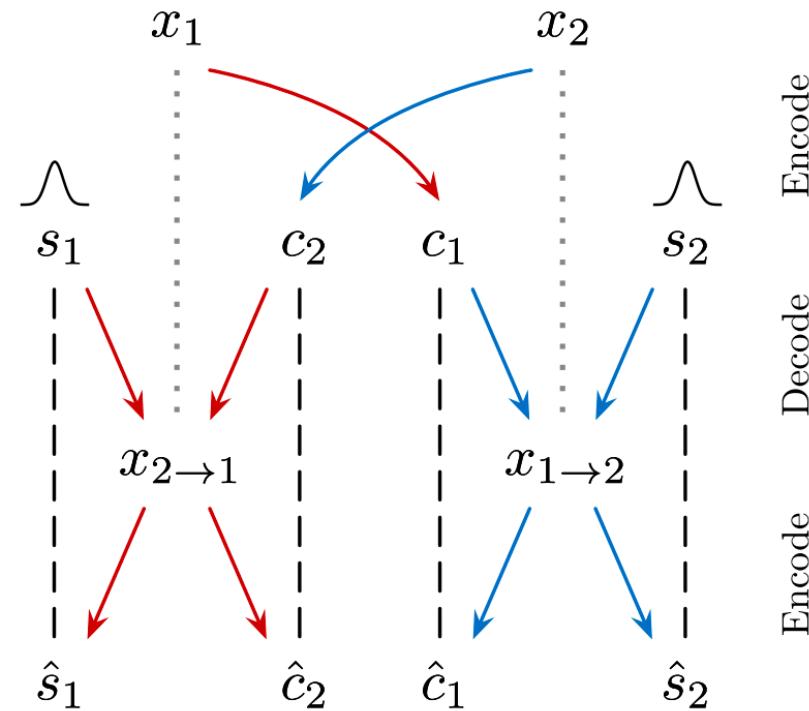


MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Latent reconstruction + Adversarial learning



(a) Within-domain reconstruction



(b) Cross-domain translation

MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Comparison against previous methods

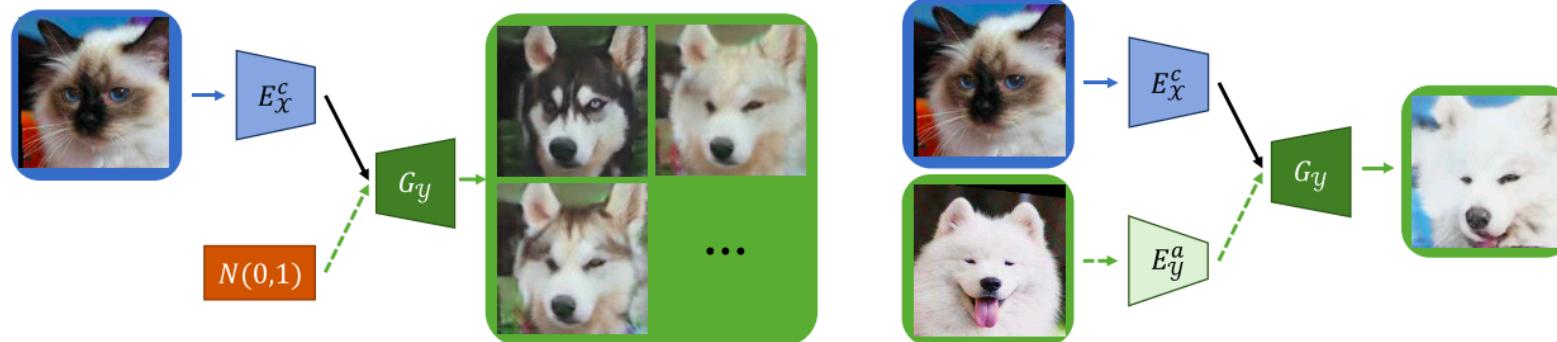


MUNIT: Multimodal Unsupervised Image-to-Image Translation. ECCV 2018.

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- **DRIT: disentangle domain-specific features**
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

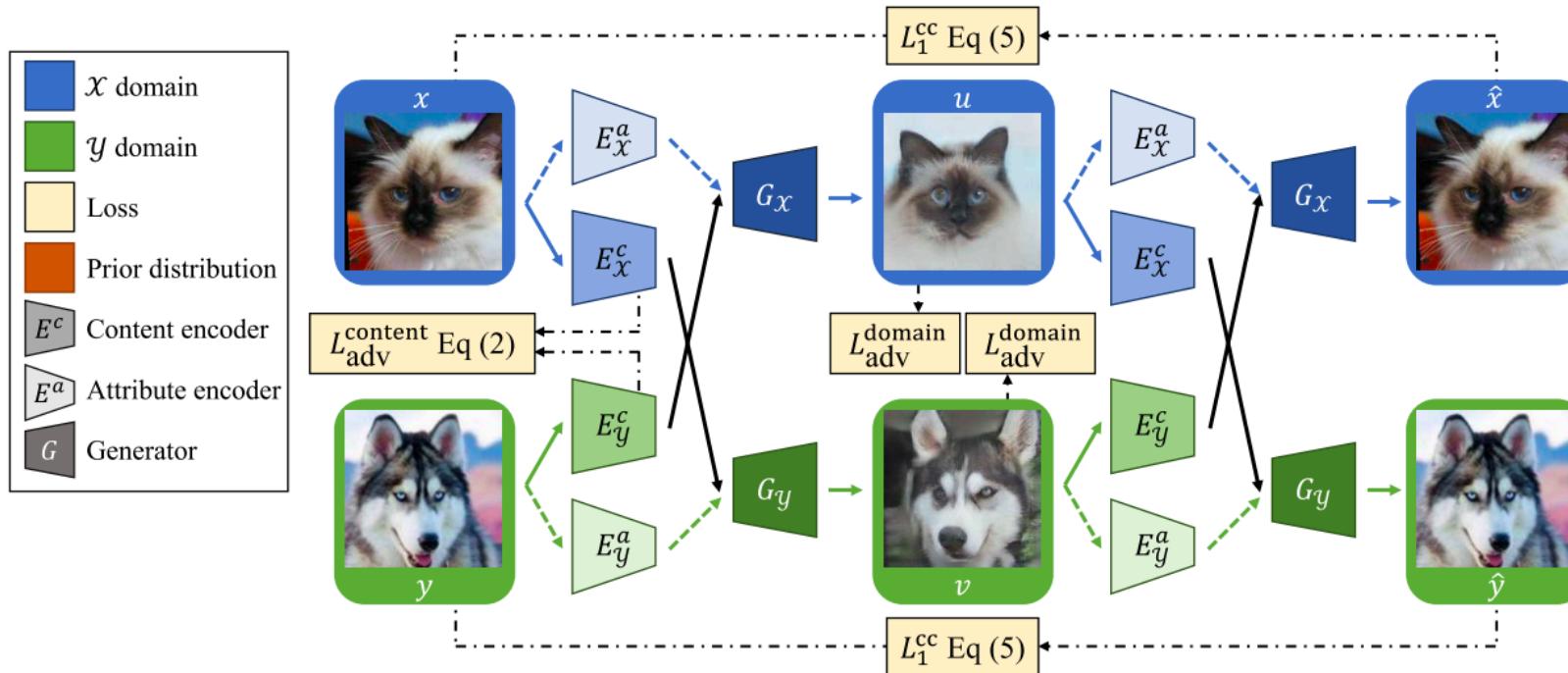
DRIT: disentangle domain-specific features

