DRIT:

Diverse Image-to-Image Translation via Disentangled Representations

Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, Ming-Hsuan Yang

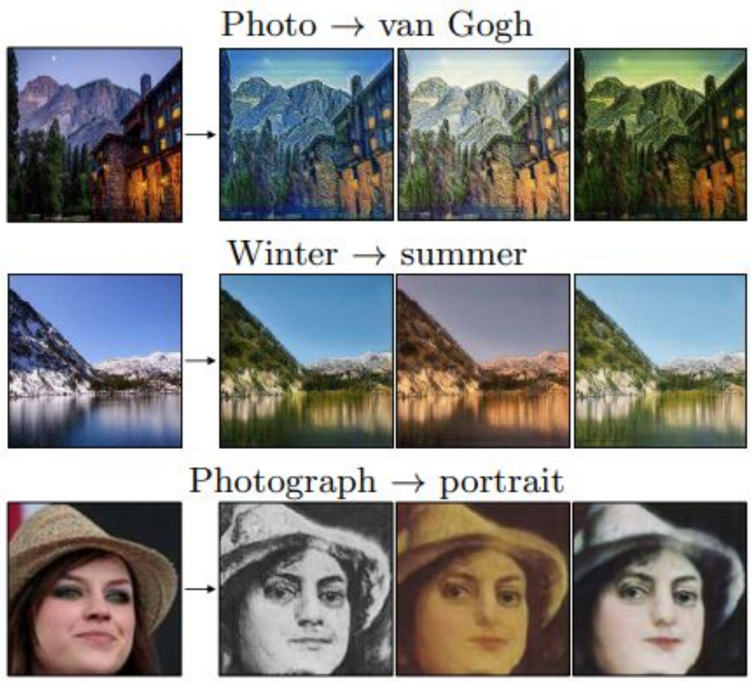
University of California, Merced. Virginia Tech. Verisk Analytics. Google Cloud.

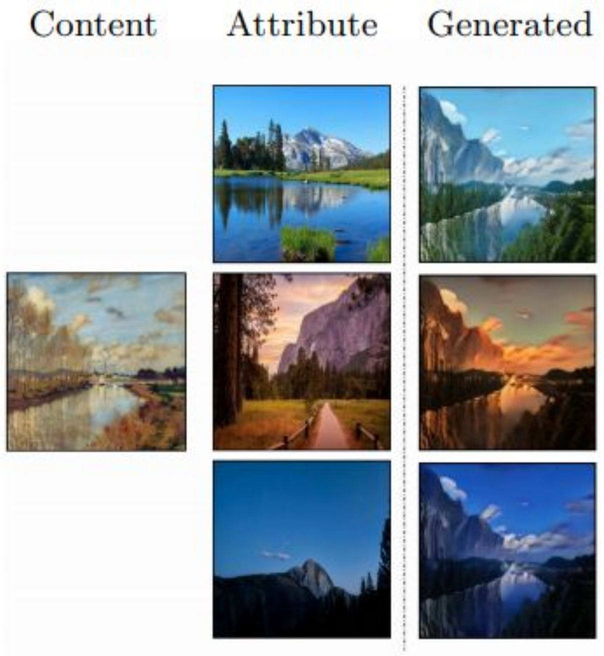
ECCV 2018

Sungman Cho

Introduction

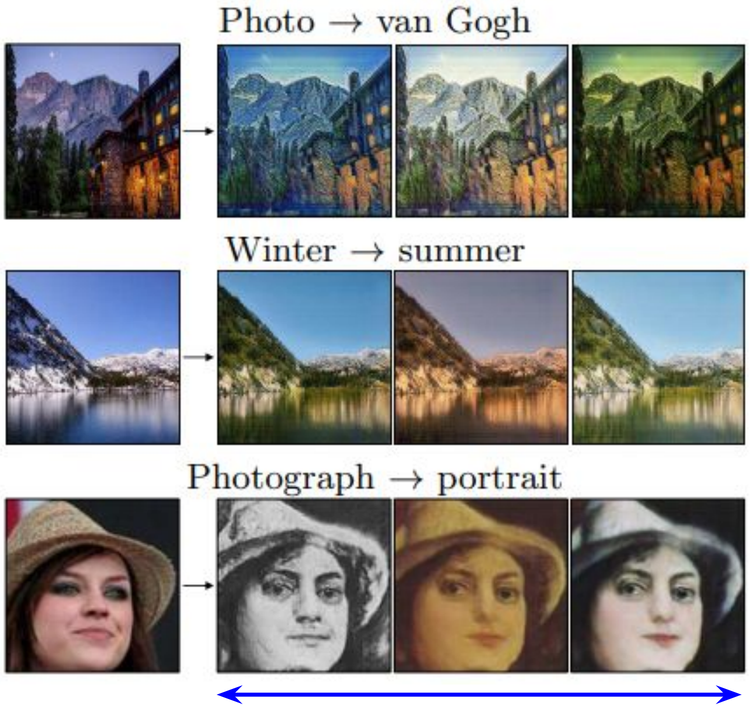
Introduction

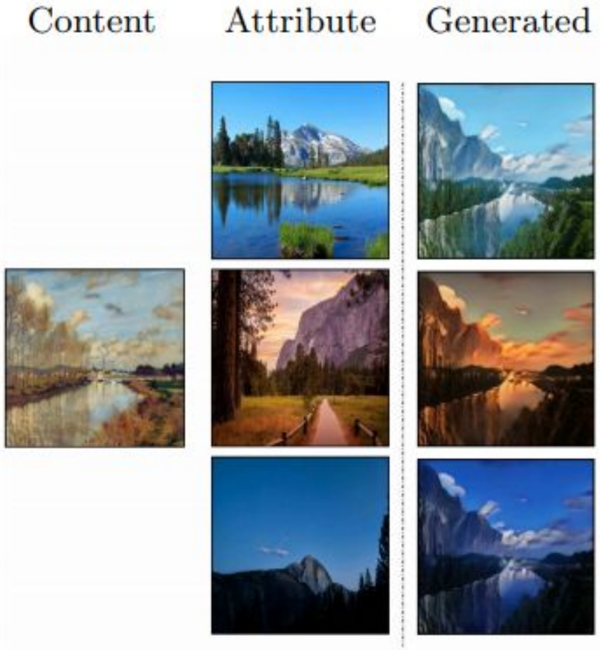




“Generate diverse outputs with unpaired training data.”

Introduction





“Generate diverse outputs with unpaired training data.”

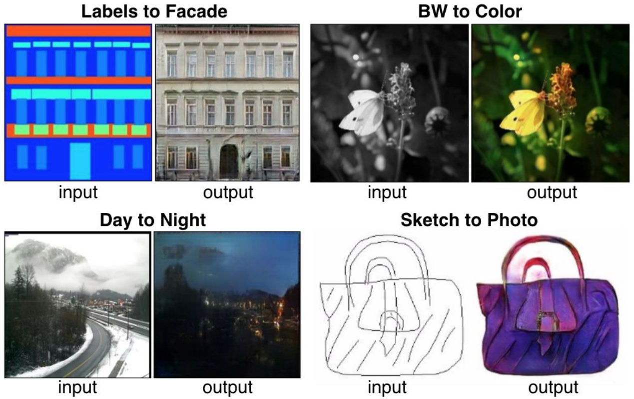
Challenges

- **Aligned training image pairs are either difficult to collect or do not exist. (Pix2Pix)**

Pair

- **Many such mappings are inherently multimodal.**
  A single input may correspond to multiple possible outputs.
  (CycleGAN, DiscoGAN, UNIT)

