Application of Generative Models: Image-to-Image Translation

Hao Dong

Peking University

Application of Generative Models: Image-to-Image Translation

PEKING UNIVERSITY

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

北京大學 PEKING UNIVERSITY

- 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

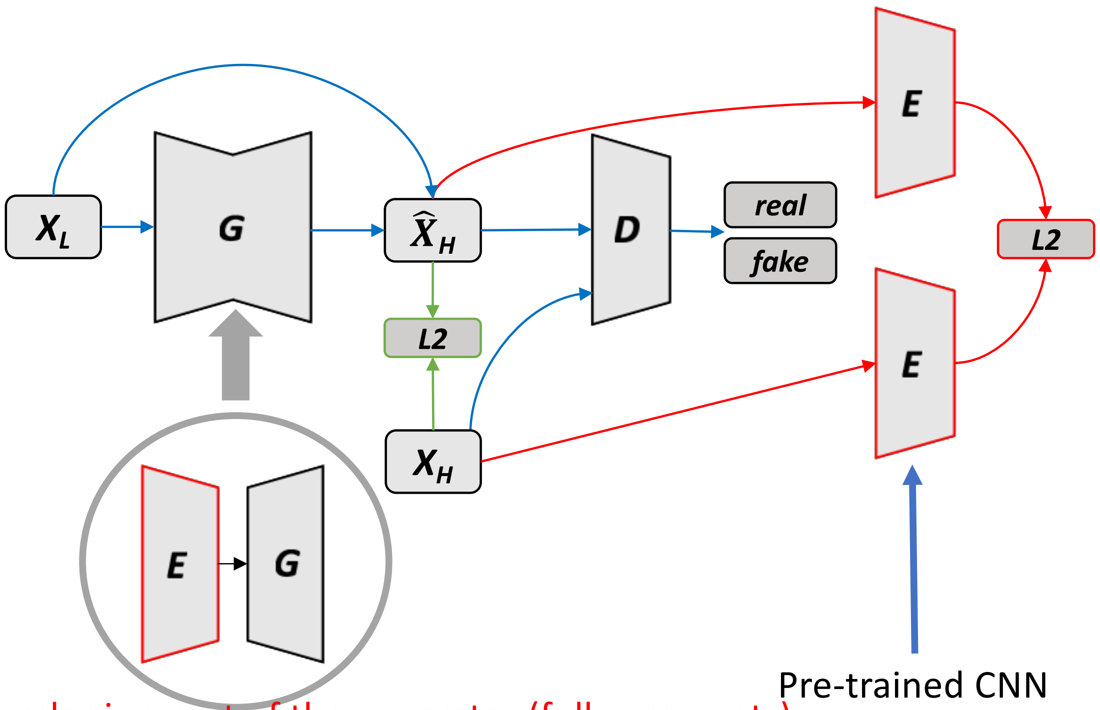
Problem Definition

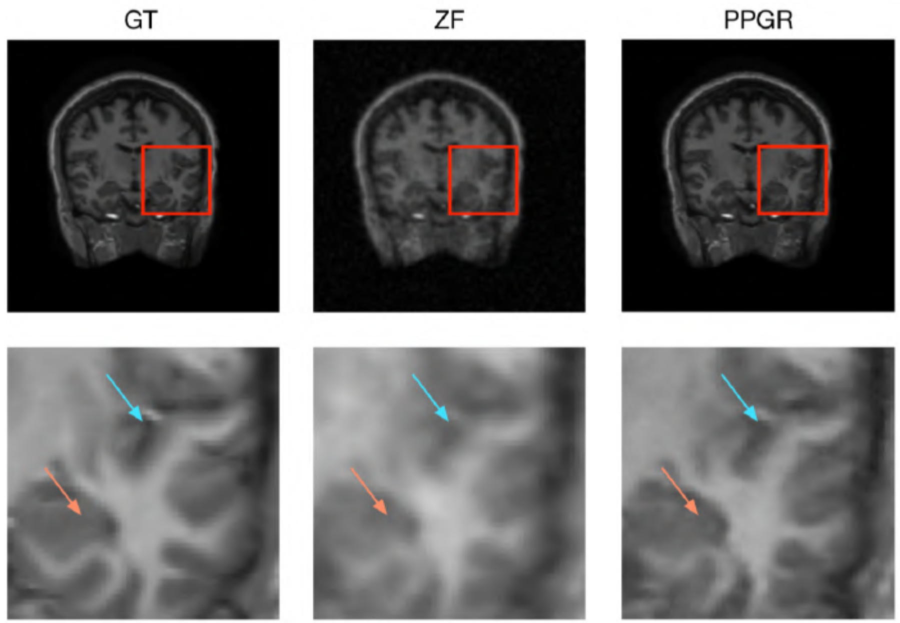
- Supervised/Paired image-to-image translation
- Unsupervised/Unpaired image-to-image translation

- 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





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

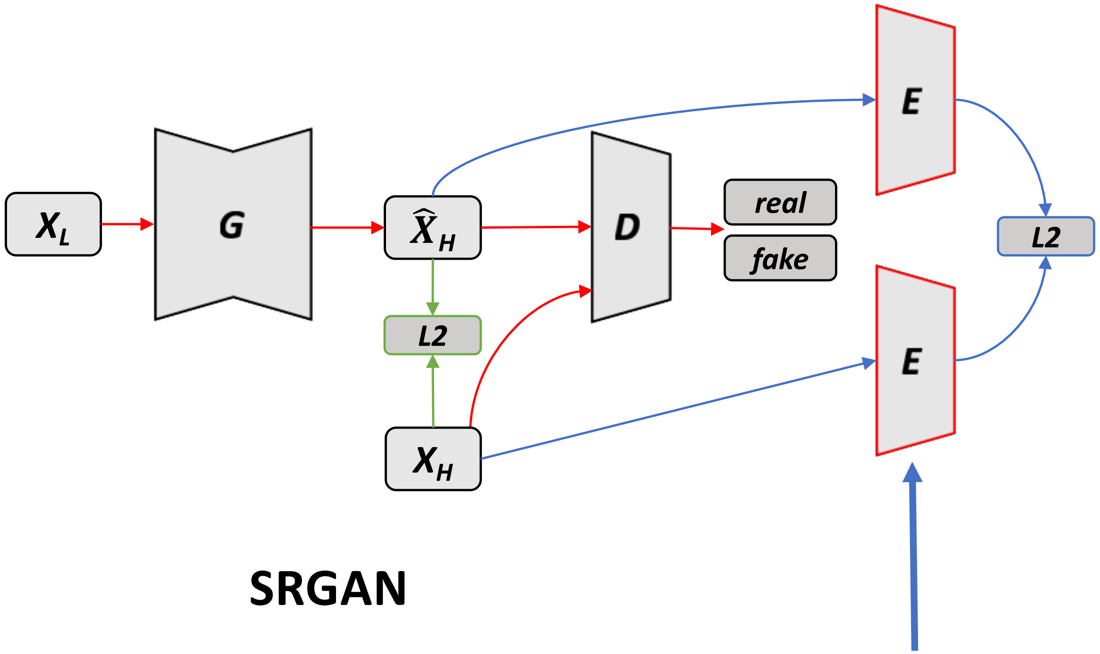
Deep De-Aliasing for Fast Compressive Sensing MRI. *S. Yu, H. Dong, G. Yang et al. arXiv:1705.07137 2017.*

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



SRGAN

Pre-trained VGG

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *C. Ledig, L. Theis et al. CVPR 2017.*

Image Inpainting / Reconstruction / Super Resolution

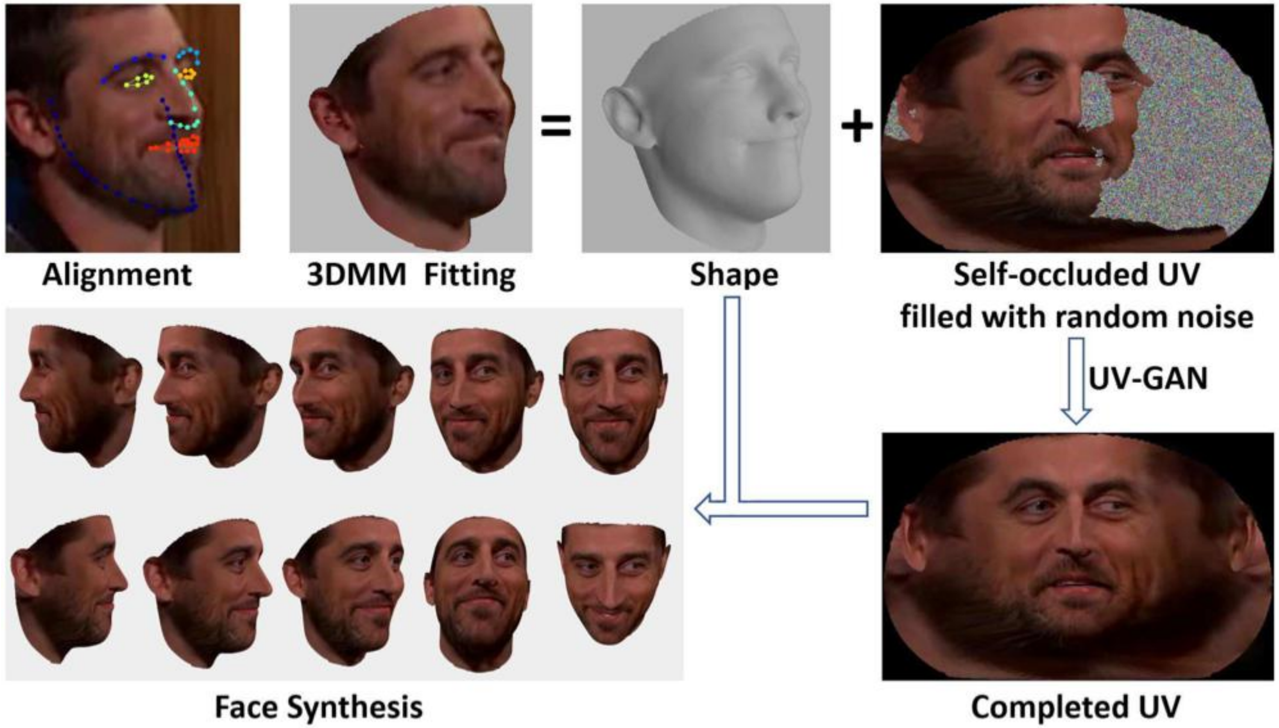
- Supervised image super resolution

Input Output

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. C. Ledig, L. Theis et al. CVPR 2017.*

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?

$$x_{grey} \rightarrow \textbf{Neural} \rightarrow \hat{x}_{rgb} \rightarrow \mathcal{L}_2 \leftarrow x_{rbg}$$

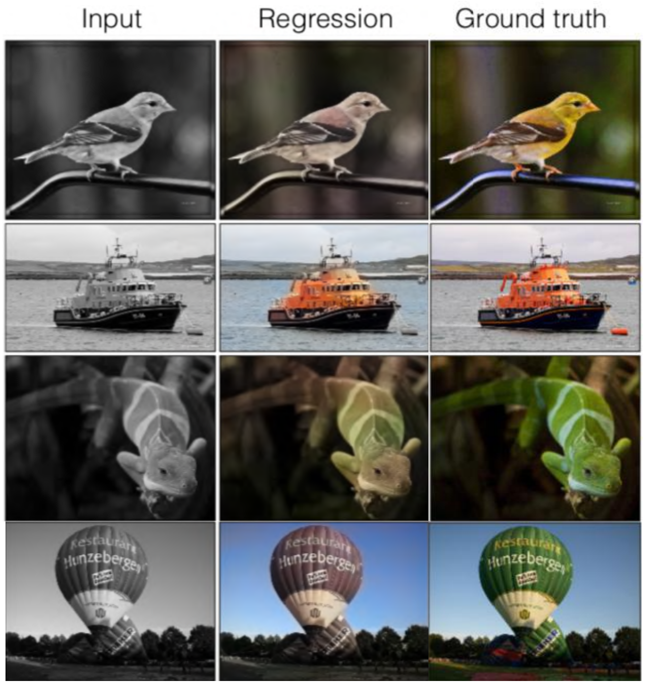
What is the problem??