- Goal: Multi-modal results + Disentanglement



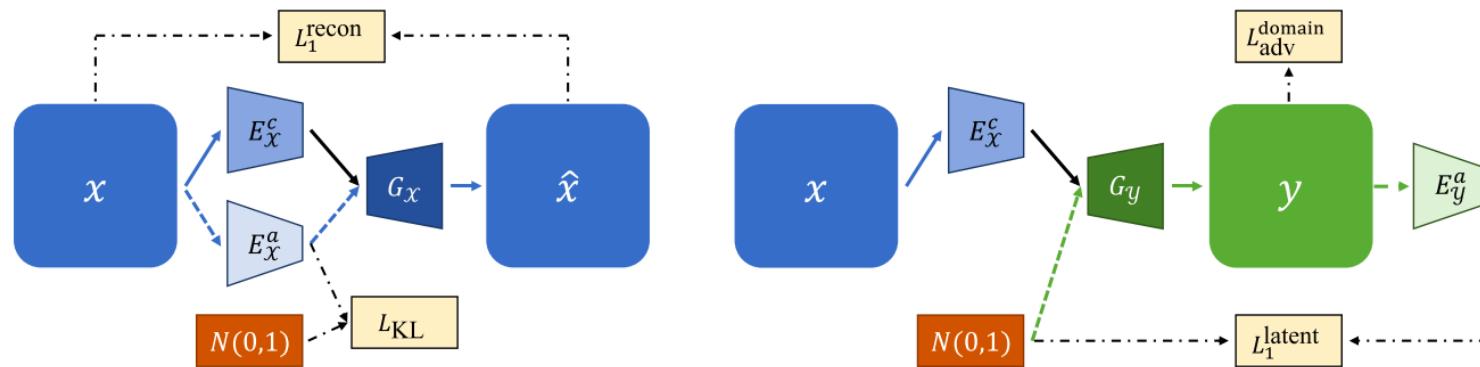
DRIT: disentangle domain-specific features

- Network bottleneck + Adversarial learning



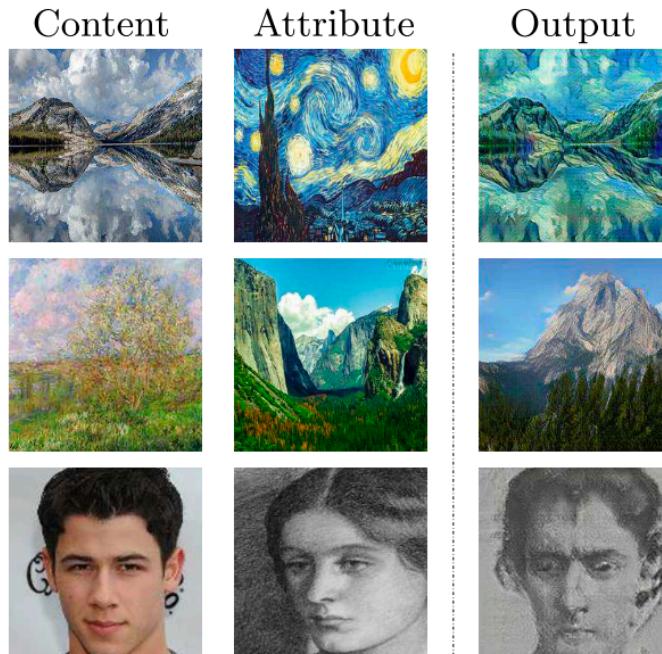
DRIT: disentangle domain-specific features

- Additional losses for better disentanglement

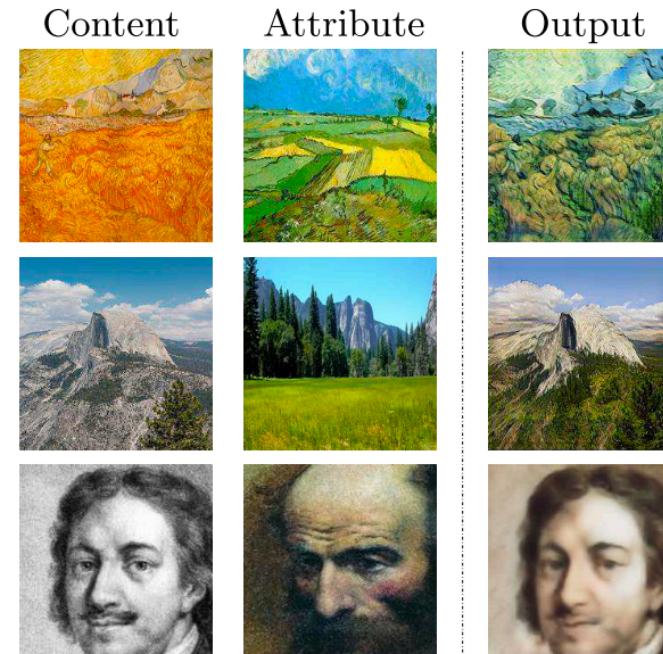


DRIT: disentangle domain-specific features

- Results



(a) Inter-domain attribute transfer

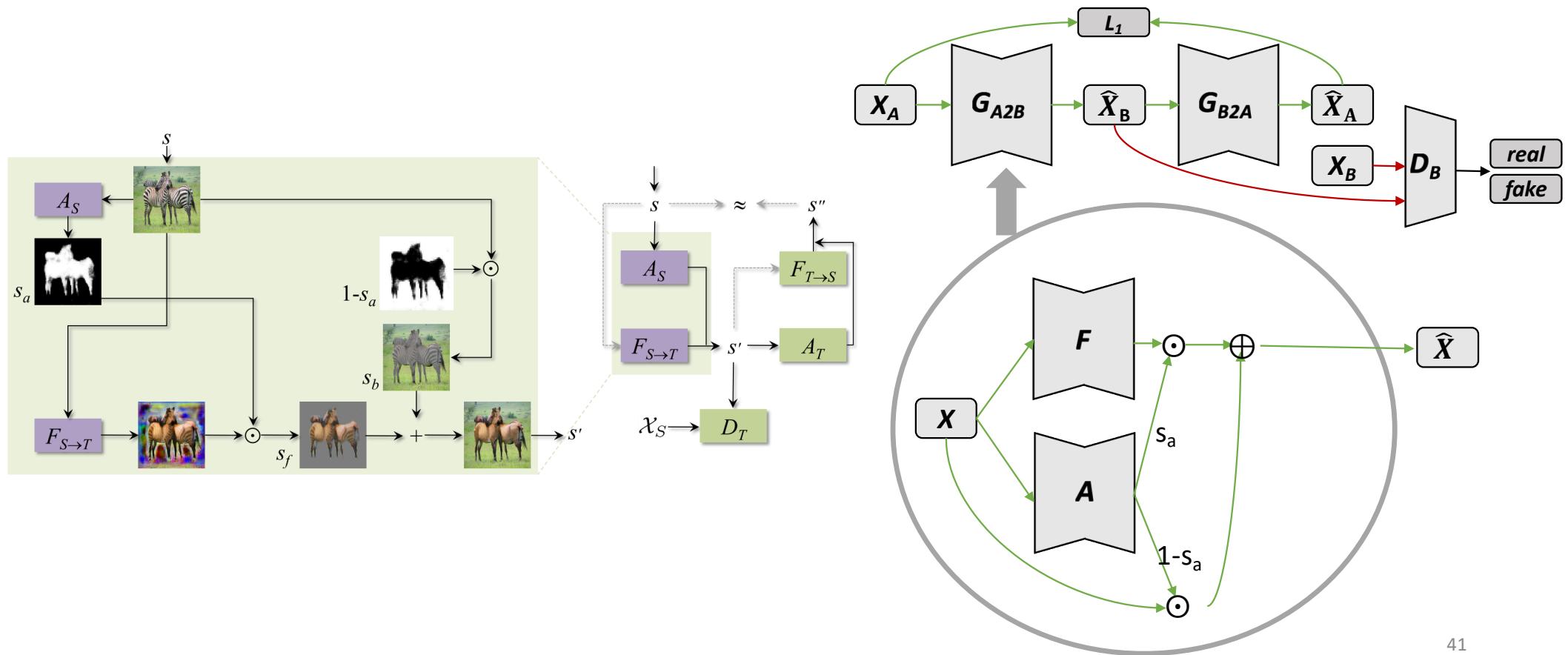


(b) Intra-domain attribute transfer

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- **Attention CycleGAN: maintain background**
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

