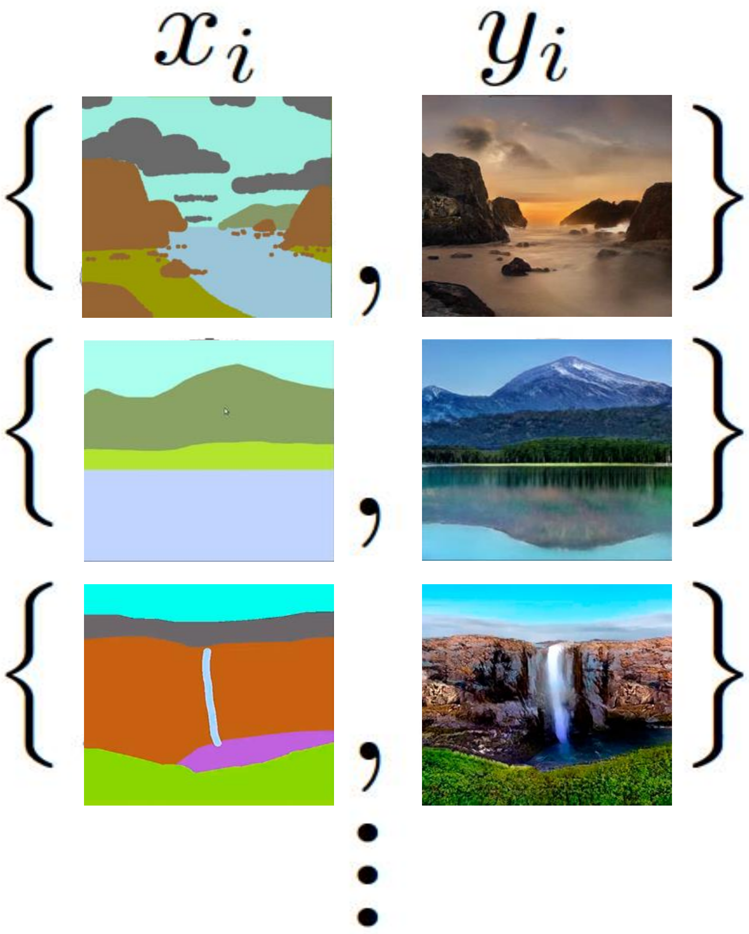
Multimodal Unsupervised Image-to-Image Translation

