 → 

Conditional GANs, Image-to-Image Translation

Jun-Yan Zhu

16-726, Spring 2023

Many slides from Phillip Isola, Ming-Yu Liu, Xun Huang, etc.

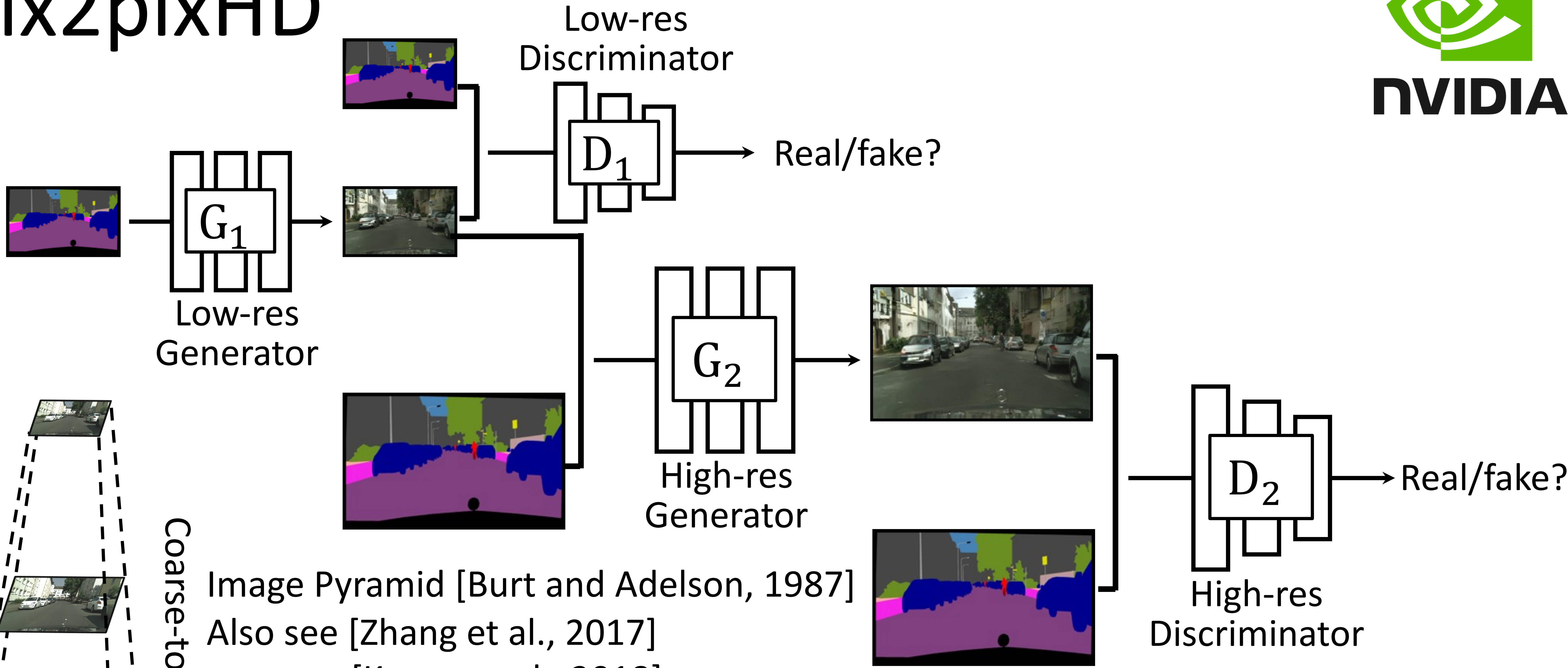
Improving Conditional GANs

- Multimodal synthesis.
- **High-resolution synthesis.**
- Model training without pairs

The Curse of Dimensionality

Pix2pix output

pix2pixHD



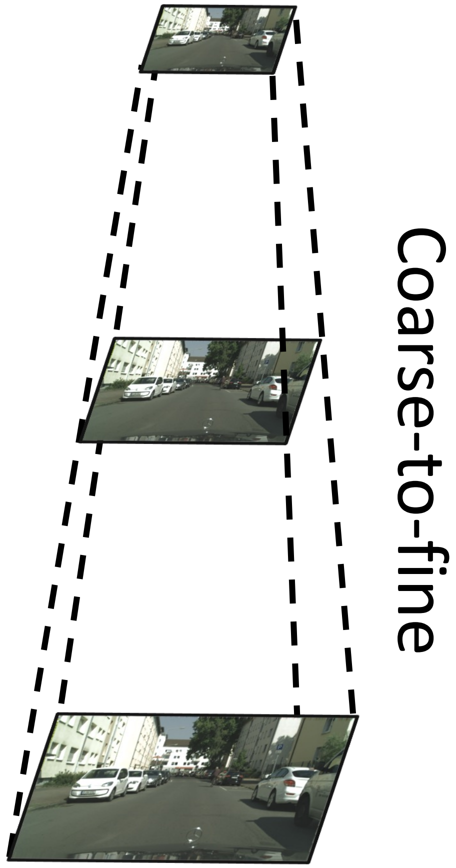
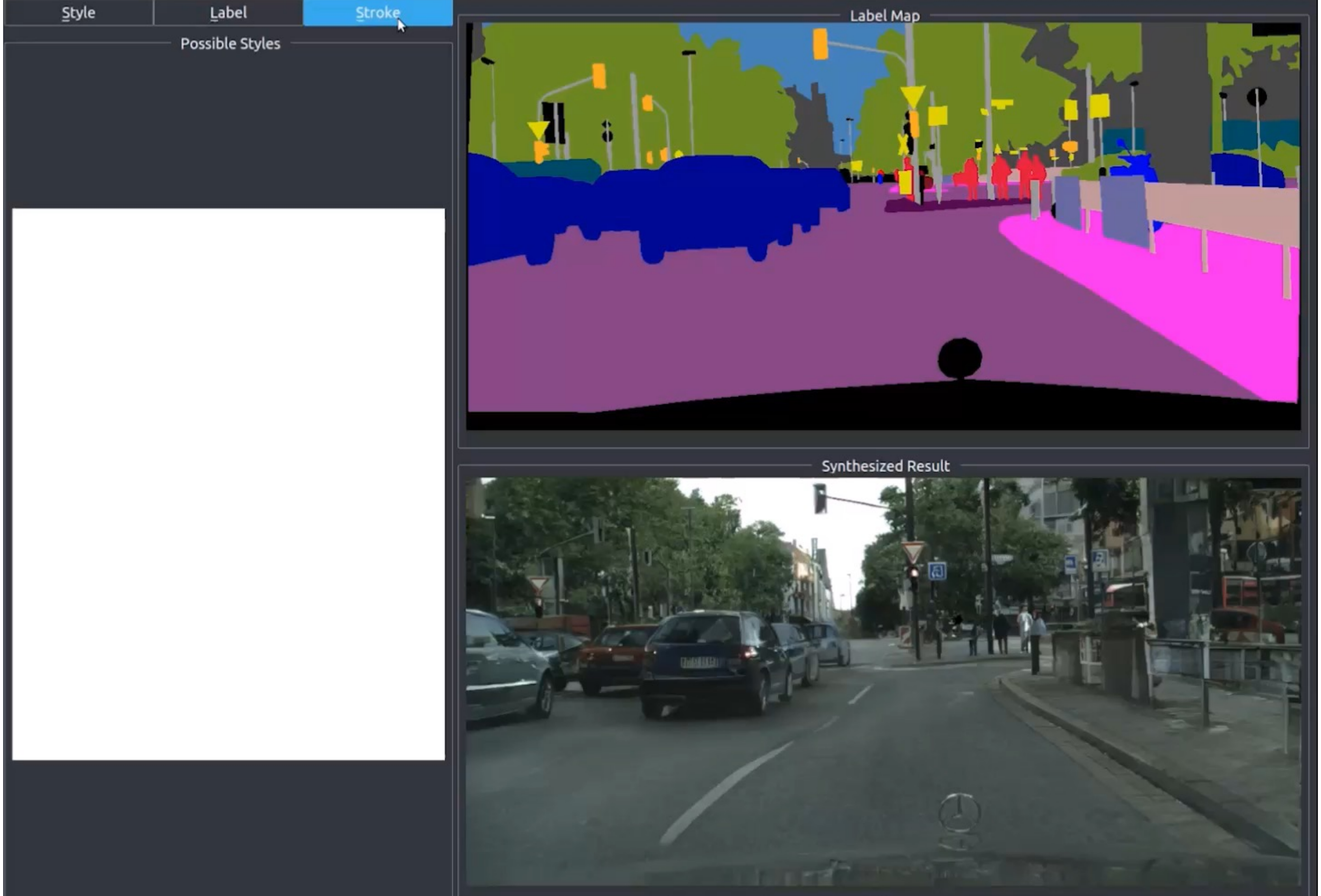


Image Pyramid [Burt and Adelson, 1987]
Also see [Zhang et al., 2017]
[Karras et al., 2018]

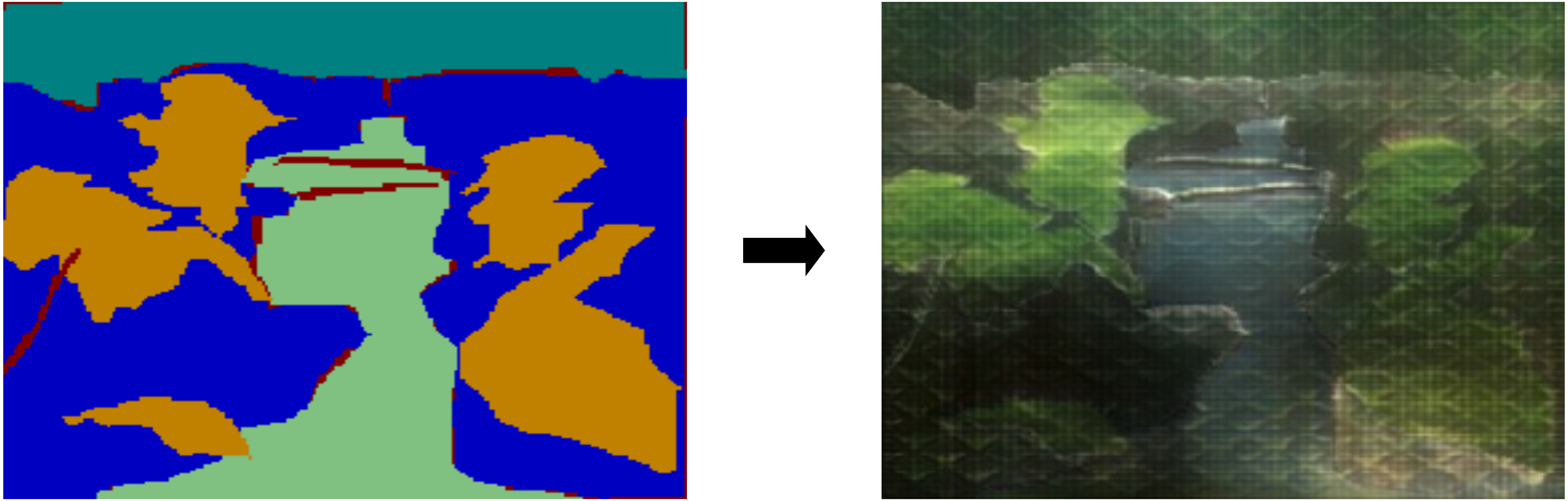
Objective: Multi-scale GANs loss + Perceptual Loss
+ Feature Matching Loss (with Discriminator's features)

pix2pixHD [Wang et al., 2018]

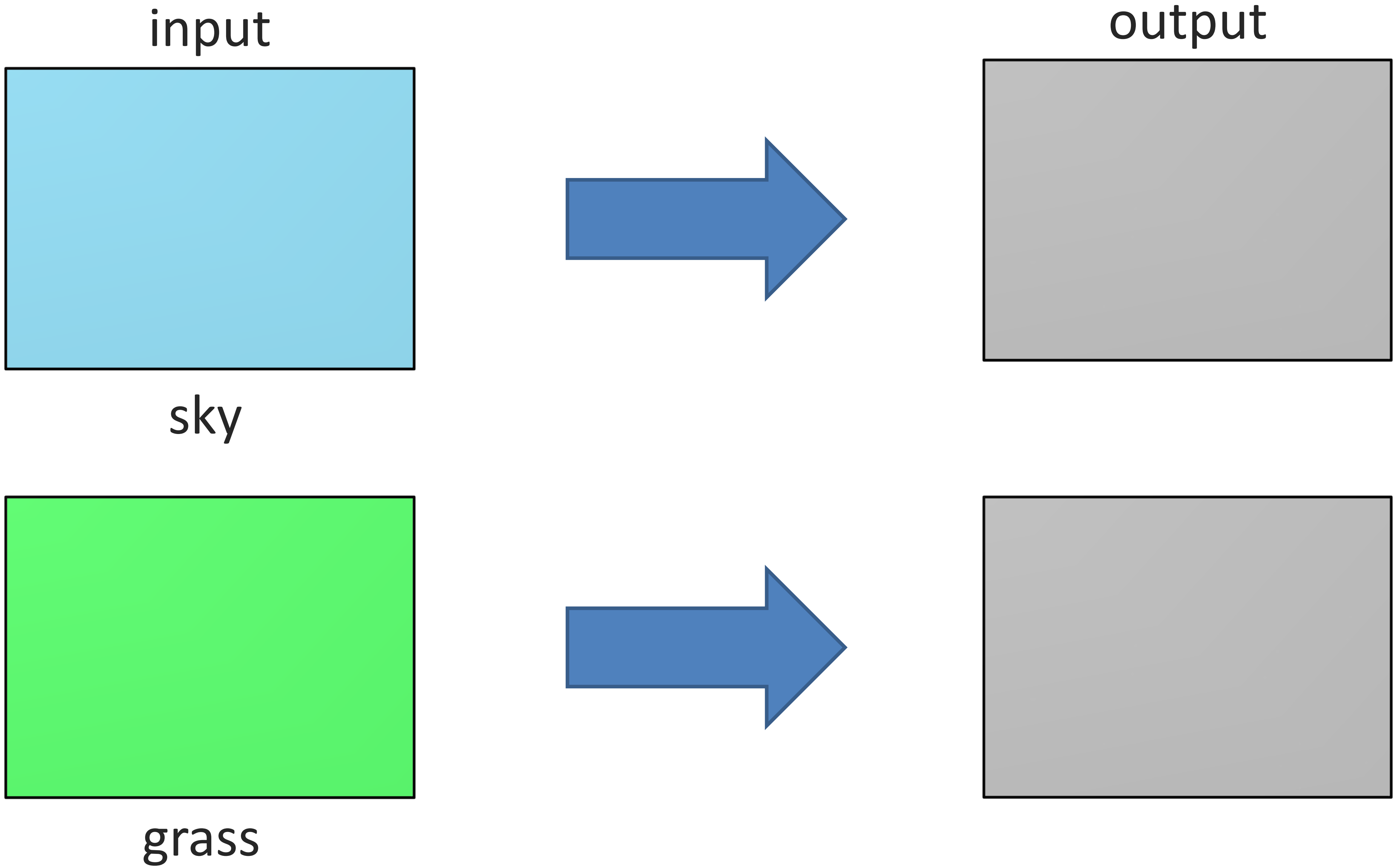
pix2pixHD: 2048×1024


Style
Label
Stroke
Possible Styles
Label Map
Synthesized Result

Conditional Image Synthesis in the Wild

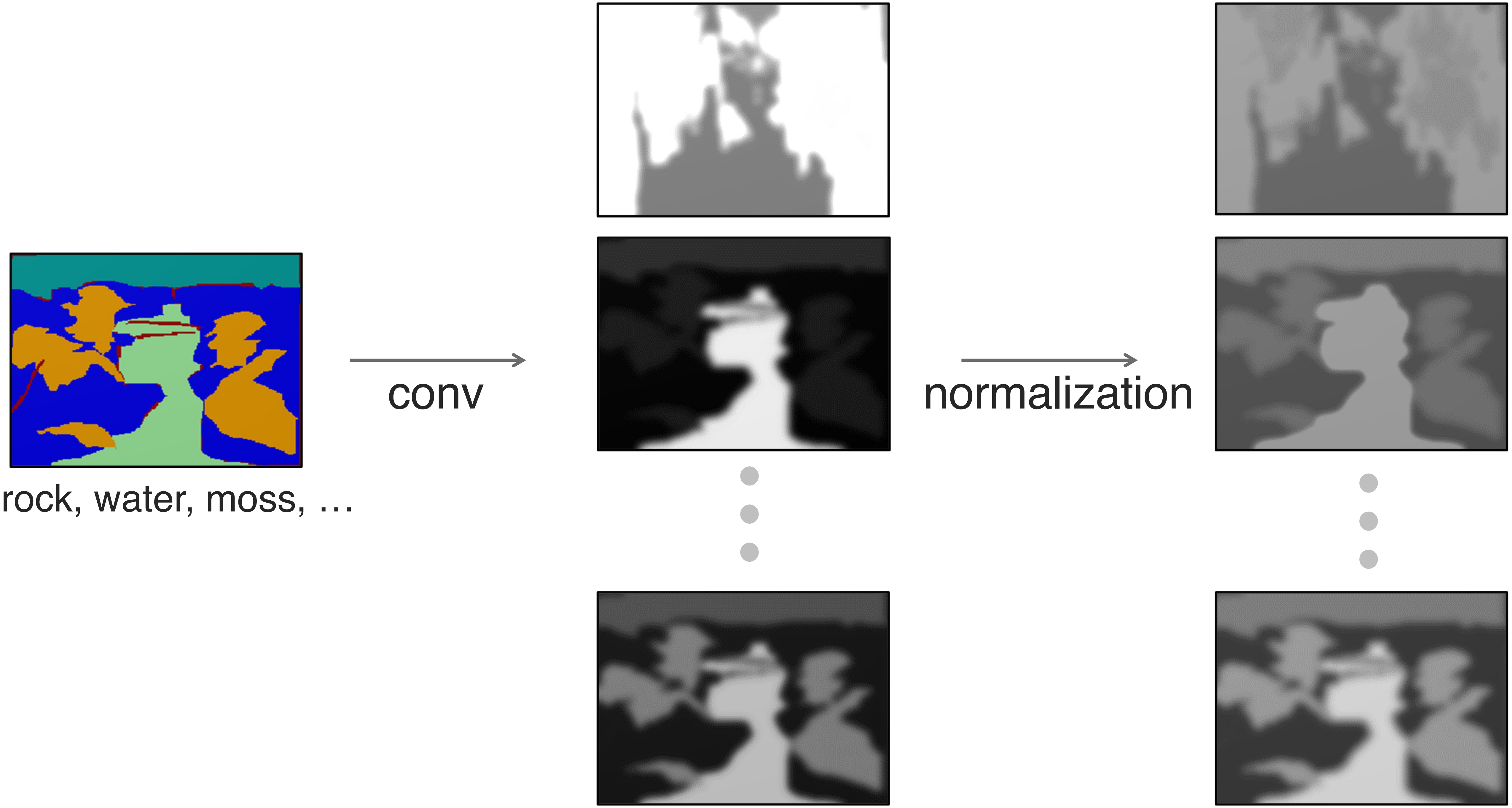


pix2pixHD [Wang et al., 2018]