Deterministic



Pix2Pix

CycleGAN

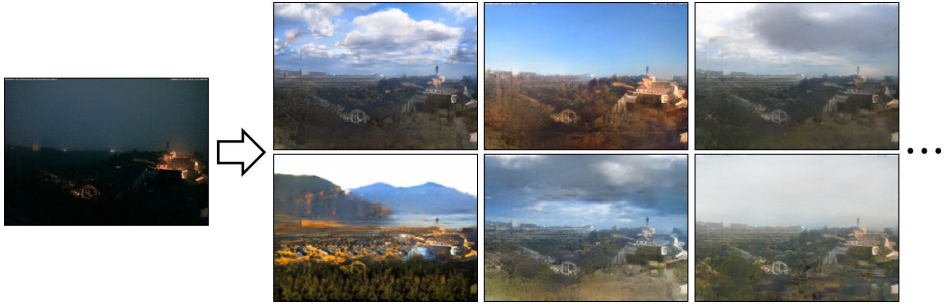
Challenges

- Aligned training image pairs are either difficult to collect or do not exist. (Pix2Pix)

Pair

- Many such mappings are inherently multimodal.
  A single input may correspond to multiple possible outputs.
  (CycleGAN, DiscoGAN, UNIT)

Deterministic

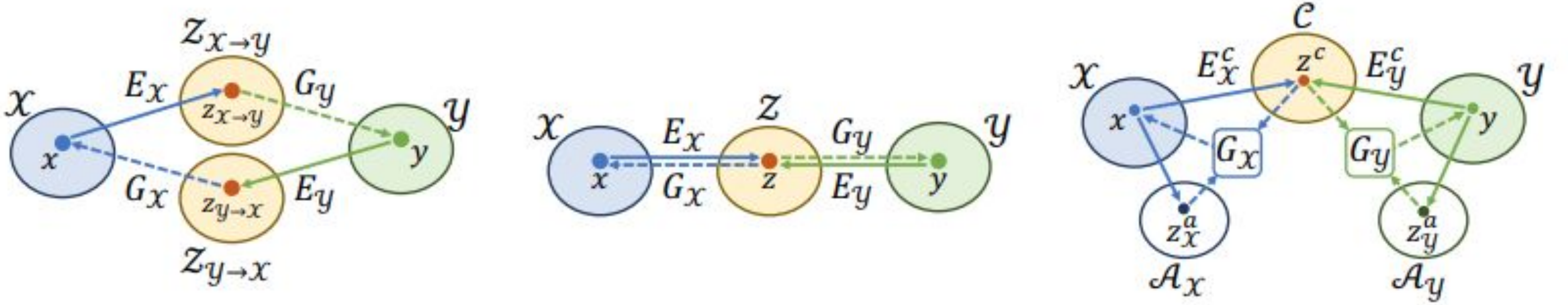


BicycleGAN

CycleGAN

Related Works

Method	Pix2Pix [18]	CycleGAN [46]	UNIT [26]	BicycleGAN [47]	Ours
Unpaired	-	✓	✓	-	✓
Multimodal	-	-	-	✓	✓



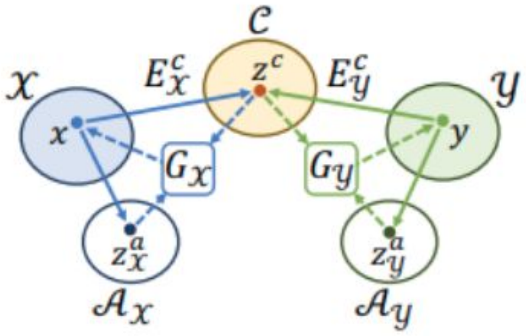
(a) CycleGAN [46] (b) UNIT [26] (c) Ours

Introduction

- We propose to **embed images onto two spaces**:

  1) A domain-invariant content space

  2) A domain-specific attribute space



(c) Ours

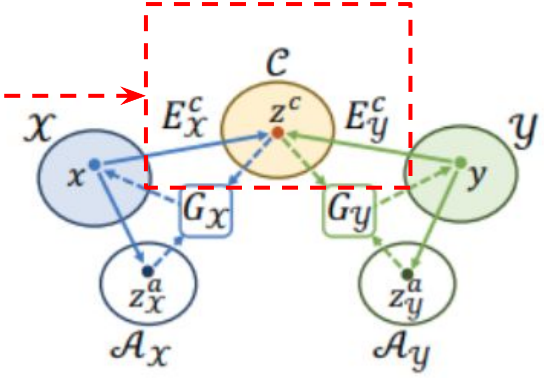
“Generate diverse outputs with unpaired training data.”

Introduction

- We propose to **embed images onto two spaces**:

  1) A domain-invariant content space

  2) A domain-specific attribute space



(c) Ours

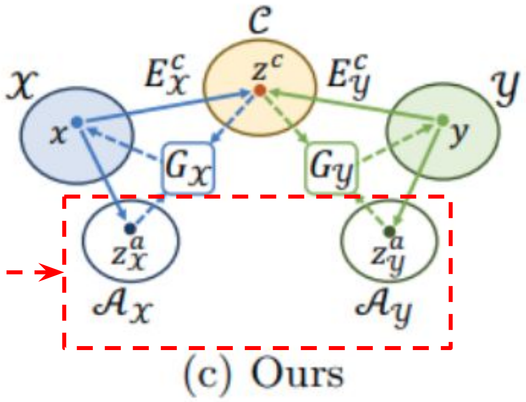
“Generate diverse outputs with unpaired training data.”

Introduction

- We propose to **embed images onto two spaces**:

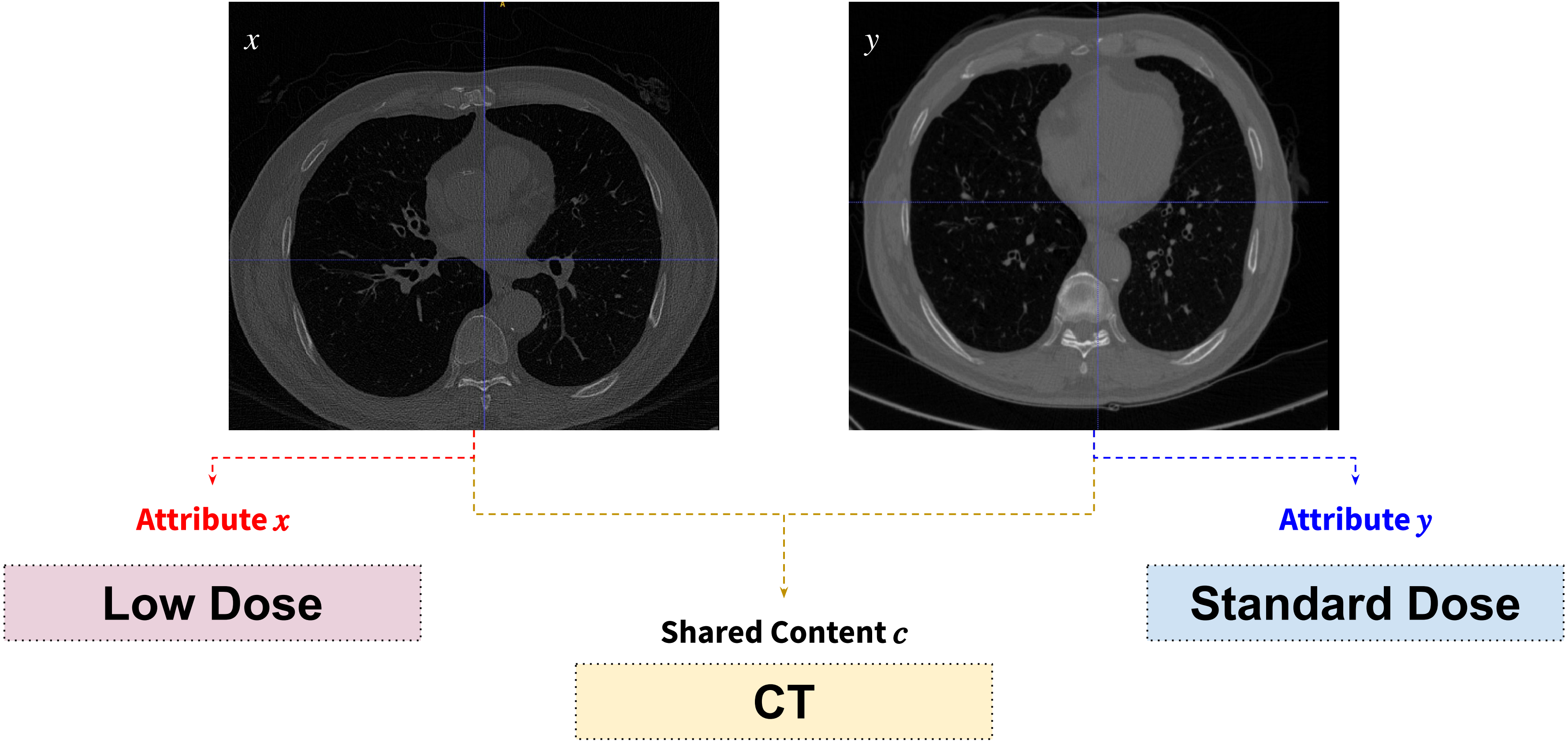
  1) A domain-invariant content space

  2) A domain-specific attribute space

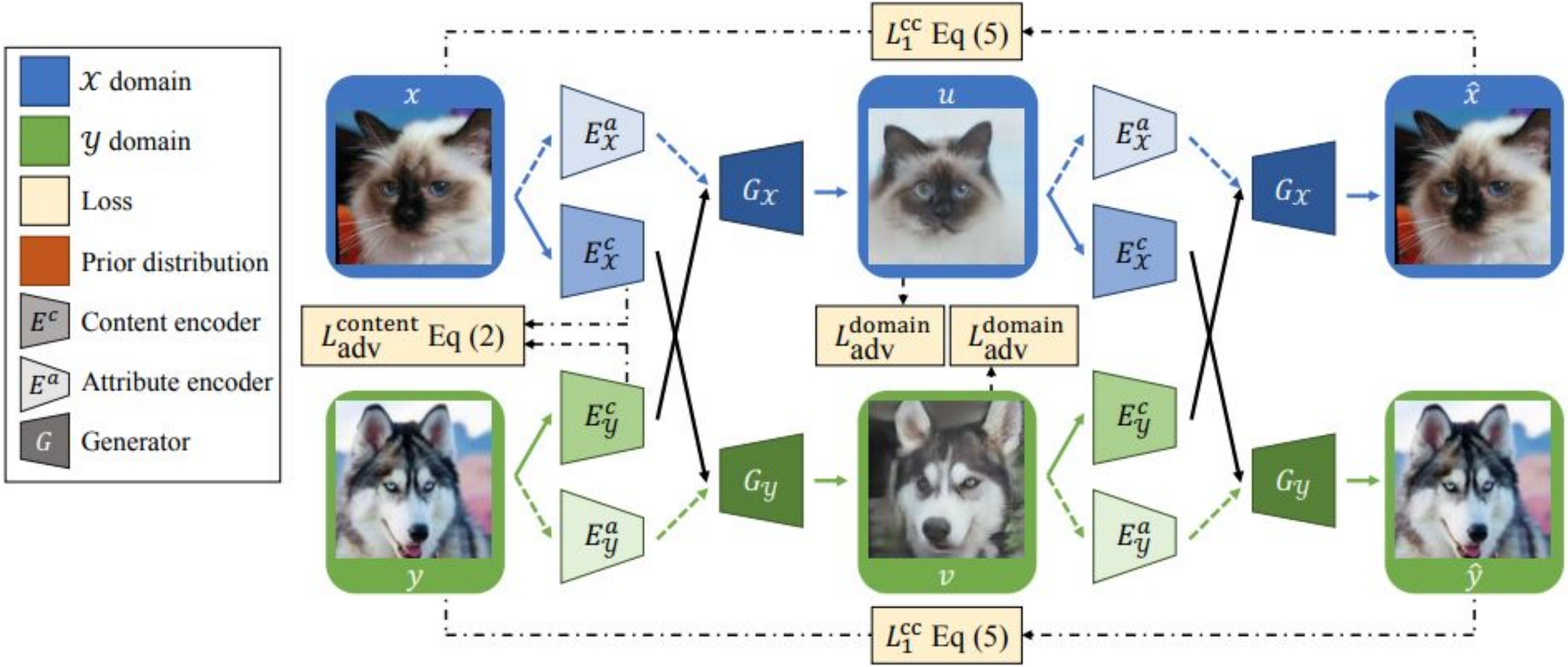


“Generate diverse outputs with unpaired training data.”

In our case ?

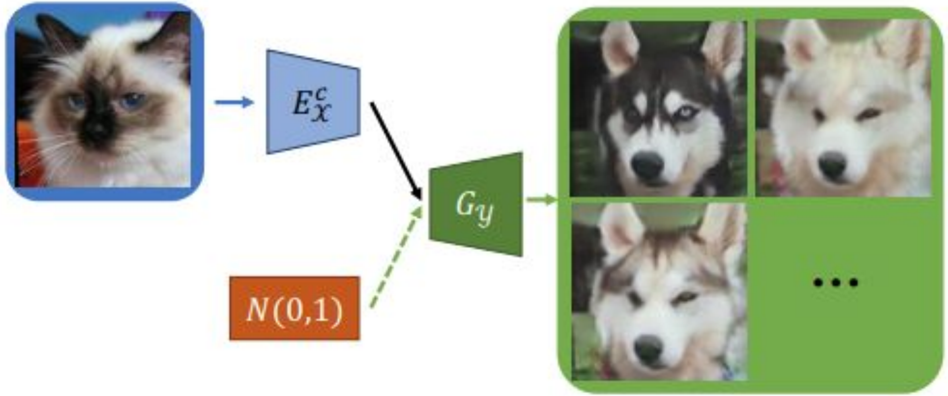

x
y
Attribute x
Low Dose
Shared Content c
CT
Attribute y
Standard Dose

DRIT (Training Phase)

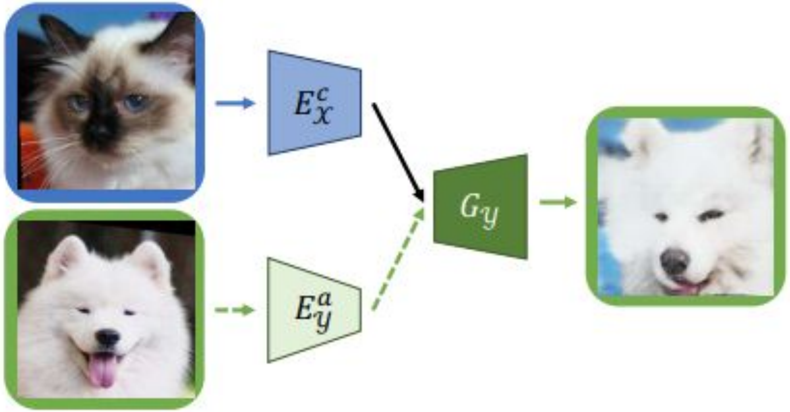


(a) Training with unpaired images

DRIT (Test Phase)