Ming-Yu Liu

NVIDIA

Supervised vs Unsupervised

Supervised/Paired/Aligned/Registered



Unsupervised/Unpaired/Unaligned/Unregistered

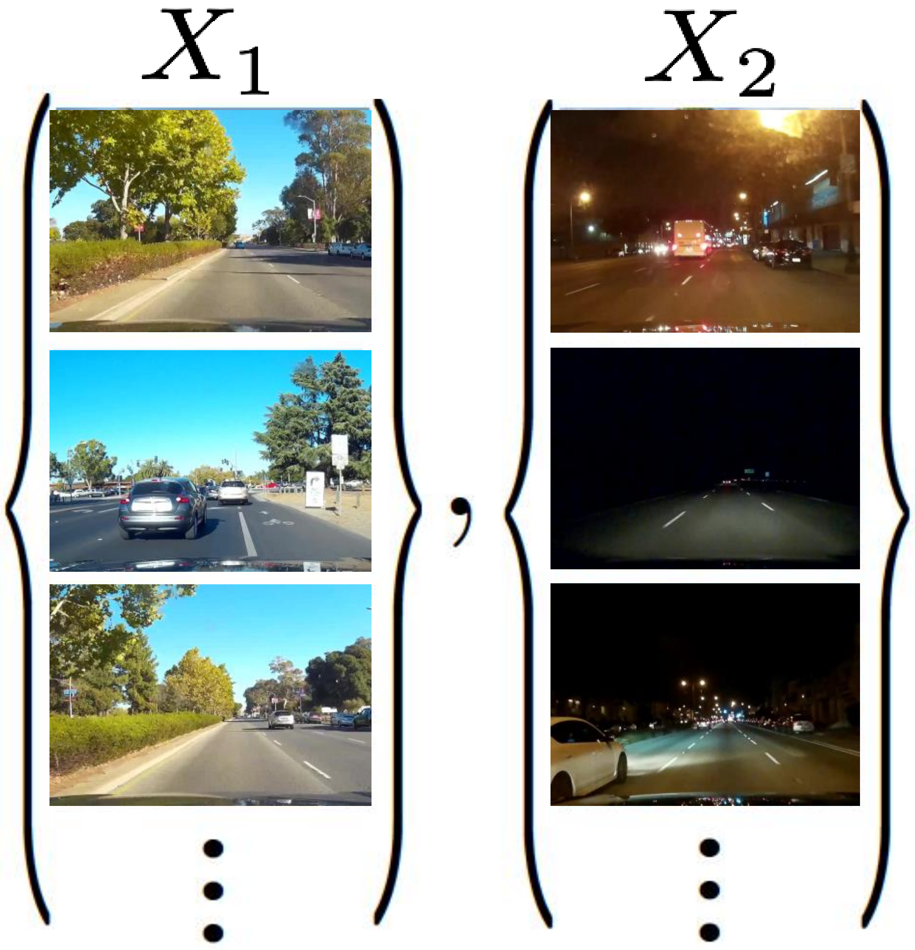


Image Domain Transfer

Given an input image
in one domain

Summer image domain

Image
Translator

F

Output a corresponding image
in a differerent domain

Winter image domain

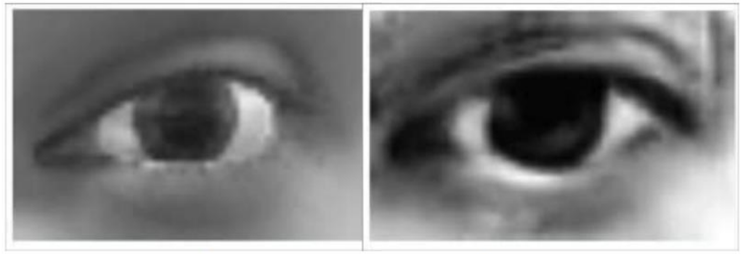
Example Applications


Low-res to high-res


Blurry to sharp


Image to painting

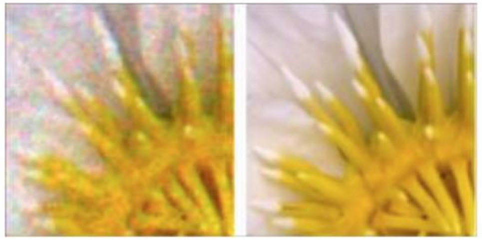

LDR to HDR


Synthetic to real

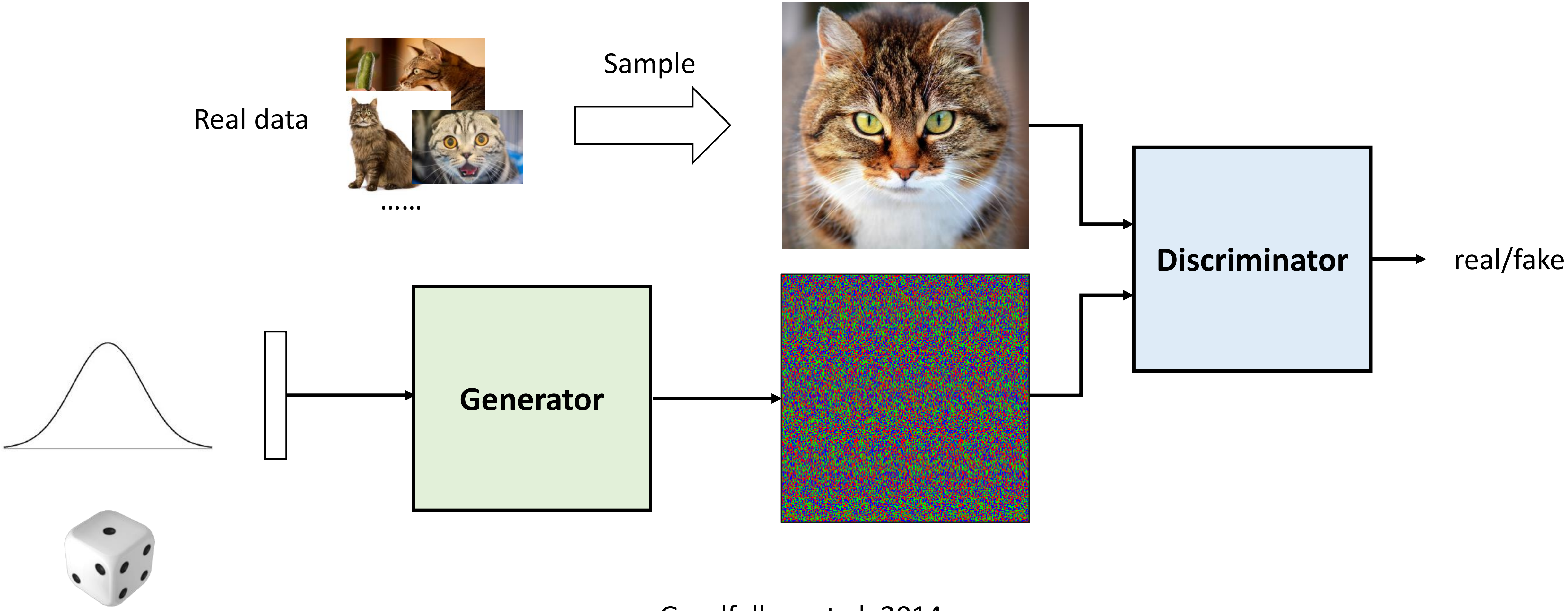

Thermal to color


Day to night


Summer to winter

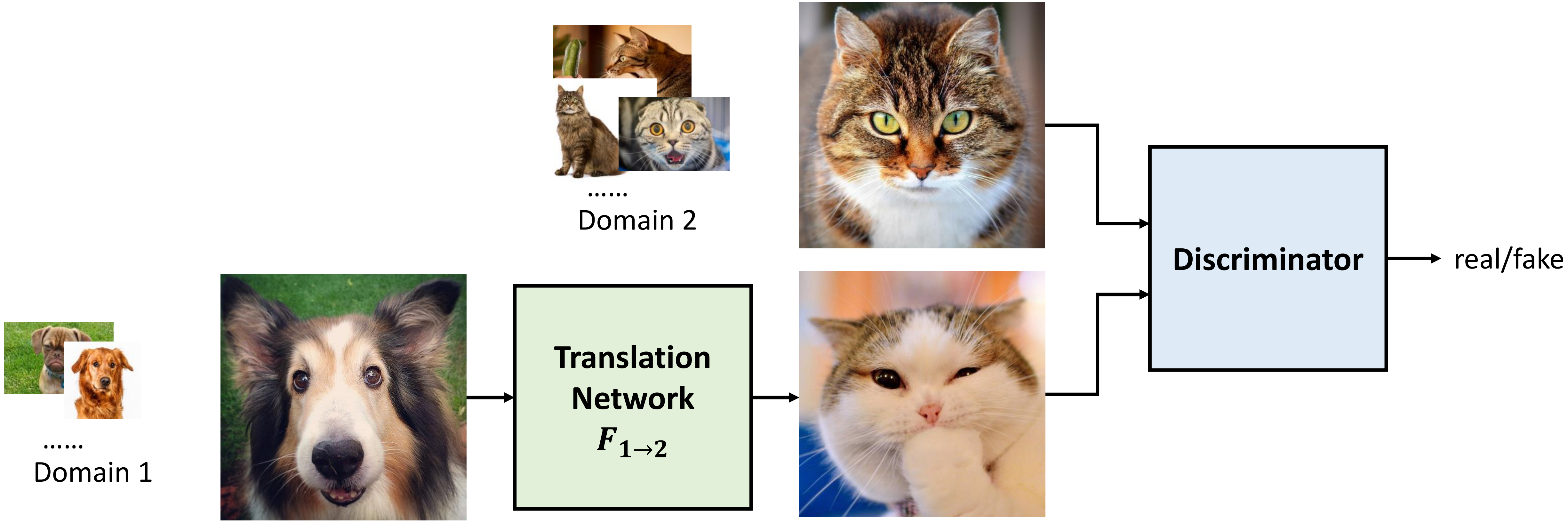

Noisy to clean

Generative Adversarial Networks (GANs)



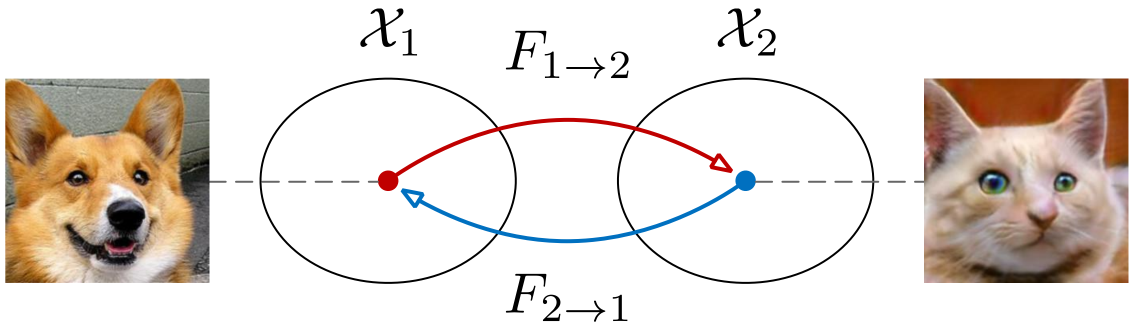
Goodfellow et al. 2014

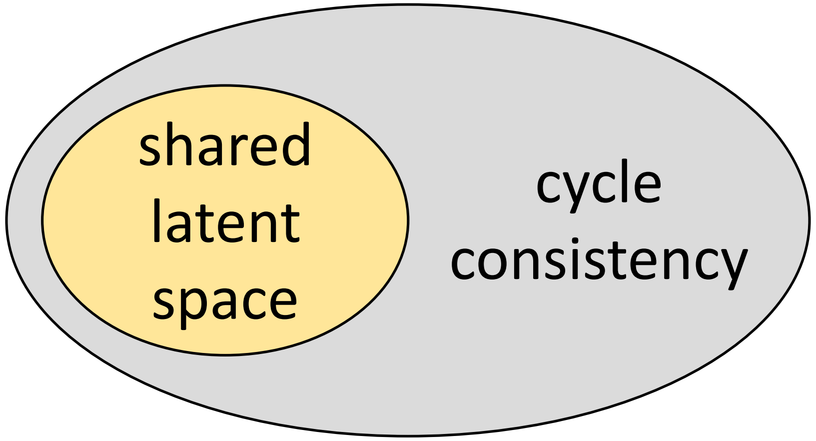
Plain GAN for Unsupervised Image-to-Image Translation



CycleGAN and UNIT

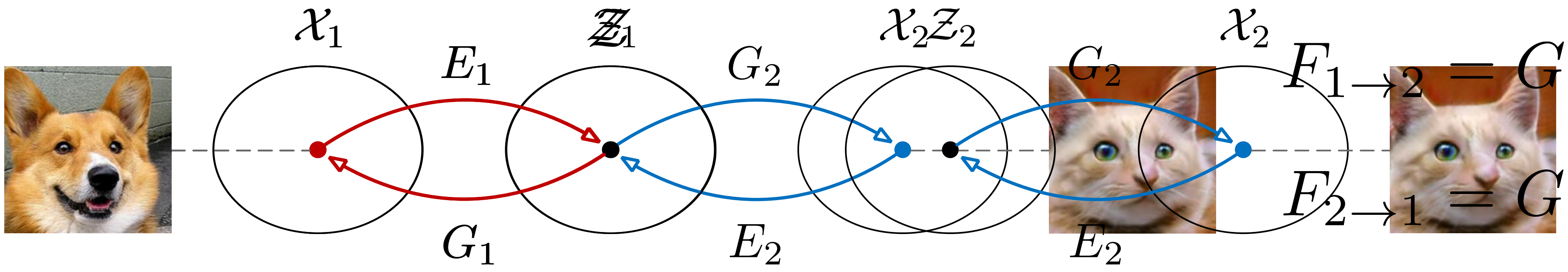
- CycleGAN (**cycle consistency**) [Zhu et al. 2017]





- UNIT (**shared latent space**) [Liu et al. 2017]

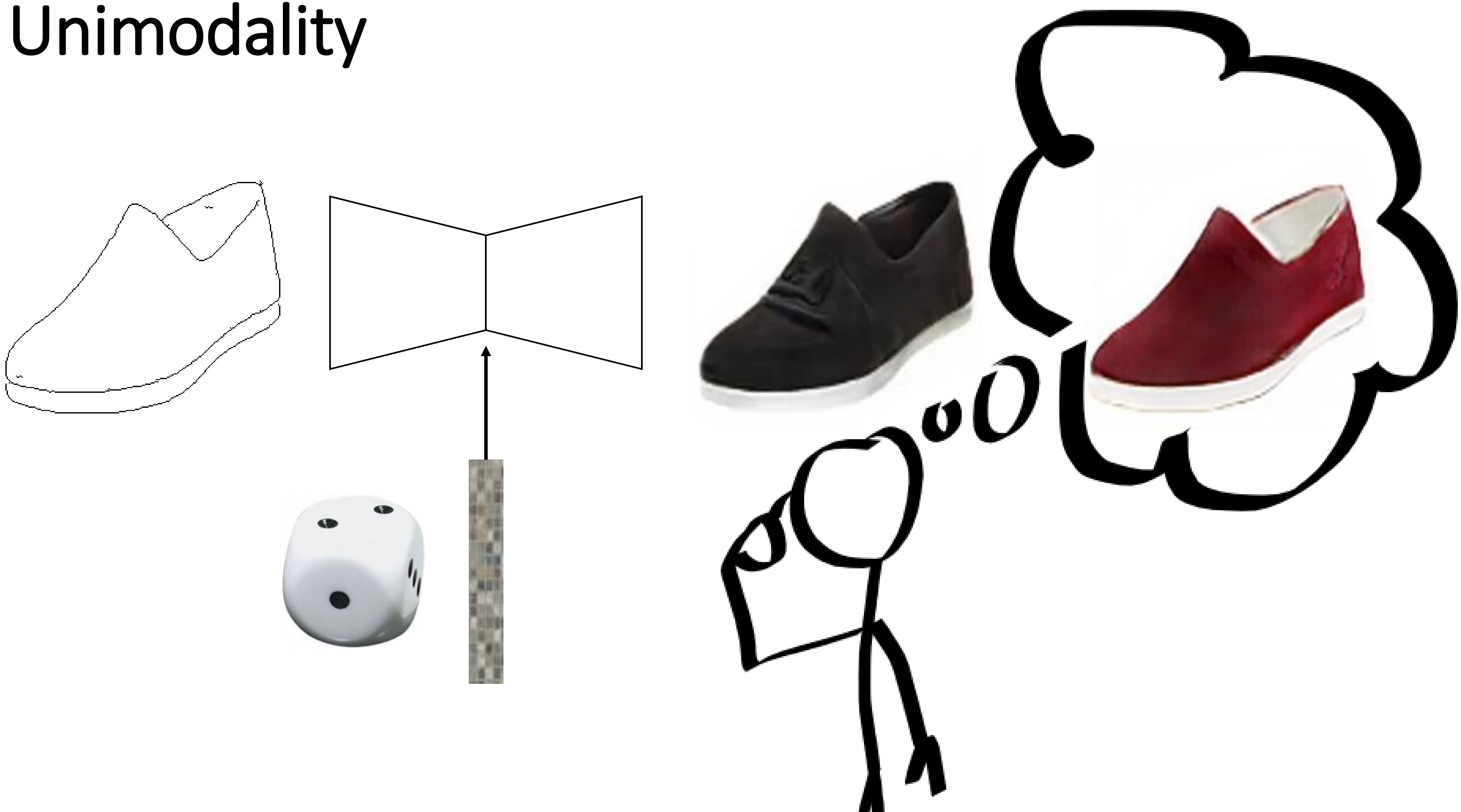
shared latent space $\Longrightarrow$ cycle consistency



$$F_{1\rightarrow 2} = G_2 \circ E_1$$

$$F_{2\rightarrow 1} = G_1 \circ E_2$$

Unimodality



Towards Multimodality

...