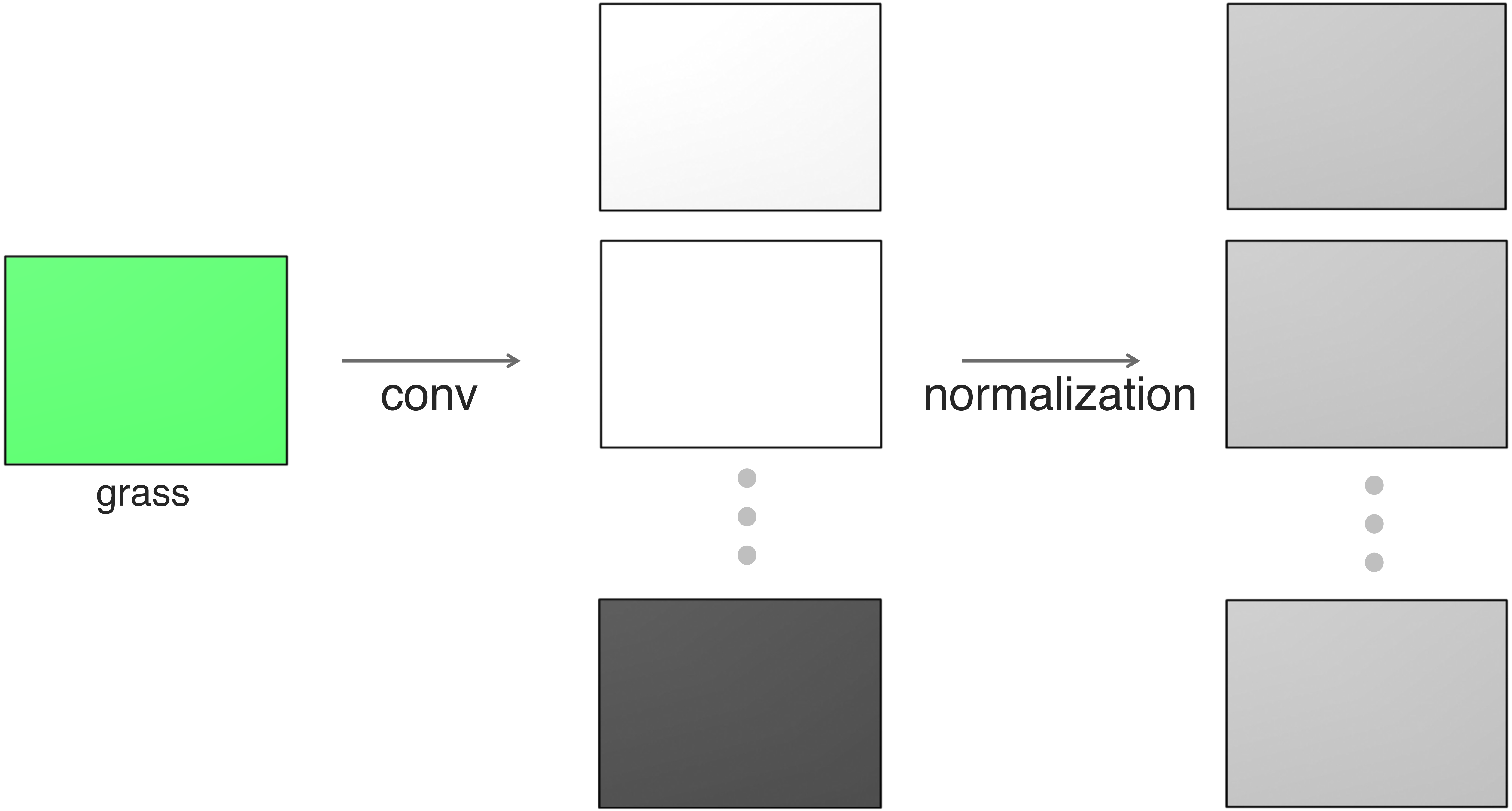


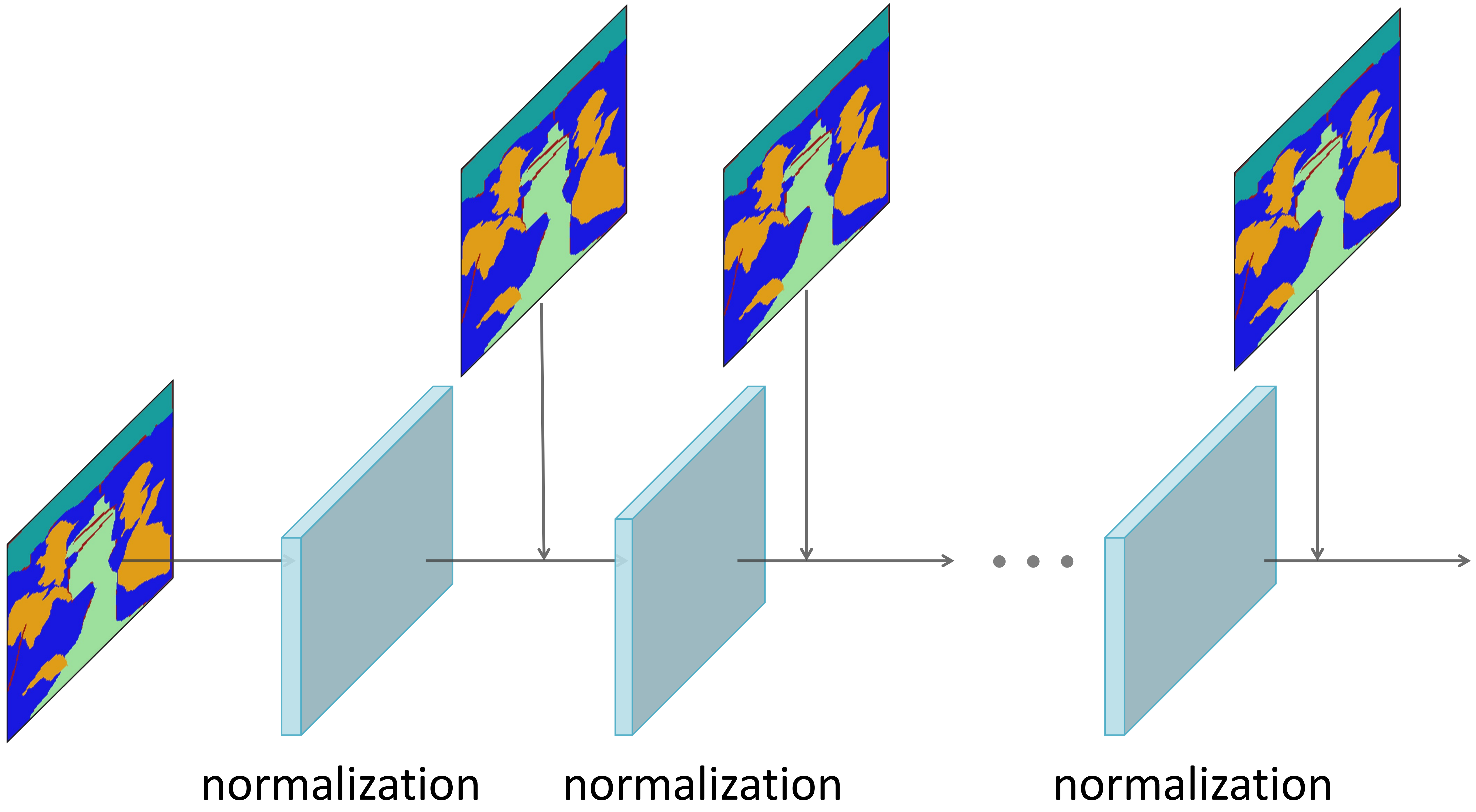
pix2pixHD [Wang et al., 2018]

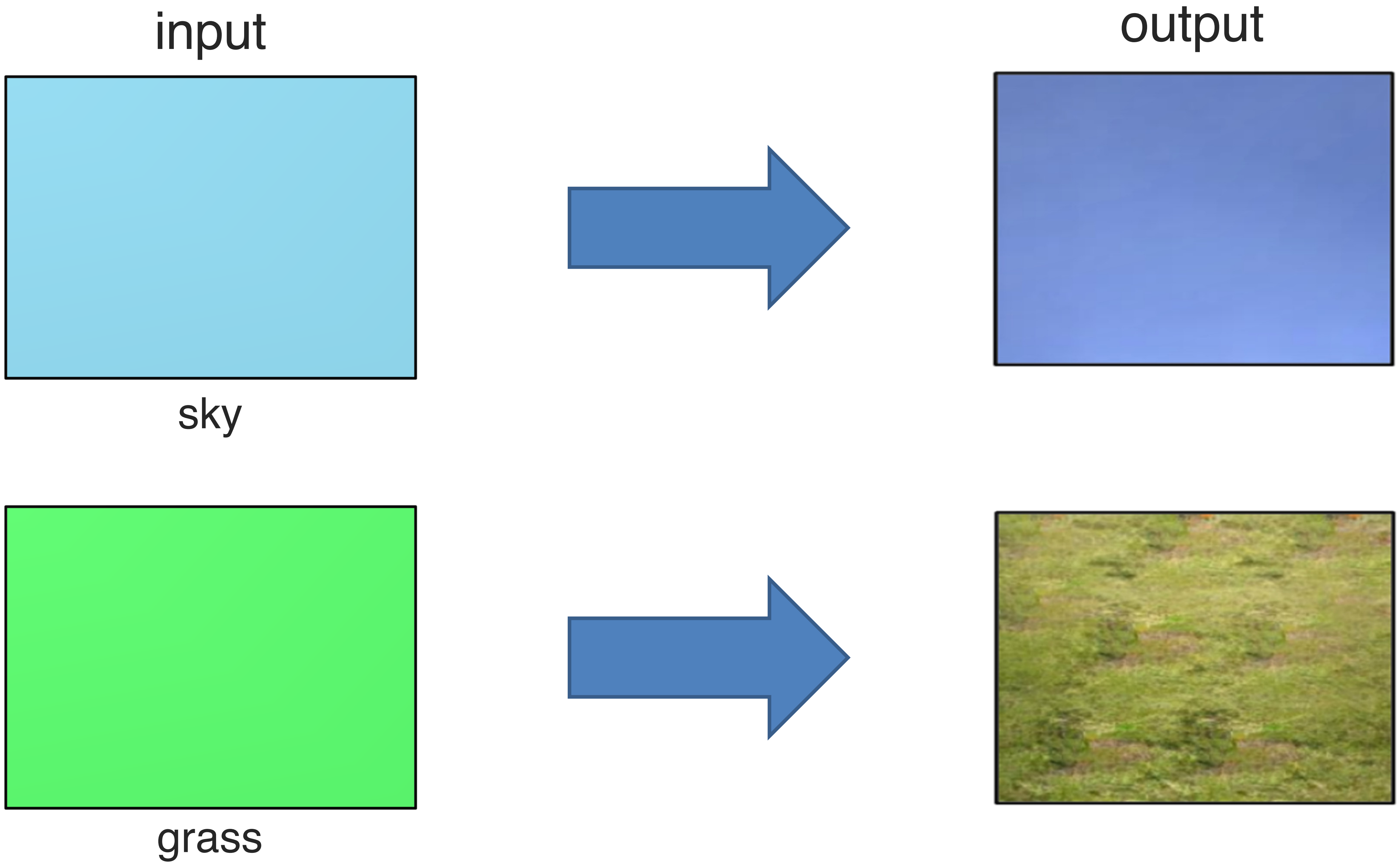
Problem with standard networks



Problem with standard networks

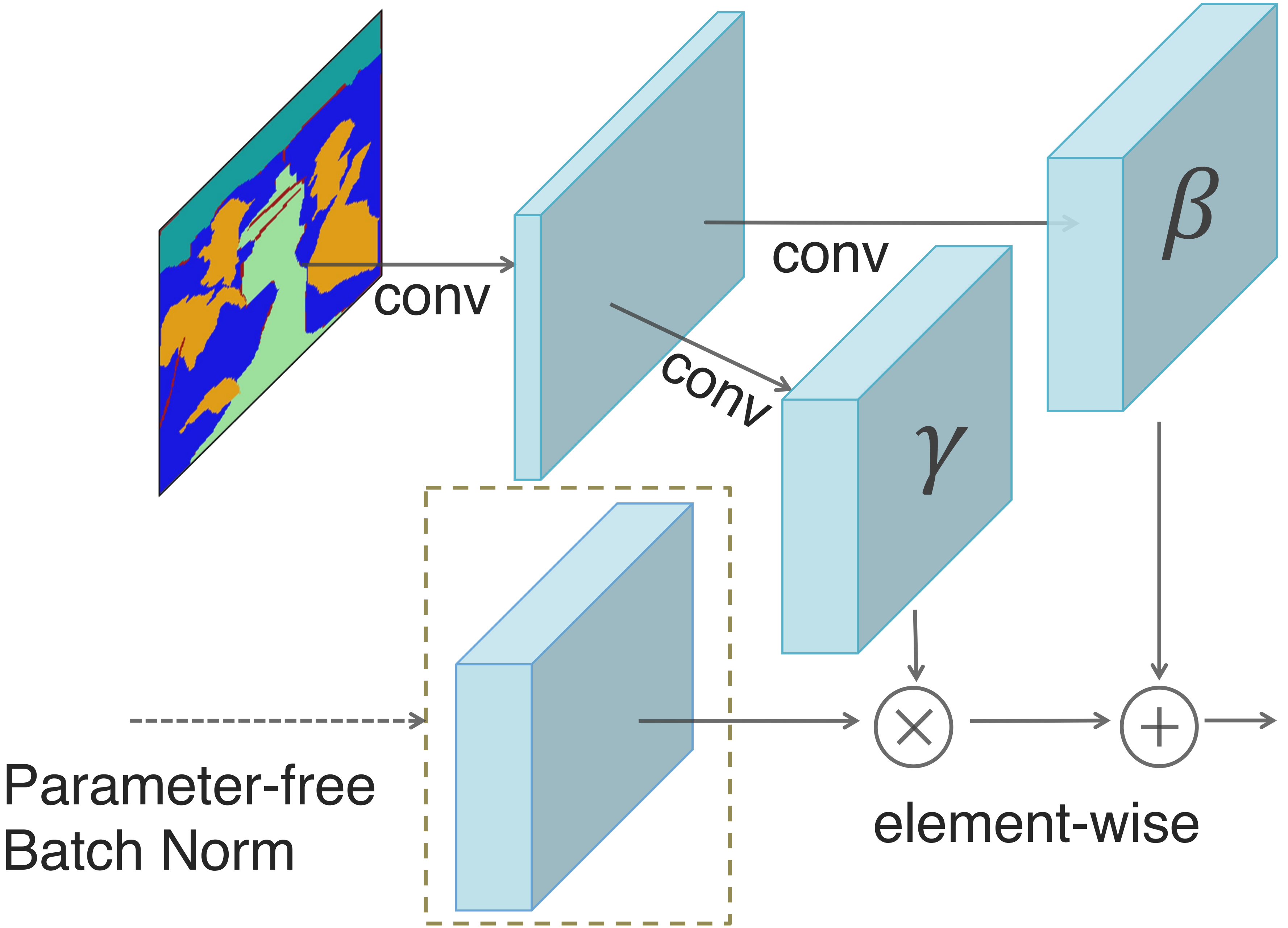



normalization
normalization
normalization



SPADE (ours)

SPADE(SPAtially ADaptive DEnormalization)



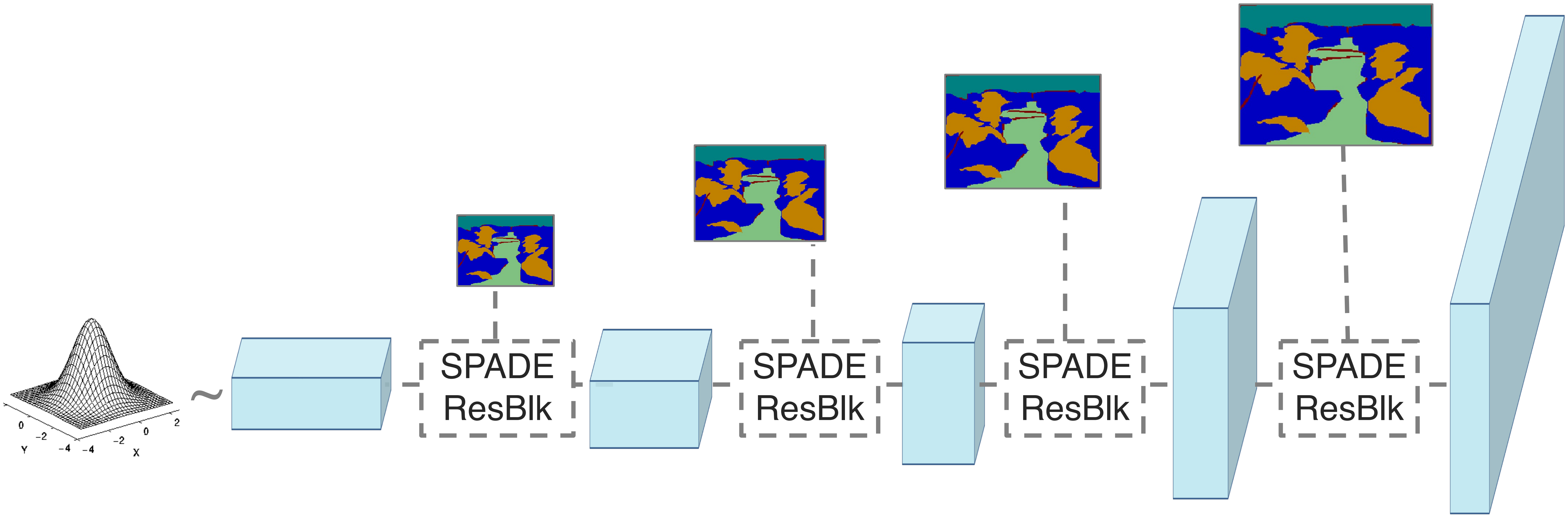
SPADE (SPAtially ADaptive DEnormalization)

Batch Norm (Ioffe et al. 2015)

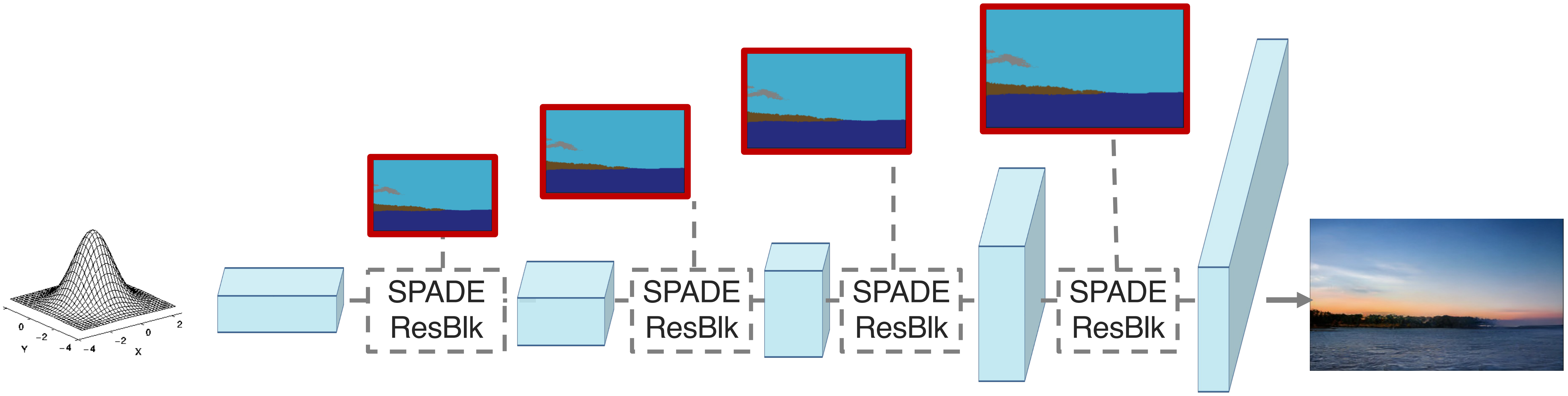
$$y = \underbrace{\frac{x - \mu}{\sigma}}_{\text{normalization}} \cdot \underbrace{\gamma + \beta}_{\text{affine transform}}$$

See other adaptive/conditional normalization: conditional BN (Dumoulin et al.), AdaIN (Huang and Belongie), SFT (Wang et al.)