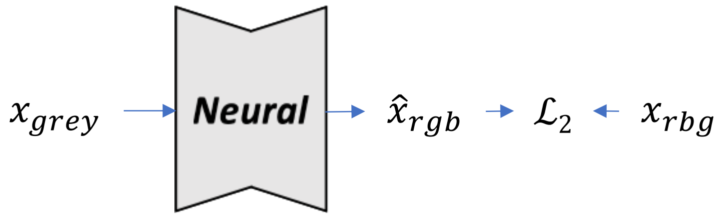
Colorful Image Colorization. *R. Zhang, P. Isola, A.A. Efros. ECCV. 2016.*

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,

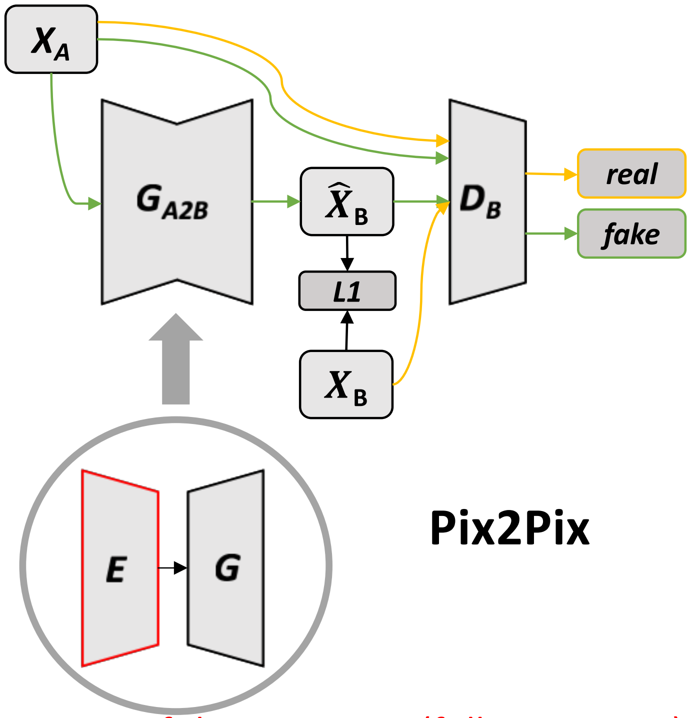
(some want red, some want blue, …)

resulting “grey” outputs

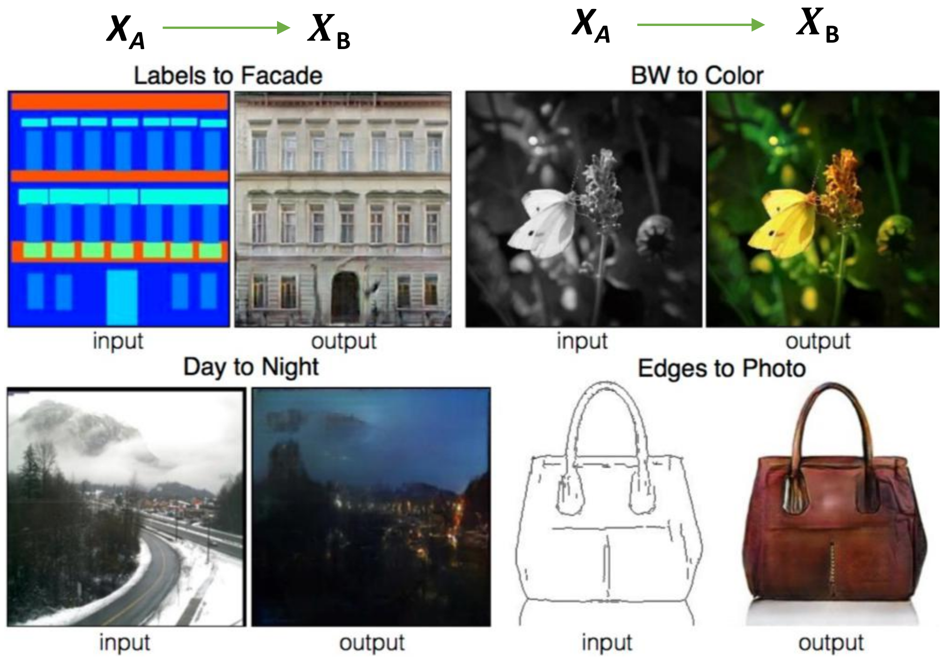
Colorful Image Colorization. *R. Zhang, P. Isola, A.A. Efros. ECCV. 2016.*
Image-to-Image Translation with Conditional Adversarial Networks. *P. Isola, J. Zhu et al. CVPR 2017.*

Pix2Pix: paired data

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



Pix2Pix



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

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

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

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

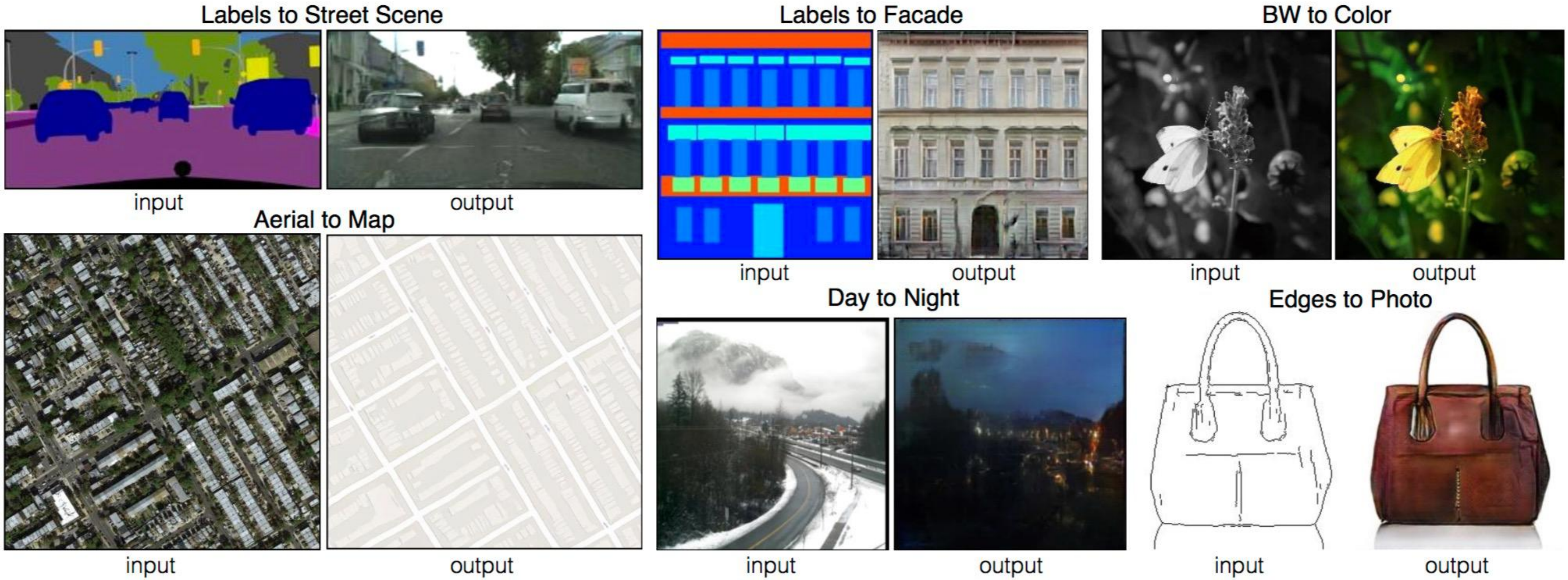


Image-to-Image Translation with Conditional Adversarial Networks. *P. Isola, J. Zhu et al. CVPR 2017.*

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.*

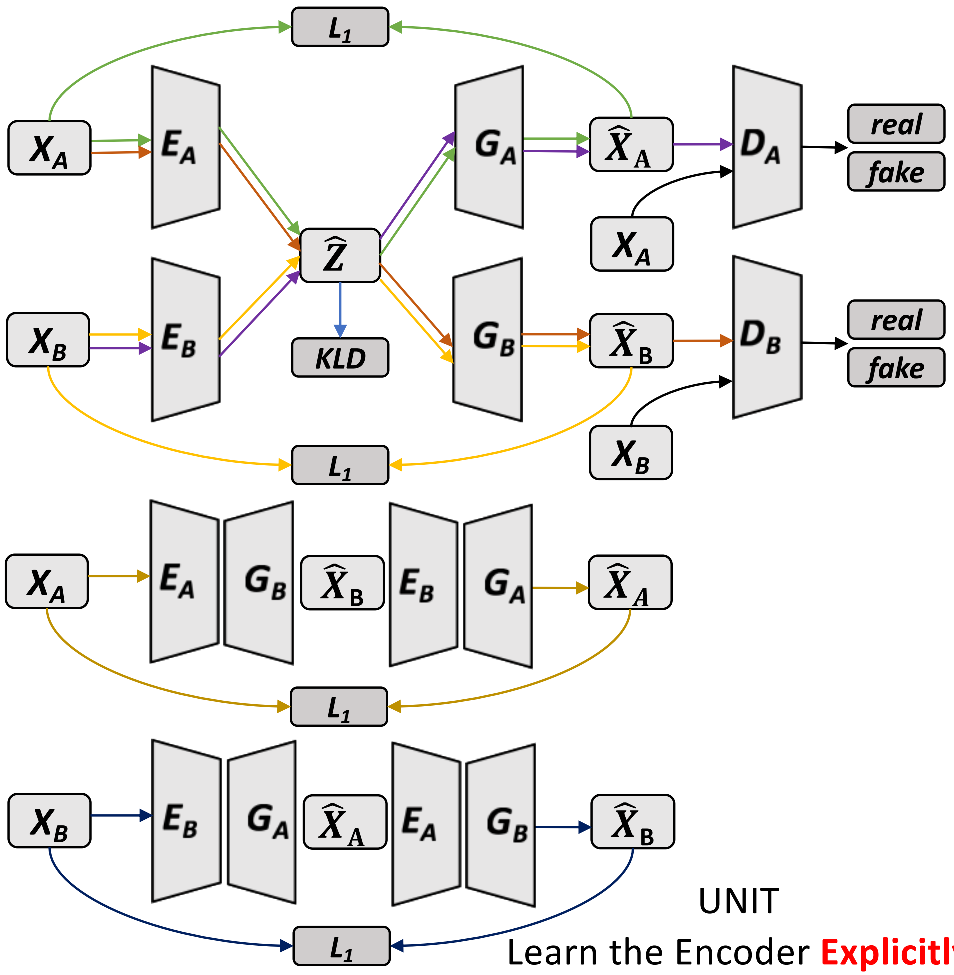
- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention 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

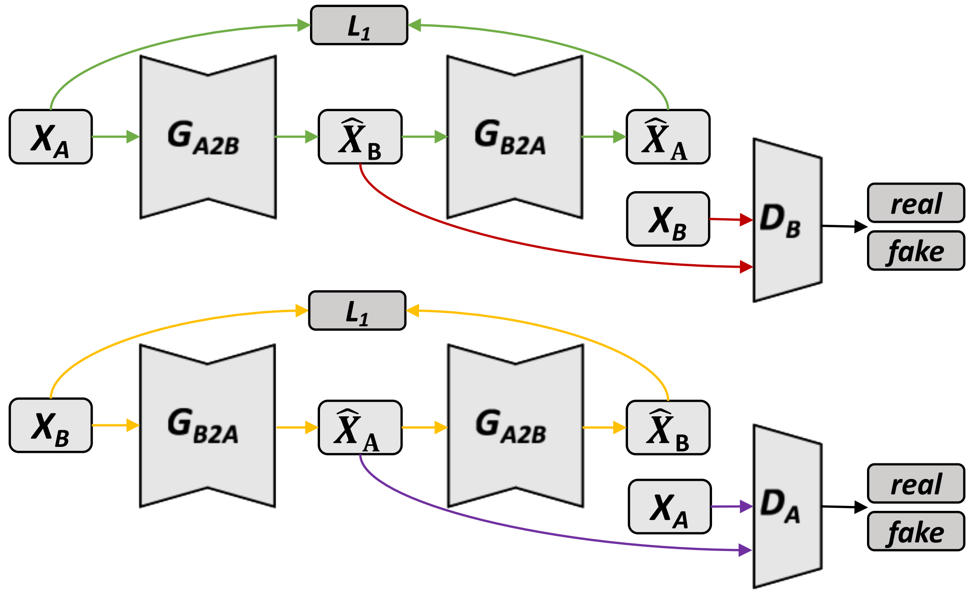
- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention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