Shared latent space does not allow multimodality

Cycle consistency does not allow multimodality

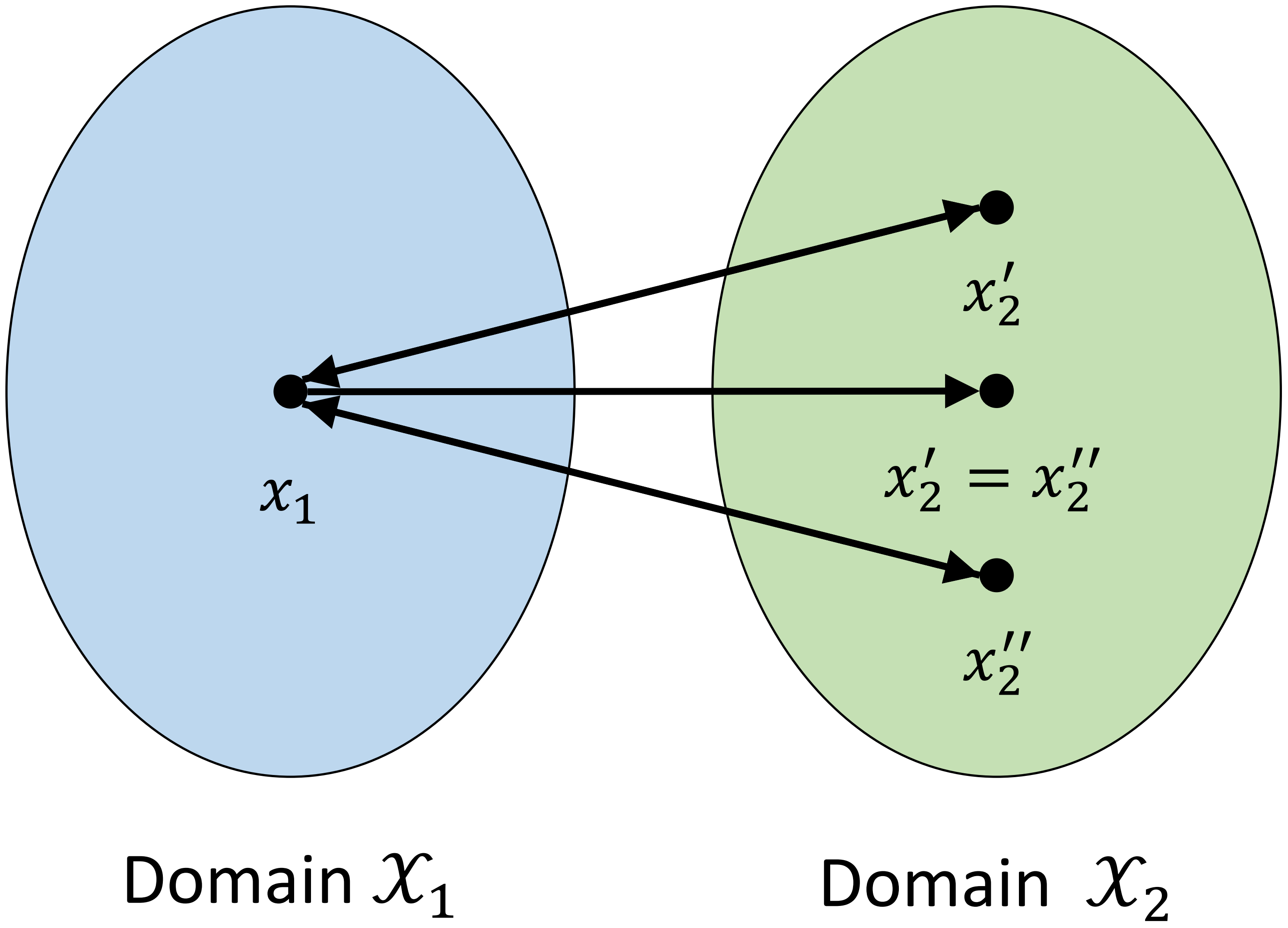
Cycle consistency:

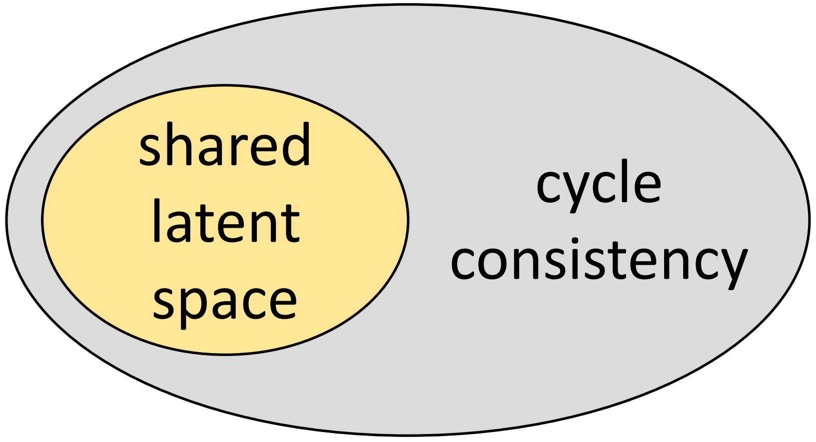
$\mathcal{X}_1 \to \mathcal{X}_2 \to \mathcal{X}_1$