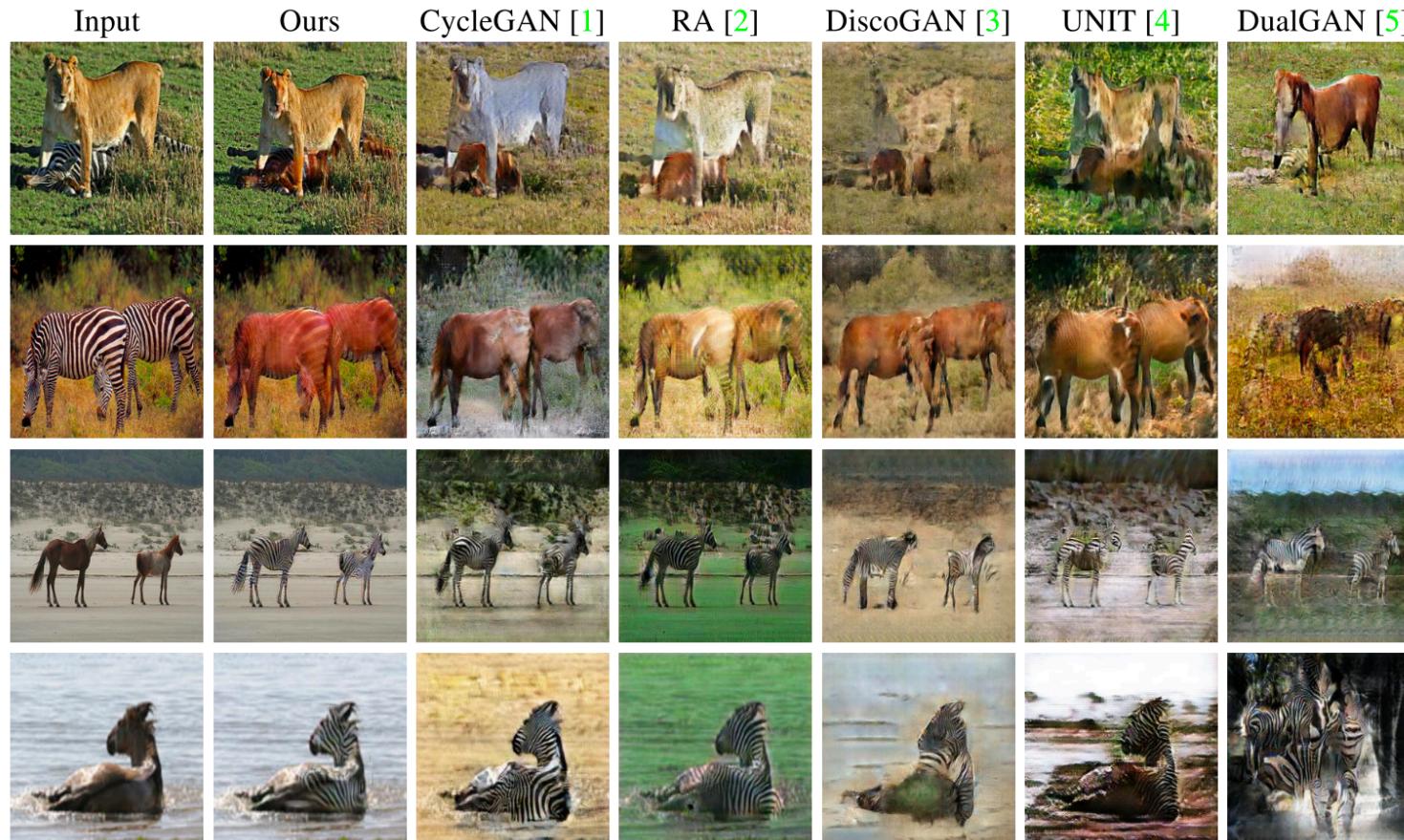
Attention CycleGAN: maintain background

- Learn the segmentation via synthesis



Attention CycleGAN: maintain background

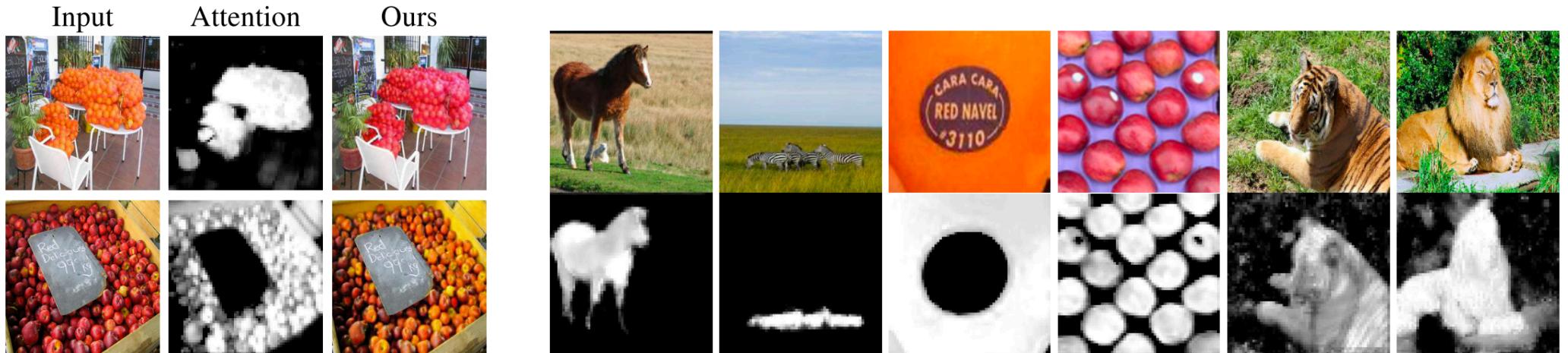
- Main background better



Unsupervised Attention-guided Image-to-Image Translation. Mejjati, Y. A., Richardt, C., & Cosker, D. NIPS, 2018

Attention CycleGAN: maintain background

- Learn the segmentation without segmentation masks





- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- **StarGAN: label condition**
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition



StarGAN: label condition

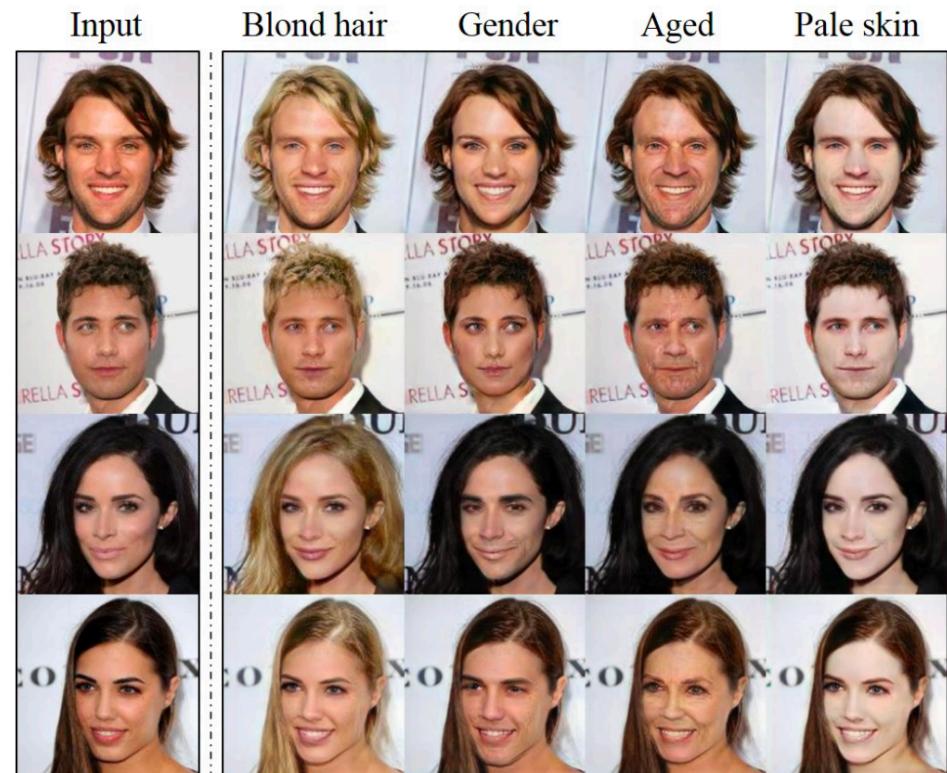
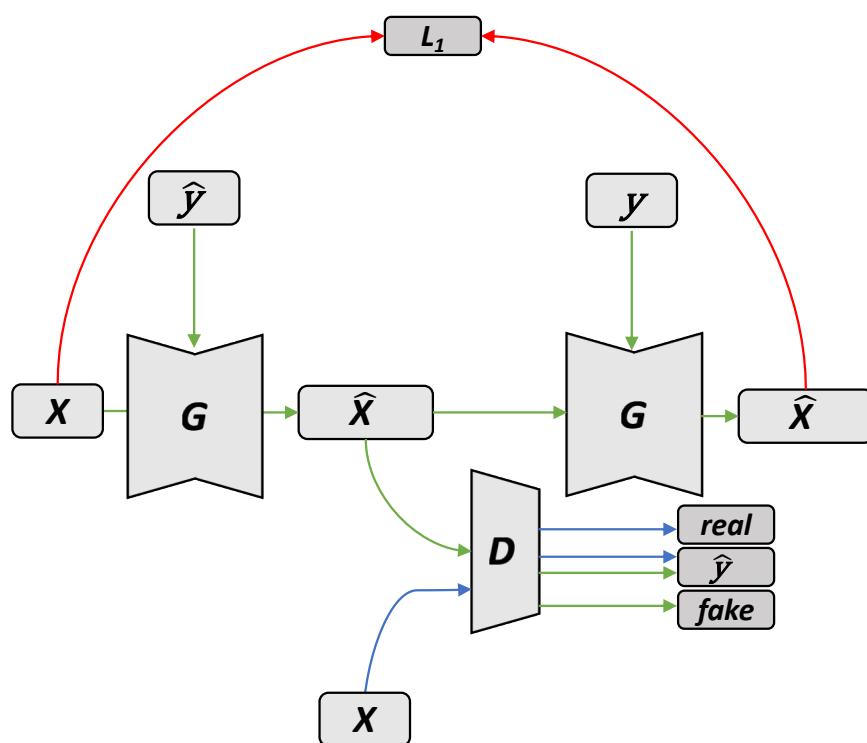
- Limitation of CycleGAN

Translations between N domains require $N(N-1)$ models

StarGAN: One model to rule them all !!

StarGAN: label condition

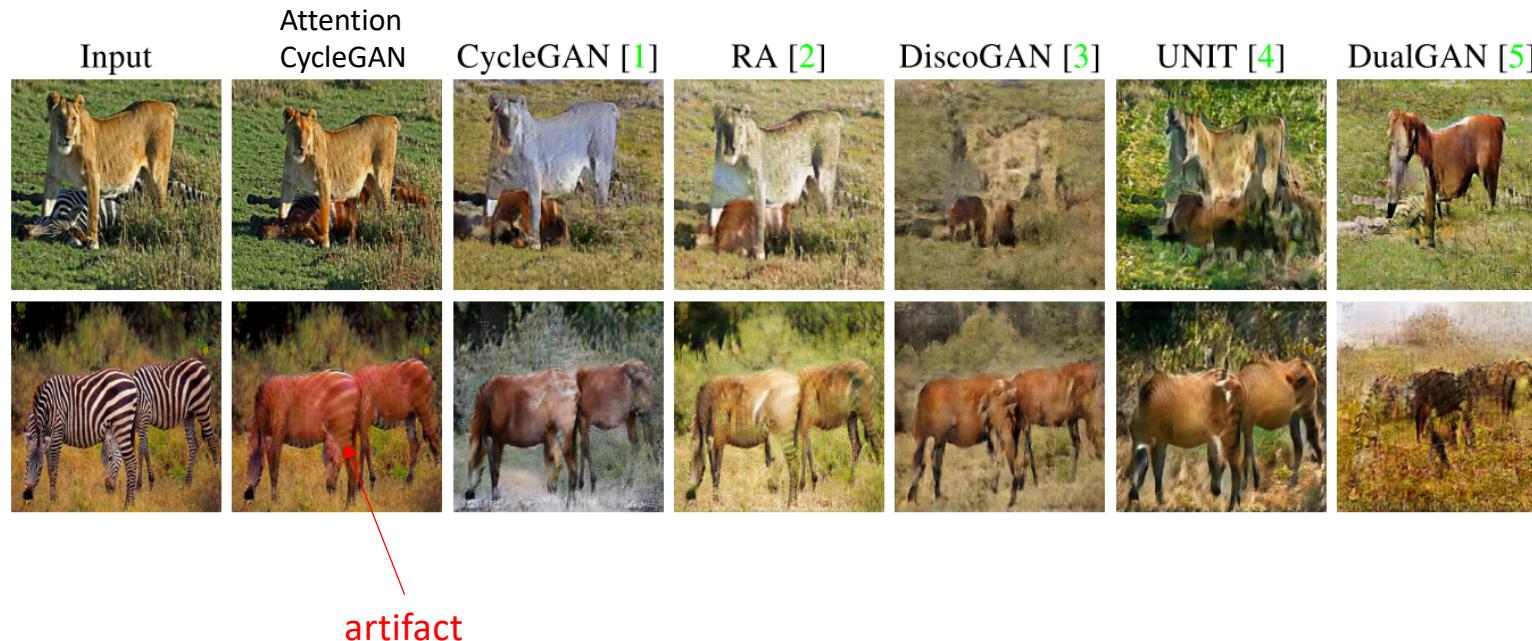
- Add class condition into the generator and the output of the discriminator



- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- **Breaking the Cycle**
- GAN-CLS and SisGAN: text condition

Breaking the Cycle

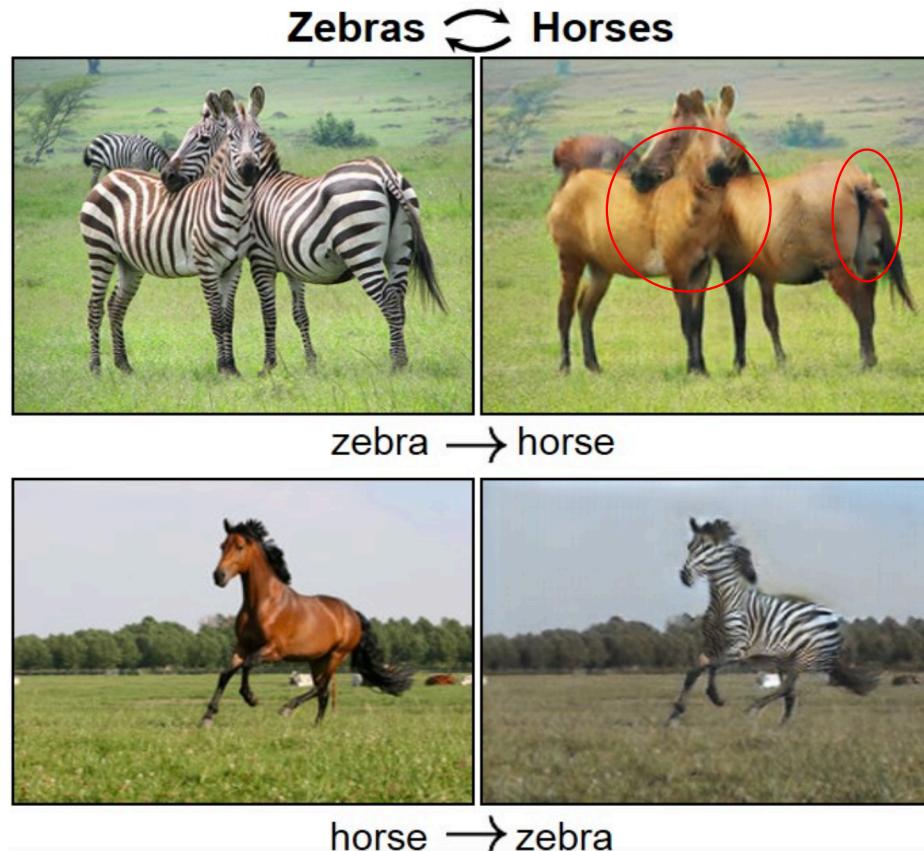
- Limitation of cycle-consistency loss



Cycle loss enforces the constraint that translating an image to the target domain and back, should obtain the original image