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



UNIT

Learn the Encoder **Explicitly**



CycleGAN

Learn the Encoder **Implicitly**

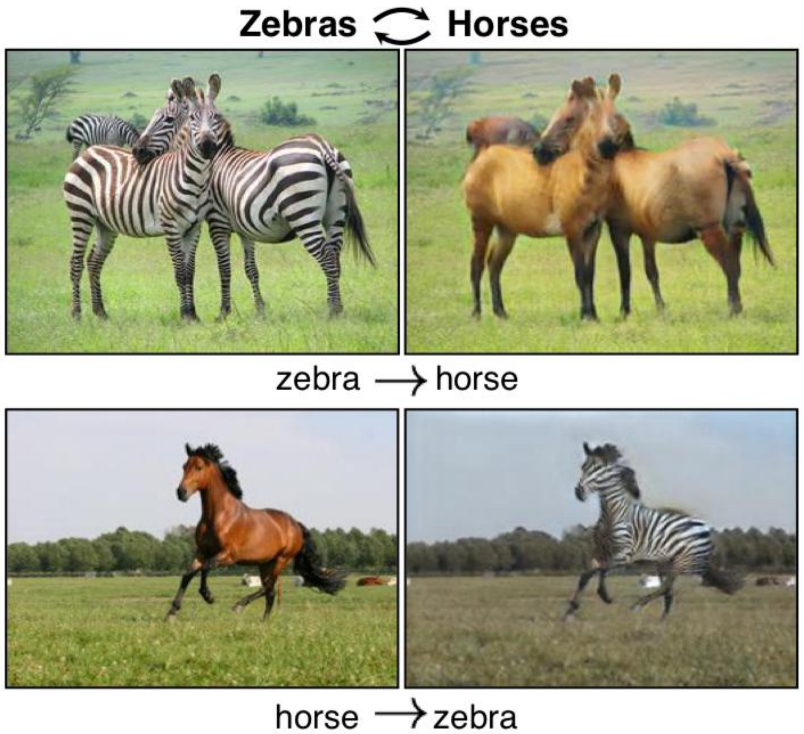
Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*

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**



CycleGAN
Learn the Encoder **Implicitly**

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

Input GTA5 CG

Output image with German street view style

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.

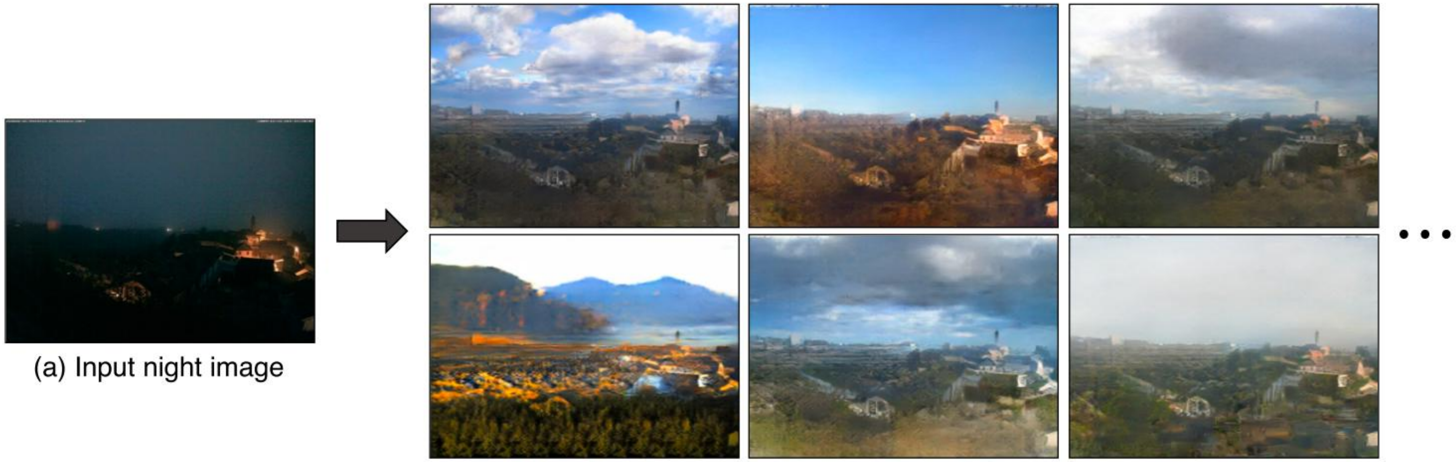
- Why the background / shape can be maintained?
  - Fully convolutional networks→ inductive bias
  - Cycle-consistency loss

- 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

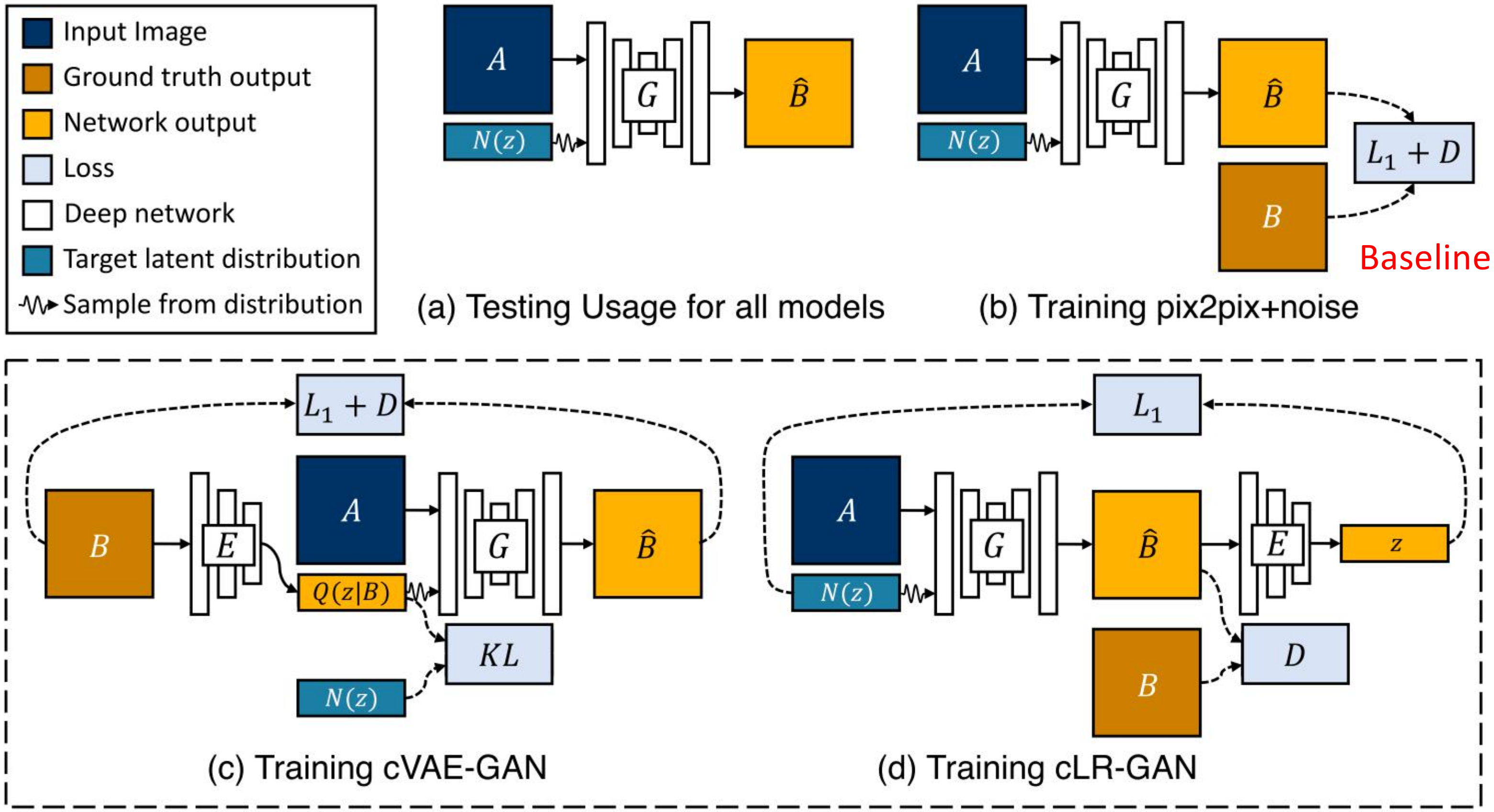


(a) Input night image

(b) Diverse day images sampled by our model

BiCycleGAN: multi-modality

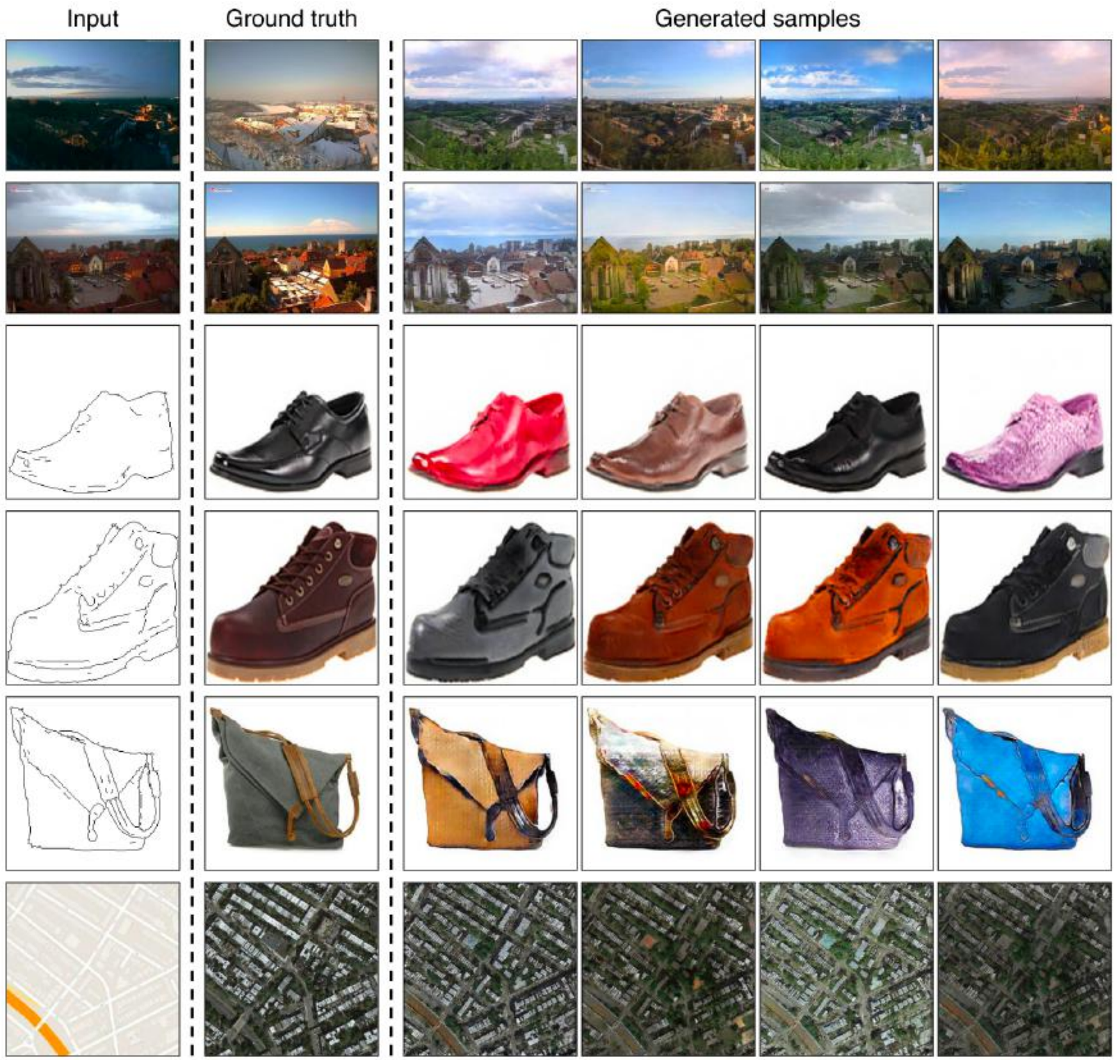
- Cycle on latent noises + Cycle on translated images