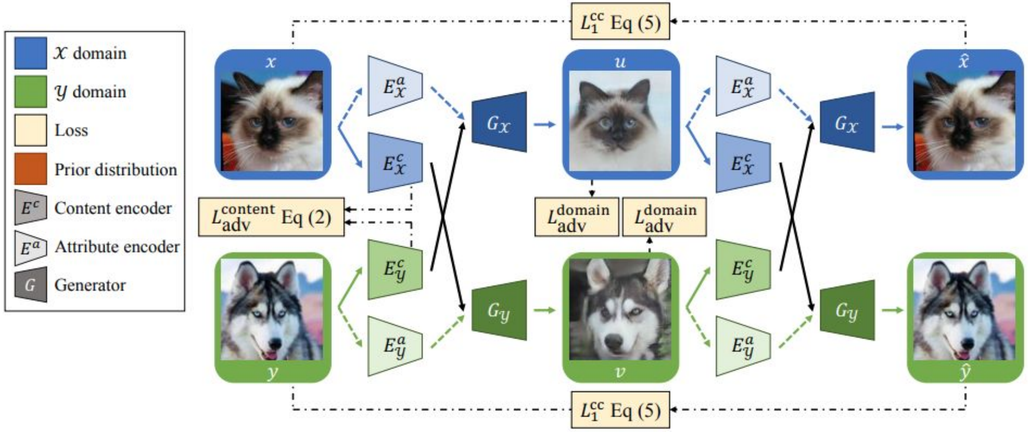
(b) Testing with random attributes



(c) Testing with a given attribute

Methods

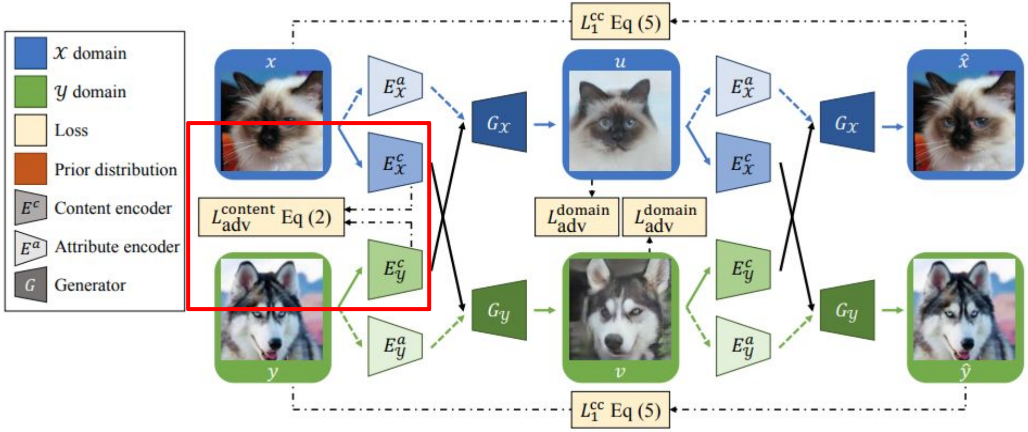
DRIT



$$\min_{G,E^c,E^a} \max_{D,D^c} \quad \lambda_{adv}^{content} L_{adv}^{c} + \lambda_1^{cc} L_1^{cc} + \lambda_{adv}^{domain} L_{adv}^{domain} + \lambda_1^{recon} L_1^{recon}$$
$$+ \lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL}$$

- **Content adversarial loss**
- **Cross-cycle consistency loss**
- **Domain adversarial loss**
- **Self-reconstruction loss**
- **KL loss**
- **Latent regression loss**

DRIT : Content adversarial loss

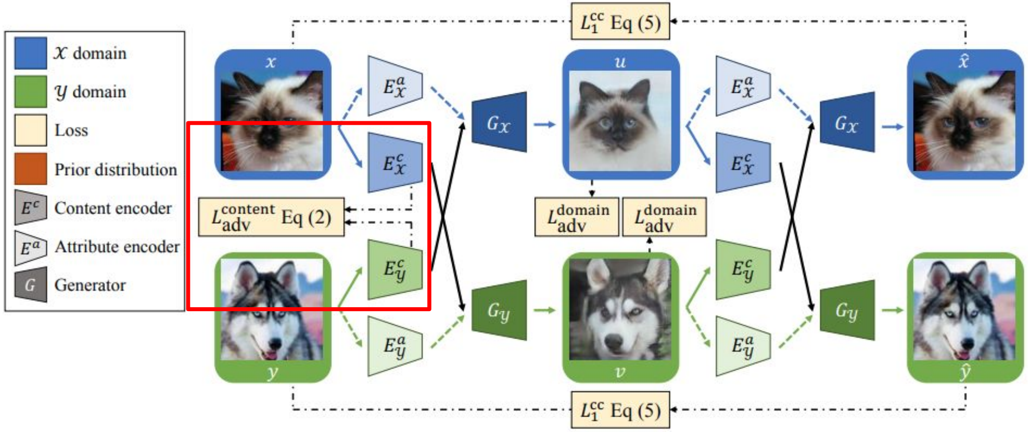


$$\min_{G,E^c,E^a} \max_{D,D^c} \boxed{\lambda^{content}_{adv} L^{c}_{adv}} + \lambda_1^{cc} L_1^{cc} + \lambda^{domain}_{adv} L^{domain}_{adv} + \lambda_1^{recon} L_1^{recon}$$
$$+ \lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL}$$

- **Content adversarial loss**
- **Cross-cycle consistency loss**
- **Domain adversarial loss**
- **Self-reconstruction loss**
- **KL loss**
- **Latent regression loss**

"Disentangle Content and Attribute Representations"

DRIT : Content adversarial loss



- Content adversarial loss
- Cross-cycle consistency loss
- Domain adversarial loss
- Self-reconstruction loss
- KL loss
- Latent regression loss

$$\min_{G,E^c,E^a} \max_{D,D^c} \boxed{\lambda_{adv}^{content} L_{adv}^{c}} + \lambda_1^{cc} L_1^{cc} + \lambda_{adv}^{domain} L_{adv}^{domain} + \lambda_1^{recon} L_1^{recon} + \lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL}$$

$$L_{adv}^{content}(E_{\mathcal{X}}^c, E_{\mathcal{Y}}^c, D^c) = \mathbb{E}_x[\frac{1}{2}\log D^c(E_{\mathcal{X}}^c(x)) + \frac{1}{2}\log(1 - D^c(E_{\mathcal{X}}^c(x)))] + \mathbb{E}_y[\frac{1}{2}\log D^c(E_{\mathcal{Y}}^c(y)) + \frac{1}{2}\log(1 - D^c(E_{\mathcal{Y}}^c(y)))]$$

DRIT : Content adversarial loss

- **Content Discriminator**

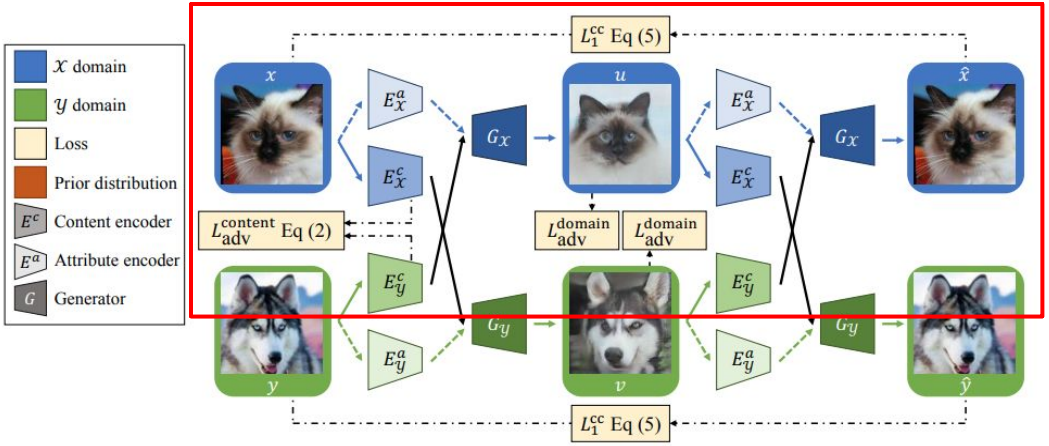
  aims to distinguish the domain of the encoded features

- **Content Encoder**

  learn to produce encoded content whose domain can't be distinguished
  by discriminator

$$L_{\mathrm{adv}}^{\mathrm{content}}(E_{\mathcal{X}}^c, E_{\mathcal{Y}}^c, D^c) = \mathbb{E}_x[\frac{1}{2}\log D^c(E_{\mathcal{X}}^c(x)) + \frac{1}{2}\log(1 - D^c(E_{\mathcal{X}}^c(x)))] + \mathbb{E}_y[\frac{1}{2}\log D^c(E_{\mathcal{Y}}^c(y)) + \frac{1}{2}\log(1 - D^c(E_{\mathcal{Y}}^c(y)))]$$