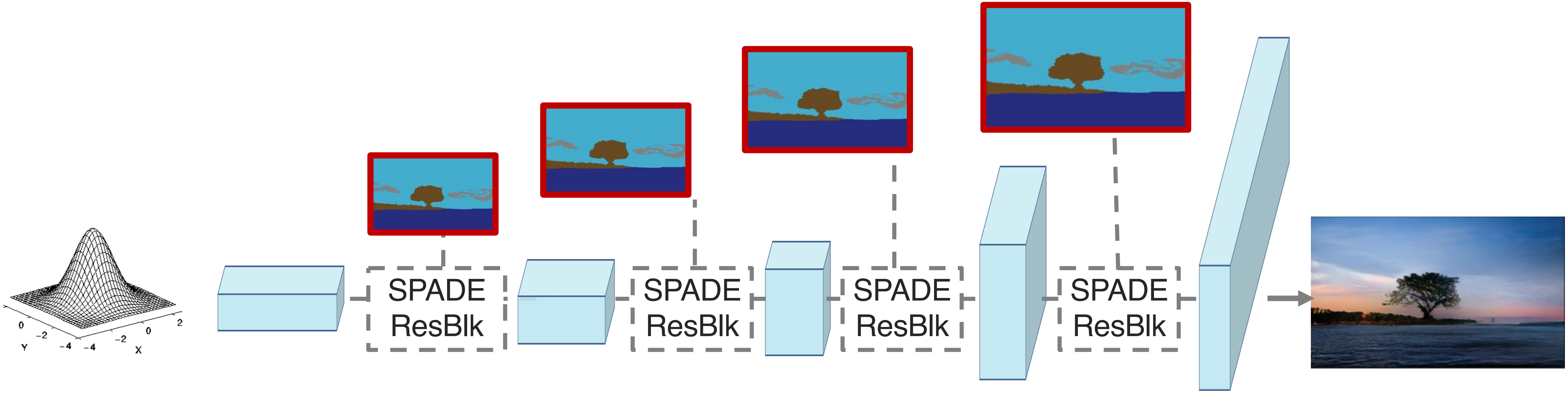
Generator



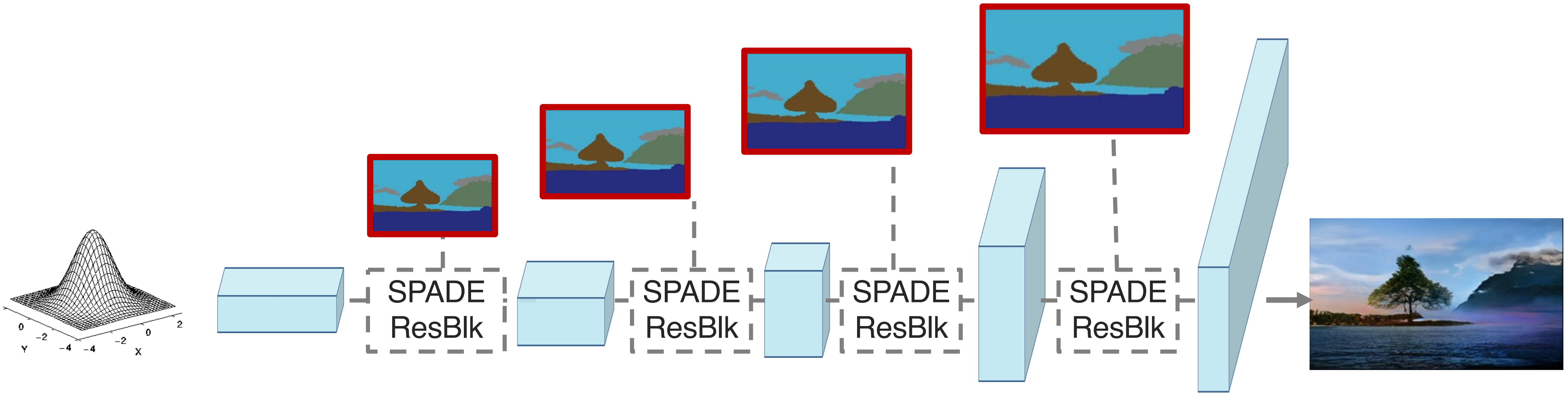
Semantic Control



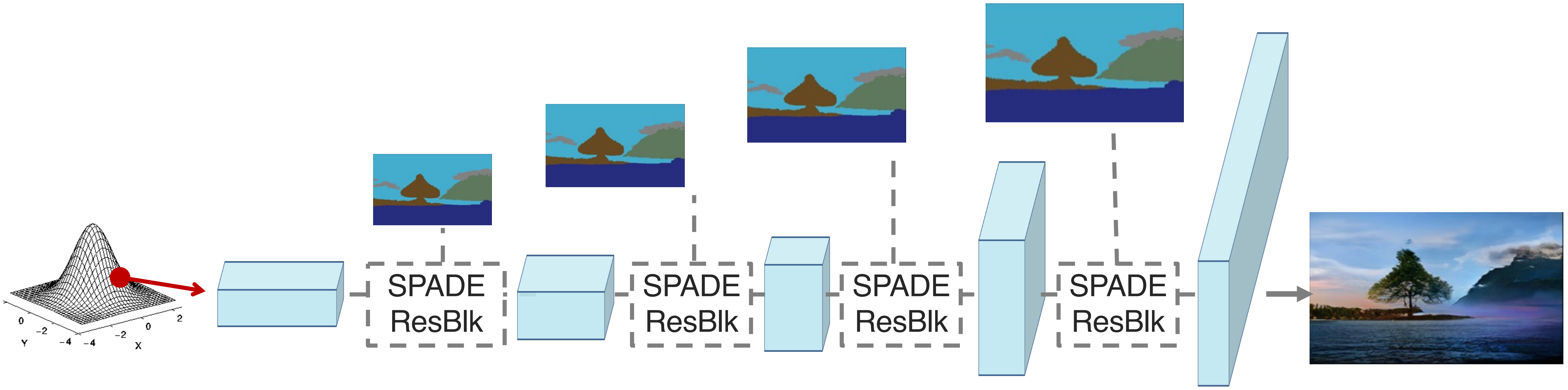
Semantic Control



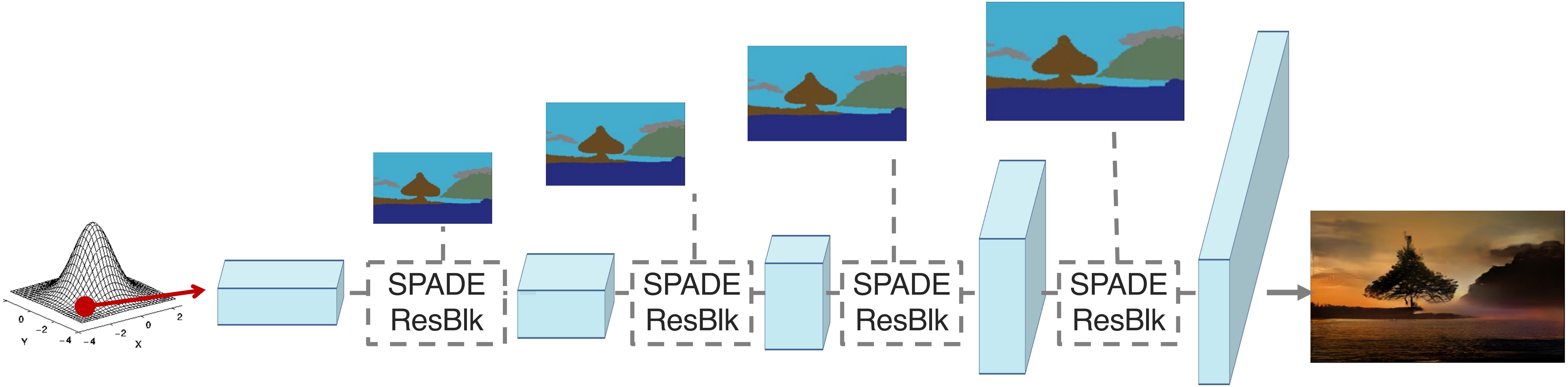
Semantic Control



Style Control

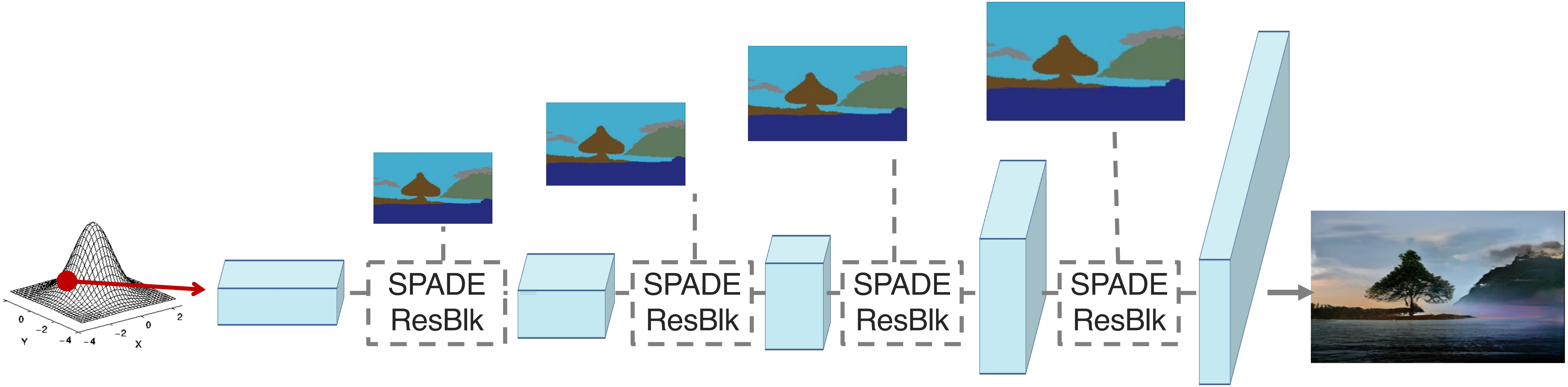


Style Control

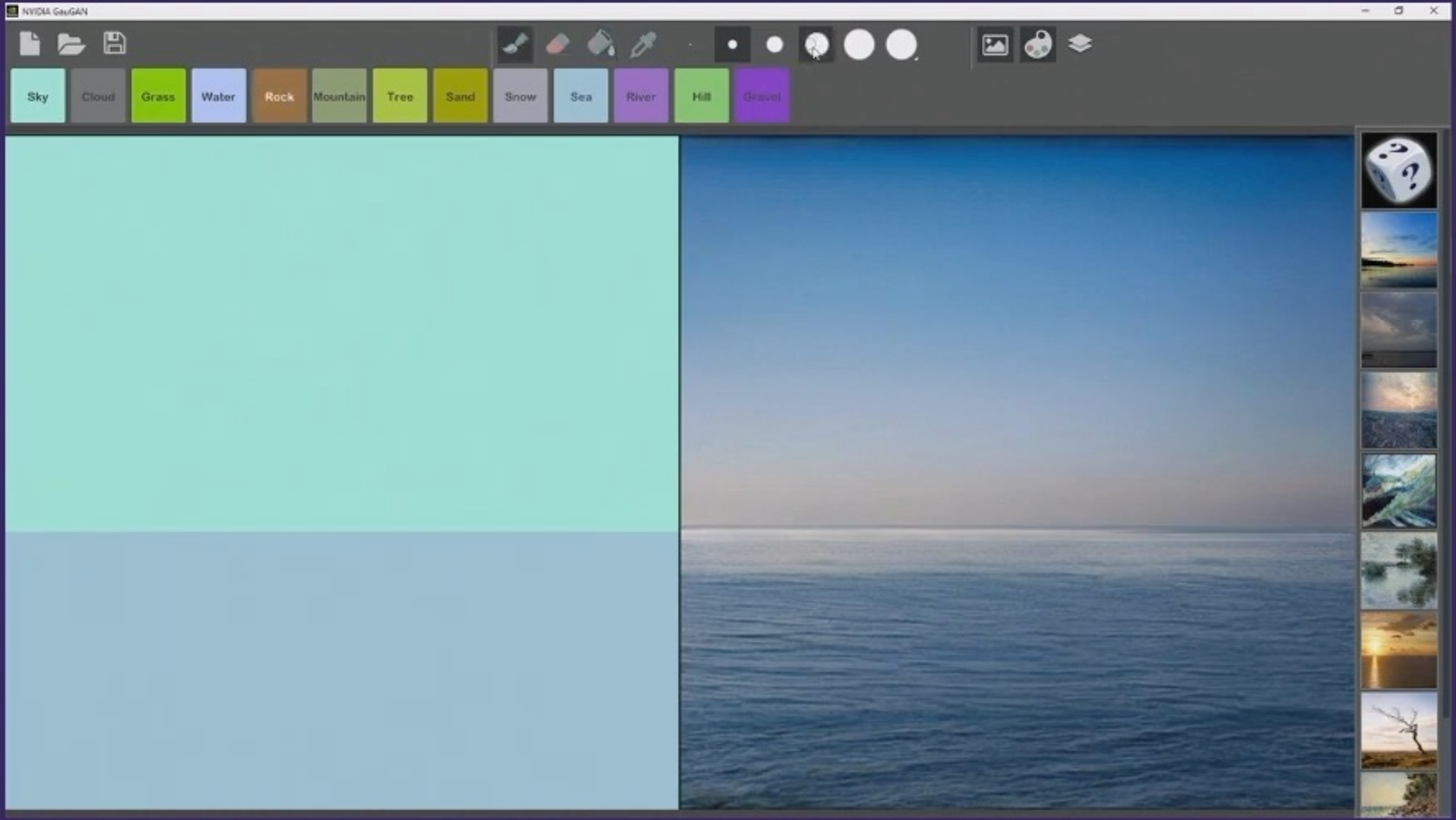


Style Manipulation

Style Control



Style Manipulation

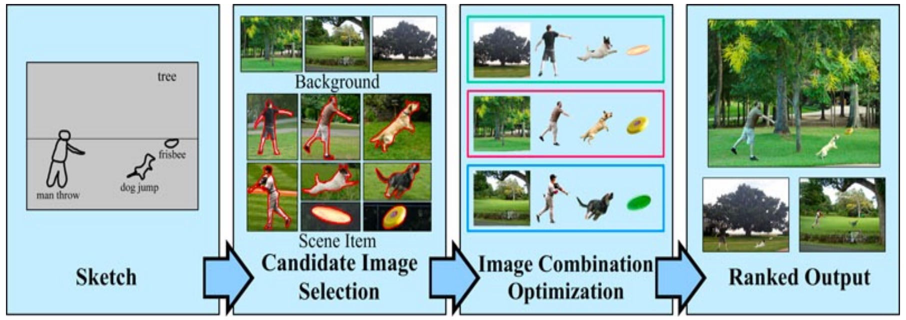


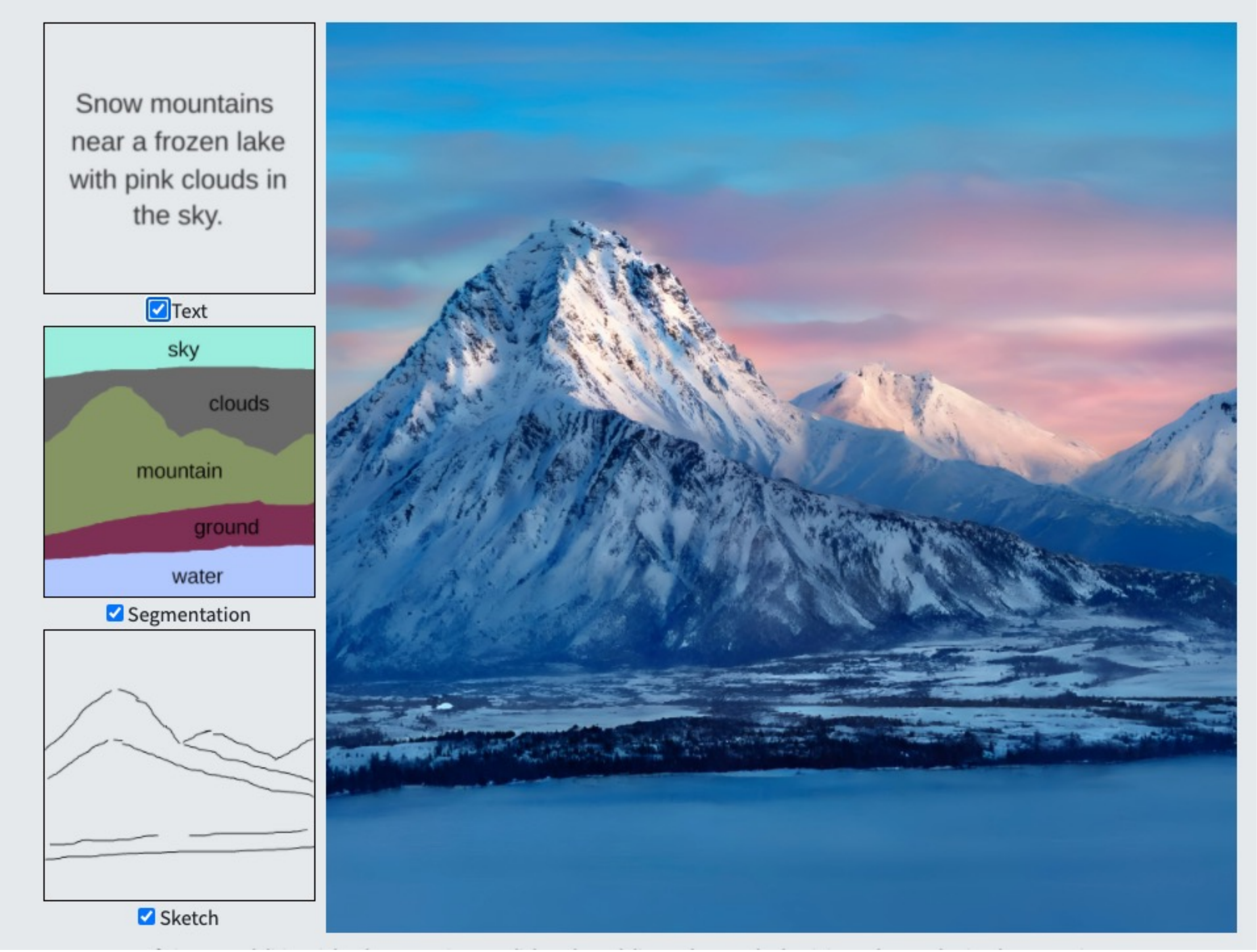
thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY - 1 AUGUST