(e) Training BicycleGAN

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

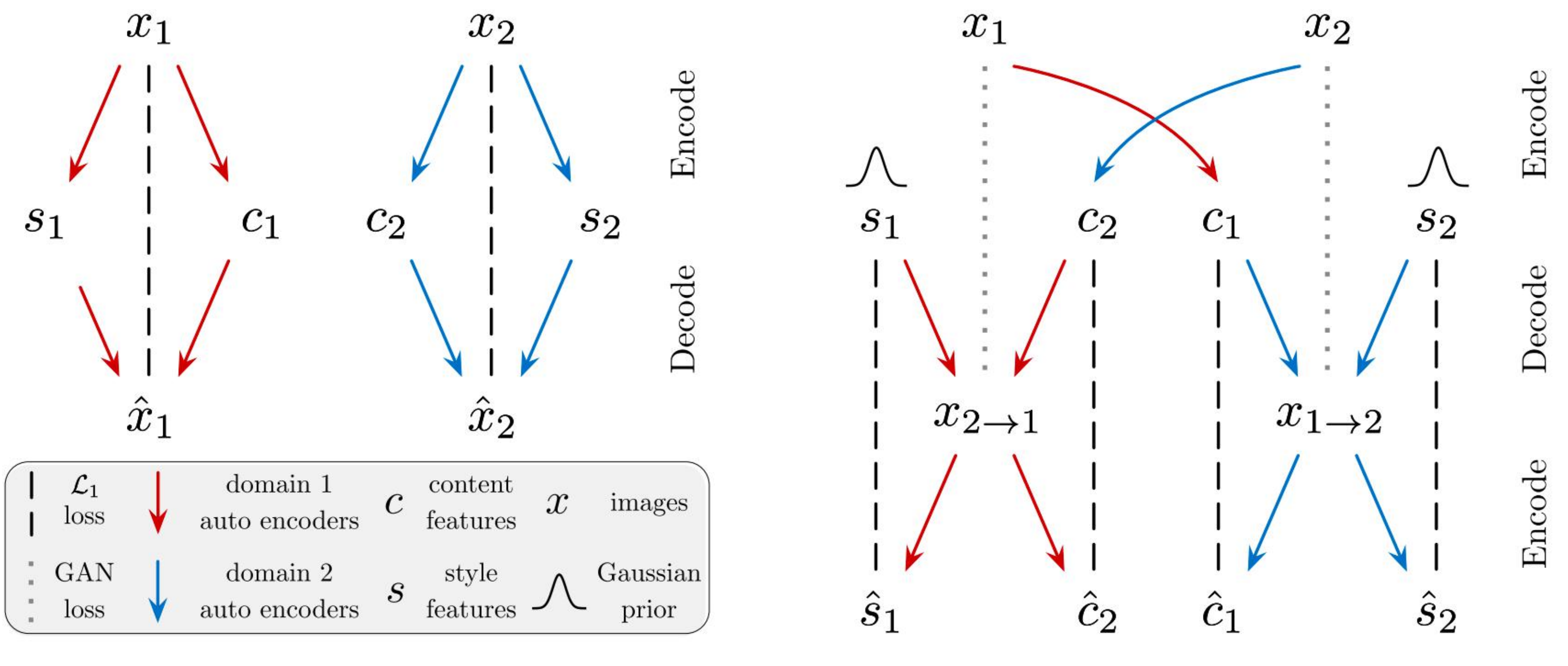
MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Goal: unpaired + multi-modal results

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

MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Latent reconstruction + Adversarial learning



(a) Within-domain reconstruction (b) Cross-domain translation

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

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

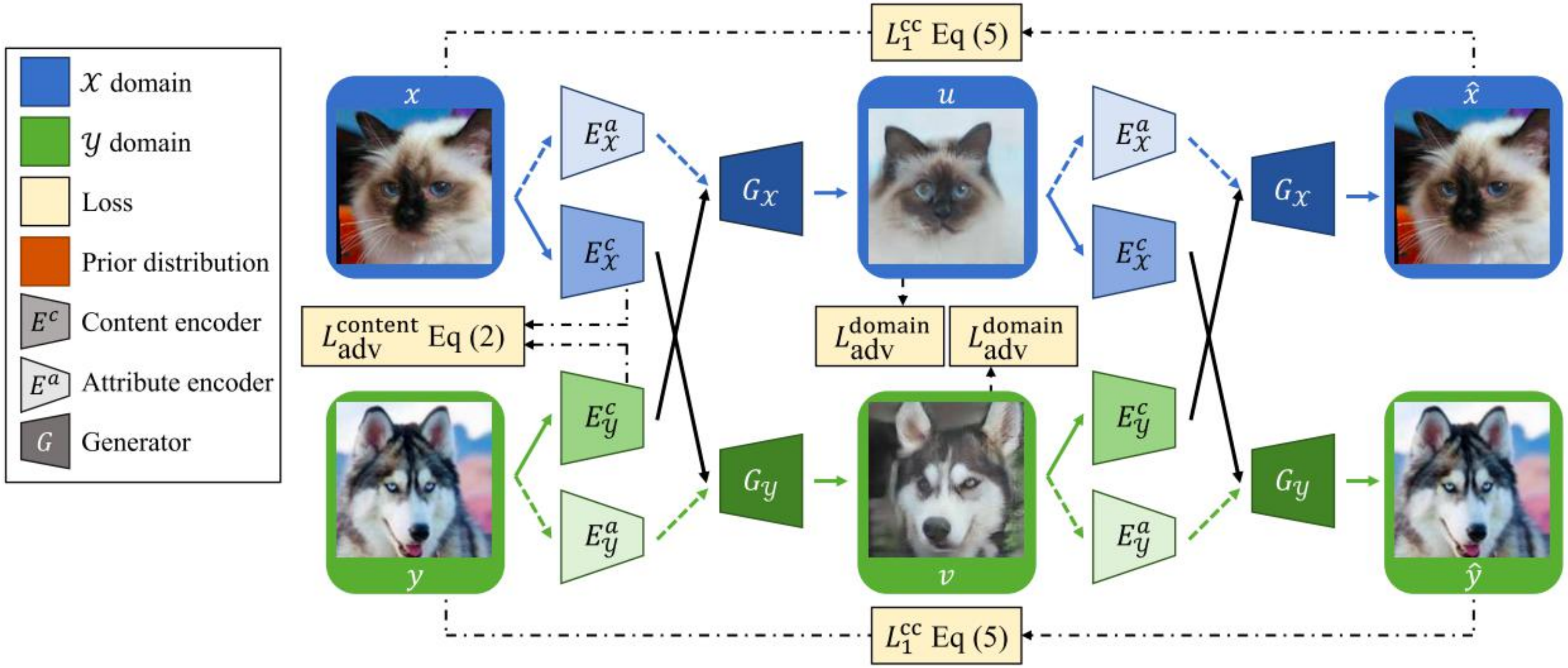
DRIT: disentangle domain-specific features

- Goal: Multi-modal results + Disentanglement

Diverse Image-to-Image Translation via Disentangled Representations. ECCV. 2018.

DRIT: disentangle domain-specific features

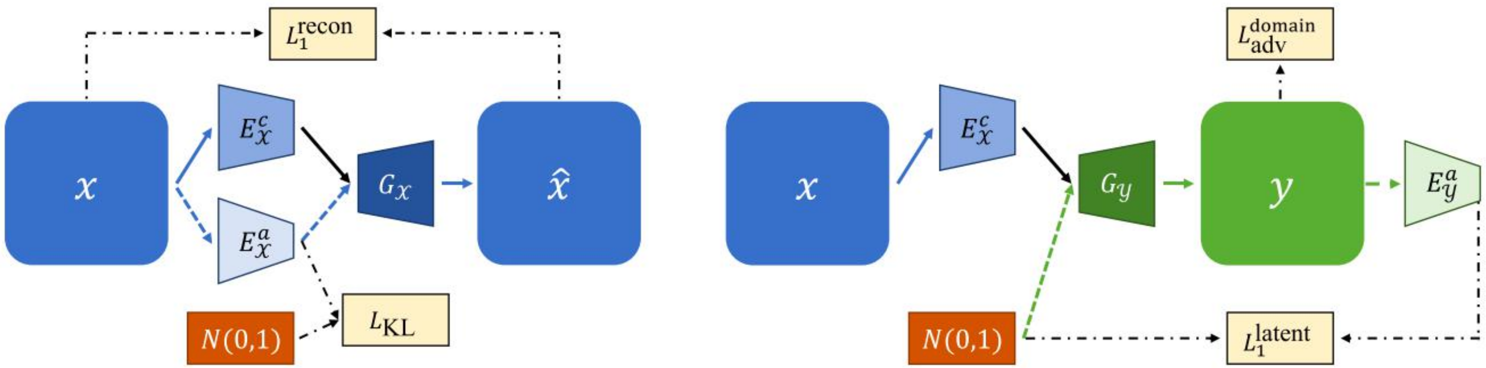
- Network bottleneck + Adversarial learning



Diverse Image-to-Image Translation via Disentangled Representations. ECCV. 2018.

DRIT: disentangle domain-specific features

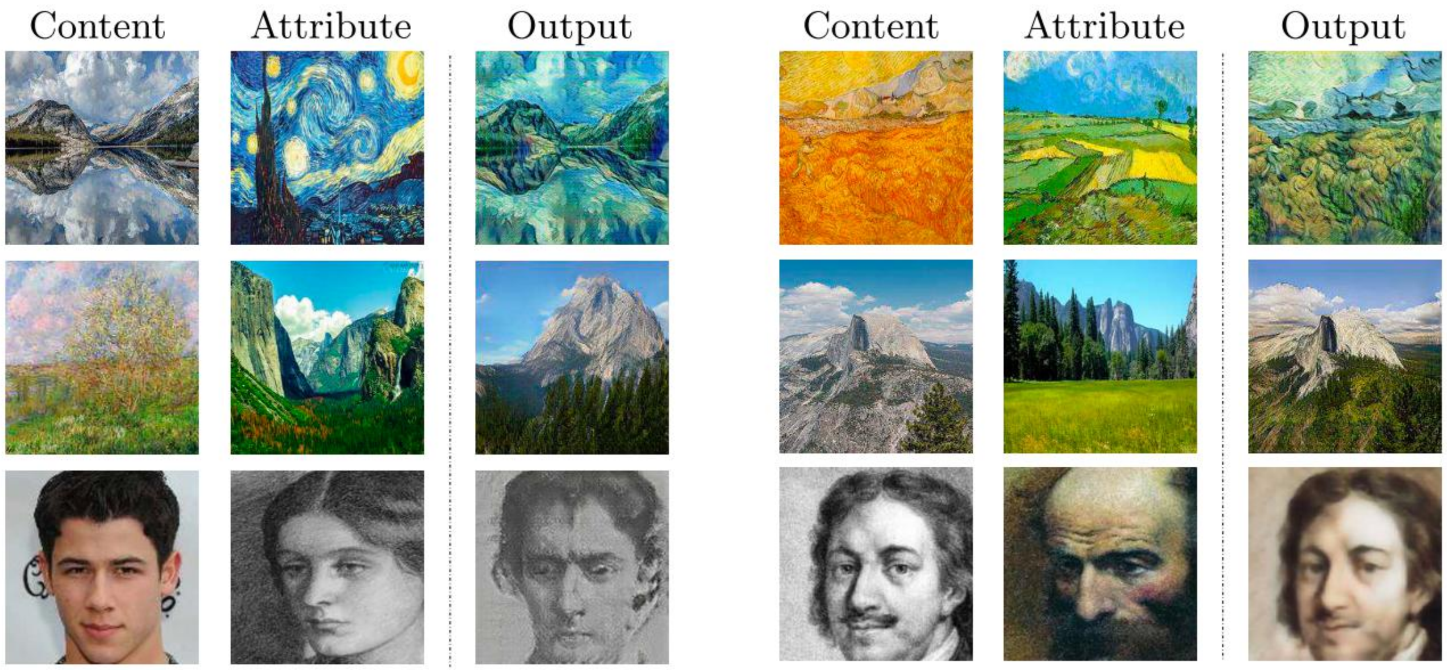
- Additional losses for better disentanglement



Diverse Image-to-Image Translation via Disentangled Representations. ECCV. 2018.