Breaking the Cycle

- Limitation of cycle-consistency loss

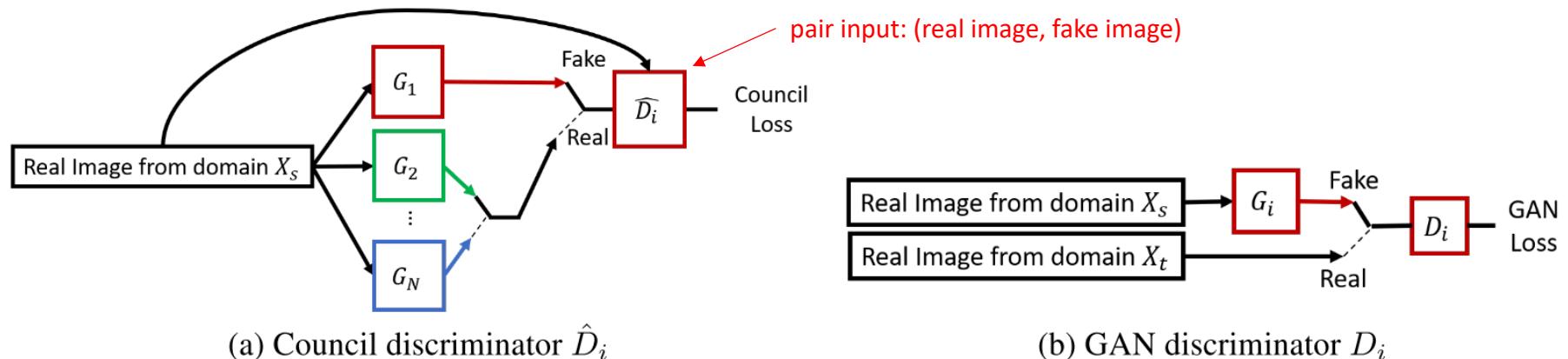


zebra -> horse Hidden Info

horse -> zebra OK

Breaking the Cycle

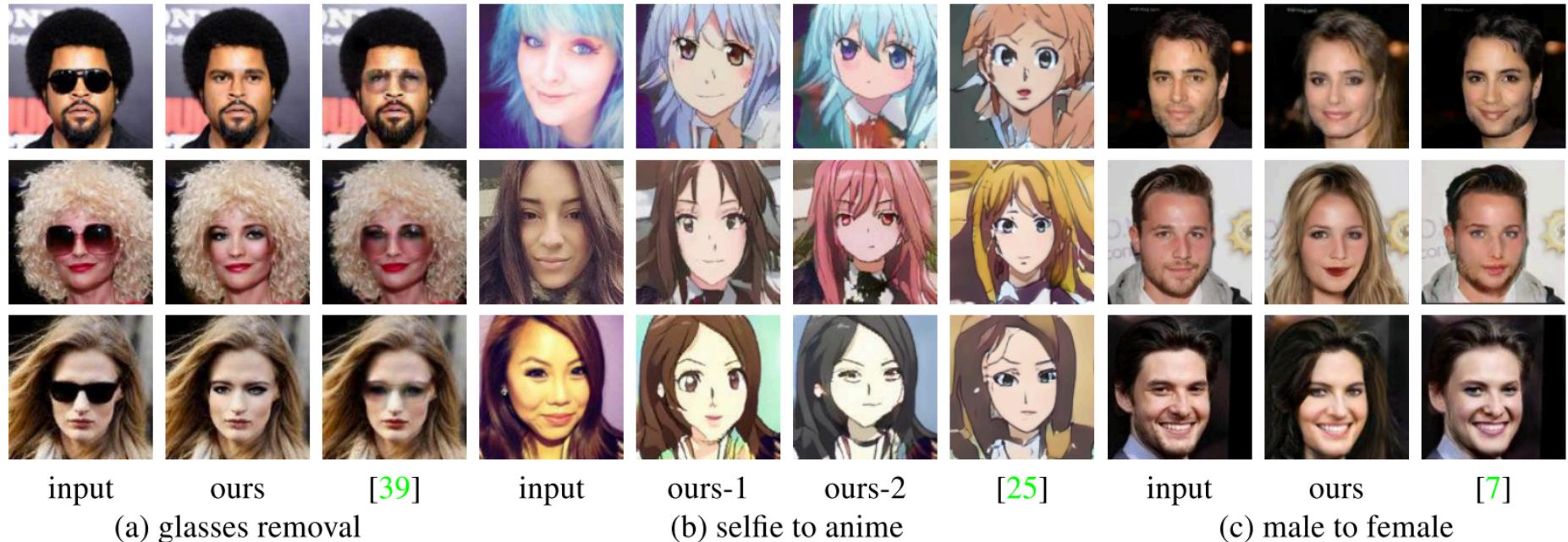
- Colleagues are all you need



- Each member of the council is a triplet (**Red indicates one council**), whose components are **one generator and two discriminators**.
- The task of discriminator D_i is to distinguish between the generator's G_i output and real examples.
- The goal of discriminator \hat{D}_i is to distinguish between images produced by G_i and images produced by the other generators in the council. This discriminator is the core of the model and this is what differentiates the model from the classical GAN model. It enforces the generator to converge to images that could be acknowledged by all council.

Breaking the Cycle

- Colleagues are all you need





Breaking the Cycle

- Discussion: Why it works?
- It is not only for im2im, other distribution transformations may benefit from it
- Better Methods:
 - ACL-GAN
 - XDCycleGAN

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- **GAN-CLS and SisGAN: text condition**

GAN-CLS and SisGAN

- Text-to-image synthesis

this small bird has a pink breast and crown, and black primaries and secondaries.



the flower has petals that are bright pinkish purple with white stigma



this magnificent fellow is almost all black with a red crest, and white cheek patch.