DRIT : Cross-cycle Consistency Loss

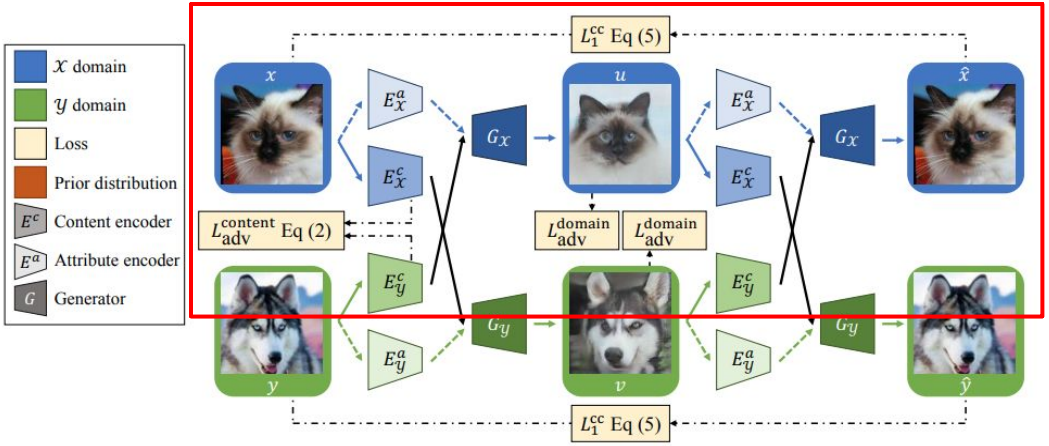


$$\min_{G,E^c,E^a}\max_{D,D^c} \quad \boxed{\lambda_1^{cc}L_1^{cc}} + \lambda_{adv}^{domain}L_{adv}^{domain} + \lambda_1^{recon}L_1^{recon}$$
$$+\lambda_1^{latent}L_1^{latent} + \lambda_{KL}L_{KL}$$

- Content adversarial loss
- Cross-cycle consistency loss
- Domain adversarial loss
- Self-reconstruction loss
- KL loss
- Latent regression loss

"Combining a content representation from an arbitrary image and an attribute representation from an image of target domain"

DRIT : Cross-cycle Consistency Loss



- Content adversarial loss
- Cross-cycle consistency loss
- Domain adversarial loss
- Self-reconstruction loss
- KL loss
- Latent regression loss

$$\min_{G,E^c,E^a}\ \max_{D,D^c} \quad \boxed{\lambda_1^{cc} L_1^{cc}} + \lambda_{adv}^{domain} L_{adv}^{domain} + \lambda_1^{recon} L_1^{recon}$$
$$+\lambda_1^{latent} L_1^{latent} + \lambda_{KL} L_{KL}$$

$$L_1^{cc}(G_{\mathcal{X}}, G_{\mathcal{Y}}, E^c_{\mathcal{X}}, E^c_{\mathcal{Y}}, E^a_{\mathcal{X}}, E^a_{\mathcal{Y}}) = \mathbb{E}_{x,y}[\|G_{\mathcal{X}}(E^c_{\mathcal{Y}}(v), E^a_{\mathcal{X}}(u)) - x\|_1$$
$$+\|G_{\mathcal{Y}}(E^c_{\mathcal{X}}(u), E^a_{\mathcal{Y}}(v)) - y\|_1],$$

where $u = G_{\mathcal{X}}(E^c_{\mathcal{Y}}(y)), E^a_{\mathcal{X}}(x))$ and $v = G_{\mathcal{Y}}(E^c_{\mathcal{X}}(x)), E^a_{\mathcal{Y}}(y))$.

DRIT : Cross-cycle Consistency Loss

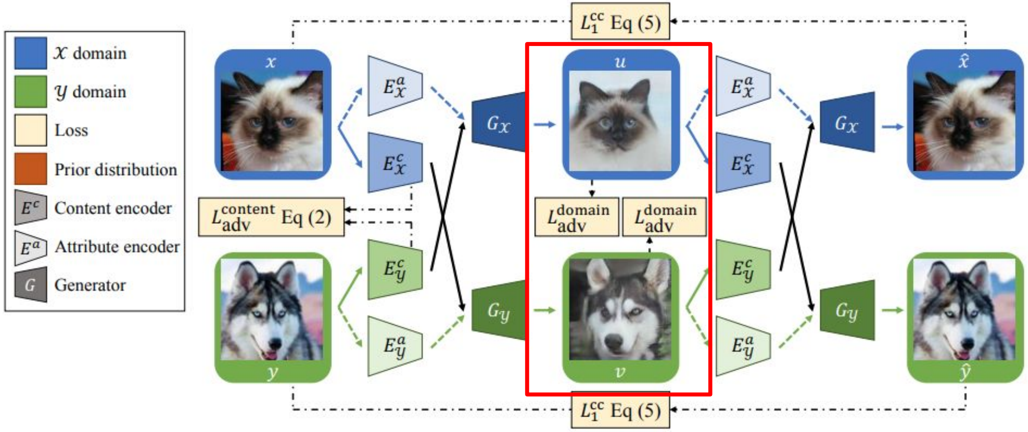
- **Forward Translation & Backward translation.**

  Exploit the disentangled content and attribute representation

$$L_1^{cc}(G_{\mathcal{X}}, G_{\mathcal{Y}}, E_{\mathcal{X}}^c, E_{\mathcal{Y}}^c, E_{\mathcal{X}}^a, E_{\mathcal{Y}}^a) = \mathbb{E}_{x,y}[\|G_{\mathcal{X}}(E_{\mathcal{Y}}^c(v), E_{\mathcal{X}}^a(u)) - x\|_1 + \|G_{\mathcal{Y}}(E_{\mathcal{X}}^c(u), E_{\mathcal{Y}}^a(v)) - y\|_1],$$

where $u = G_{\mathcal{X}}(E_{\mathcal{Y}}^c(y)), E_{\mathcal{X}}^a(x))$ and $v = G_{\mathcal{Y}}(E_{\mathcal{X}}^c(x)), E_{\mathcal{Y}}^a(y))$.

DRIT : Others



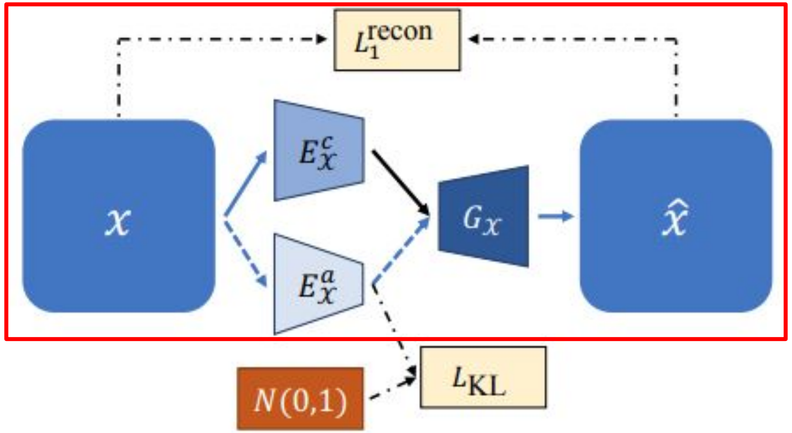
$$\min_{G,E^c,E^a}\ \max_{D,D^c} \quad + \boxed{\lambda^{\text{domain}}_{\text{adv}} L^{\text{domain}}_{\text{adv}}} + \lambda^{\text{recon}}_1 L^{\text{recon}}_1$$

$$+\lambda^{\text{latent}}_1 L^{\text{latent}}_1 + \lambda_{\text{KL}} L_{\text{KL}}$$

- **Content adversarial loss**
- **Cross-cycle consistency loss**
- **Domain adversarial loss**
- **Self-reconstruction loss**
- **KL loss**
- **Latent regression loss**

“Generator attempt to generate realistic images”