DRIT: disentangle domain-specific features

- Results



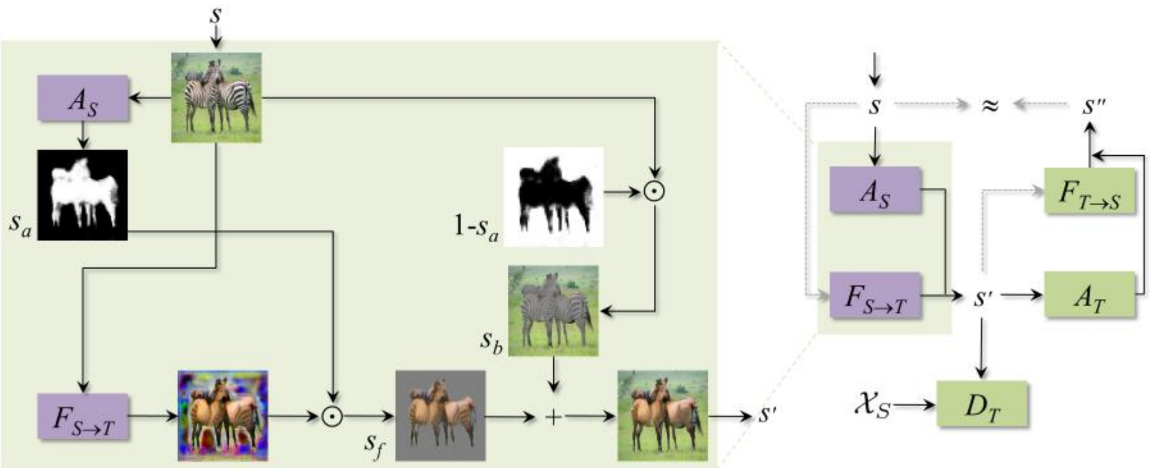
(a) Inter-domain attribute transfer (b) Intra-domain attribute transfer

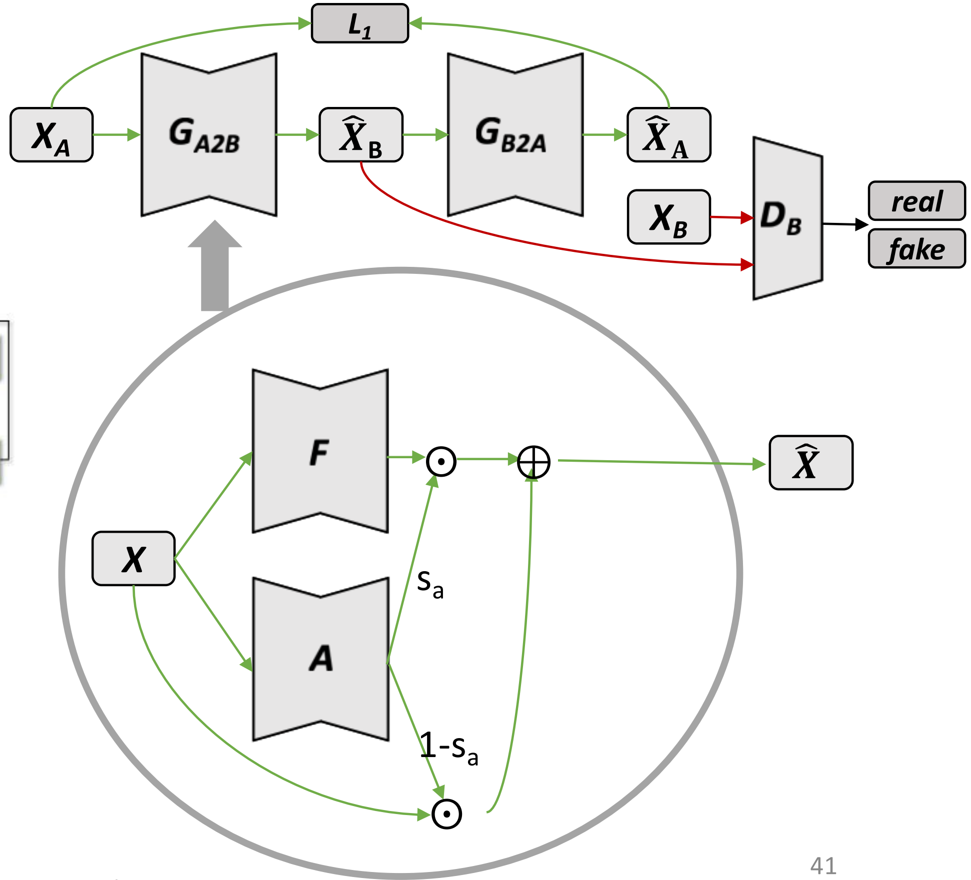
Diverse Image-to-Image Translation via Disentangled Representations. ECCV. 2018.

- 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

Attention CycleGAN: maintain background

- Learn the segmentation via synthesis

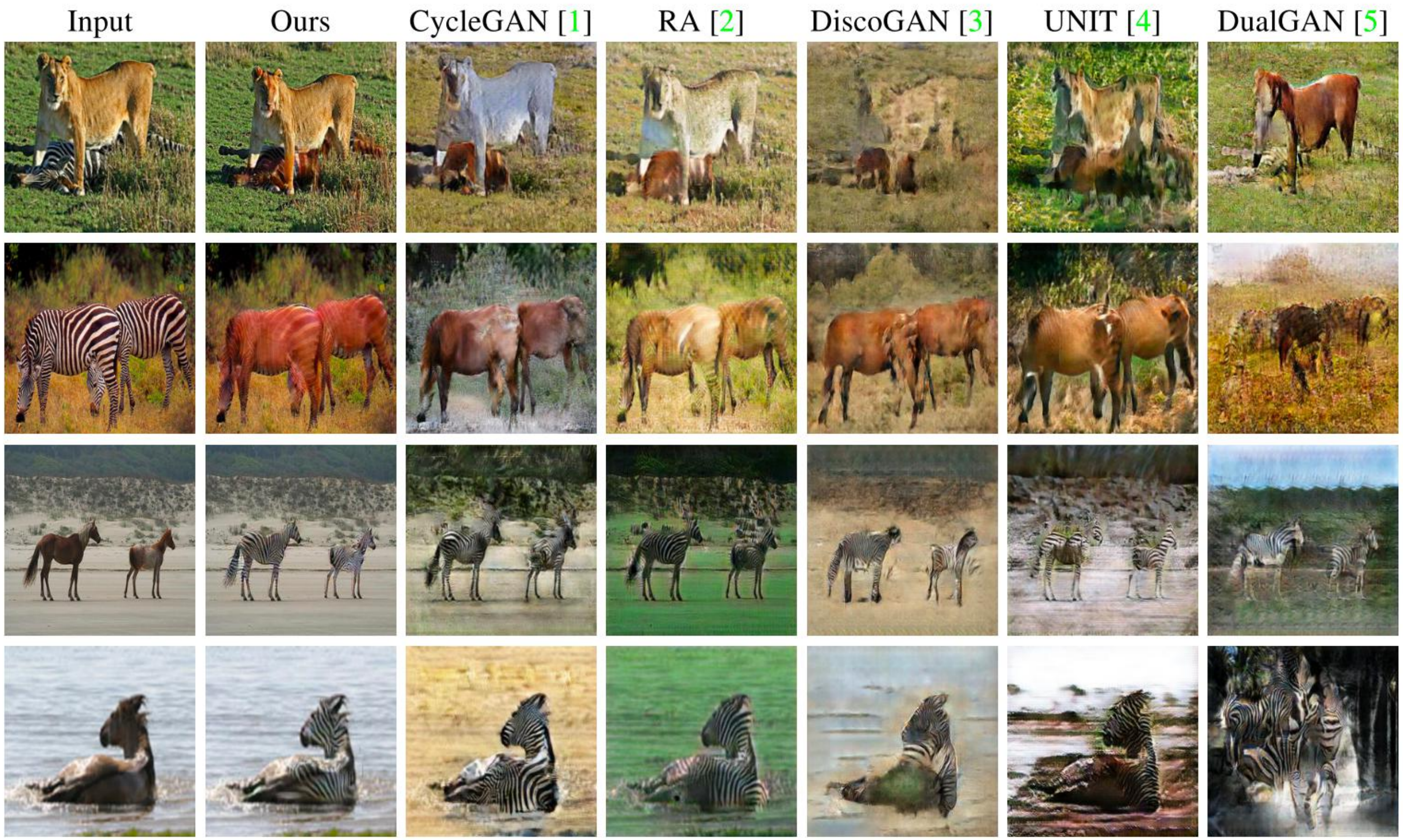




Unsupervised Attention-guided Image-to-Image Translation. *Mejjati, Y. A., Richardt, C., & Cosker, D.* NIPS, 2018

Attention CycleGAN: maintain backgrounds

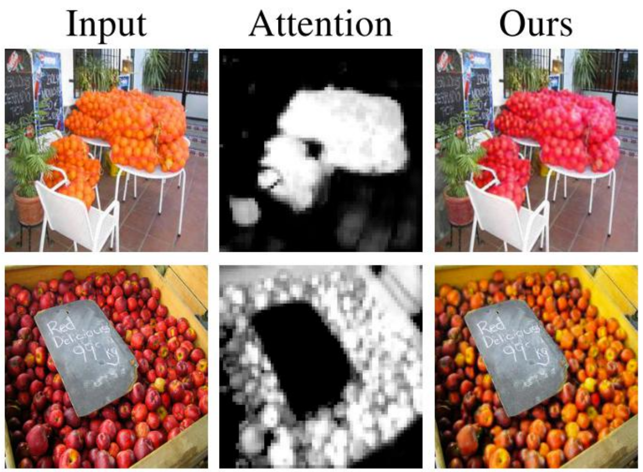
- Maintain backgrounds better



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

Attention CycleGAN: maintain backgrounds

- Learn the segmentation without segmentation labels



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

StarGAN: label condition

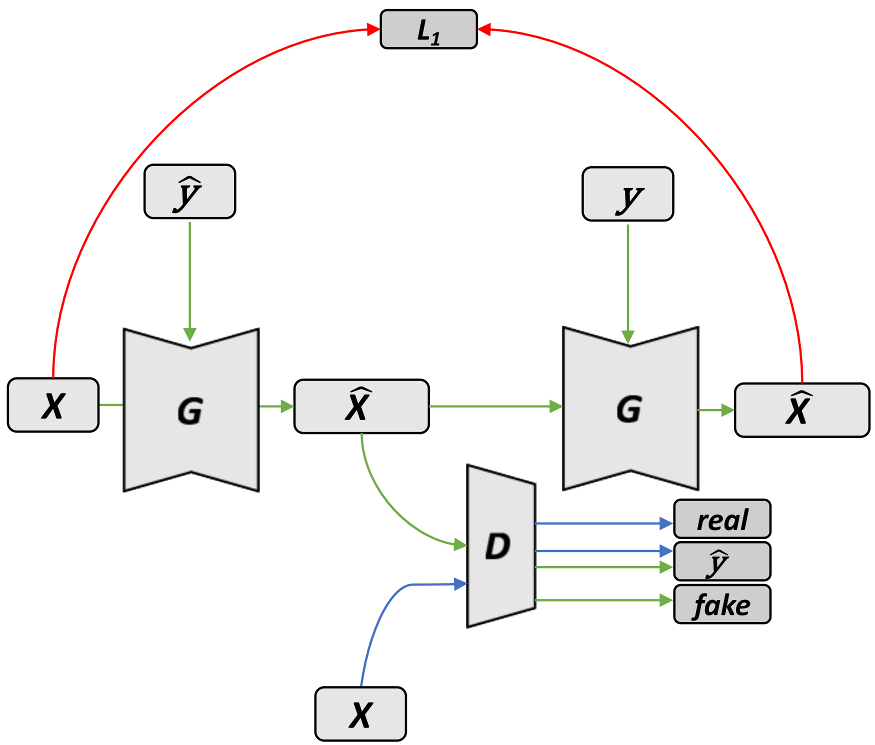
- Limitation of CycleGAN

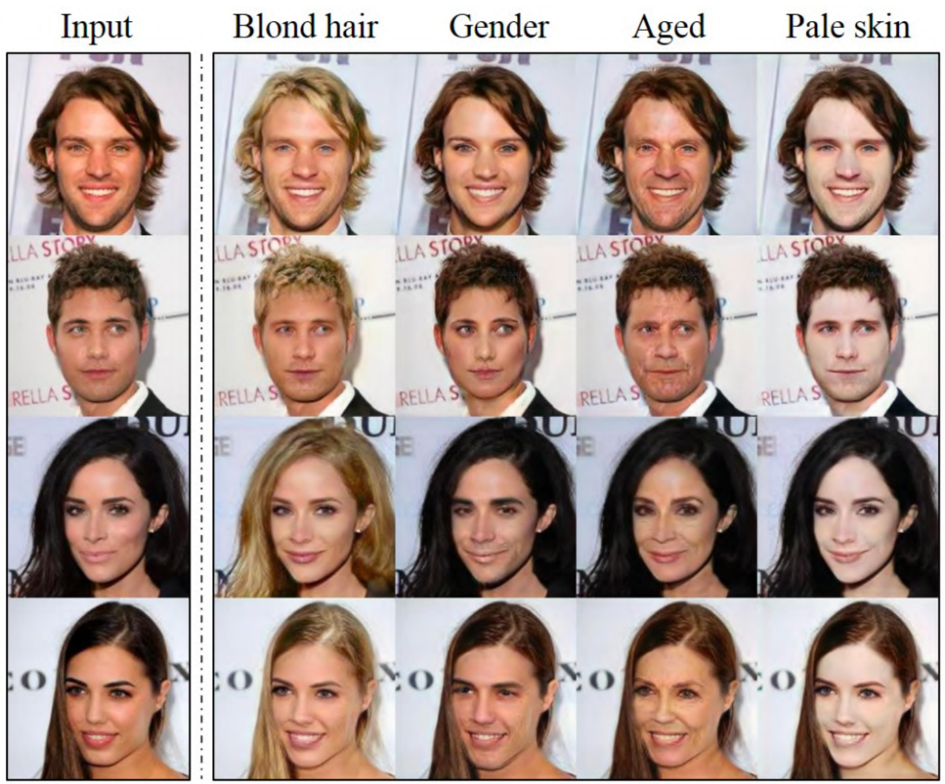
Translations between N domains require $N(N\text{-}1)$ models

StarGAN: One model to rule them all !!