SIGGRAPH 2019 Real-time Live! “Best of Show Award" and "Audience Choice Award"

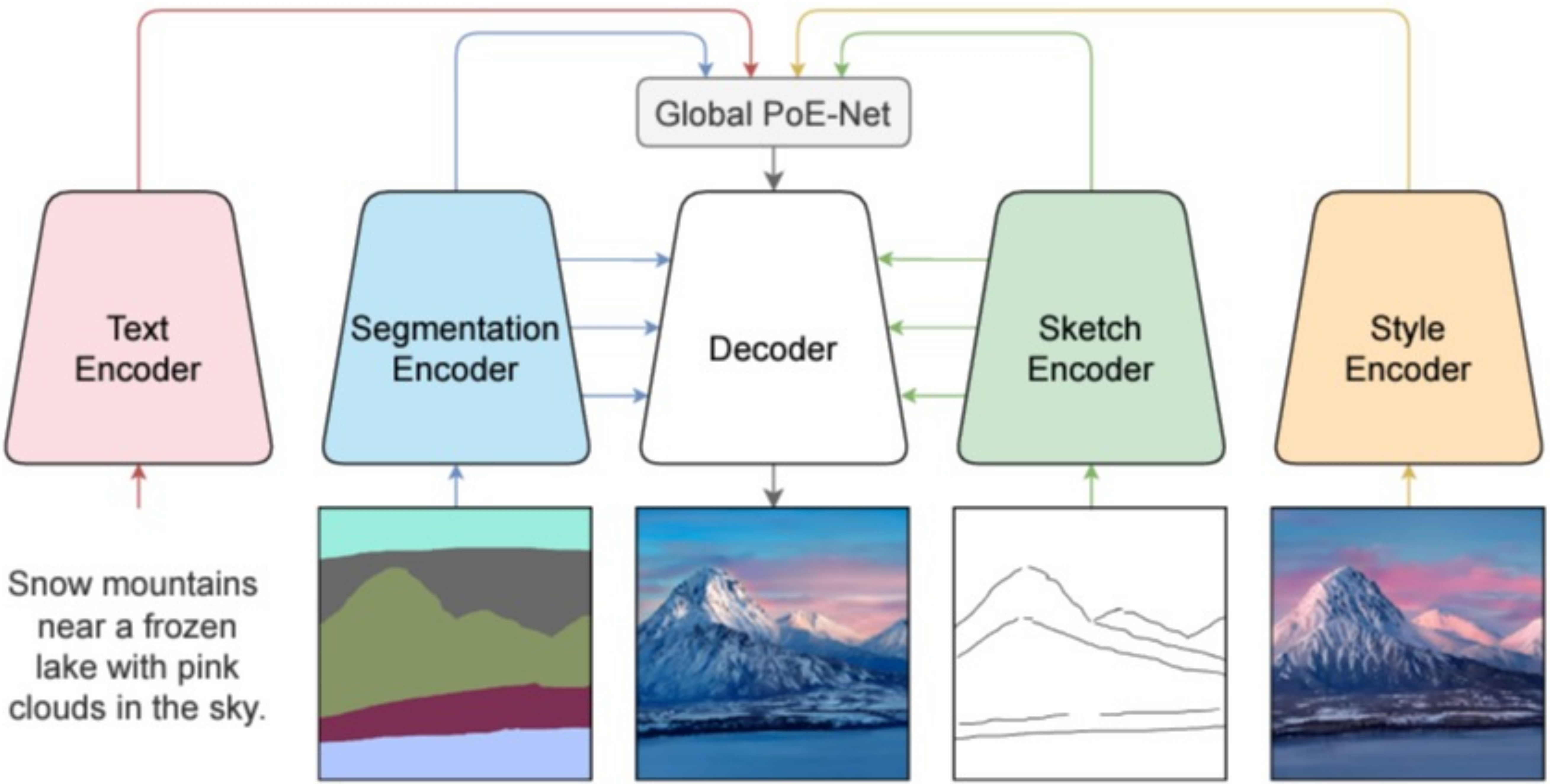
By Darek Zabrocki, Concept Designer and Illustrator

Learning vs. Exemplar-based

	Learning-based	Exemplar-based
		
	[Isola et al], [Wang et al] [Park et al], SEAN [Zhu et al]	[Johnson et al], [Lalonde et al] [Tao et al], [Bansal et al]
Speed	😊	☹
Local realism	😐	😊
Global realism	😊	☹
Match Input	😊	😐



Multimodal Conditional Image Synthesis with Product-of-Experts GANs [Huang et al., 2021]

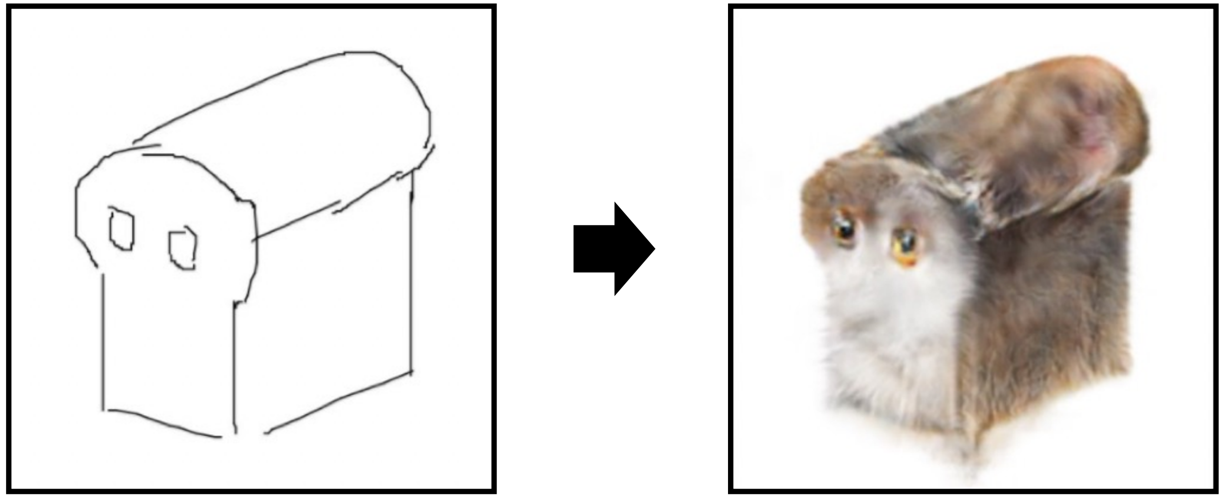


Multimodal Conditional Image Synthesis with Product-of-Experts GANs [Huang et al., 2021]

GauGAN2 Demo

http://gaugan.org/gaugan2/

Supervised Learning Approach



Edges2cats

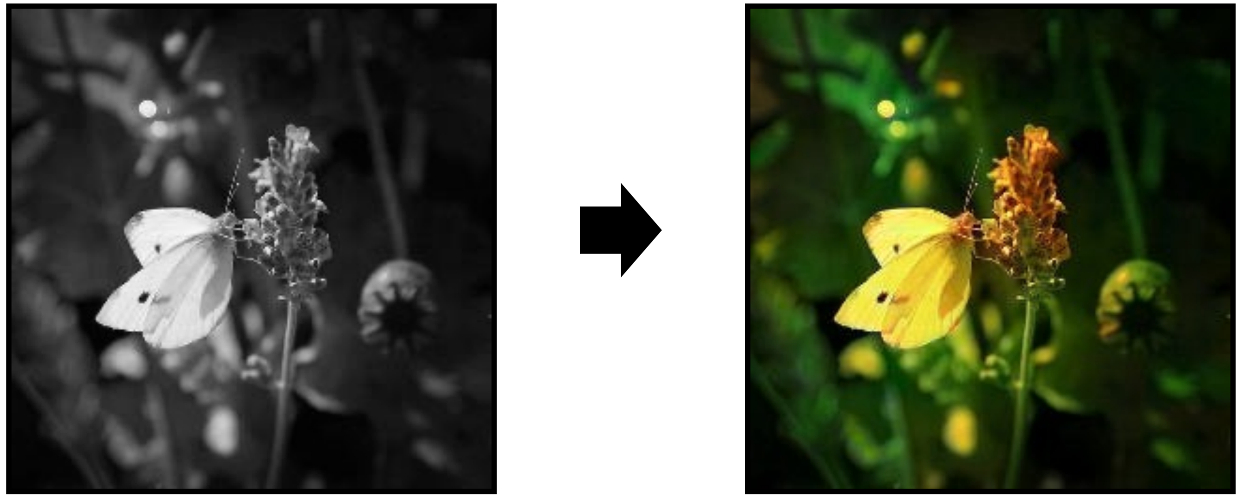
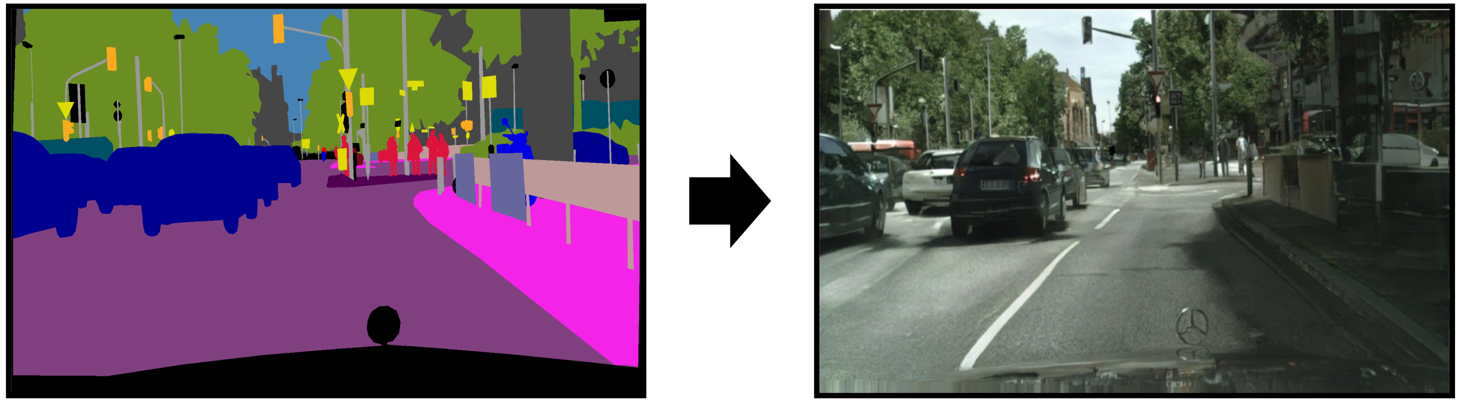
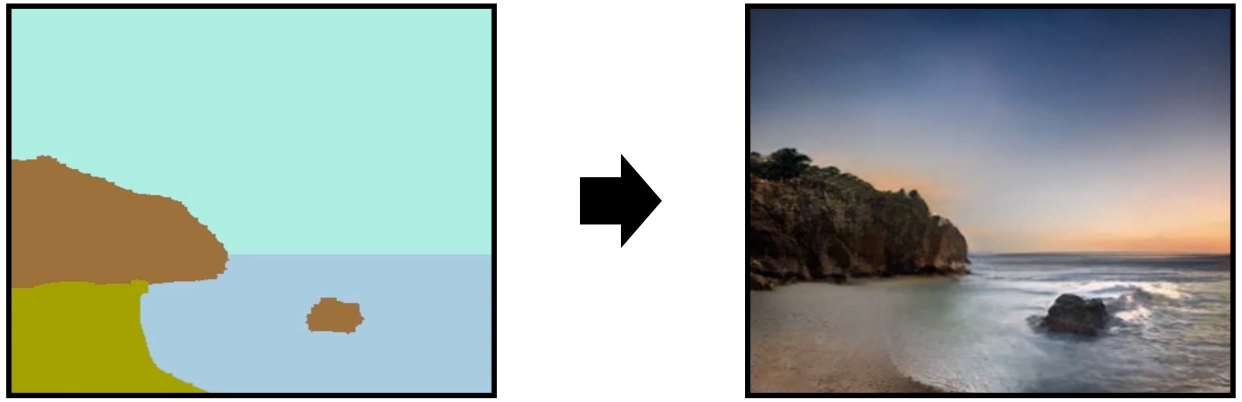


Image colorization

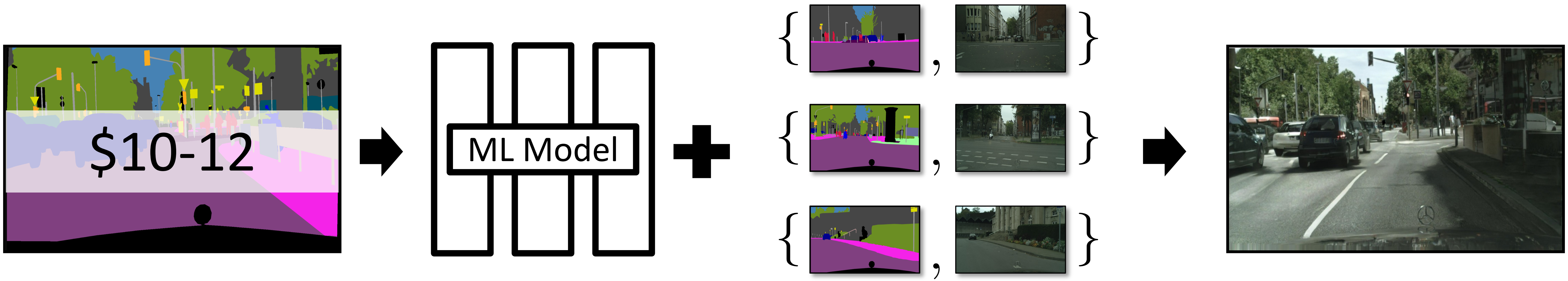


Street view images

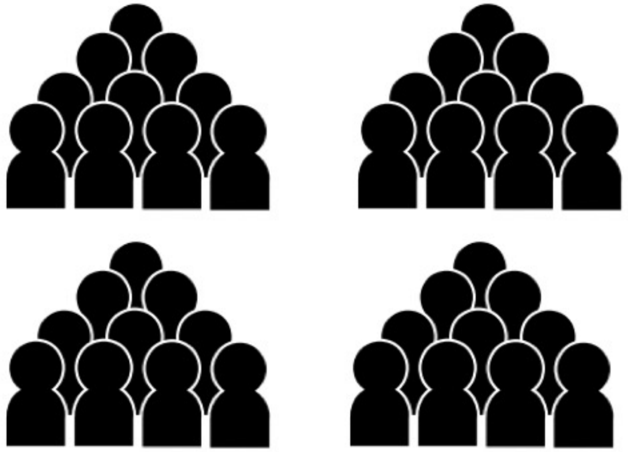


Natural outdoor images

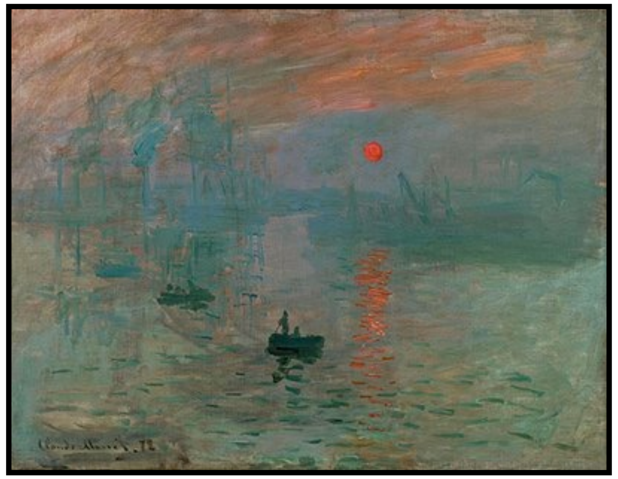
Supervised Learning Approach



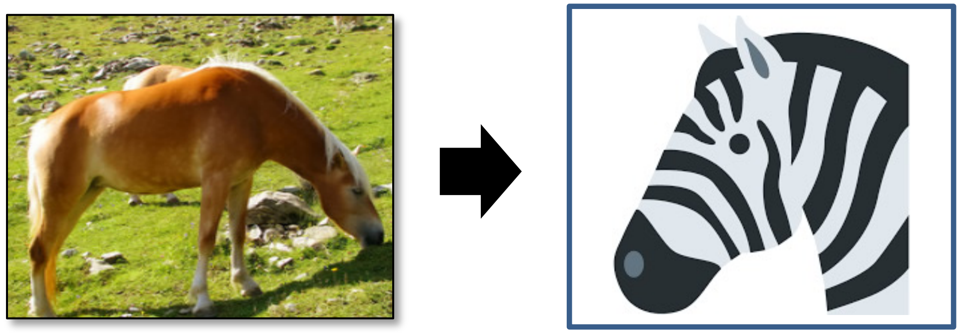
User Input | Learning algorithm | Labeled data | Visual Content