$\mathcal{X}_2 \to \mathcal{X}_1 \to \mathcal{X}_2$

$p(x_2|x_1) = \delta(x_2 - x_2')$

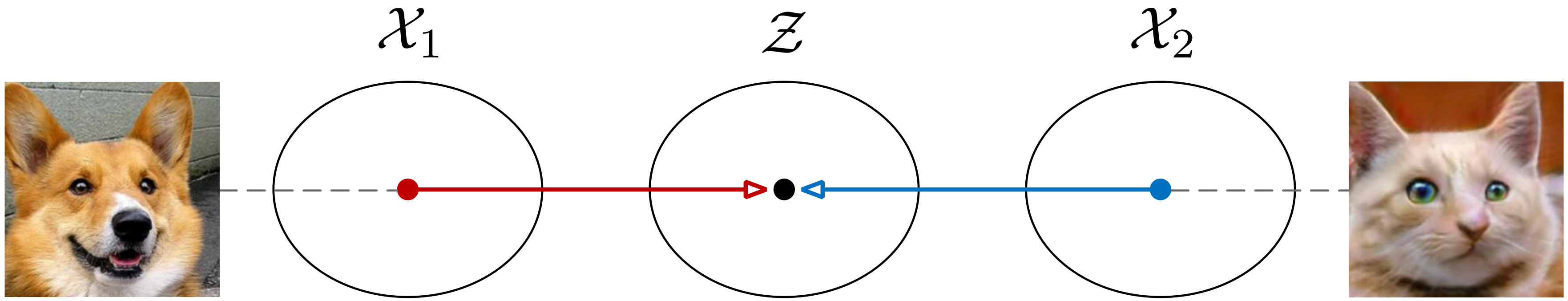
$\delta(x_2 - x_2'')$





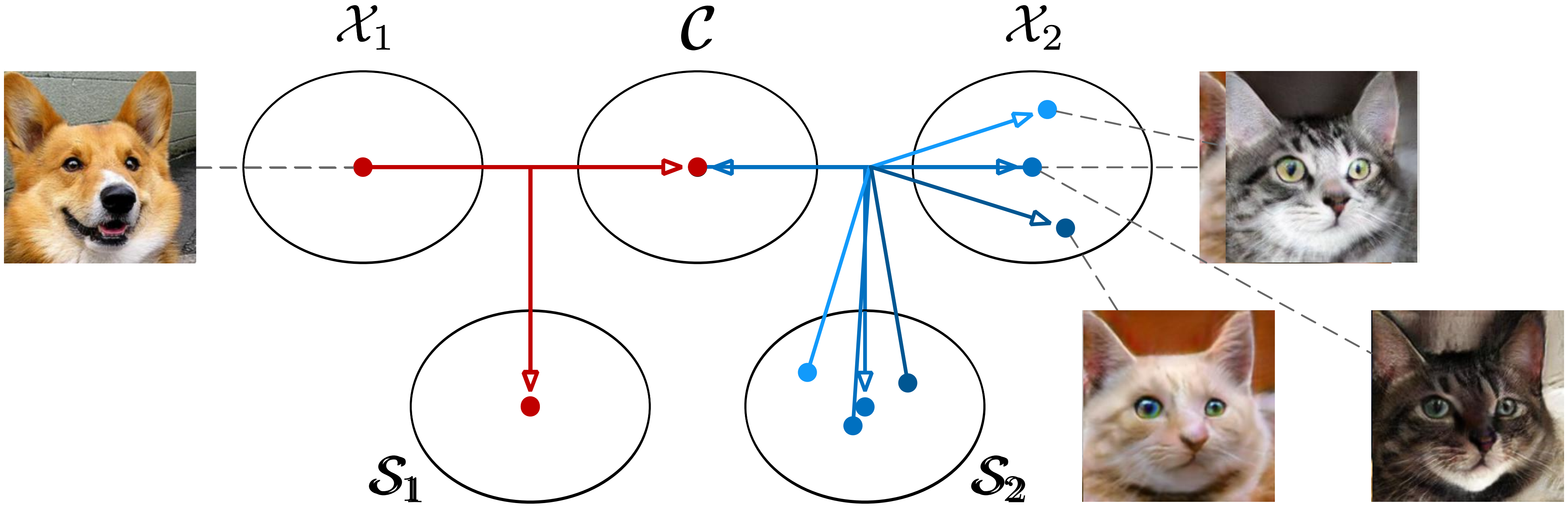
Disentangling the Latent Space

- UNIT
  - A single **shared**, **domain-invariant** latent space $\mathcal{Z}$



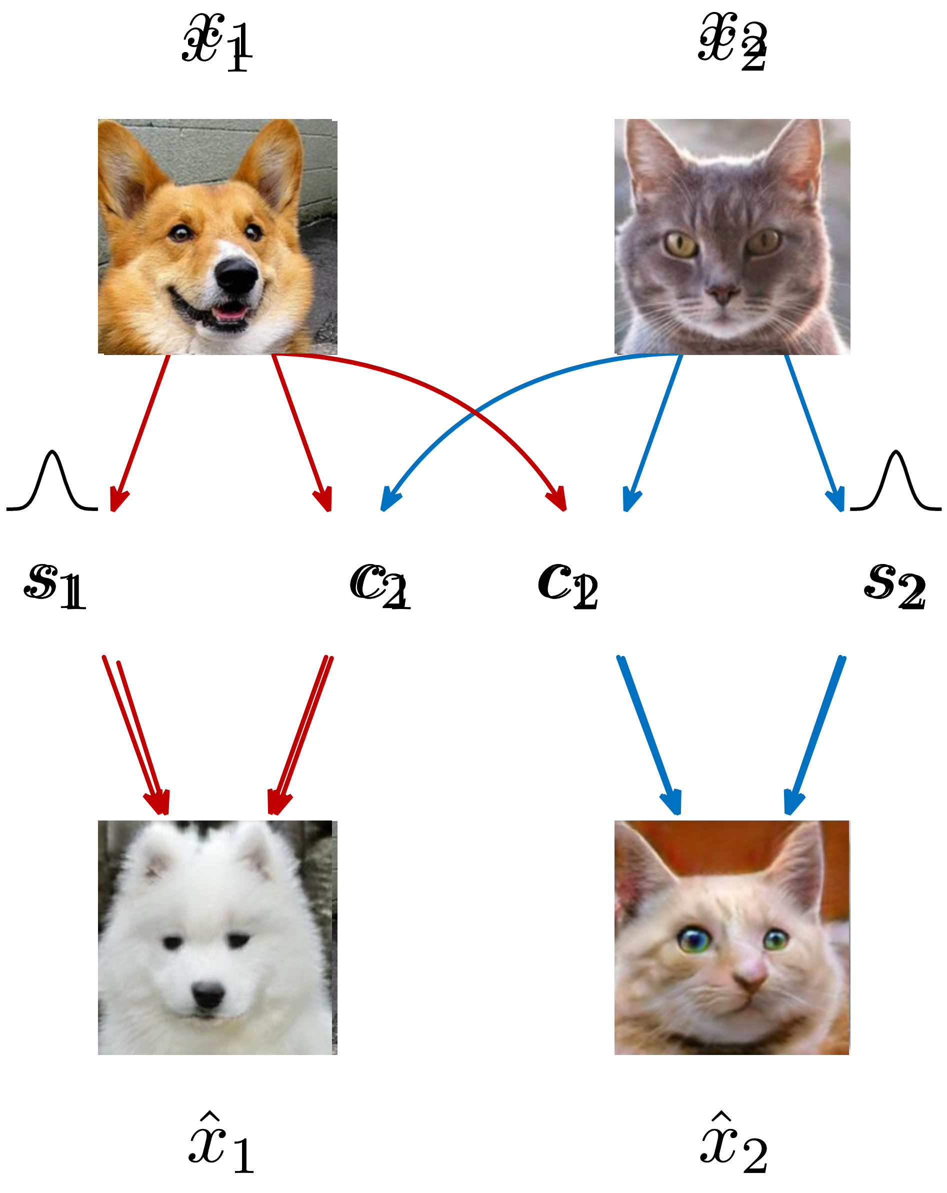
Disentangling the Latent Space

- Multimodal UNIT (MUNIT)
  - A **content** space $\mathcal{C}$ that is **shared, domain-invariant**
  - Two **style** spaces $\mathcal{S}_1, \mathcal{S}_2$ that are **unshared, domain-specific**



Training

- Notations:
  - x: images
  - c: content
  - s: style
- Loss:
  - Bidirectional reconstruction loss
    - Image reconstruction loss
    - Latent reconstruction loss
  - GAN loss



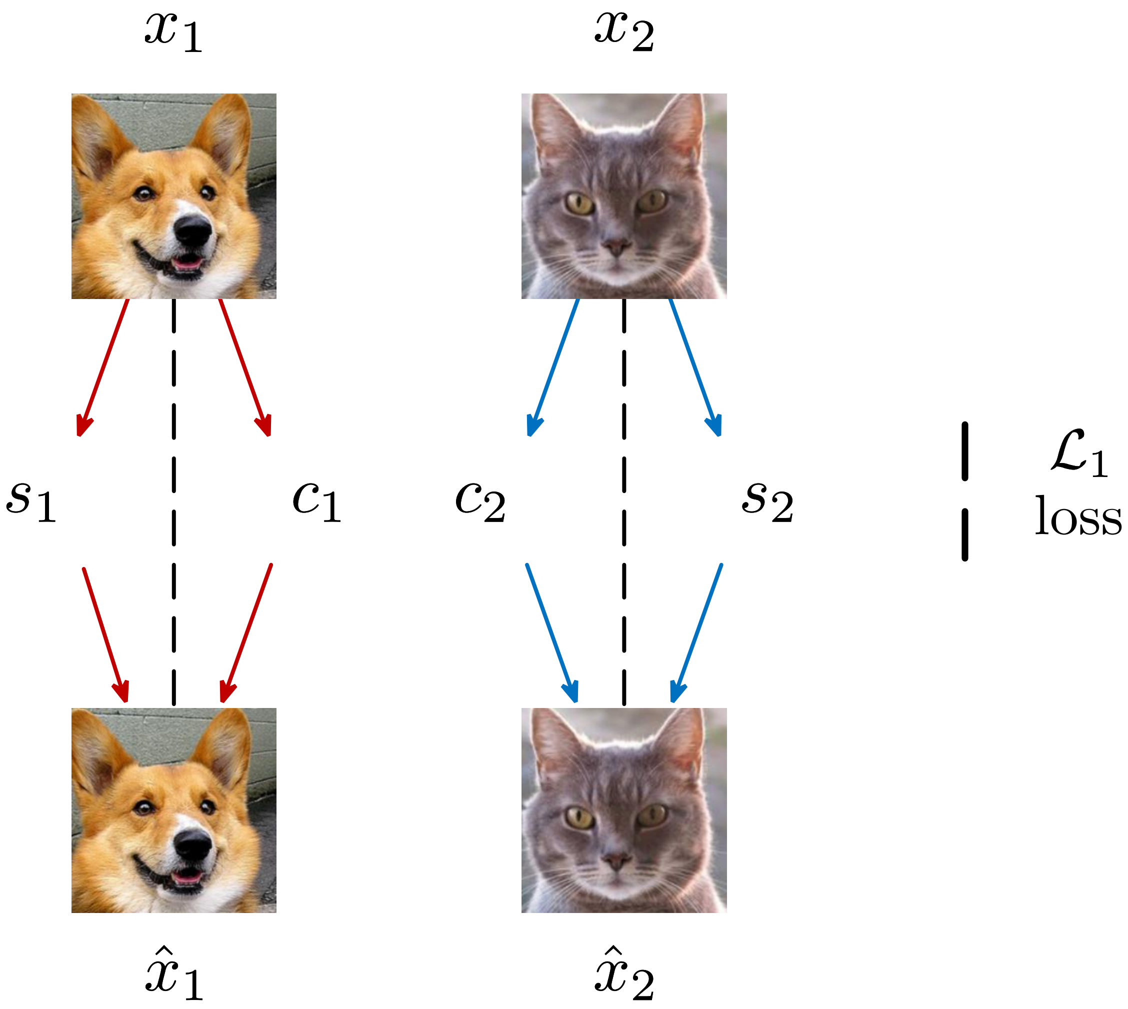
Within-domain reconstruction

Cross-domain translation

Bidirectional Reconstruction Loss: Image Reconstruction

Notations:

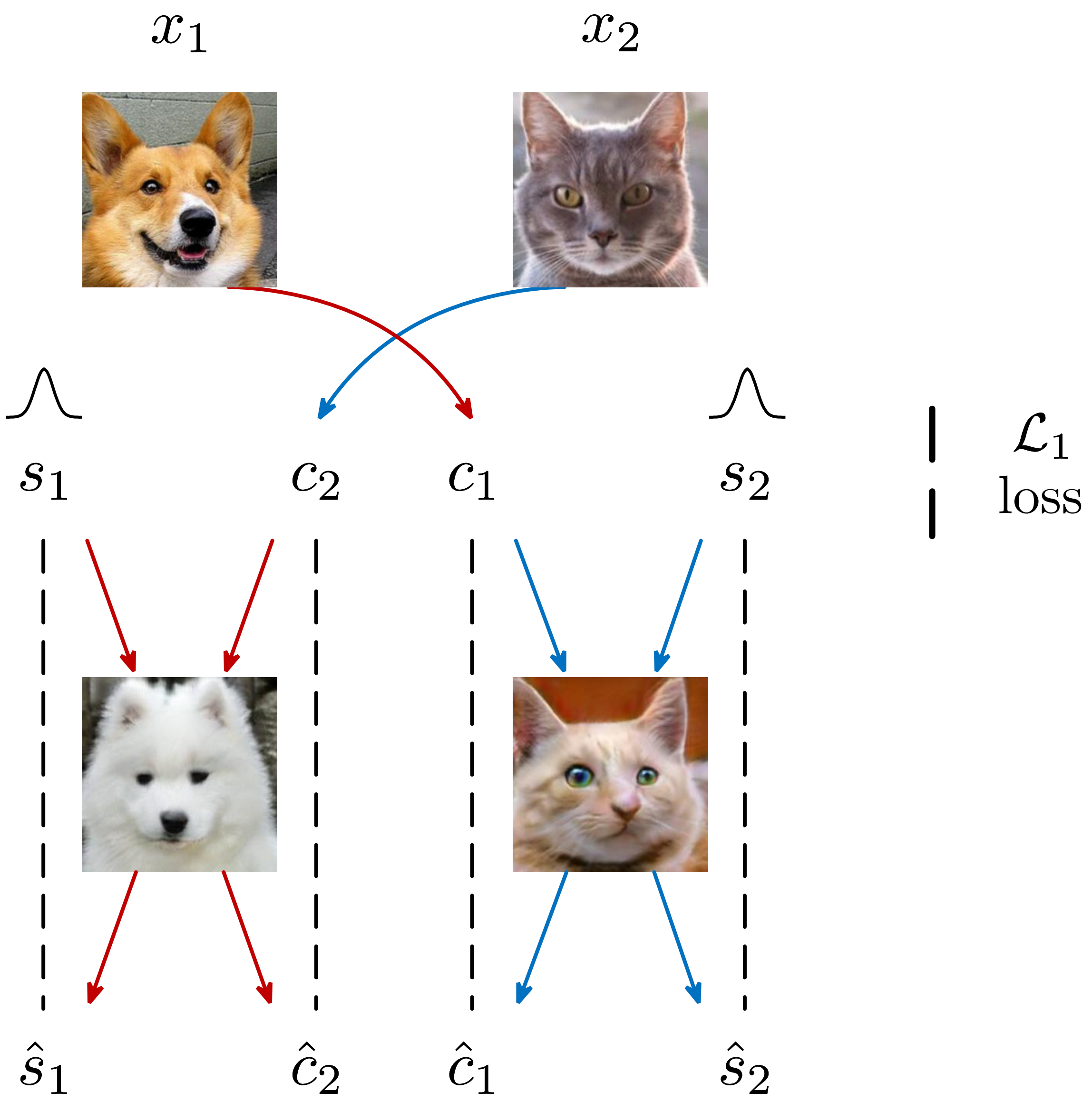
- x: images
- c: content
- s: style



Bidirectional Reconstruction Loss: Latent Reconstruction

Notations:

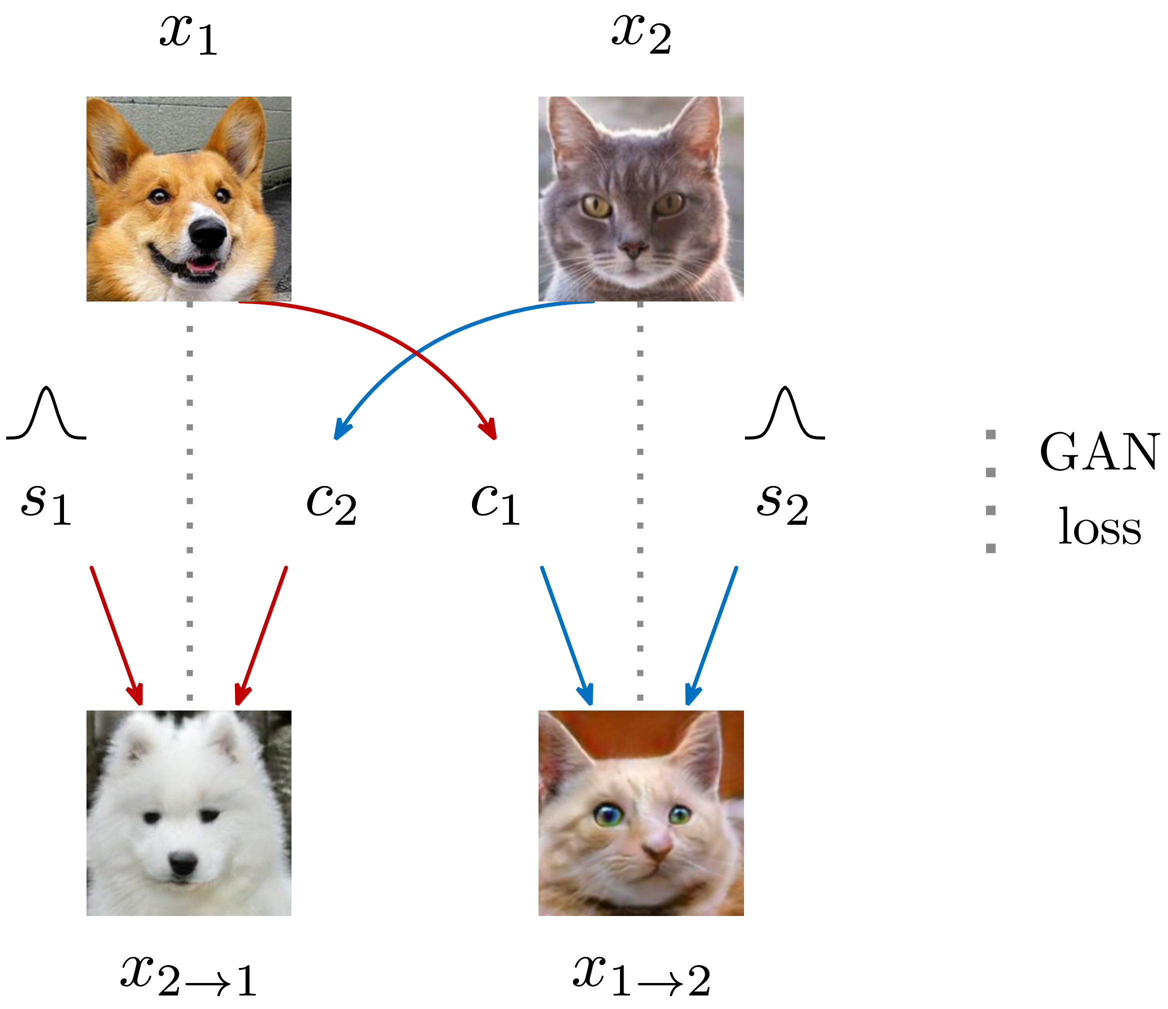
- x: images
- c: content
- s: style



GAN Loss

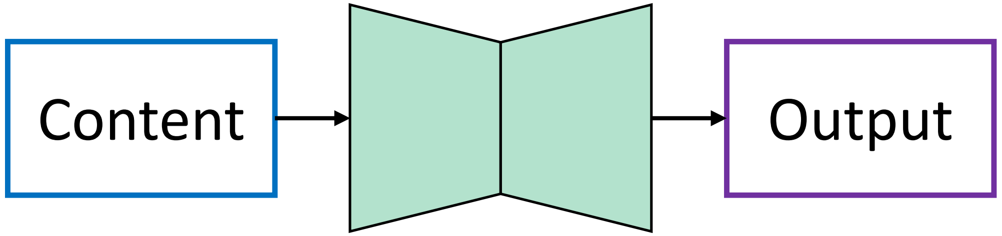
Notations:

- x: images
- c: content
- s: style

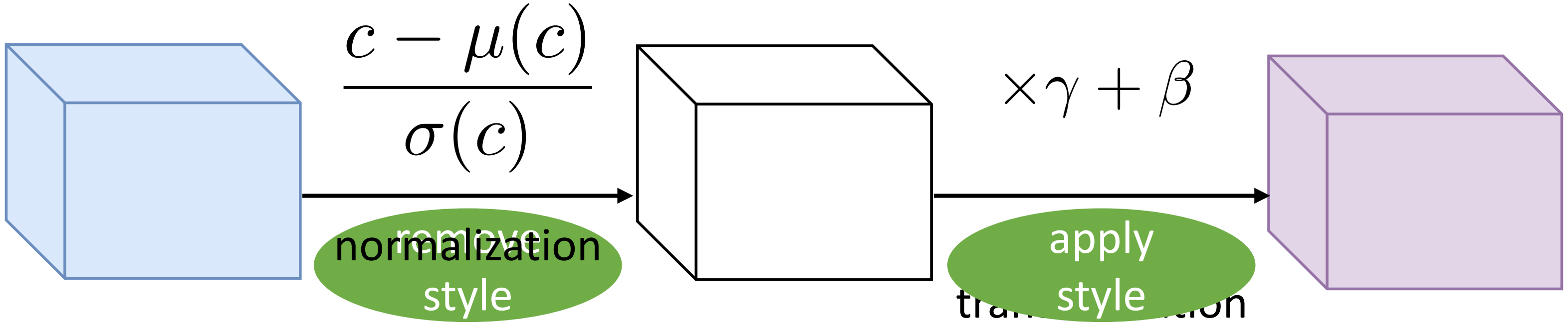


Background: Instance Normalization (IN)