StarGAN: label condition

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



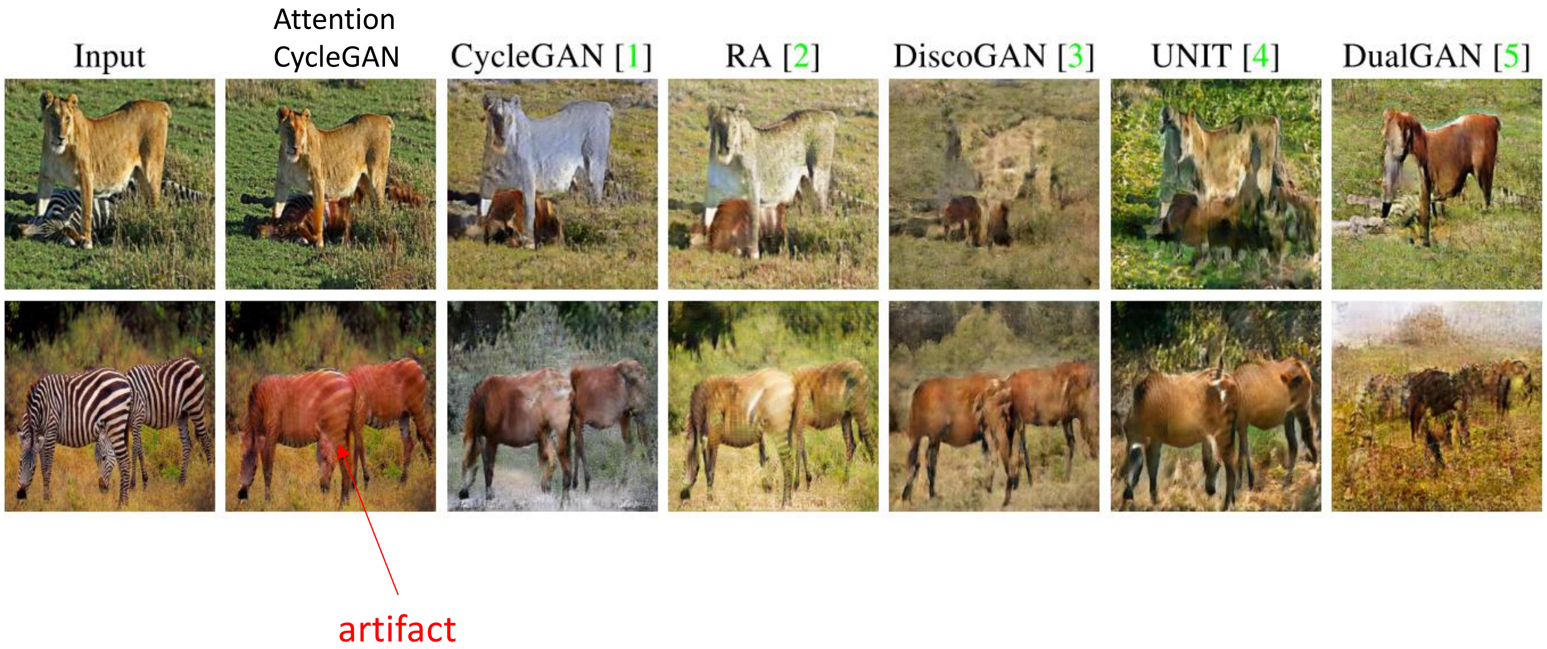


StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. CVPR. 2018.

- 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

Breaking the Cycle

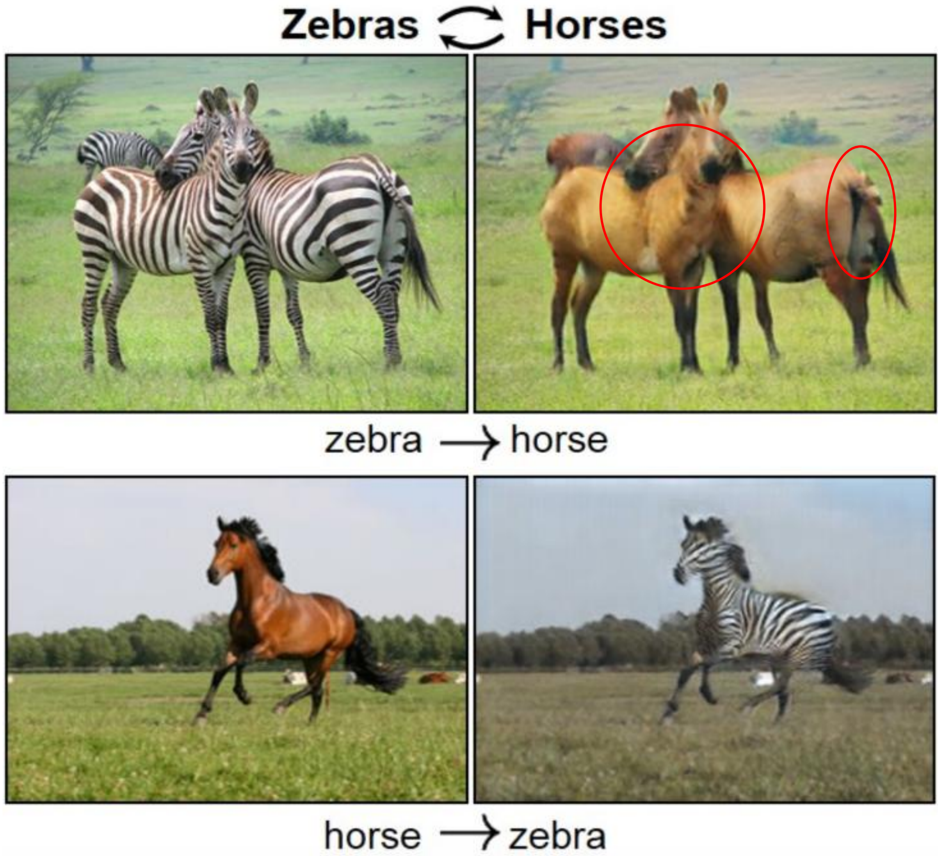
- Limitation of the cycle-consistency loss



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



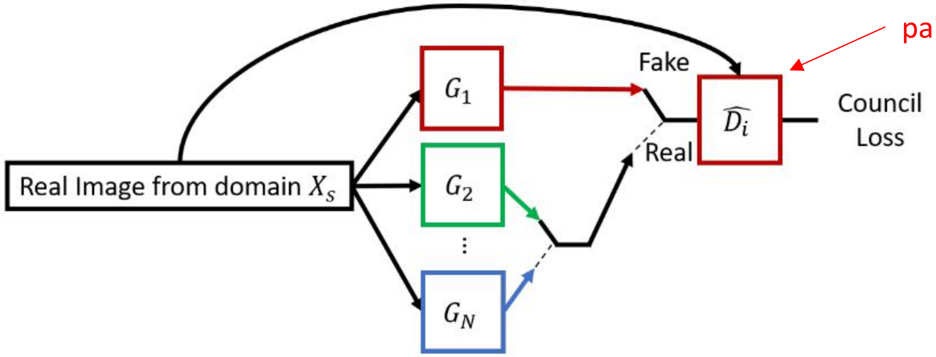
zebra -> horse Hidden Info

horse -> zebra OK

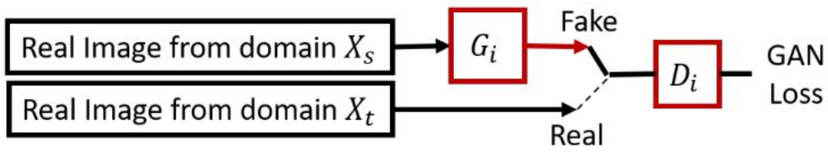
Image from: https://github.com/junyanz/CycleGAN

Breaking the Cycle

- Colleagues are all you need



(a) Council discriminator $\hat{D}_i$



(b) GAN discriminator D_i

- 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 $\widehat{D}_i$ 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. Nizan, Ori. Tal, Ayellet. arXiv 2019.11

Breaking the Cycle

- Colleagues are all you need

input ours [39] input ours-1 ours-2 [25] input ours [7]
(a) glasses removal (b) selfie to anime (c) male to female

Breaking the cycle—Colleagues are all you need. Nizan, Ori. Tal, Ayellet. arXiv 2019.11

Breaking the Cycle

- Discussion: Why it works?

- It is not only for im2im,
  other distribution transformations may also benefit from this approach

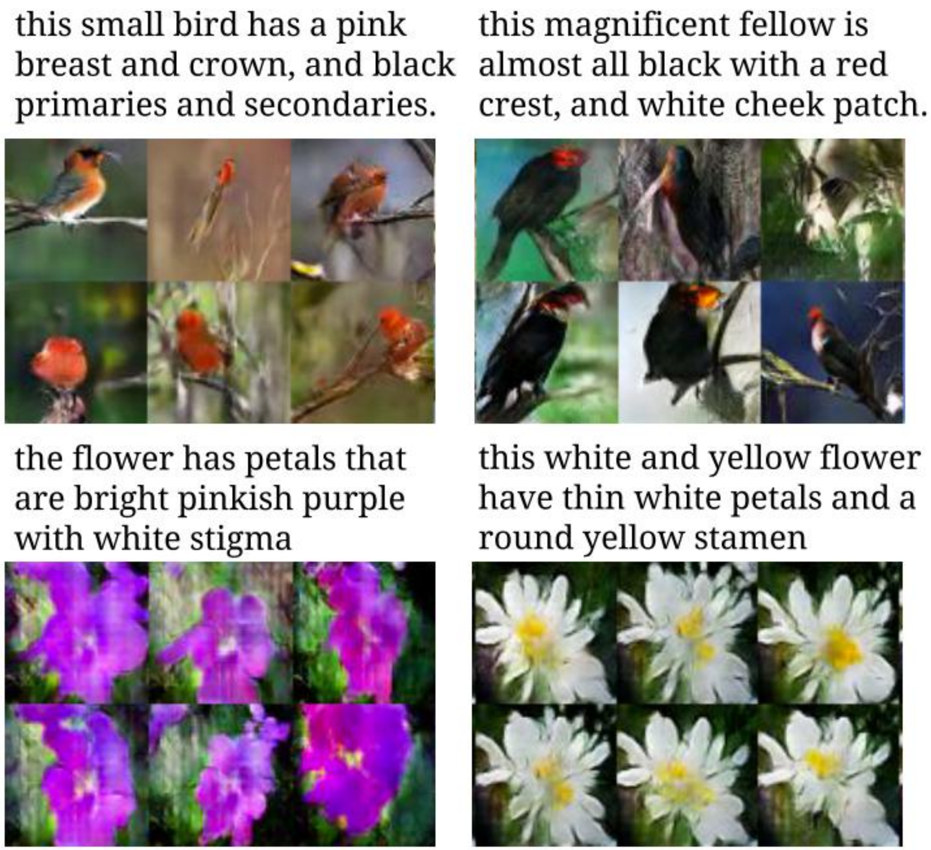
- Better Methods:
  - ACL-GAN
  - XDCycleGAN