Expensive labor



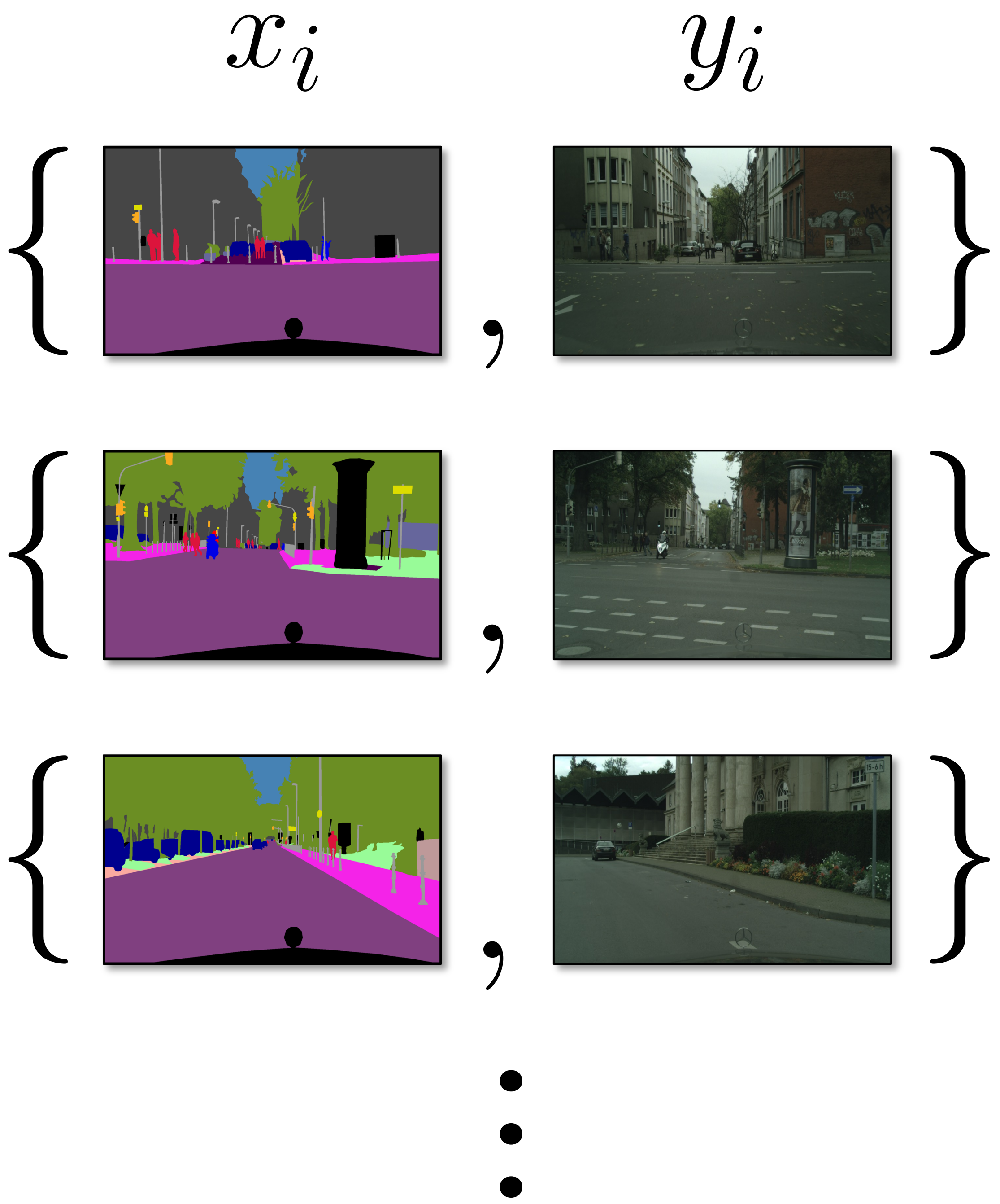
Artistic authoring



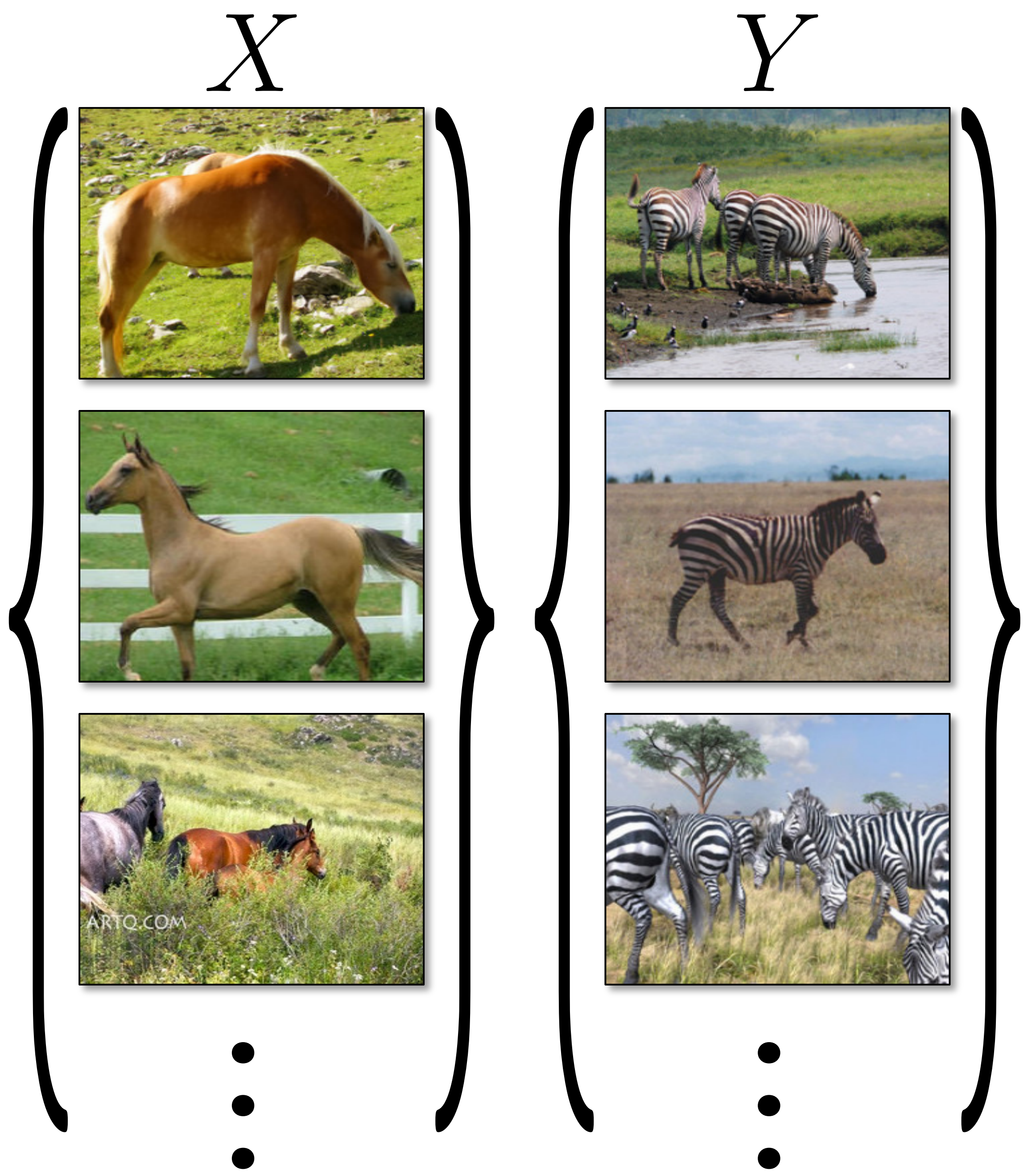
horse | zebra

Infeasible

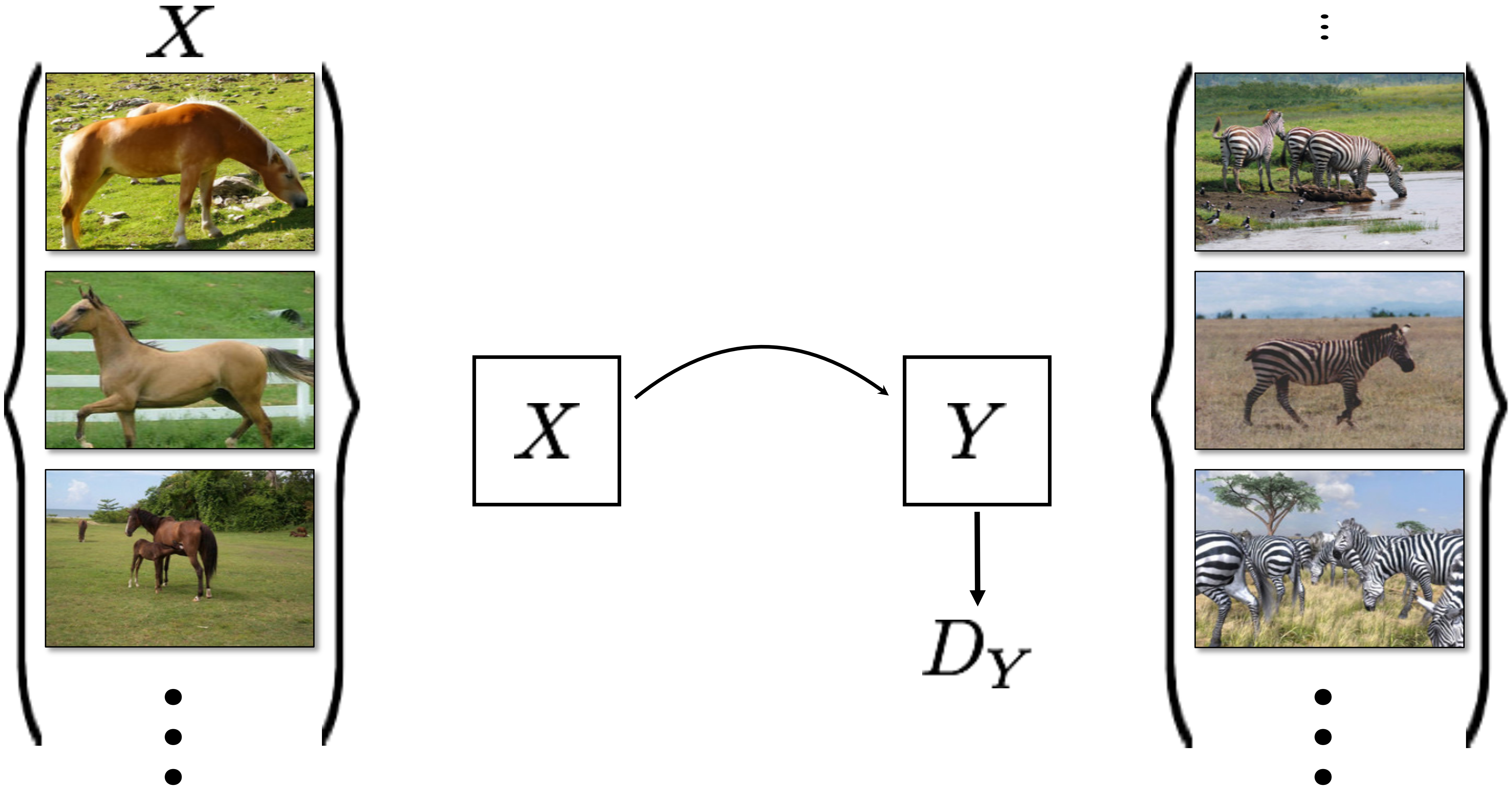
Supervised



Unsupervised



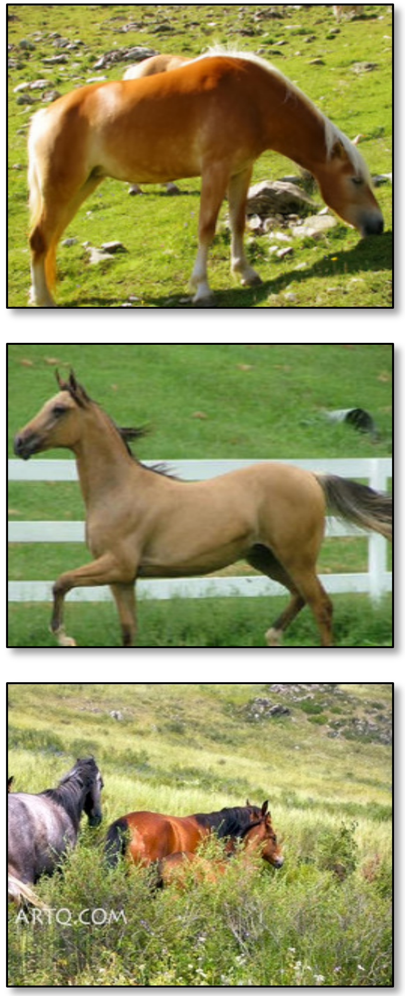
Unsupervised Learning of $p(y \mid x)$



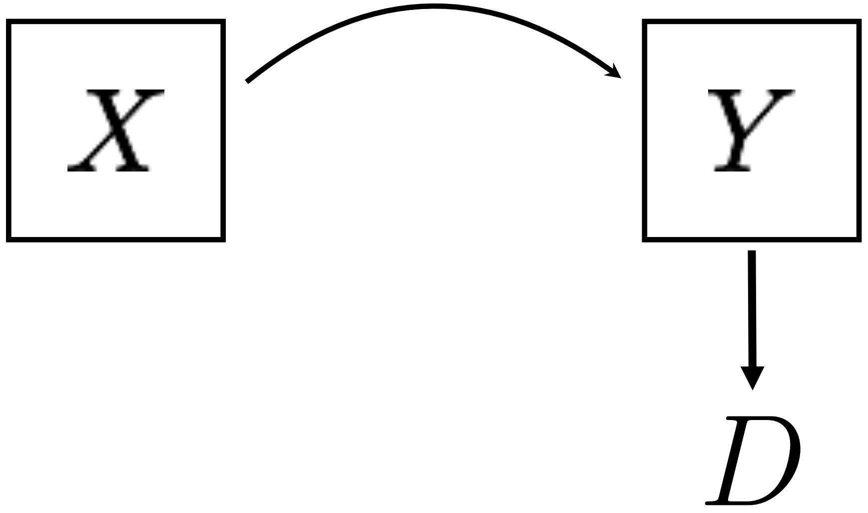
[Zhu*, Park*, Isola, and Efros, 2017]

Unsupervised Learning of $p(y \mid x)$

X

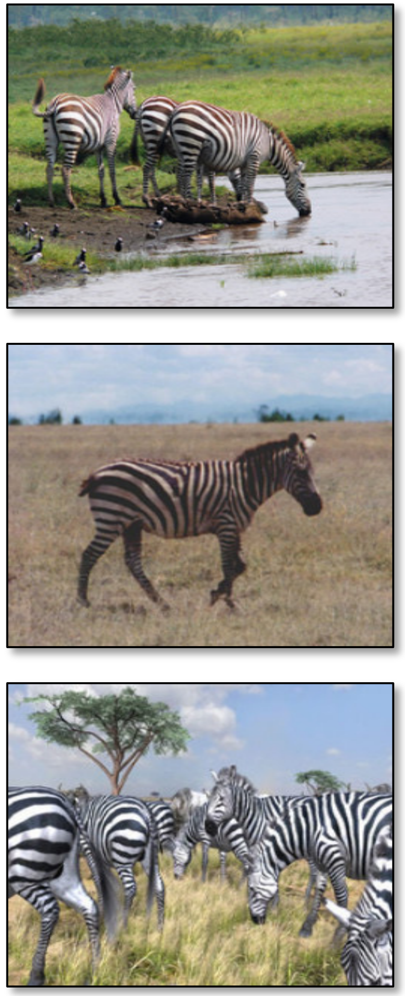


$$\mathbb{E}_x \log(1 - D(\overset{\text{fake zebra}}{\underline{G(x)}})) + \mathbb{E}_y \log D(\overset{\text{real zebra}}{\underline{y}})$$