DRIT : Others



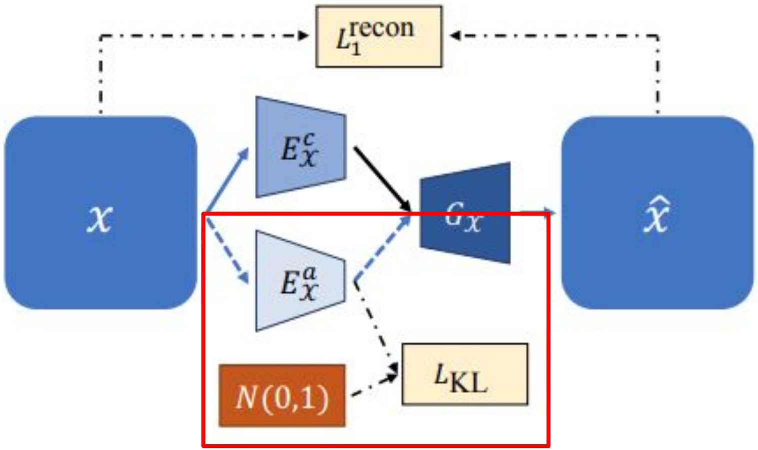
$$\min_{G,E^c,E^a} \max_{D,D^c} \qquad \boxed{\lambda_1^{\text{recon}} L_1^{\text{recon}}}$$

$$+\lambda_1^{\text{latent}} L_1^{\text{latent}} + \lambda_{\text{KL}} L_{\text{KL}}$$

- **Content adversarial loss**
- **Cross-cycle consistency loss**
- **Domain adversarial loss**
- **Self-reconstruction loss**
- **KL loss**
- **Latent regression loss**

"With encoded content/attribute features,
the decoders should decode them back to original inputs"

DRIT : Others



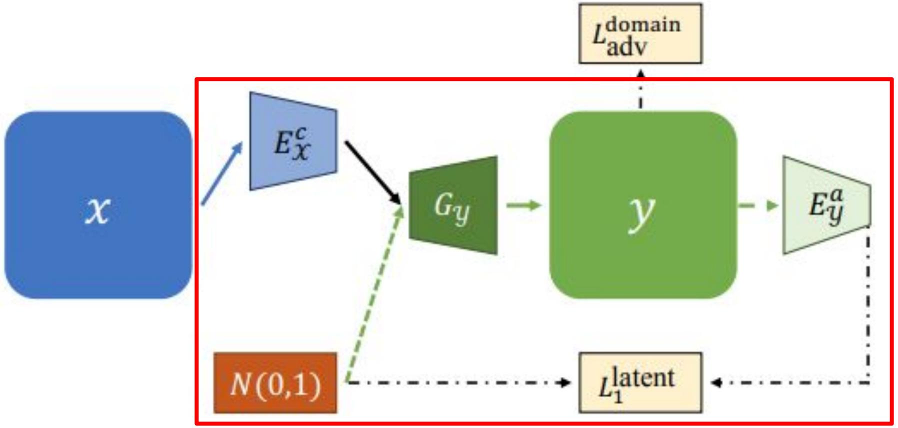
- Content adversarial loss
- Cross-cycle consistency loss
- Domain adversarial loss
- Self-reconstruction loss
- KL loss
- Latent regression loss

$$\min_{G,E^c,E^a} \max_{D,D^c}$$

$$\lambda_1^{\text{latent}} L_1^{\text{latent}} + \boxed{\lambda_{\text{KL}} L_{\text{KL}}}$$

"In order to perform stochastic sampling at test time, we encourage the attribute to be as close to a prior Gaussian distribution."

DRIT : Others



- Content adversarial loss
- Cross-cycle consistency loss
- Domain adversarial loss
- Self-reconstruction loss
- KL loss
- Latent regression loss

$$\min_{G,E^c,E^a} \max_{D,D^c}$$

$$\boxed{\lambda_1^{\text{latent}} L_1^{\text{latent}}}$$

“To encourage invertible mapping between the image and the latent space”