Breaking the cycle—Colleagues are all you need. Nizan, Ori. Tal, Ayellet. arXiv 2019.11

- 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



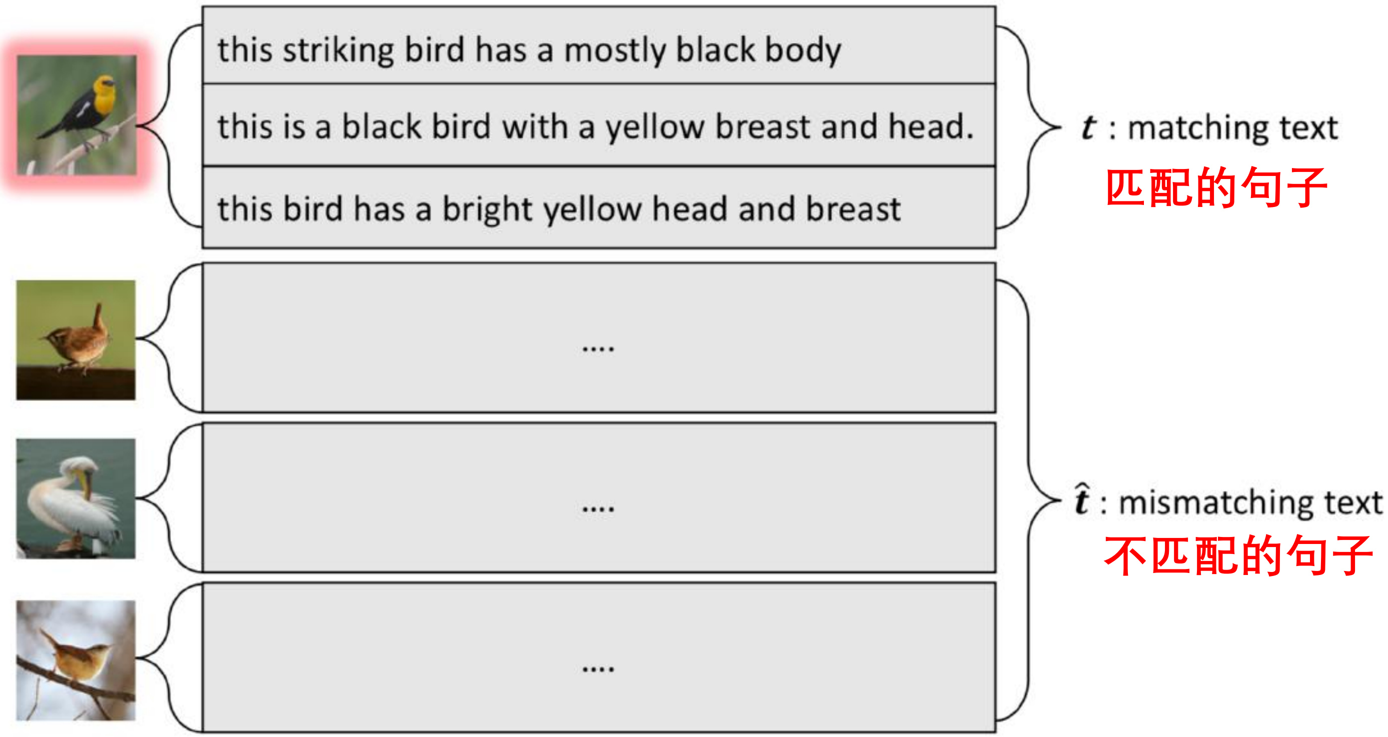
- Classic multi-modal problem

P(t, z)

Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*

GAN-CLS and SisGAN

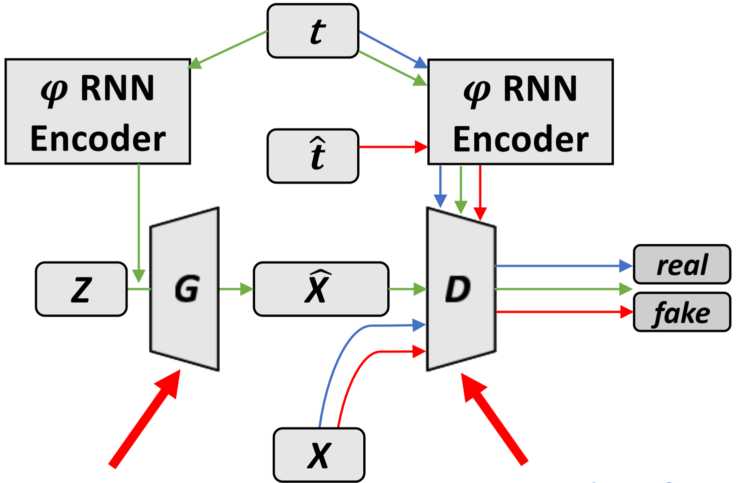
- Text-to-image synthesis



Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*

GAN-CLS and SisGAN

- Text-to-image synthesis

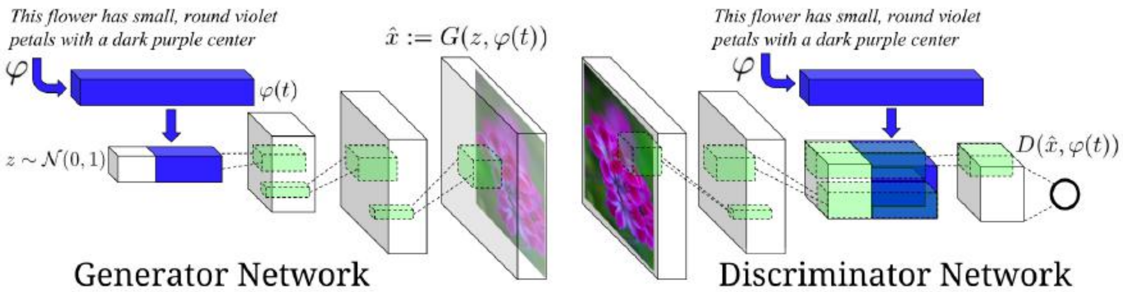


t : matching text

$\hat{t}$: mismatched text

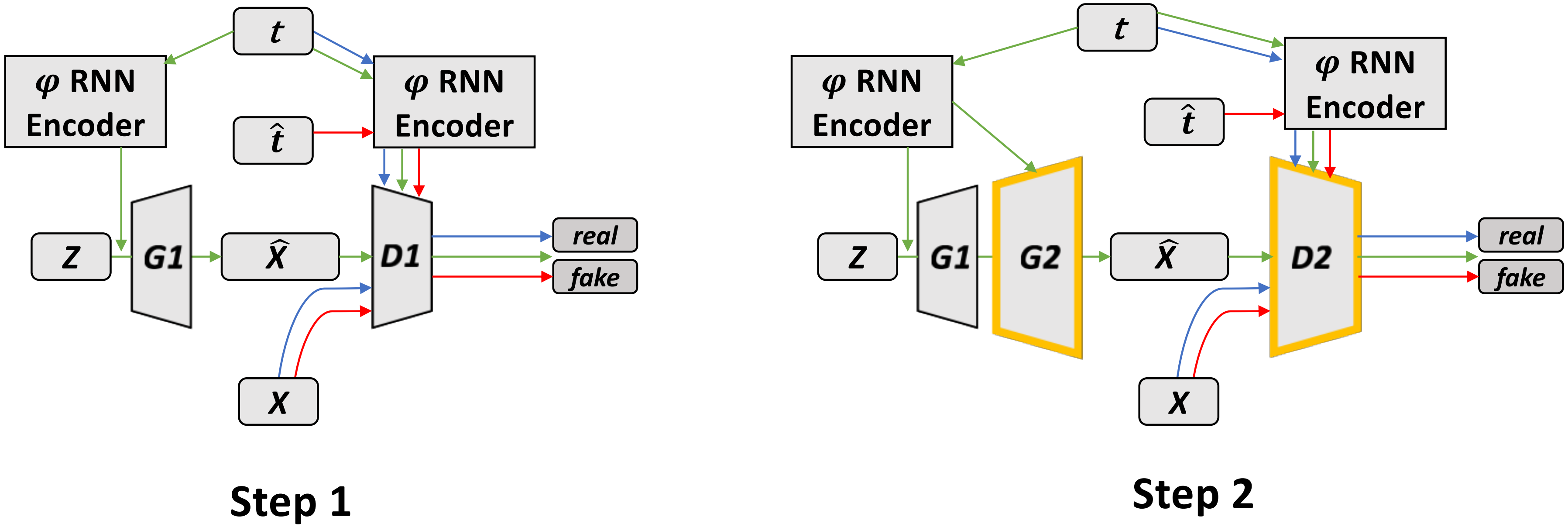
Learn to fool discriminator

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



GAN-CLS and SisGAN

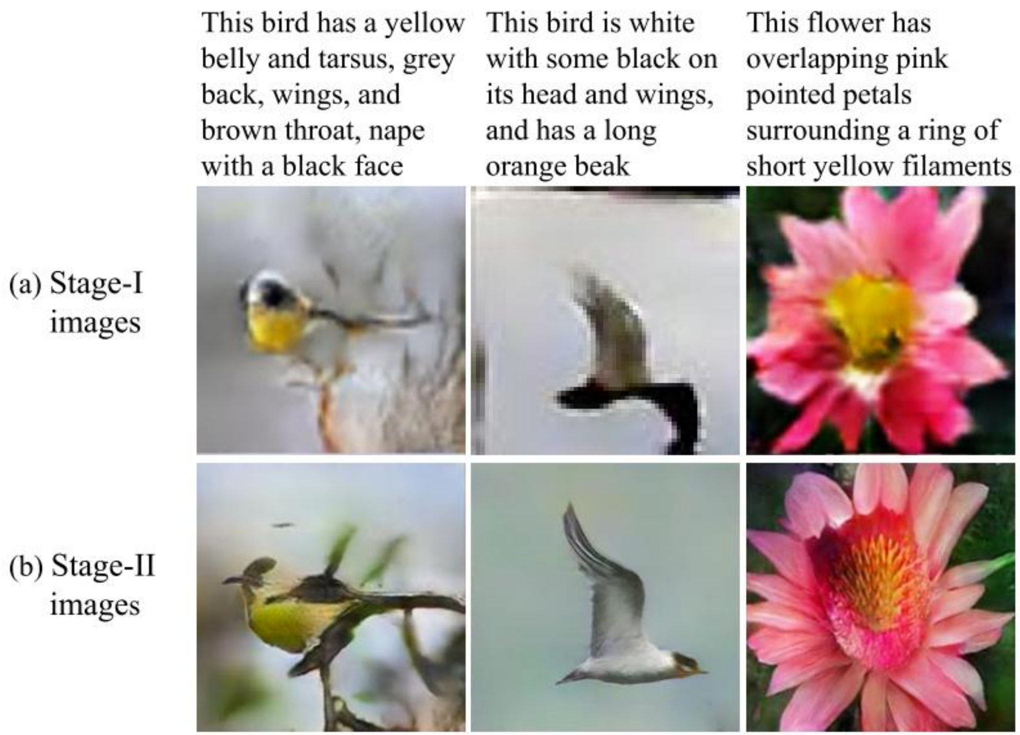
- Text-to-image synthesis + High resolution image



StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. *H. Zhang, T. Xu et al. ICCV. 2017.*