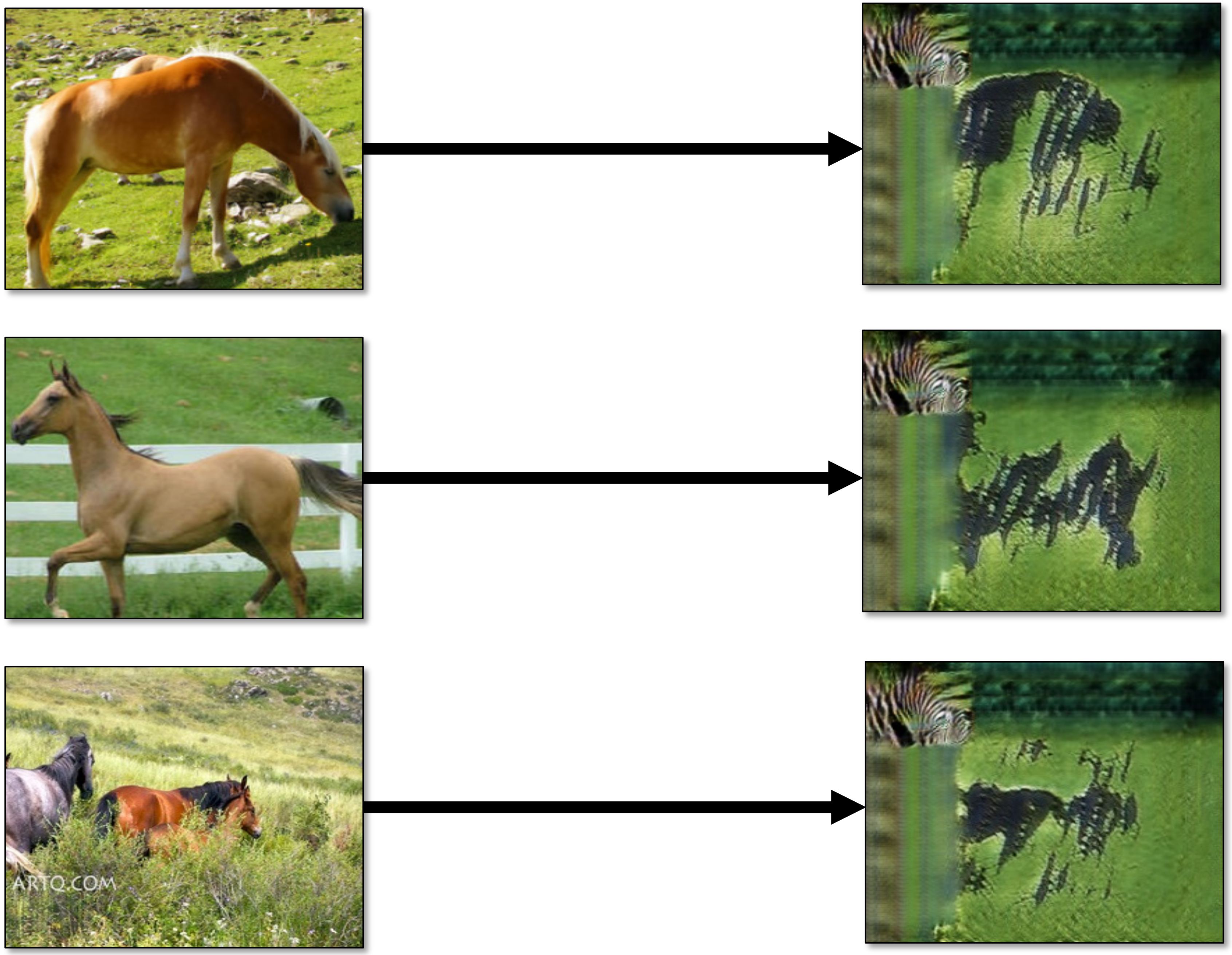


Discriminator

Y



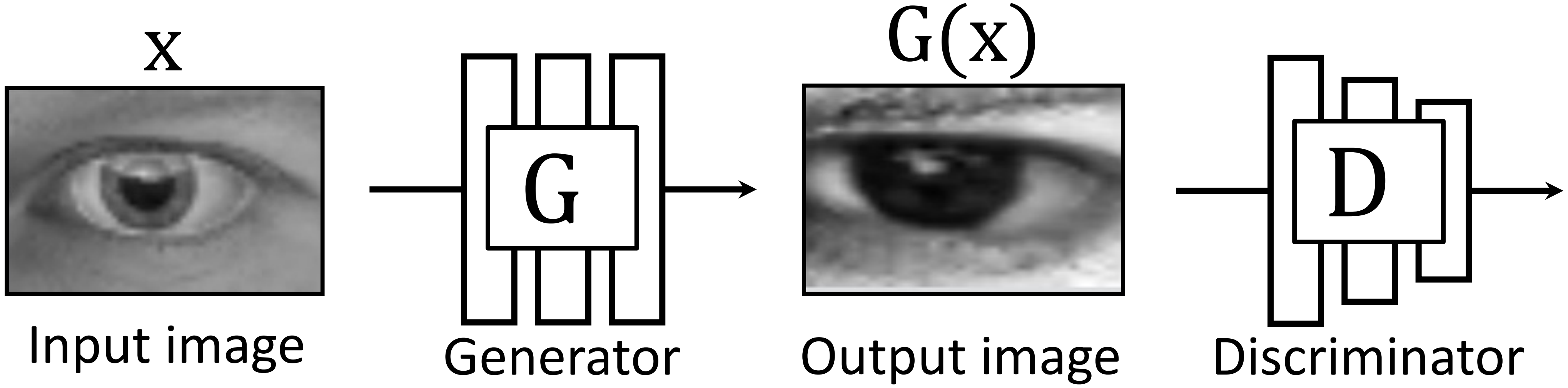
Unsupervised Learning of $p(y \mid x)$



- artifacts
- ignore inputs

[Goodfellow et al. 2014]

Additional Constraint: Identity Mapping

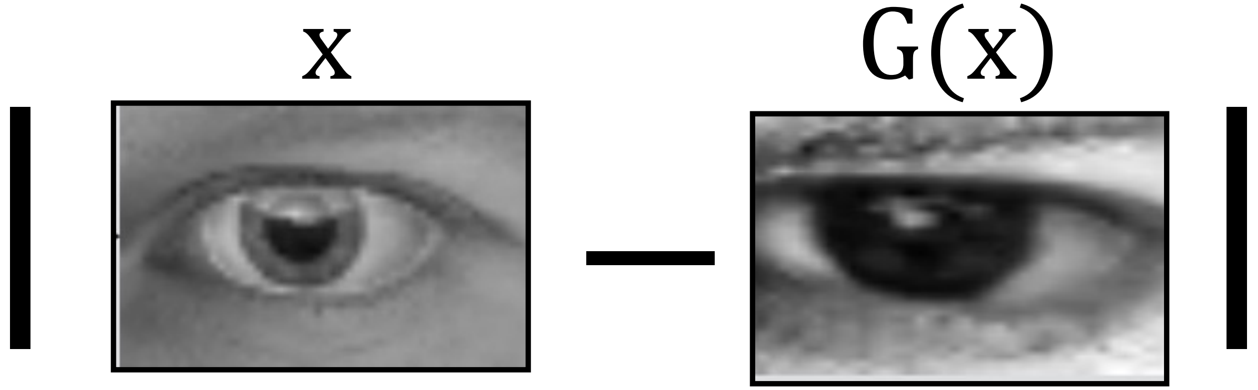


Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

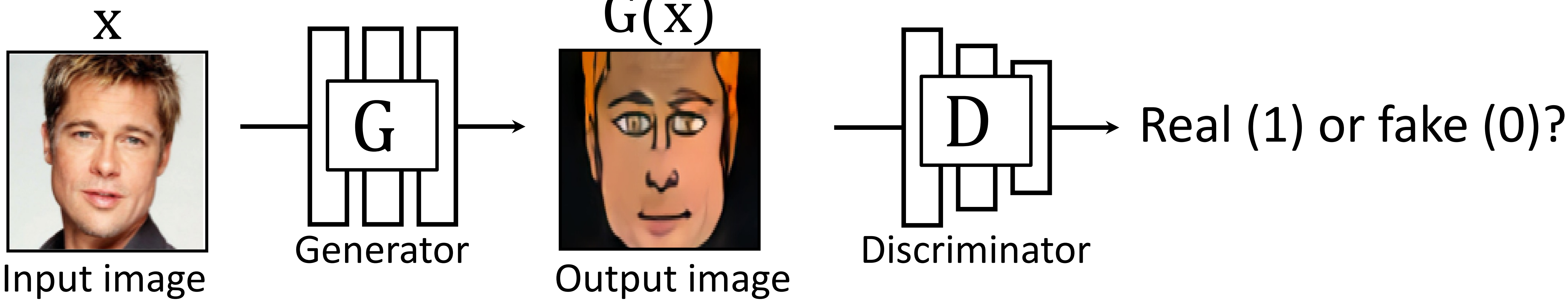
Self-Regularization loss

$$\mathbb{E}_x ||G(x) - x||_1$$



SimGAN [Shrivastava et al., 2017]

Additional Constraint: Feature Loss

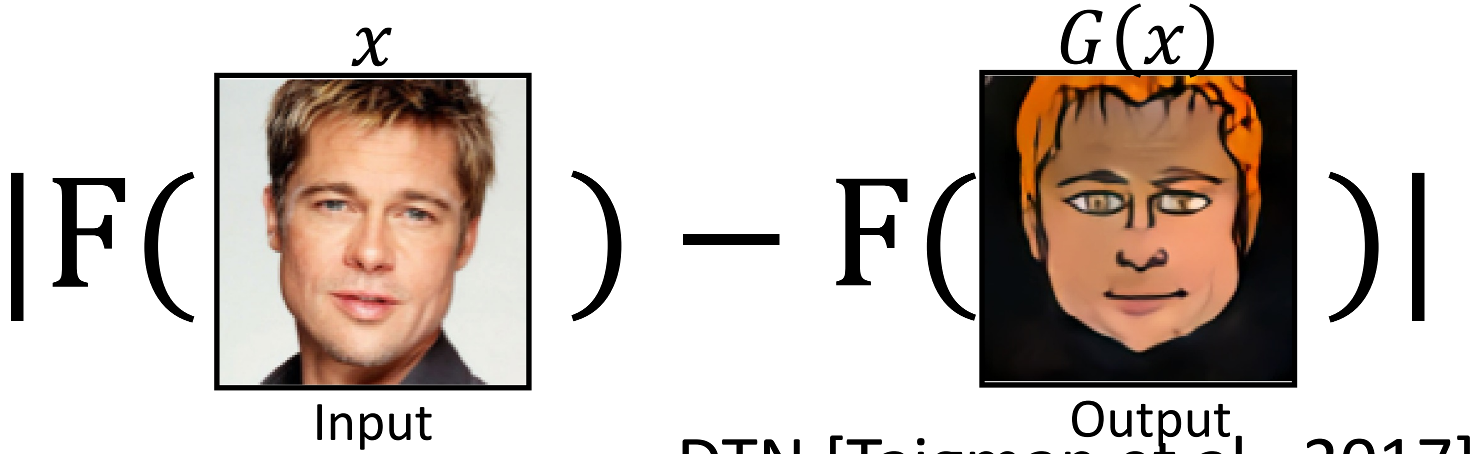


Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Feature loss

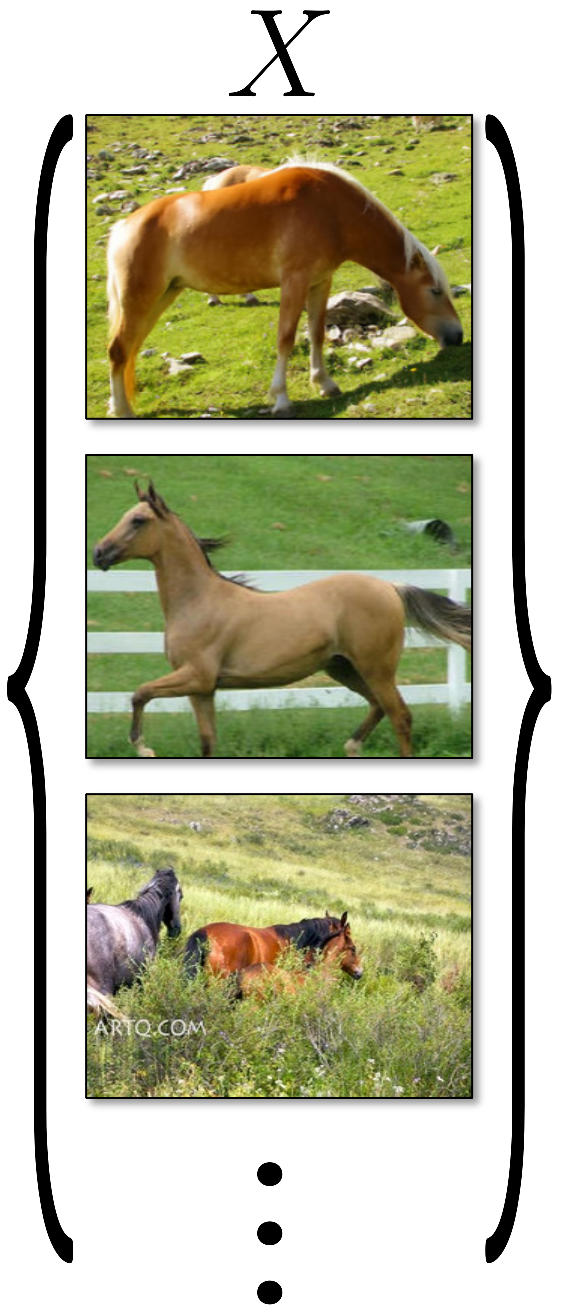
$$\mathbb{E}_x ||F(G(x)) - F(x)||$$

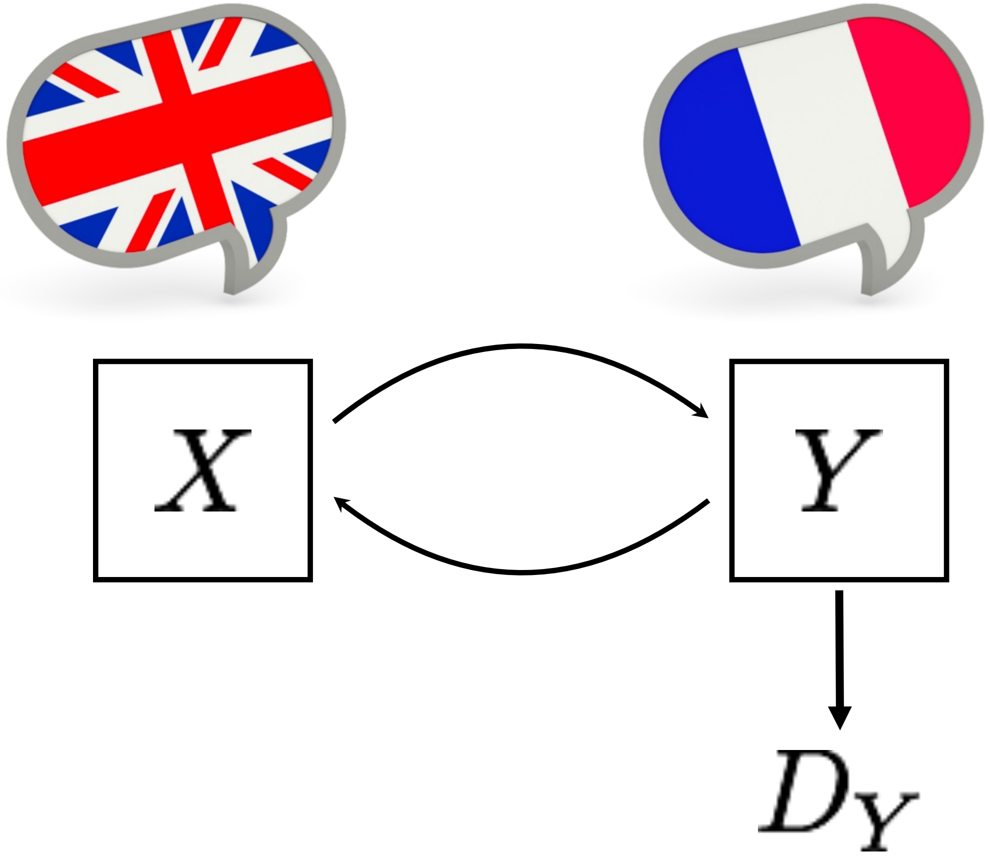


Requires F to work across two domains

DTN [Taigman et al., 2017]

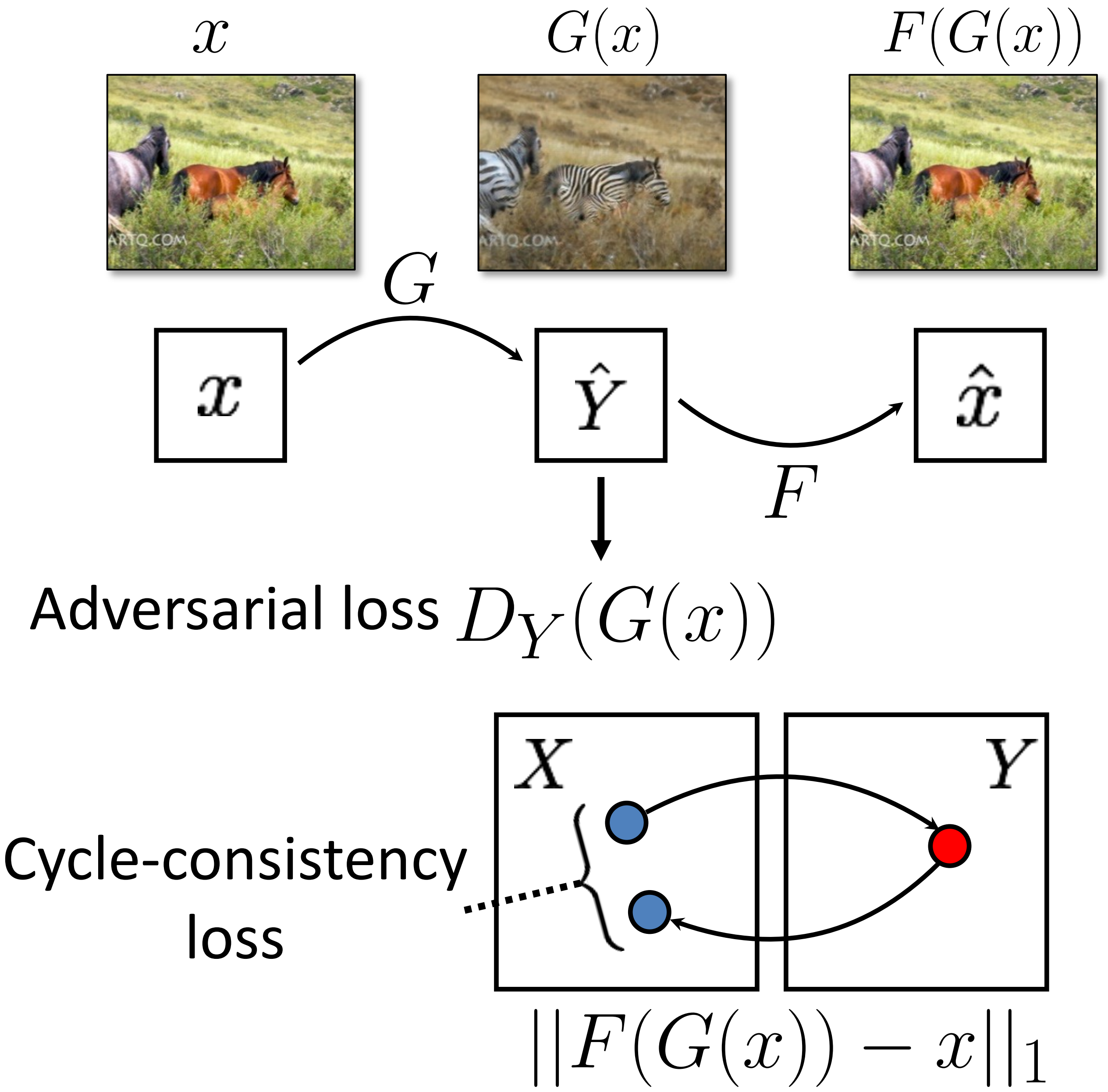
Additional Constraint: Cycle-Consistency





CycleGAN [Zhu*, Park* et al., ICCV 2017]