Feedforward transfer of a single style



Content feature: c



$$\mathrm{IN}(c) = \gamma\left(\frac{c-\mu(c)}{\sigma(c)}\right) + \beta$$

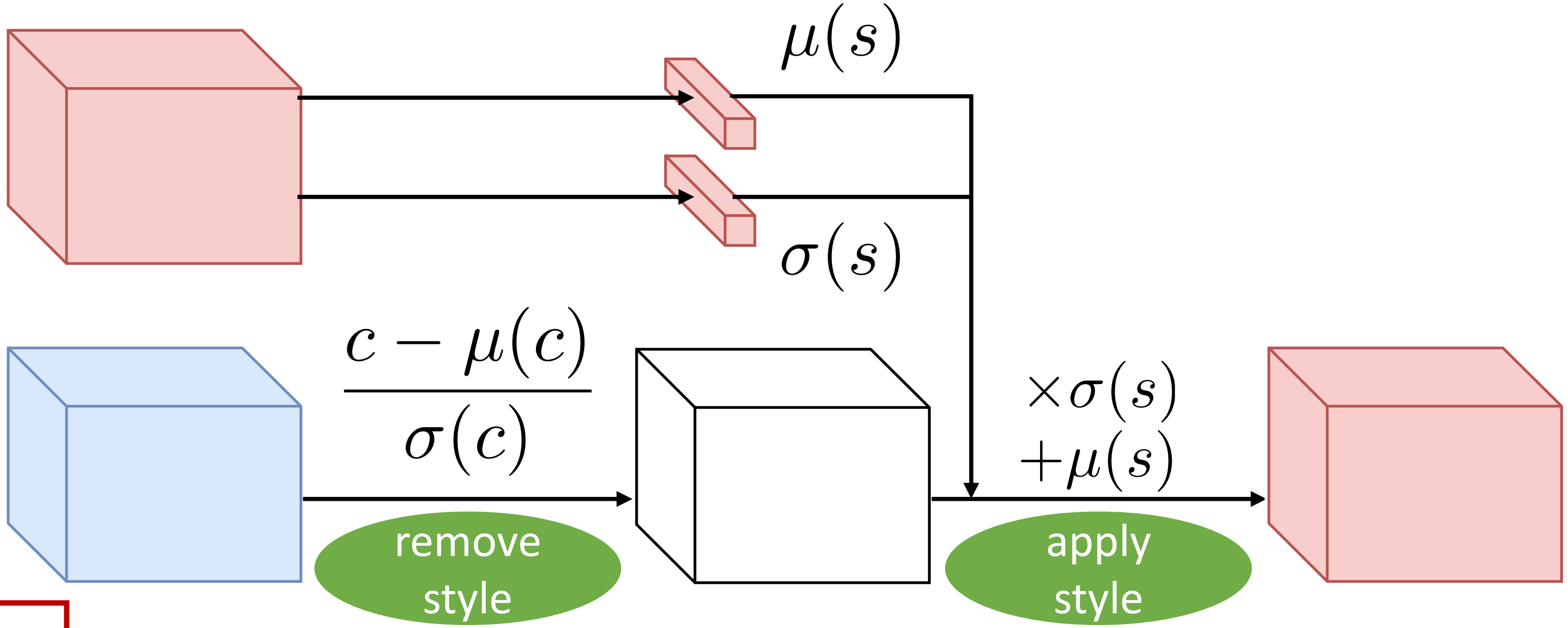
"Instance Normalization: The Missing Ingredient for Fast Stylization", Ulyanov et al. 2017

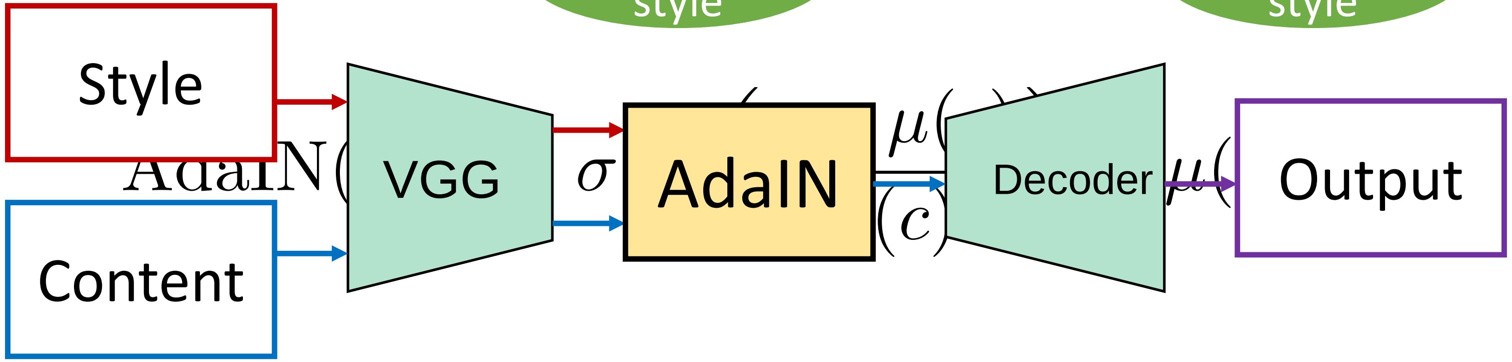
Adaptive Instance Normalization (AdaIN)

Feedforward transfer of **arbitrary** styles

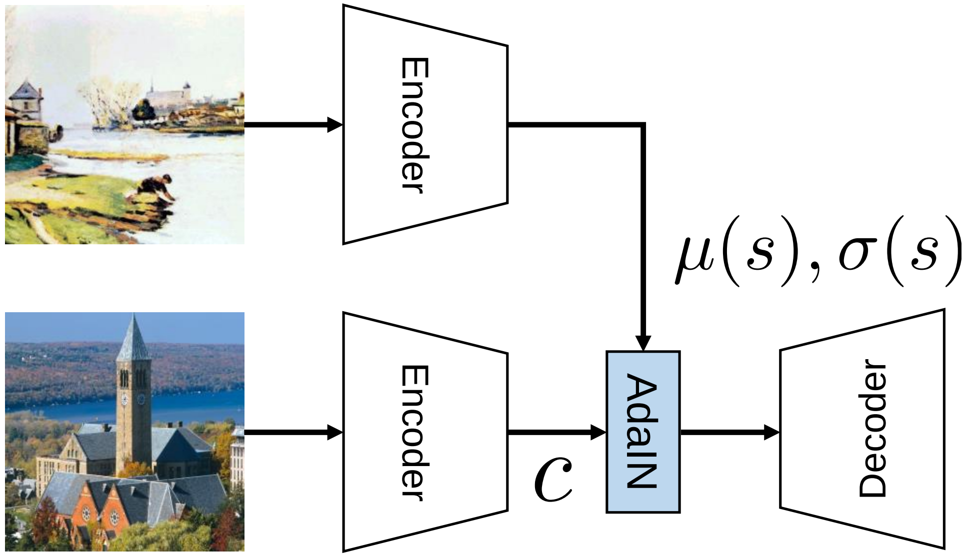
Style feature: s

Content feature: c



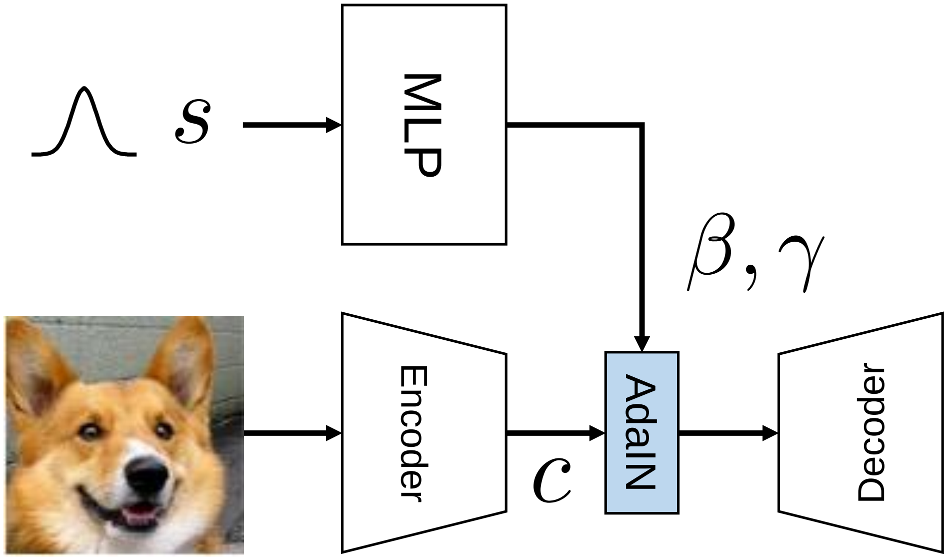


AdaIN in a Generative Network



$$\mathrm{AdaIN}(c, s) = \sigma(s)\left(\frac{c - \mu(c)}{\sigma(c)}\right) + \mu(s)$$

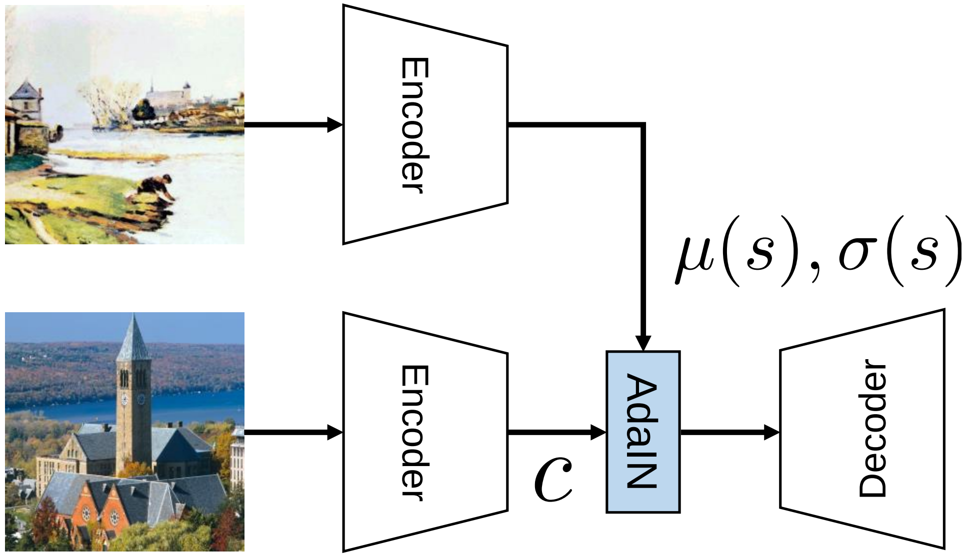
AdaIN in style transfer



$$\mathrm{AdaIN}(c, s) = \gamma\left(\frac{c - \mu(c)}{\sigma(c)}\right) + \beta$$