this white and yellow flower have thin white petals and a round yellow stamen

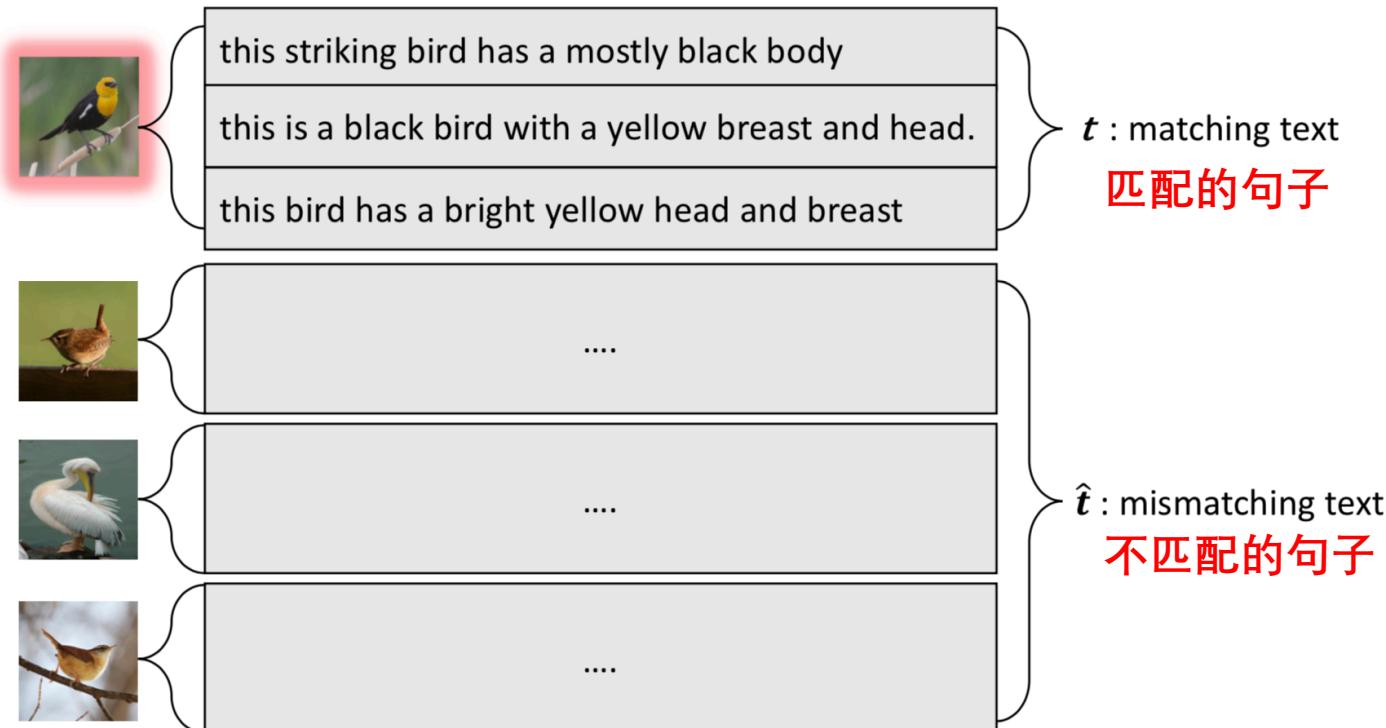


- Classic multi-modal problem

$$P(t, z)$$

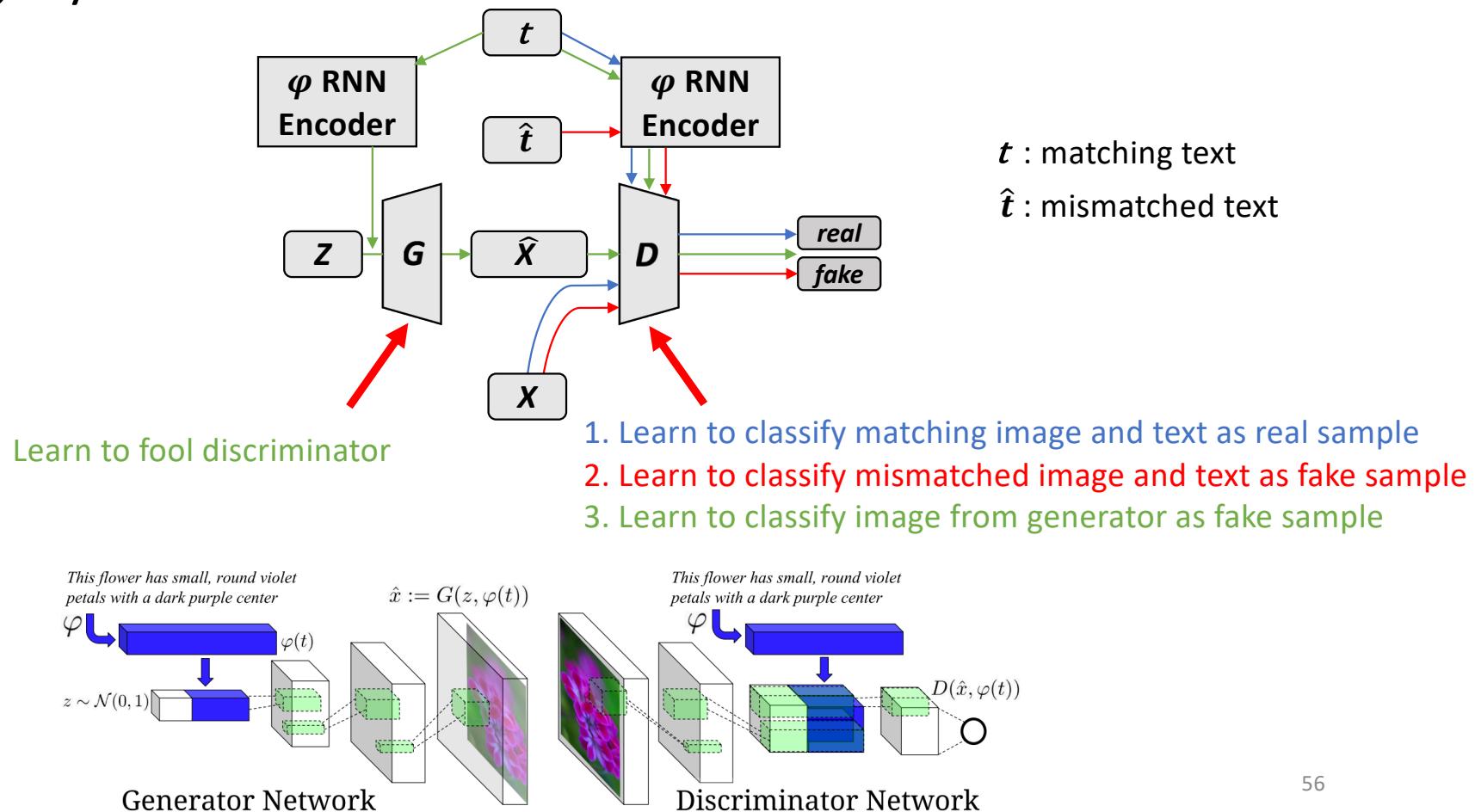
GAN-CLS and SisGAN

- Text-to-image synthesis



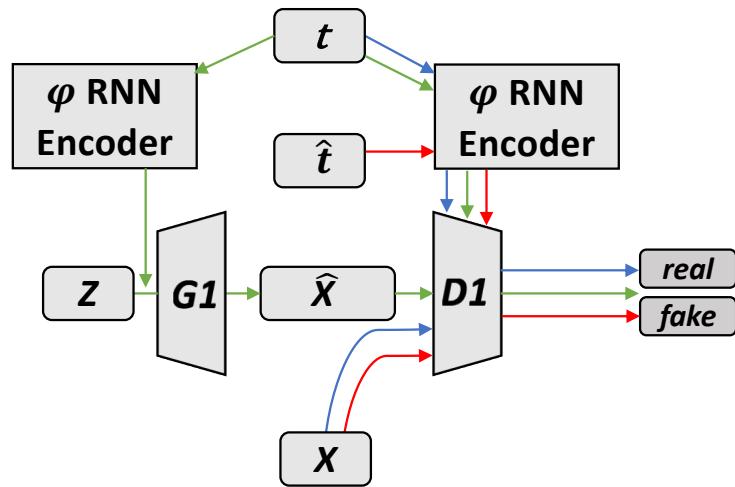
GAN-CLS and SisGAN

- Text-to-image synthesis

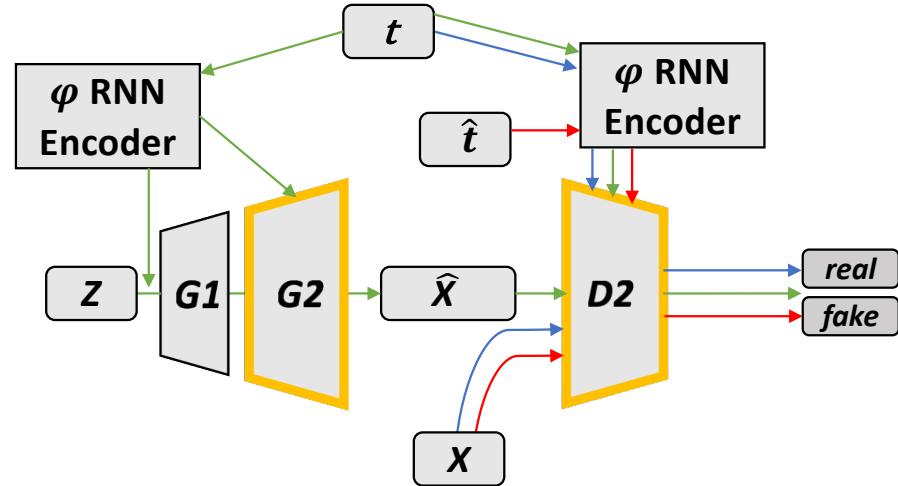