Cycle-Consistent Adversarial Networks



Adversarial loss

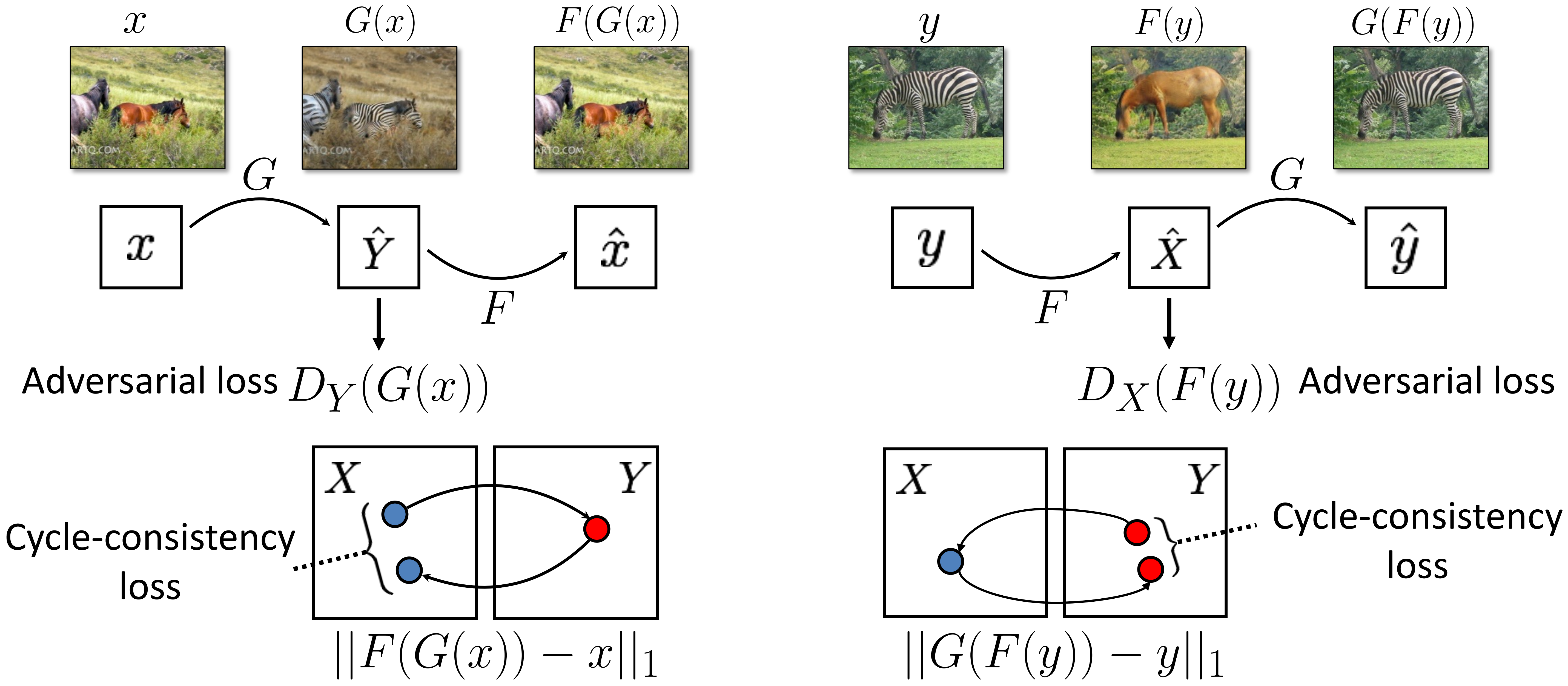
$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Cycle-consistency loss

$$\mathbb{E}_x ||F(G(x)) - x||_1$$

CycleGAN [Zhu*, Park* et al., ICCV 2017]

Cycle-Consistent Adversarial Networks



CycleGAN [Zhu*, Park* et al., ICCV 2017]

Results

Horse → Zebra

Orange → Apple

Monet's paintings → photographic style

Monet's paintings → photographic style

Collection Style Transfer

Photograph ©Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e

Improving the Realism of CG Rendering

CG Game: Grand Theft Auto

Street view images in German cities

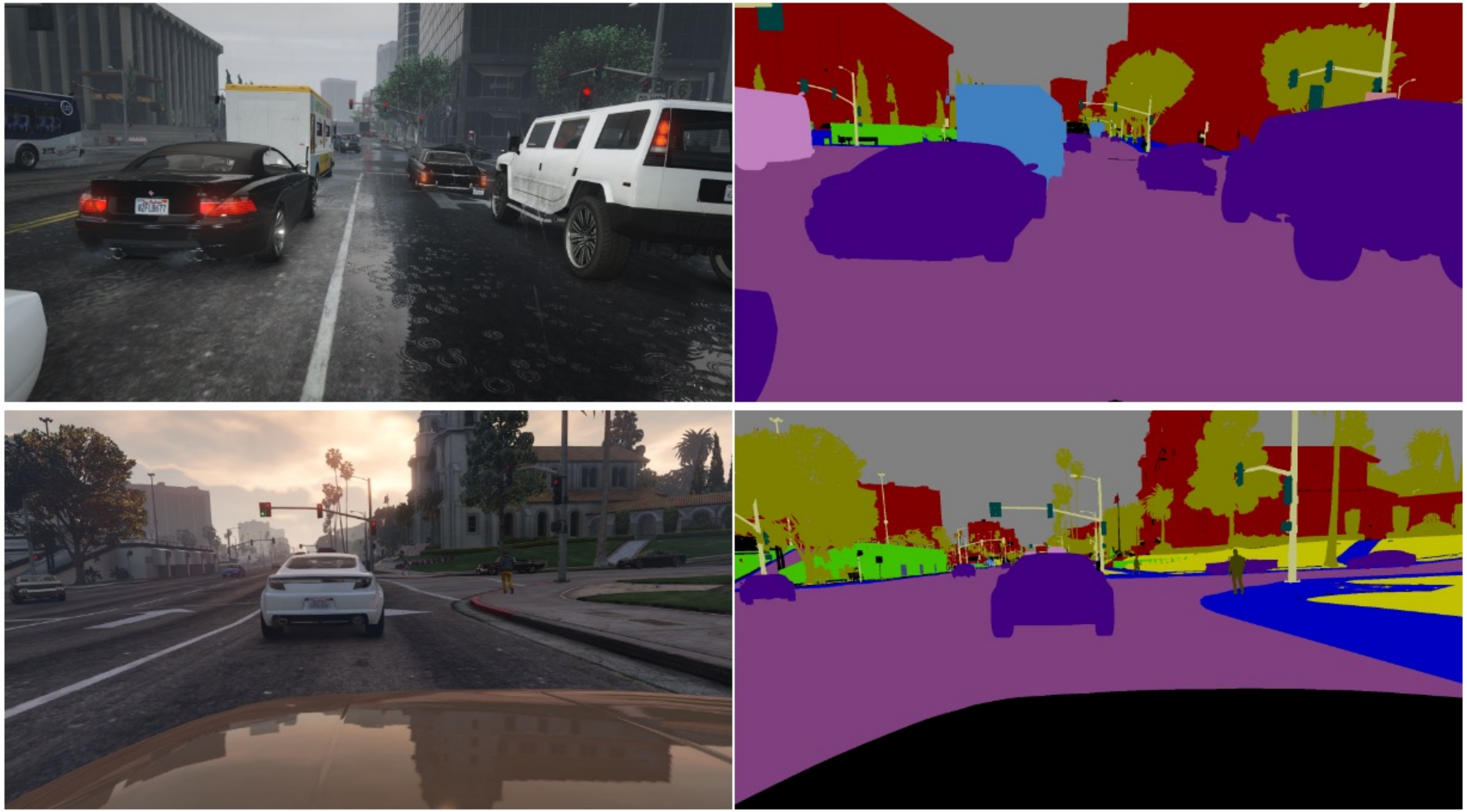
Data from [Richter et al., 2016], [Cordts et al, 2016]

Improving the Realism of CG Rendering

Output image with German street view style

Input CG image

Domain Adaptation with CycleGAN



CG images | Free segmentation labels

Data and labels from [Richter et al. 2016]

Domain Adaptation with CycleGAN

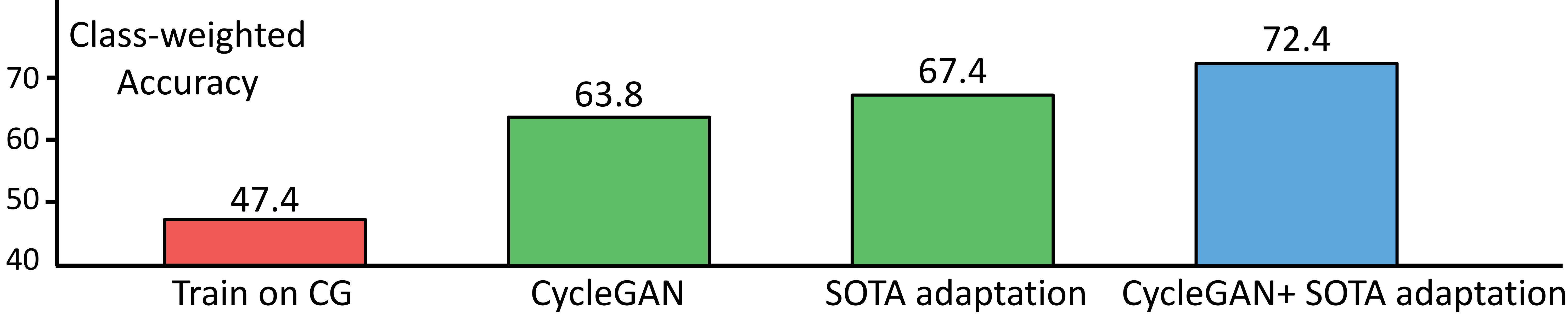
Train on CG data

Test on real images

Domain Adaptation with CycleGAN

Train on CycleGAN images

Test on real images



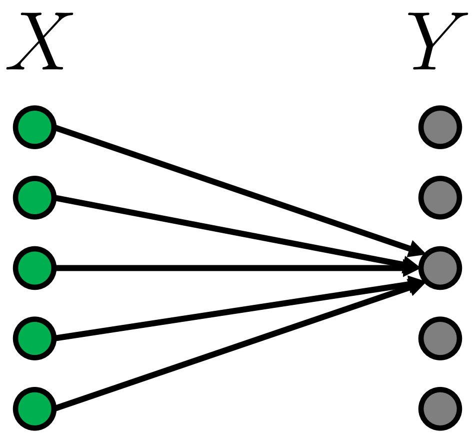
CycleGAN [Zhu*, Park* et al., ICCV 2017]
CycADA [Hoffman et al.,. ICML 2018]

Why CycleGAN works

Why CycleGAN works

Adversarial loss

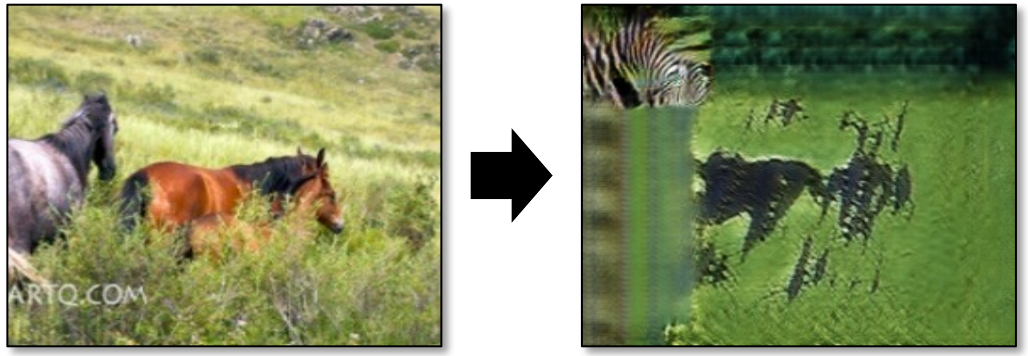
$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$



Why CycleGAN works

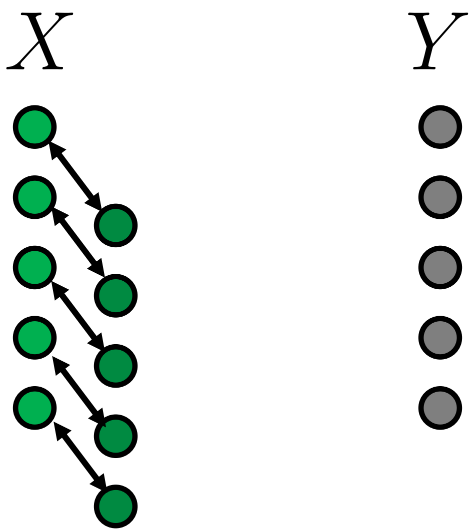
Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$



Cycle-consistency loss

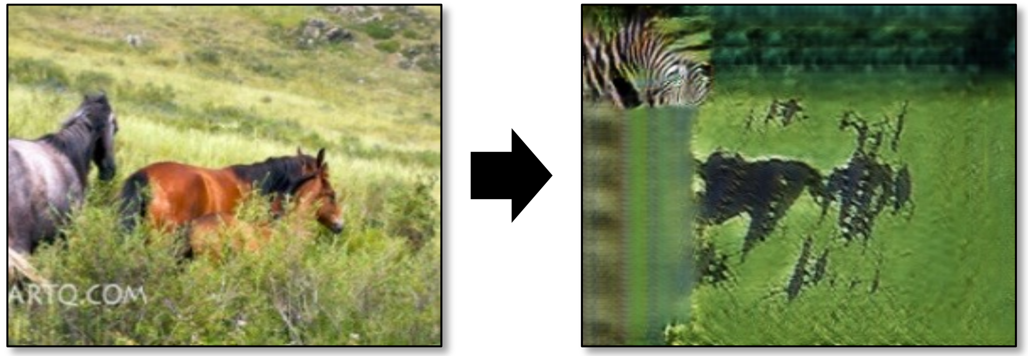
$$\mathbb{E}_x ||F(G(x)) - x||_1$$



Why CycleGAN works

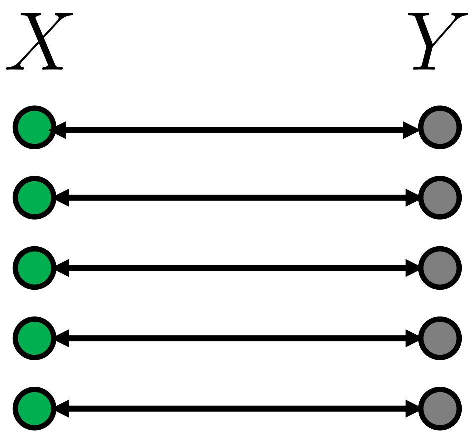
Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$



Cycle-consistency loss

$$\mathbb{E}_x ||F(G(x)) - x||_1$$



Full objective

Why CycleGAN works

Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Cycle-consistency loss

$$\mathbb{E}_x ||F(G(x)) - x||_1$$

x

$G(x)$

$F(G(x))$

Auto-encoder w/ domain prior

code

$$x \xrightarrow{\text{Encoder } G} \hat{Y} \xrightarrow{\text{Decoder } F} \hat{x}$$

Constraint: $\text{len}(x) = 30$ $G(x) \sim p_{data}(Y)$

[Hinton and Salakhutdinov. Science 2006]

Why CycleGAN works

Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Under-constrained problem

Cycle-consistency loss

$$\mathbb{E}_x ||F(G(x)) - x||_1$$

Prior of G

<u>A strong regularizer</u>

Assumption: simple invertible function

Probabilistic Interpretation : Upper bound of conditional entropy $H(y|x)$

[Li et al. 2017]

Why CycleGAN works

Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Cycle-consistency loss

$$\mathbb{E}_x ||F(G(x)) - x||_1$$

flip the image

$P \circ G$

$F \circ P^{-1}$

flip the image again

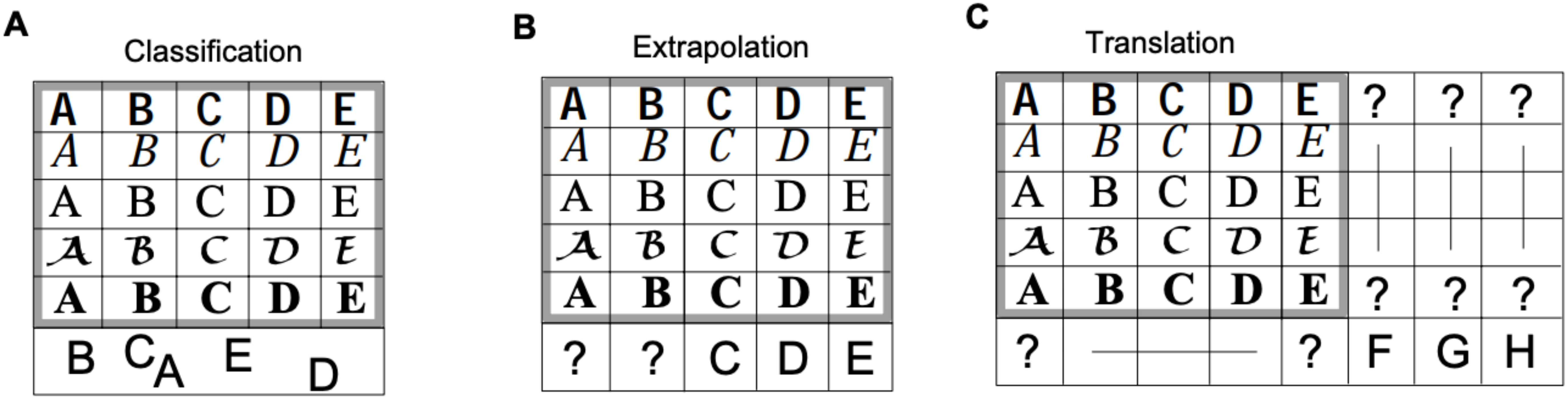
<u>Invertible Perturbation</u>

Adversarial loss: images are horizontally symmetric

Cycle-consistency loss : $||F \circ P^{-1}(P \circ G(x)) - x||$

Style and Content Disentanglement

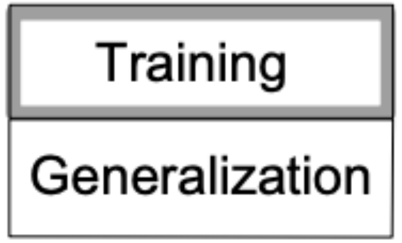
Style and Content Separation



Domain Adaptation

Paired Image-to-Image Translation

Unpaired Image-to-Image Translation



Separating Style and Content
[Tenenbaum and Freeman 1996]

$$y_k^{sc} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ijk} a_i^s b_j^c .$$