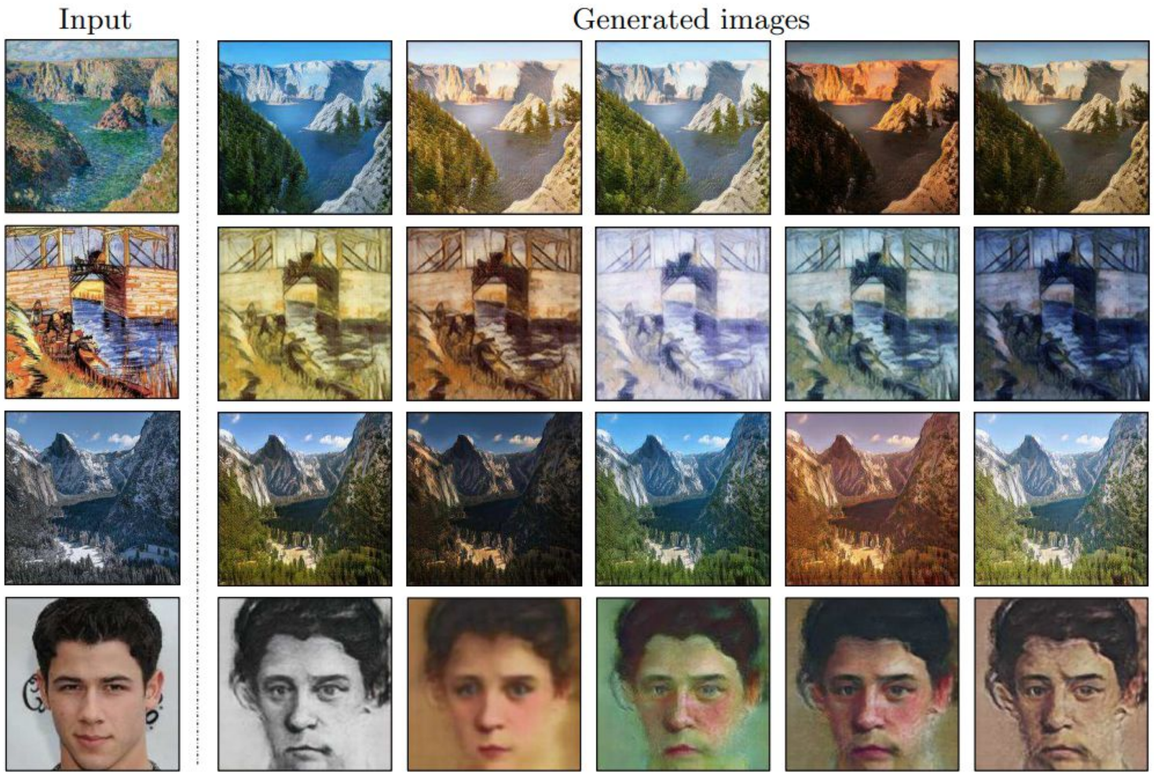
DRIT : Results

Monet → Photo

Van Gogh → Monet

Winter → Summer

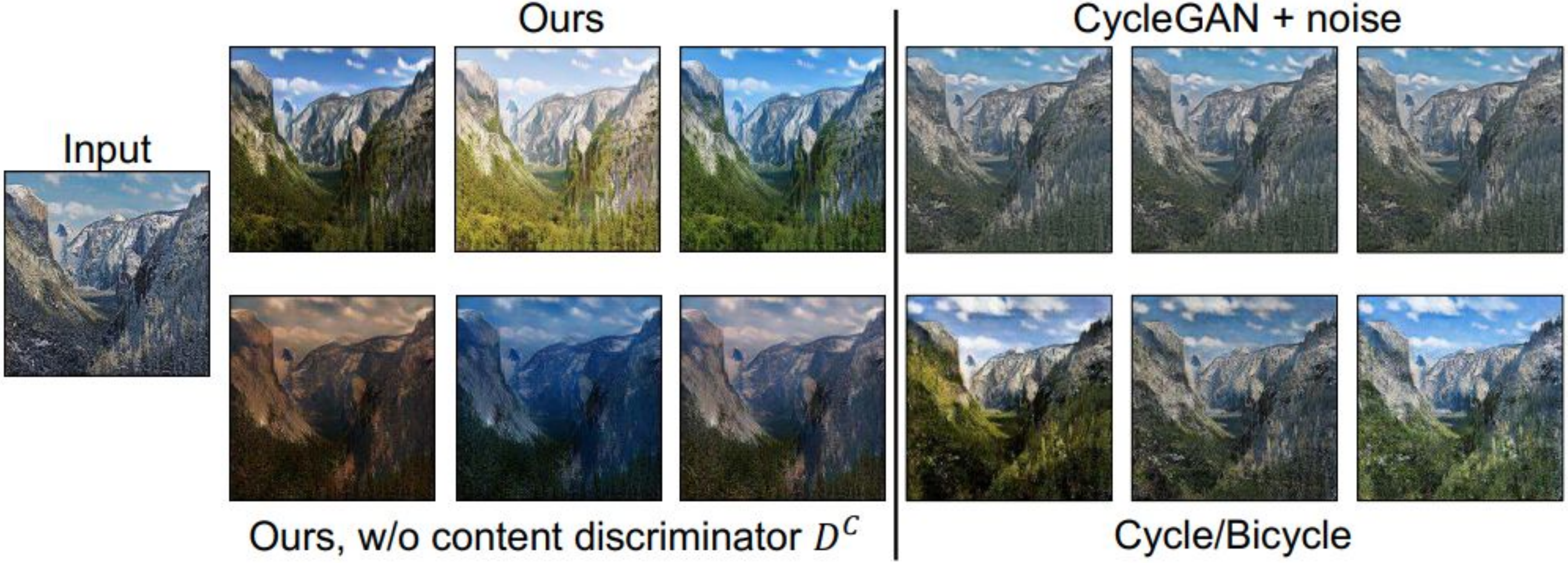
Photo → Portrait



Experiments

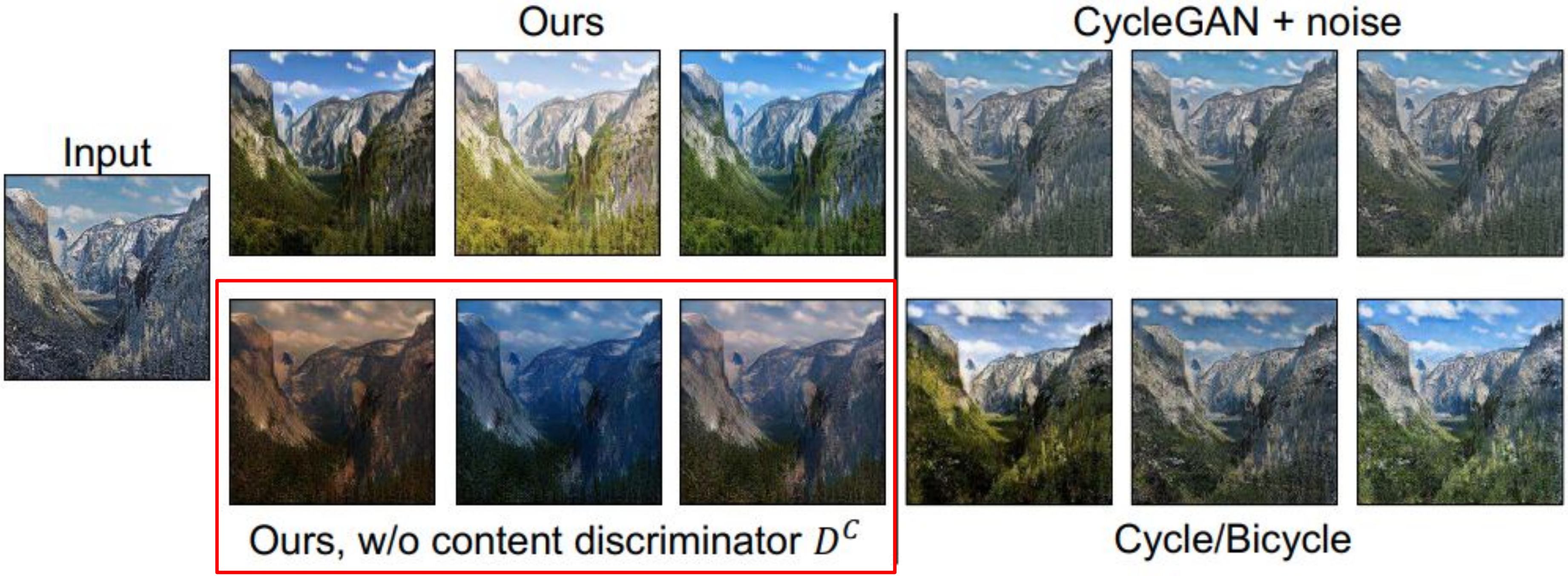
- Diversity
- Unpaired
- Disentangle: Content, Attribute

Qualitative: diversity



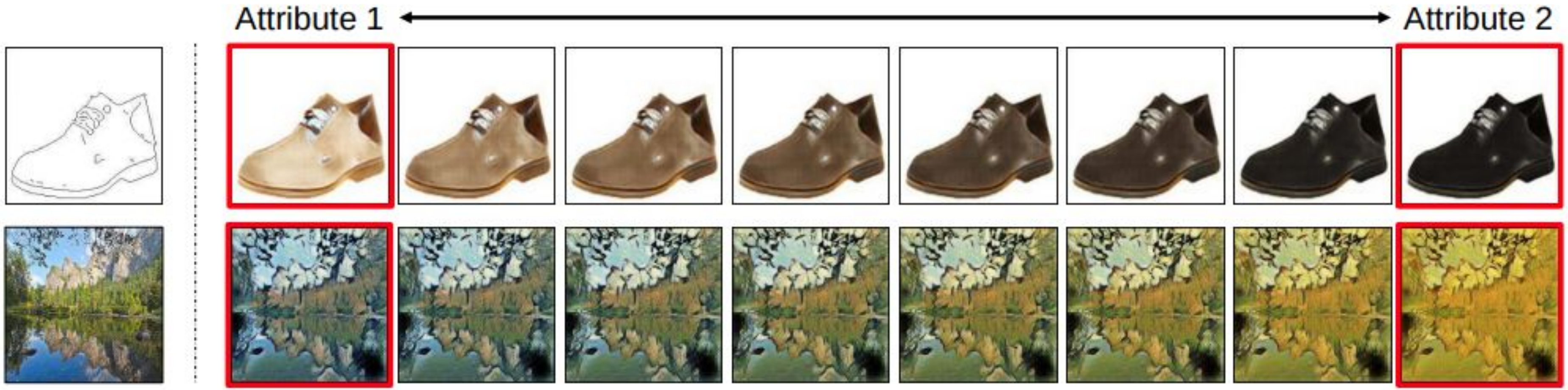
“Winter → Summer”

Qualitative: diversity



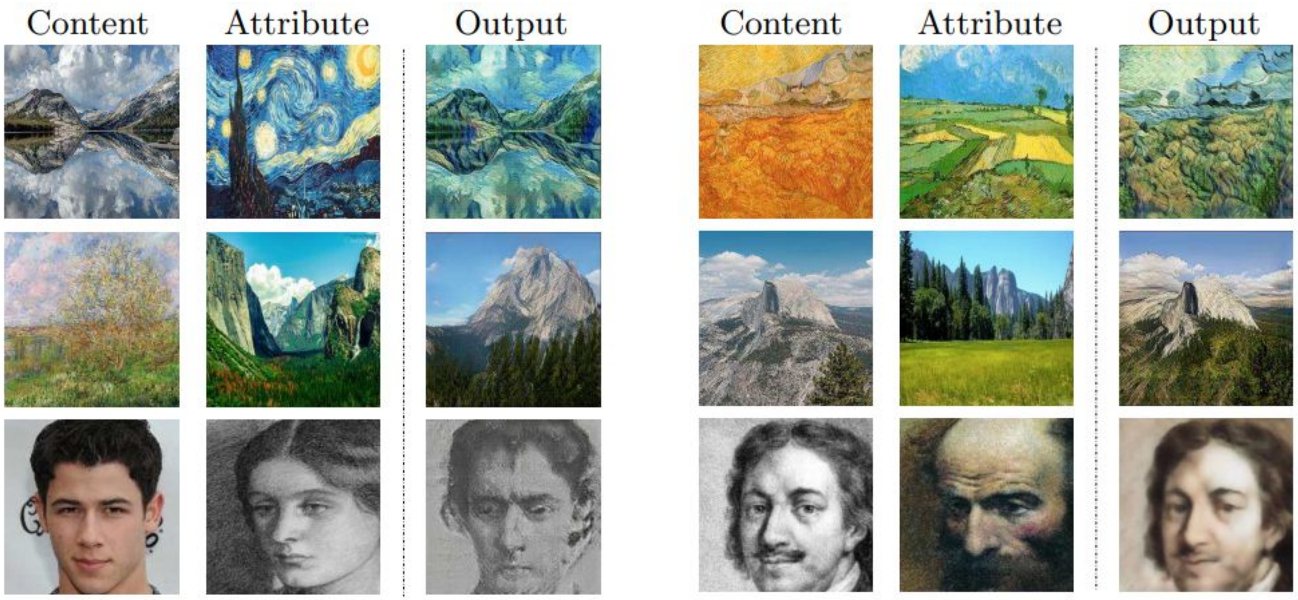
Without the content discriminator,
model fails to capture domain-related details (e.g., the color of tree and sky)