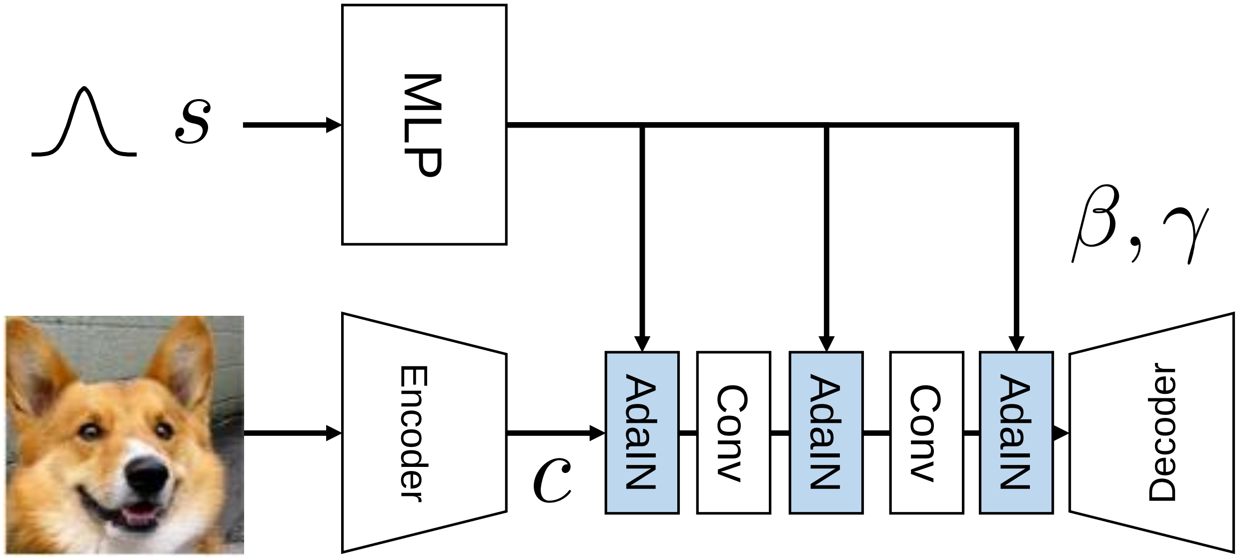
AdaIN in a generative network

AdaIN in a Generative Network



$$\mathrm{AdaIN}(c, s) = \sigma(s)\left(\frac{c - \mu(c)}{\sigma(c)}\right) + \mu(s)$$

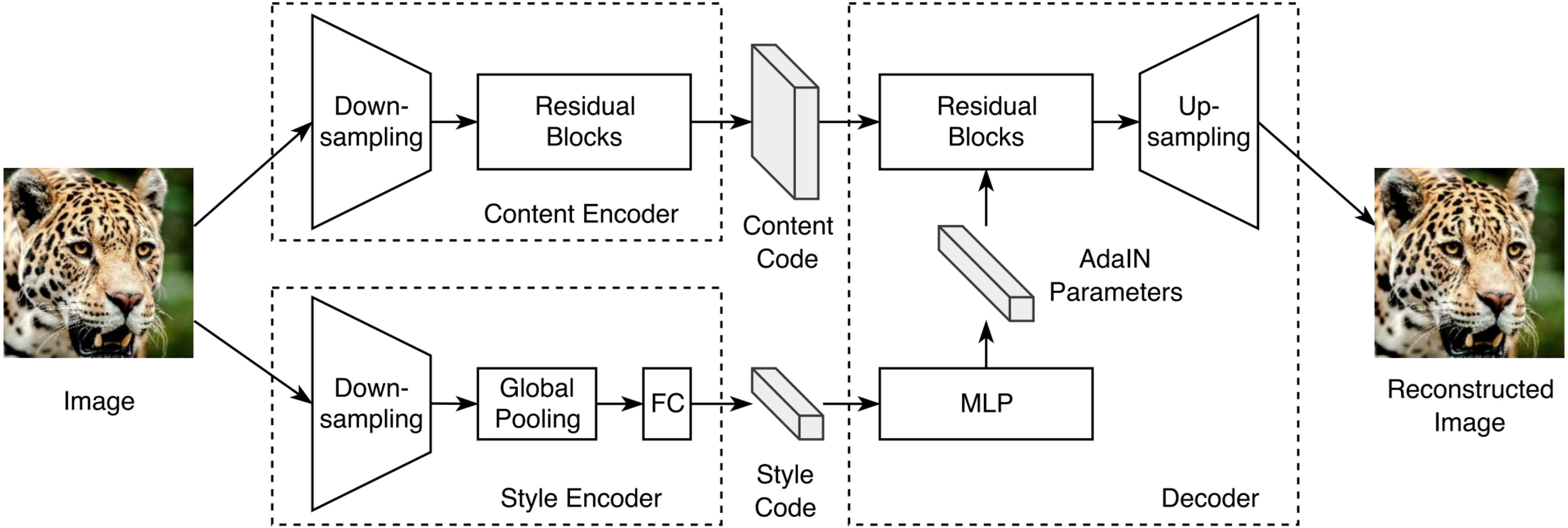
AdaIN in style transfer



$$\mathrm{AdaIN}(c, s) = \gamma\left(\frac{c - \mu(c)}{\sigma(c)}\right) + \beta$$