GAN-CLS and SisGAN

- Text-to-image synthesis + High resolution image



Step 1



Step 2

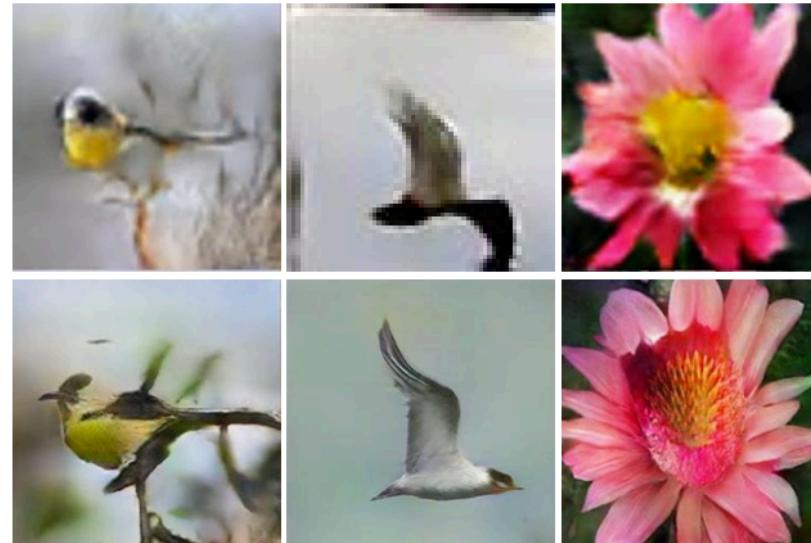
GAN-CLS and SisGAN

- Text-to-image synthesis + High resolution image

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

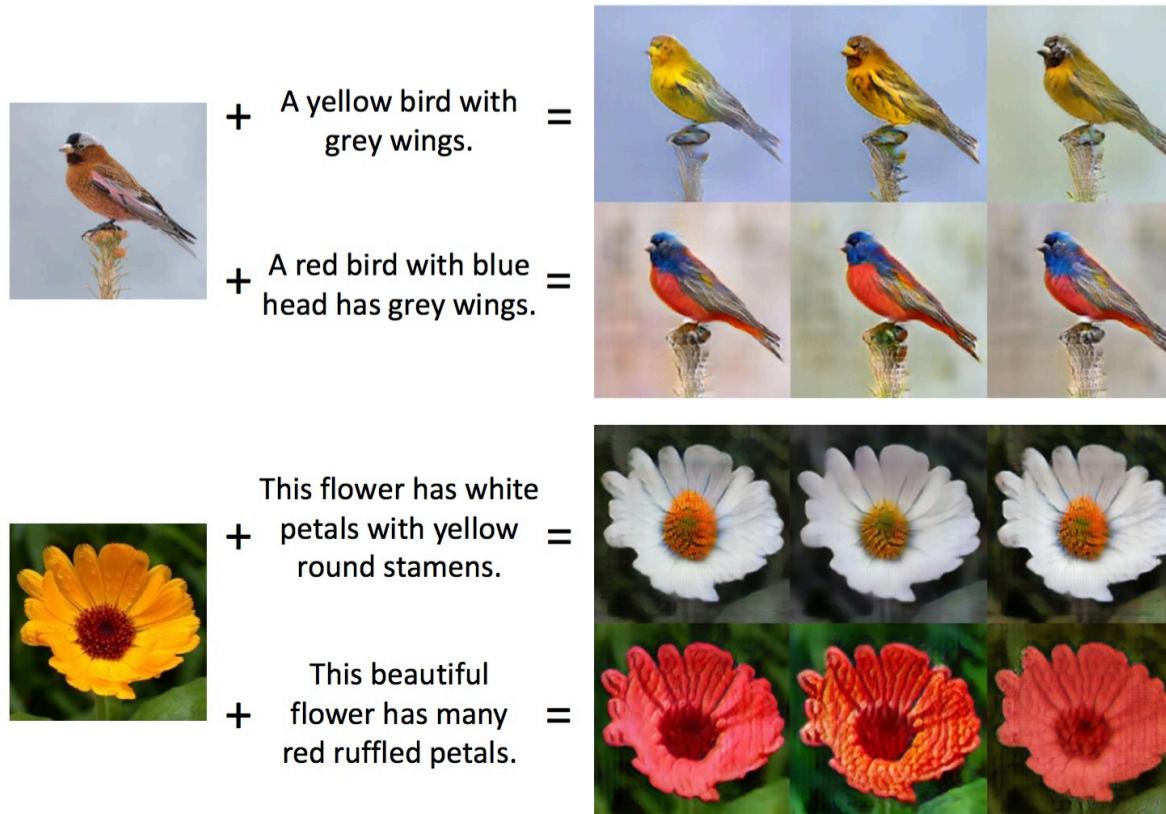
This bird is white with some black on its head and wings, and has a long orange beak