GAN-CLS and SisGAN

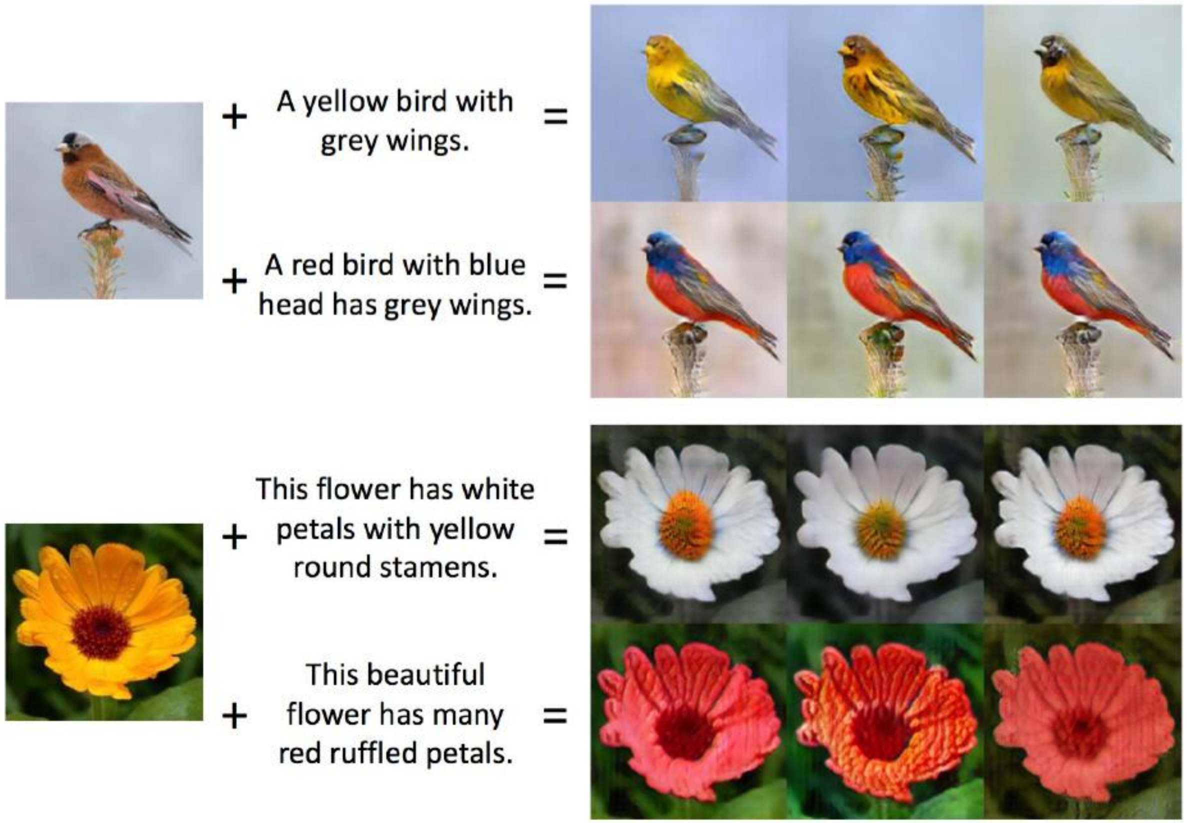
- Text-to-image synthesis + High resolution image



StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. *H. Zhang, T. Xu et al. ICCV. 2017.*

GAN-CLS and SisGAN

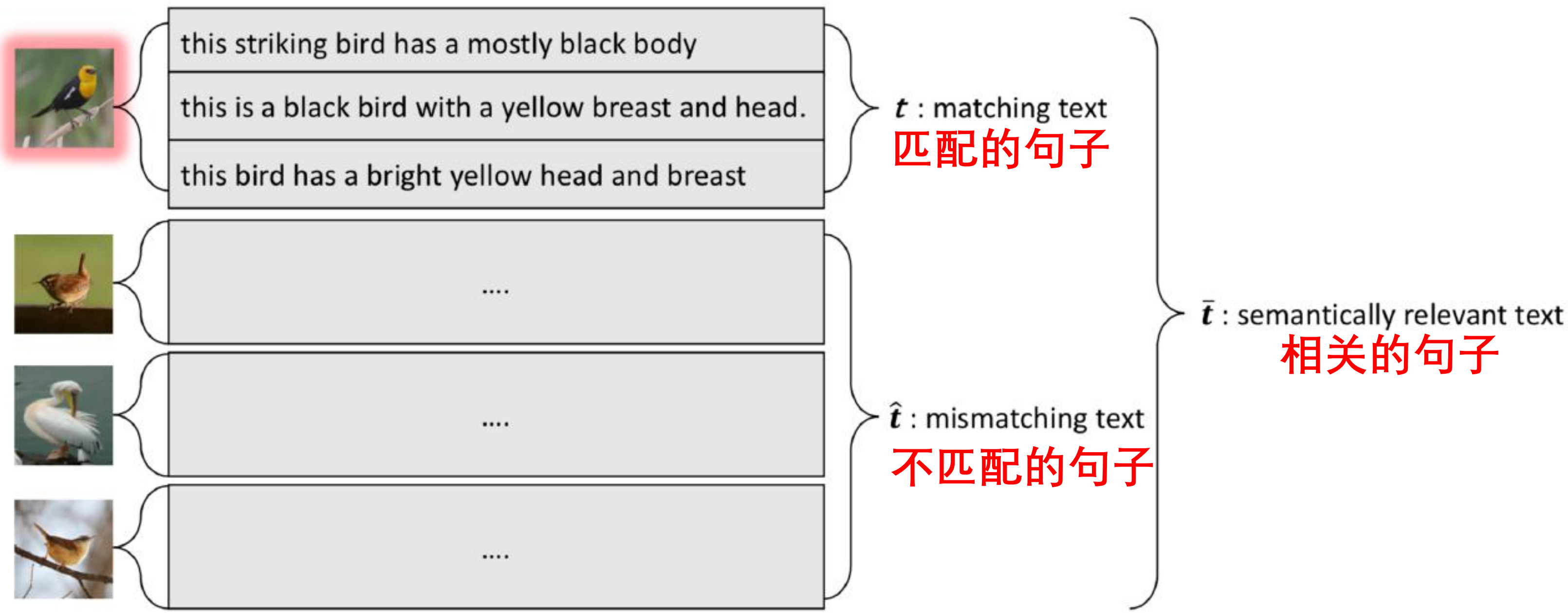
- Semantic image synthesis: image manipulation with natural language



Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

GAN-CLS and SisGAN

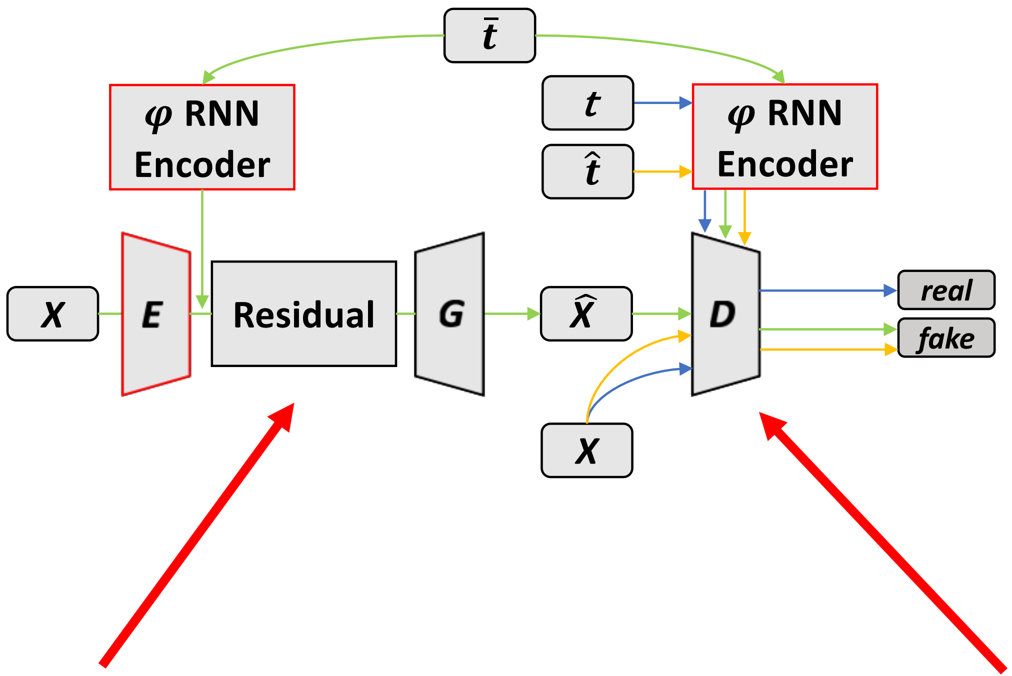
- Semantic image synthesis: image manipulation with natural language



Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

GAN-CLS and SisGAN

- Semantic image synthesis: image manipulation with natural language



t : matching text

$\hat{t}$: mismatched text

$\bar{t}$: semantically relevant text

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

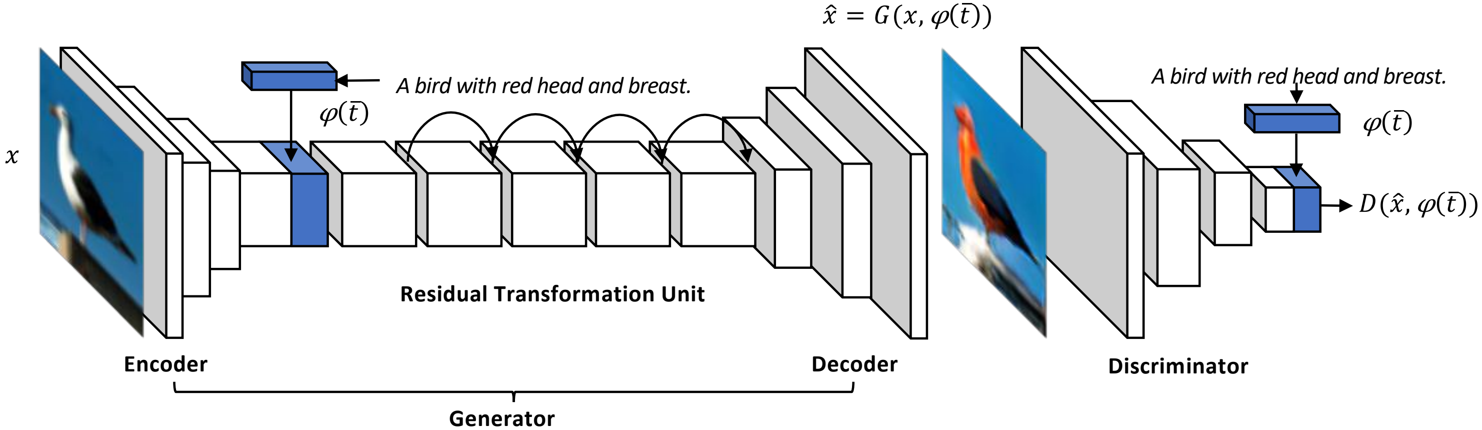
Learn to fool discriminator when inputting image with semantically relevant text

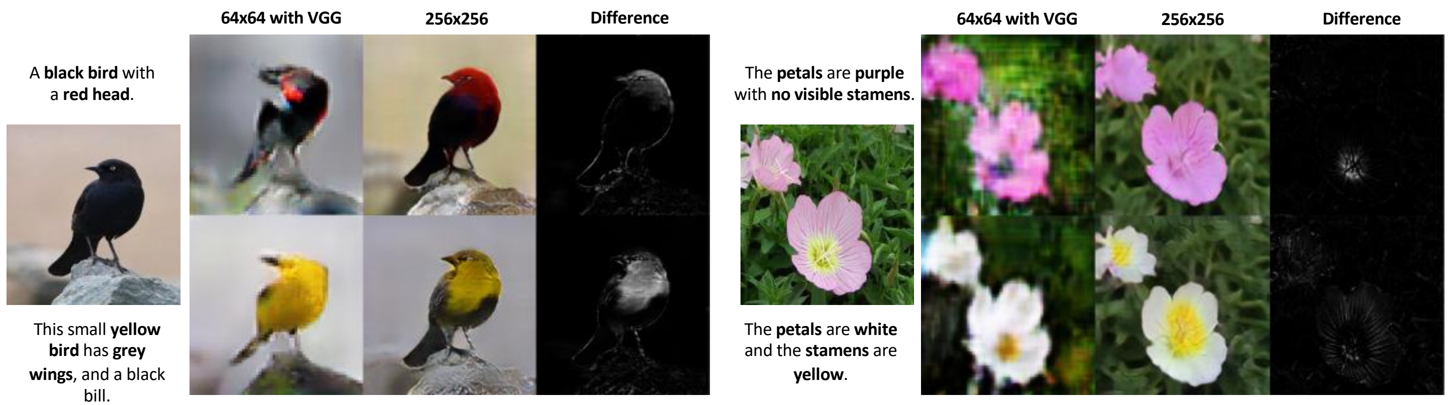
1. Learn to classify matching image and text pairs as real samples
2. Learn to classify mismatched image and text pairs as fake samples
3. Learn to classify samples from generator as fake samples

Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

GAN-CLS and SisGAN

- Semantic image synthesis: Learn the location information via synthesis





Semantic Image Synthesis via Adversarial Learning. *Dong, H., Yu, S., Wu, C., Guo, Y.* 2017. ICCV

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