AdaIN in a generative network

Architectural Implementation



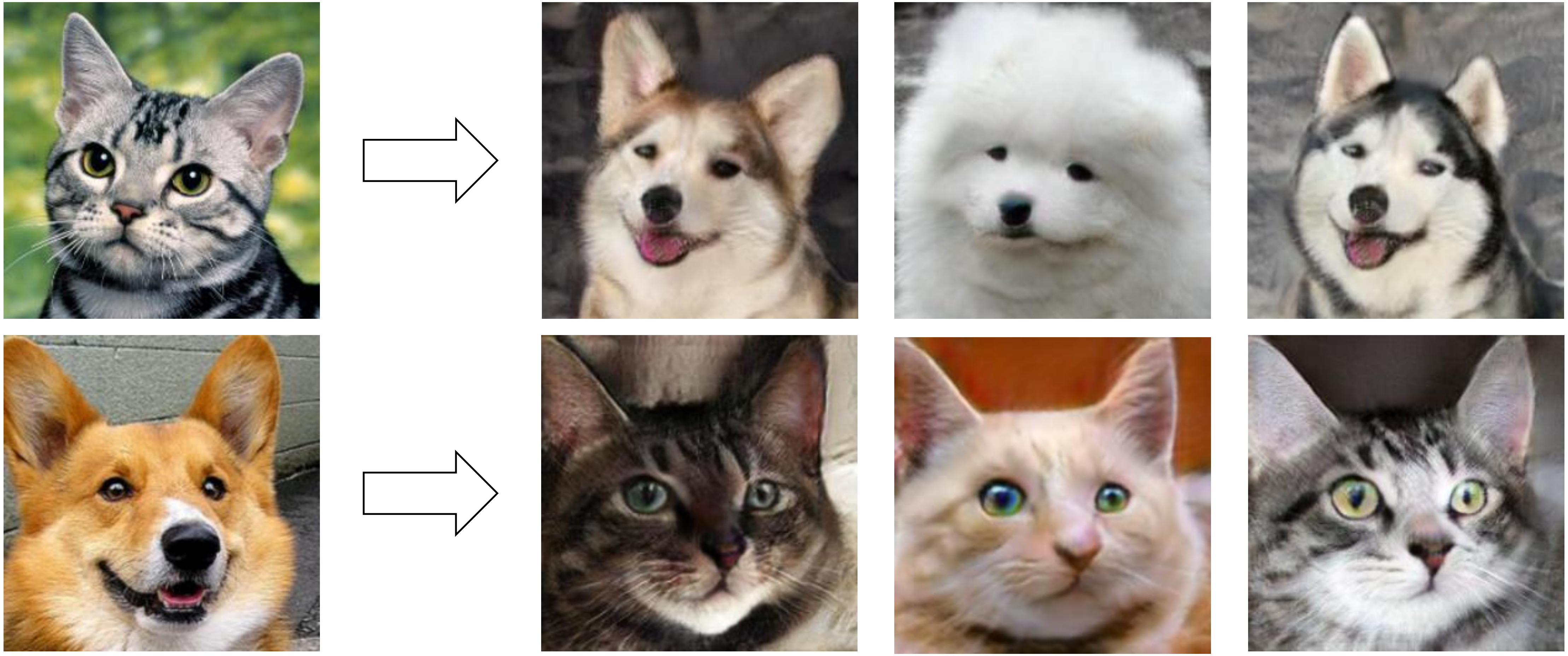
Sketches <-> Photo

Input Outputs

Cats ↔ Dogs

Input

Outputs

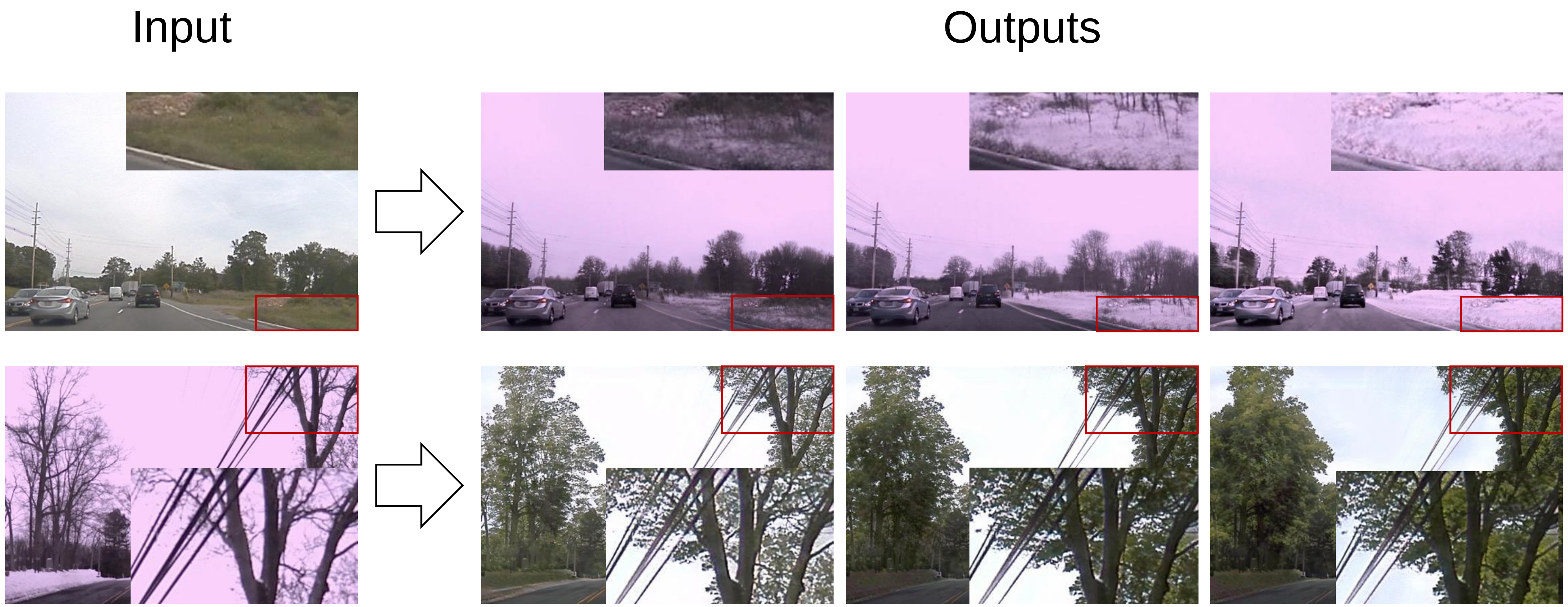


Synthetic ↔ Real

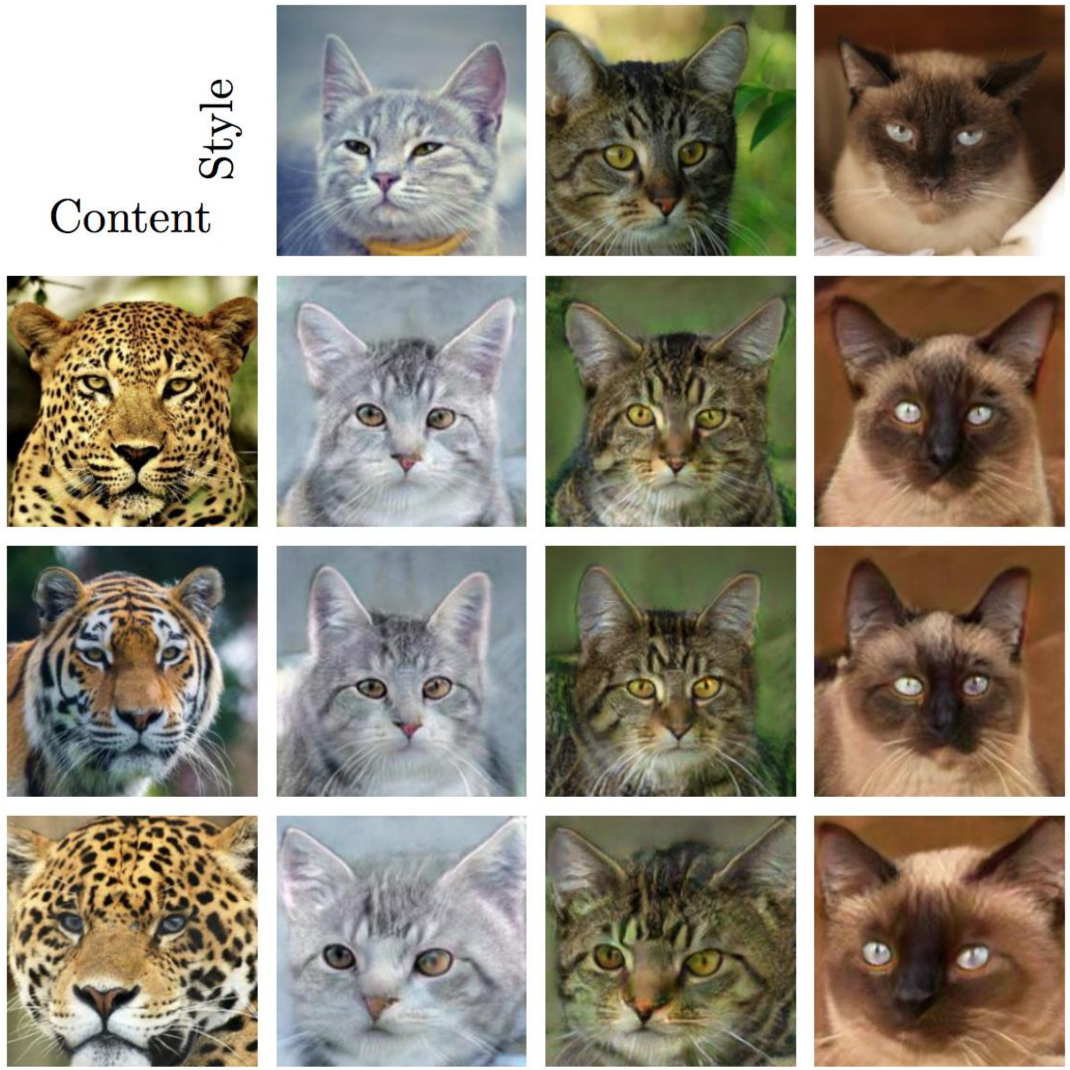
Input

Outputs

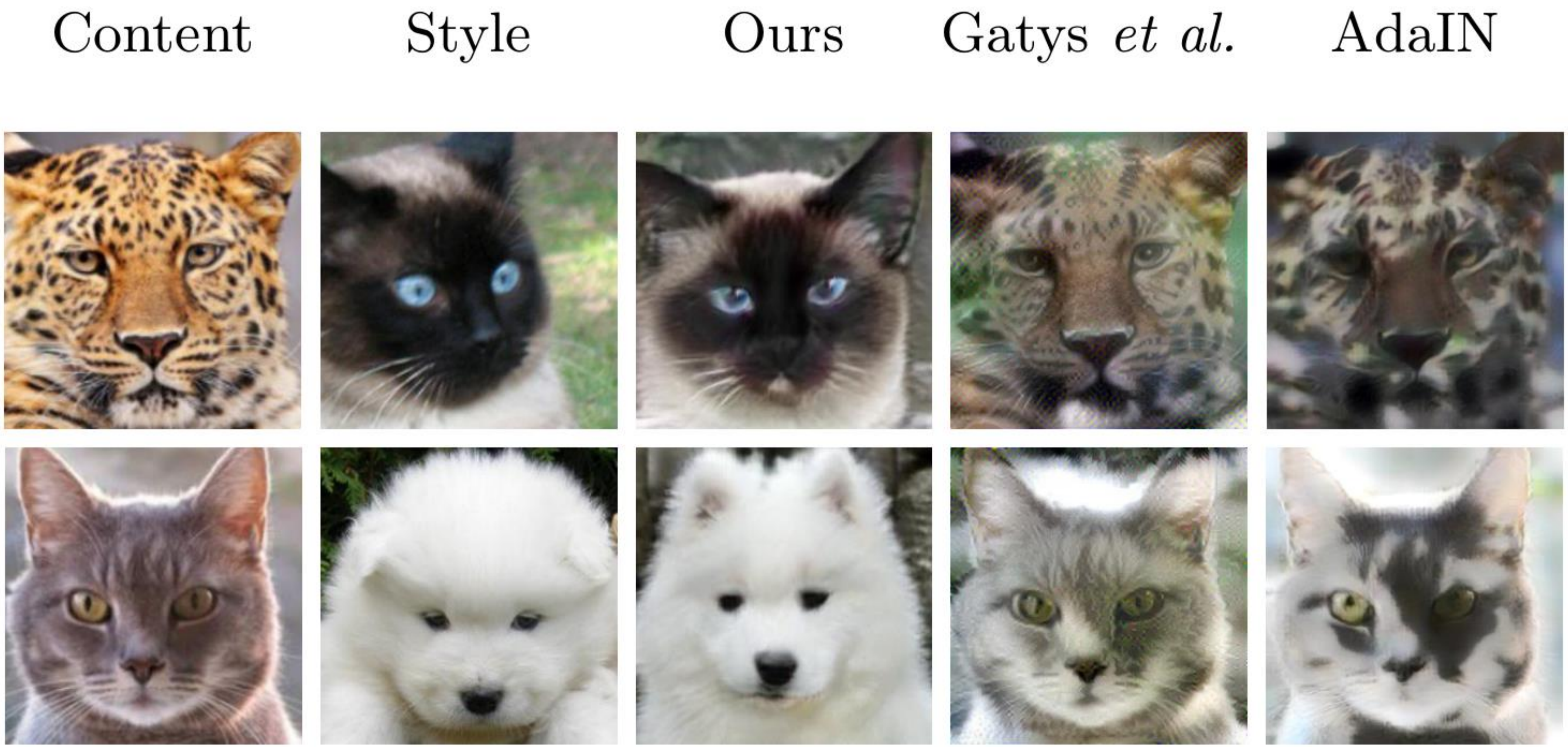
Summer ↔ Winter



Example-guided Translation

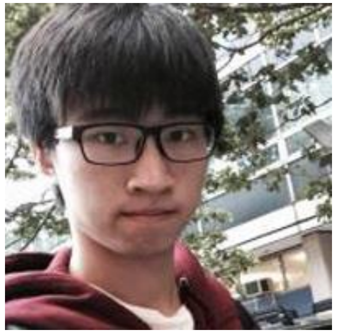


Example-guided Translation



Conclusion

- Translate one input image to multiple corresponding images in the target domain.
- Content and style decomposition via the AdaIN design
- ECCV 2018
- MUNIT code: https://github.com/nvlabs/munit/
- Paper: https://arxiv.org/abs/1804.04732



Xun Huang
NVIDIA, Cornell

Ming-Yu Liu
NVIDIA

Serge Belongie
Cornell

Jan Kautz
NVIDIA