Qualitative: attribute



Translation results with linear-interpolated attribute vectors between attributes

Qualitative: disentangle

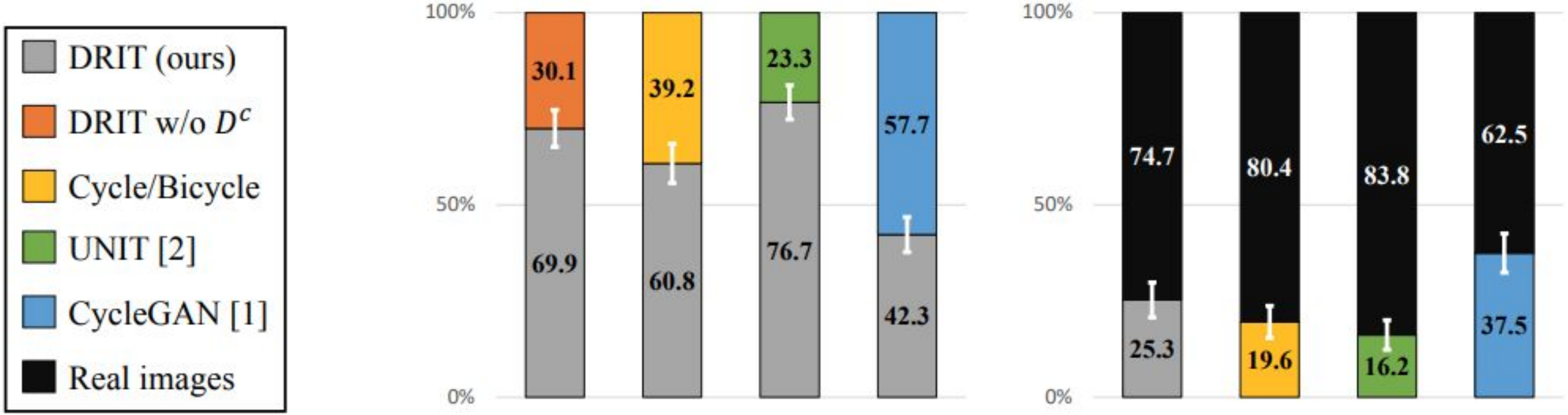


(a) Inter-domain attribute transfer

(b) Intra-domain attribute transfer

Translation results with linear-interpolated attribute vectors between attributes

Quantitative: realism preference results



Quantitative: diversity, reconstruction err

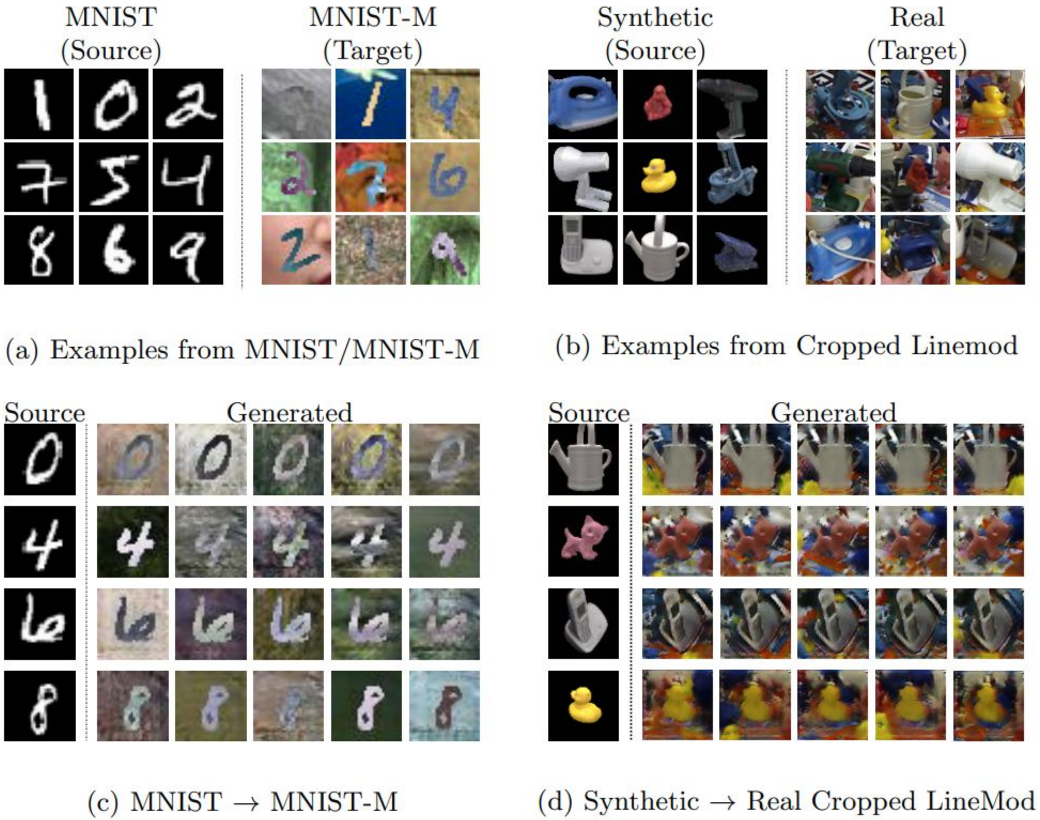
Table 2: **Diversity.** We use the LPIPS metric [45] to measure the diversity of generated images on the Yosemite dataset.

Method	Diversity
real images	$.448 \pm .012$
DRIT	$\mathbf{.424} \pm .010$
DRIT w/o D^c	$.410 \pm .016$
UNIT [26]	$.406 \pm .022$
CycleGAN [46]	$\underline{.413} \pm .008$
Cycle/Bicycle	$.399 \pm .009$

Table 3: **Reconstruct error**. We use the edge-to-shoes dataset to measure the quality of our attribute encoding. The reconstruction error is $\|y - G_{\mathcal{Y}}(E^c_{\mathcal{X}}(x), E^a_{\mathcal{Y}}(y))\|_1$. * BicycleGAN uses *paired* data for training.

Method	Reconstruct error
BicycleGAN [47]*	**0.0945**
DRIT	<u>0.1347</u>
DRIT, w/o D^c	0.2076

Domain adaptation



(a) Examples from MNIST/MNIST-M

(b) Examples from Cropped Linemod

(c) MNIST → MNIST-M

(d) Synthetic → Real Cropped LineMod

(a) MNIST-M

Model	Classification Accuracy (%)
Source-only	56.6
CycleGAN [46]	74.5
Ours, ×1	86.93
Ours, ×3	90.21
Ours, ×5	**91.54**
DANN [13]	77.4
DSN [4]	83.2
PixelDA [3]	**95.9**
Target-only	96.5

(b) Cropped LineMod

Model	Classification Accuracy (%)	Mean Angle Error (°)
Source-only	42.9 (47.33)	73.7 (89.2)
CycleGAN [46]	68.18	47.45
Ours, ×1	95.91	42.06
Ours, ×3	97.04	37.35
Ours, ×5	**98.12**	**34.4**
DANN [13]	99.9	56.58
DSN [4]	**100**	53.27
PixelDA [3]	99.98	**23.5**
Target-only	100	12.3 (6.47)

Thank You.