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments



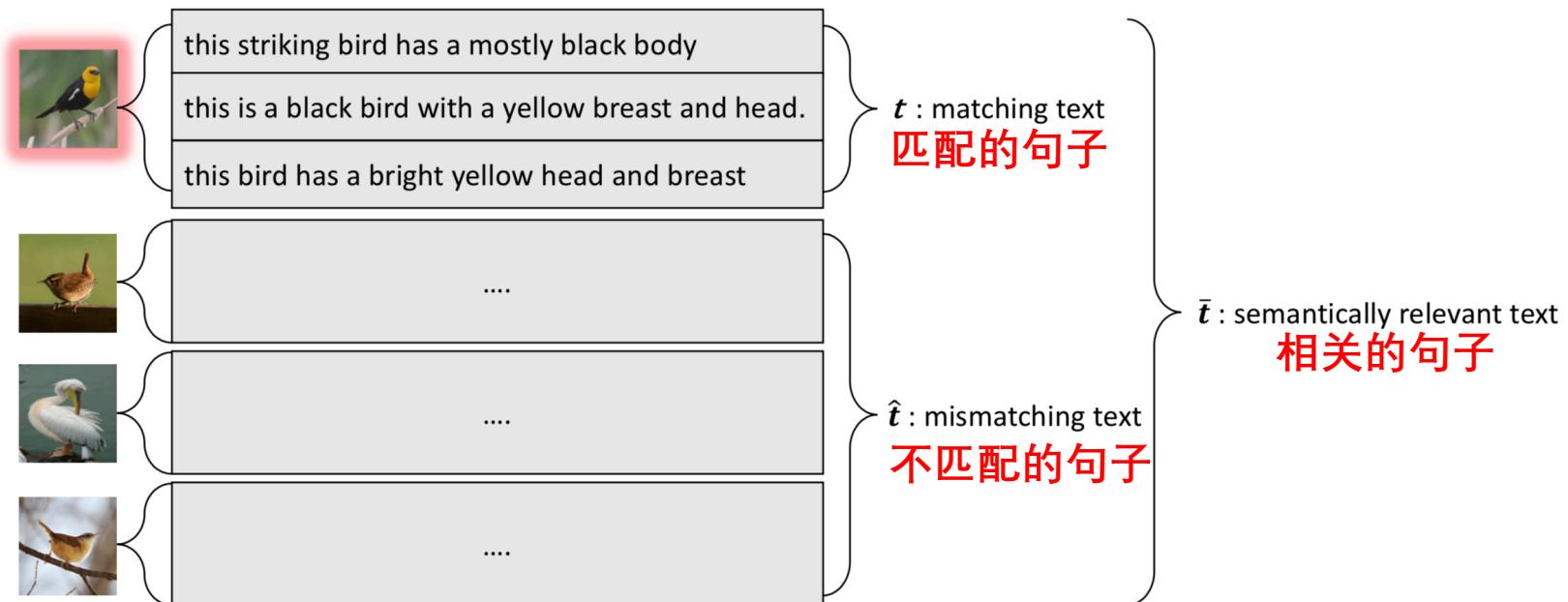
GAN-CLS and SisGAN

- Semantic image synthesis



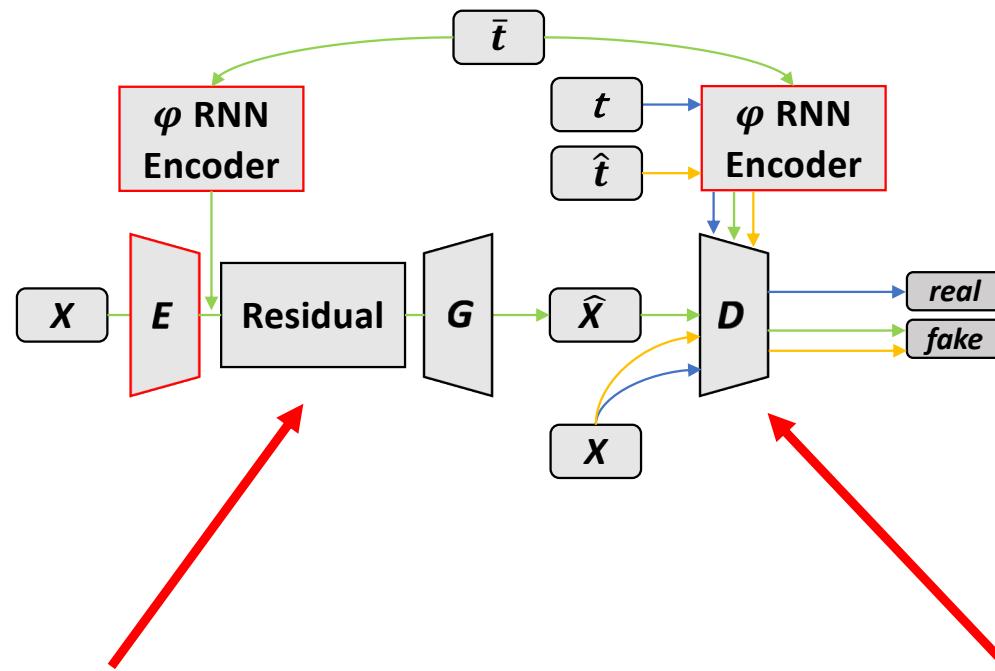
GAN-CLS and SisGAN

- Semantic image synthesis



GAN-CLS and SisGAN

- Semantic image synthesis



Learn to fool discriminator
when inputting image with
semantically relevant text

1. Learn to classify matching image and text pairs as real samples
2. Learn to classify mismatched image and text pairs as fake samples
3. Learn to classify samples from generator as fake samples

t : matching text

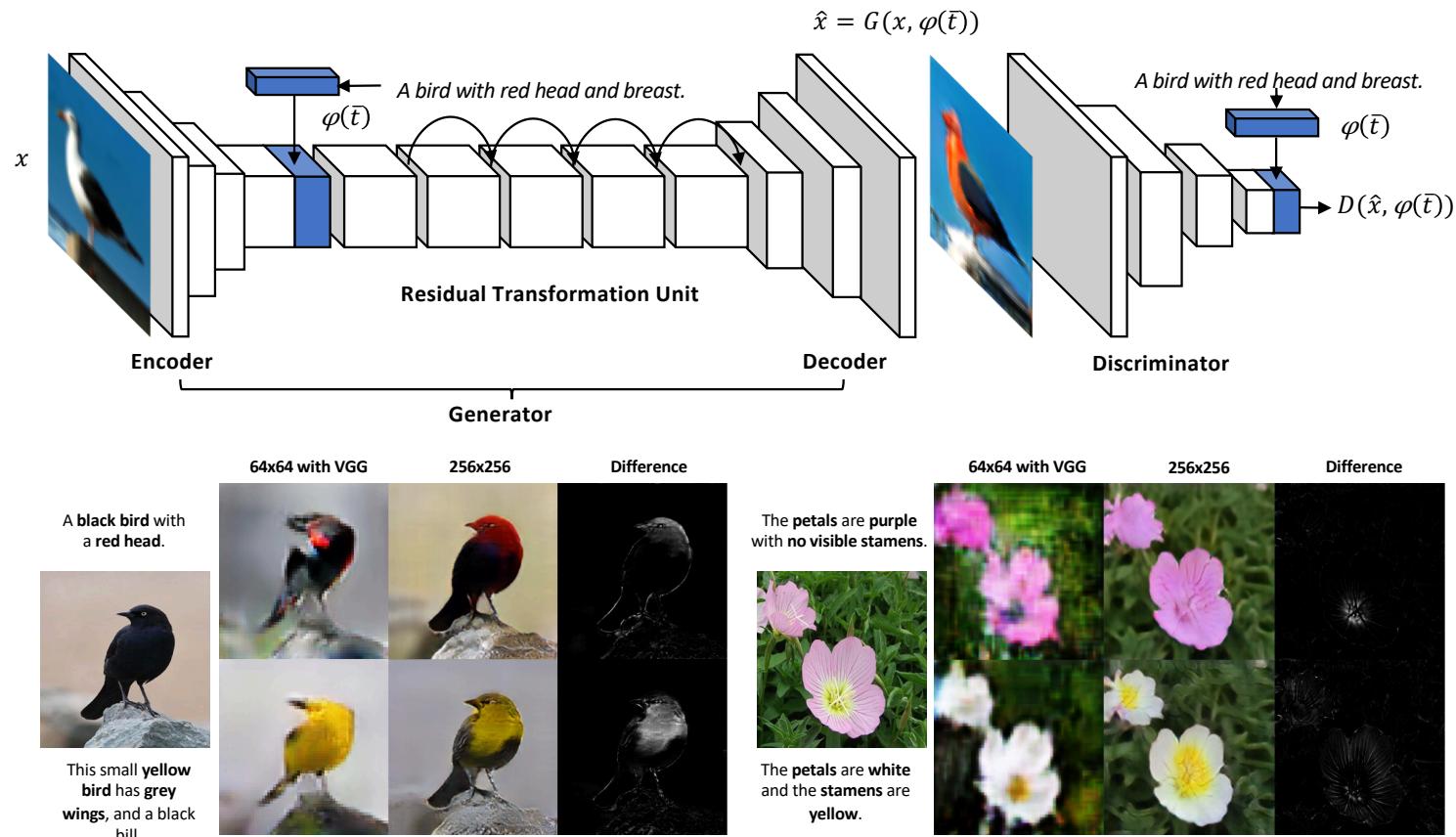
\hat{t} : mismatched text

\bar{t} : semantically relevant text

$$\begin{aligned}\mathcal{L}_D = & \mathbb{E}_{(x,t) \sim p_{data}} \log D(x, \varphi(t)) \\ & + \mathbb{E}_{(x,\hat{t}) \sim p_{data}} \log(1 - D(x, \varphi(\hat{t}))) \\ & + \mathbb{E}_{(x,\bar{t}) \sim p_{data}} \log(1 - D(G(x, \varphi(\bar{t})), \varphi(\bar{t}))) \\ \mathcal{L}_G = & \mathbb{E}_{(x,\bar{t}) \sim p_{data}} \log(D(G(x, \varphi(\bar{t})), \varphi(\bar{t})))\end{aligned}$$

GAN-CLS and SisGAN

- Semantic image synthesis: Learn the location information via synthesis



Summary

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition

Discussion: ideal im2im

- What should the ideal image-to-image translation to be?
 - Unpaired data
 - Maintain background
 - Multi-modality
 - Disentanglement
 - Multi-domain
 - Conditional translation



Thanks