Style and Content

Adversarial loss

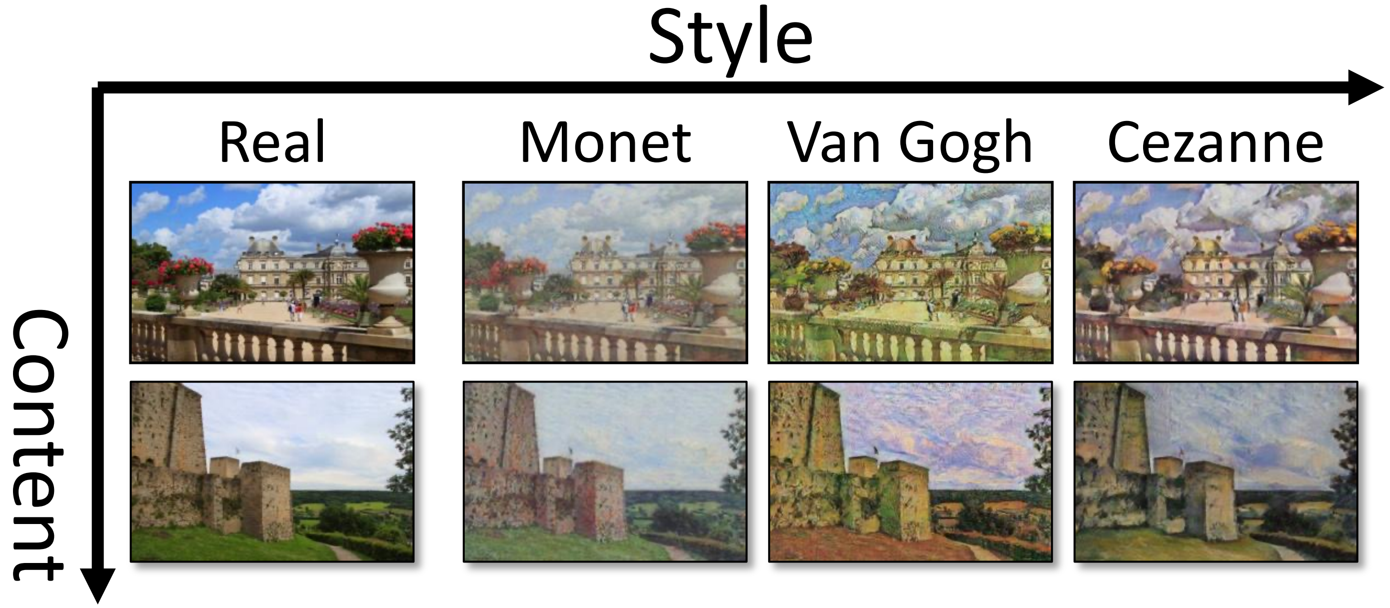
$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

$p(x) \rightarrow p(y)$ change **style**

Cycle-consistency loss

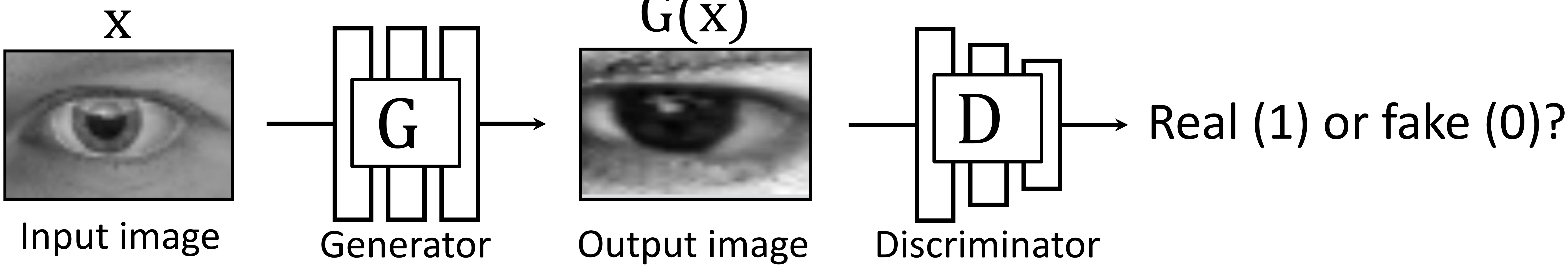
$$\mathbb{E}_x ||F(G(x)) - x||_1$$

Bidirectional: preserve **content**



Separating Style and Content
[Tenenbaum and Freeman 1996]

Style and Content

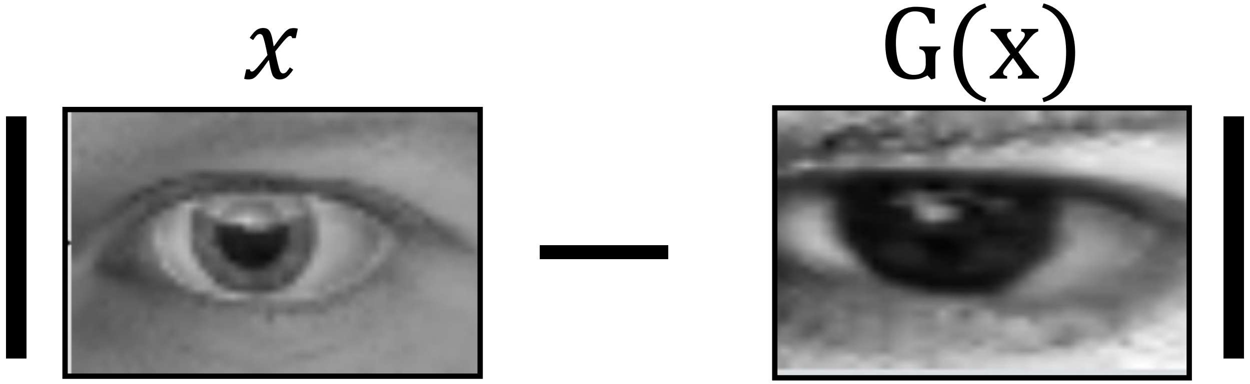


Adversarial loss (change style)

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

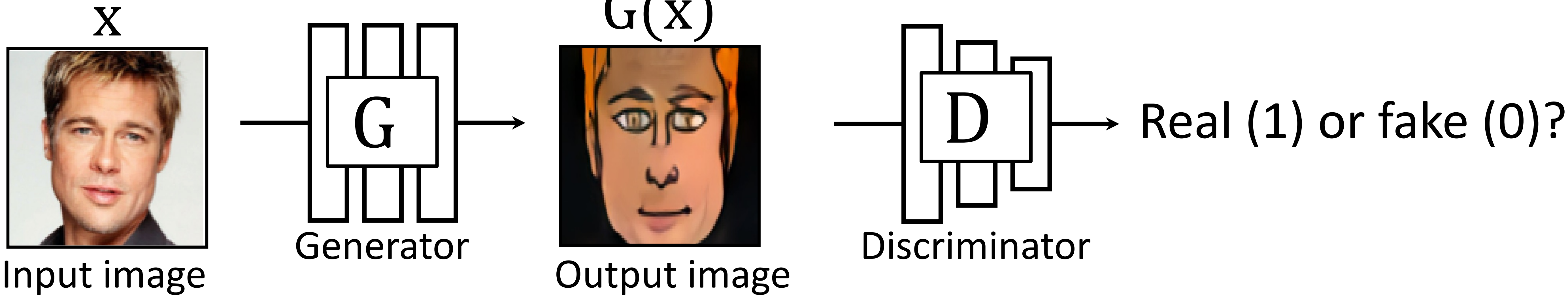
L1 loss (preserve content in pixel space)

$$\mathbb{E}_x ||G(x) - x||_1$$



SimGAN [Shrivastava et al., 2017]

Style and Content

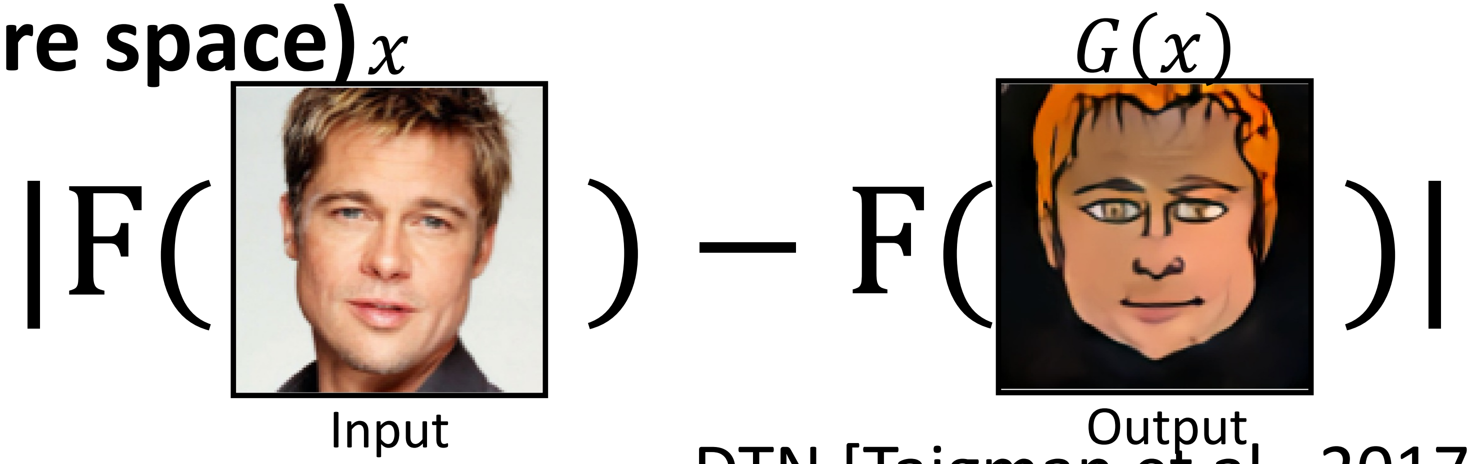


Adversarial loss (change style)

$$\mathbb{E}_x \log(1-D_Y(G(x)))+\mathbb{E}_y \log D_Y(y)$$

Feature loss (Preserve content in feature space)

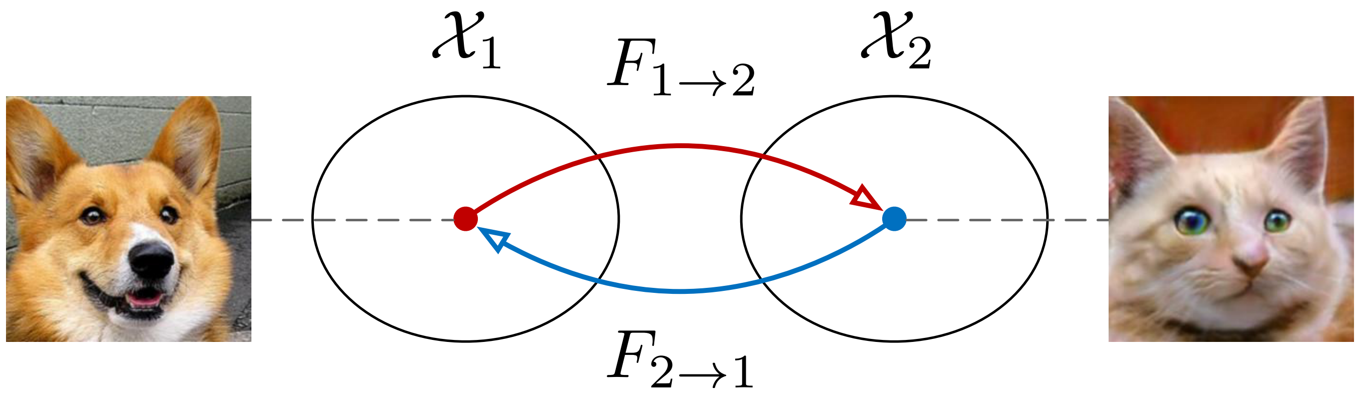
$$\mathbb{E}_x||F(G(x)) - F(x)||$$

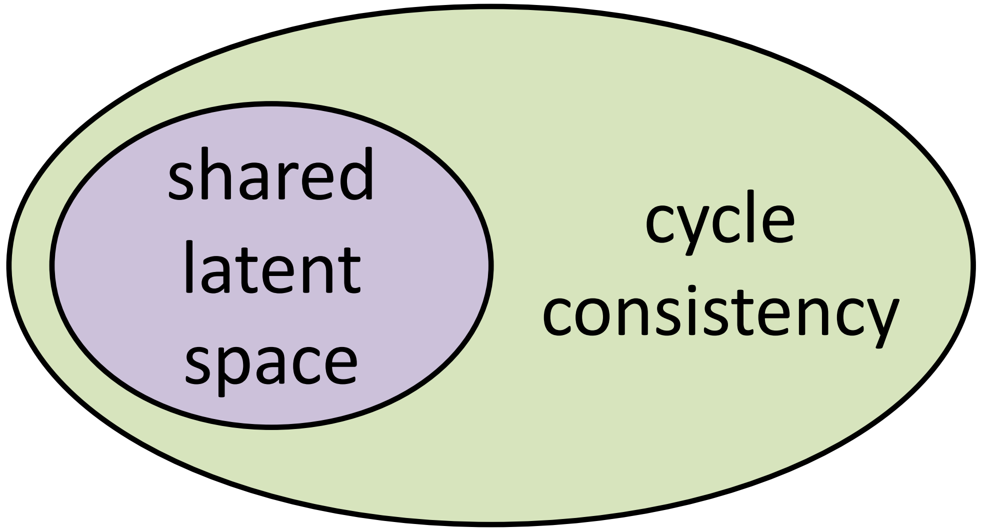


DTN [Taigman et al., 2017]

CycleGAN and UNIT

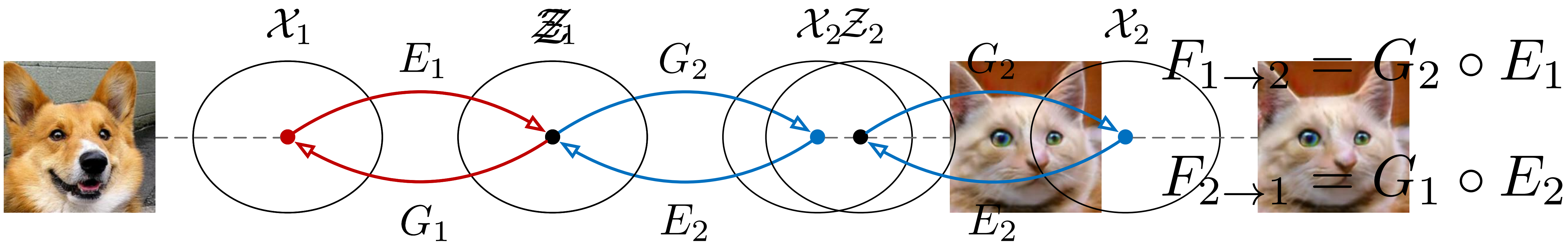
- CycleGAN (**cycle consistency**)





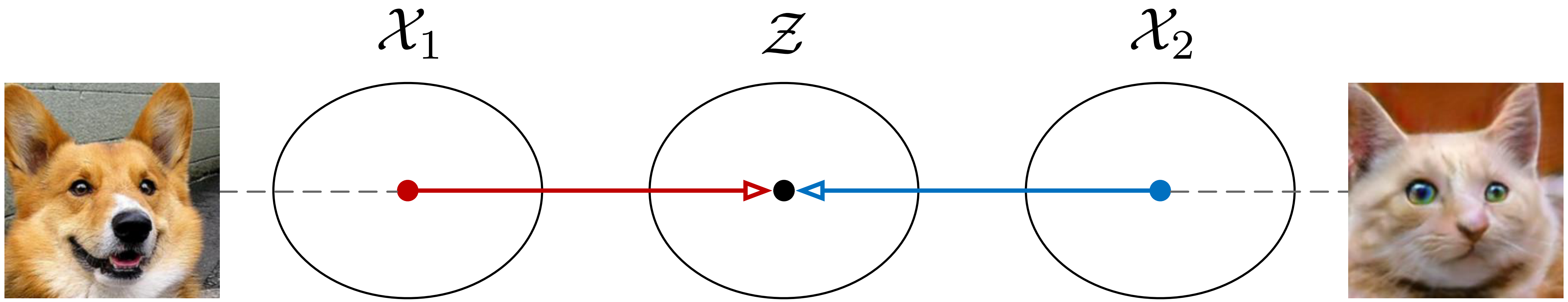
- UNIT (**shared latent space**) [Liu et al. 2017]

shared latent space $\Longrightarrow$ cycle consistency



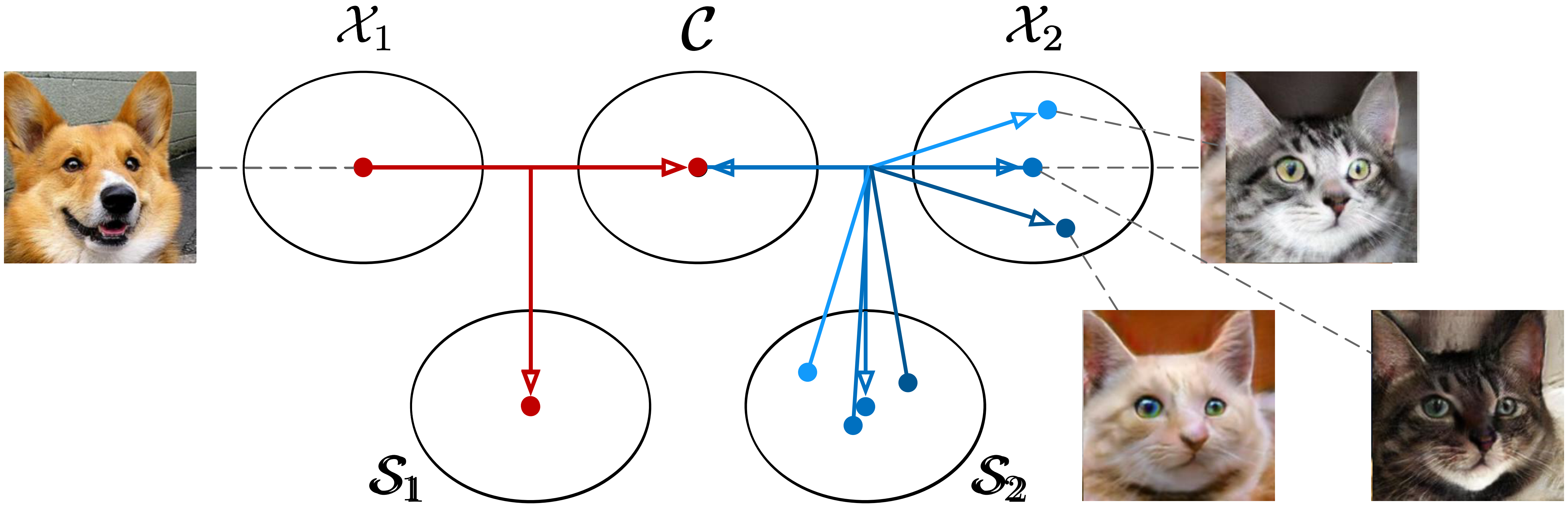
Disentangling the Latent Space

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

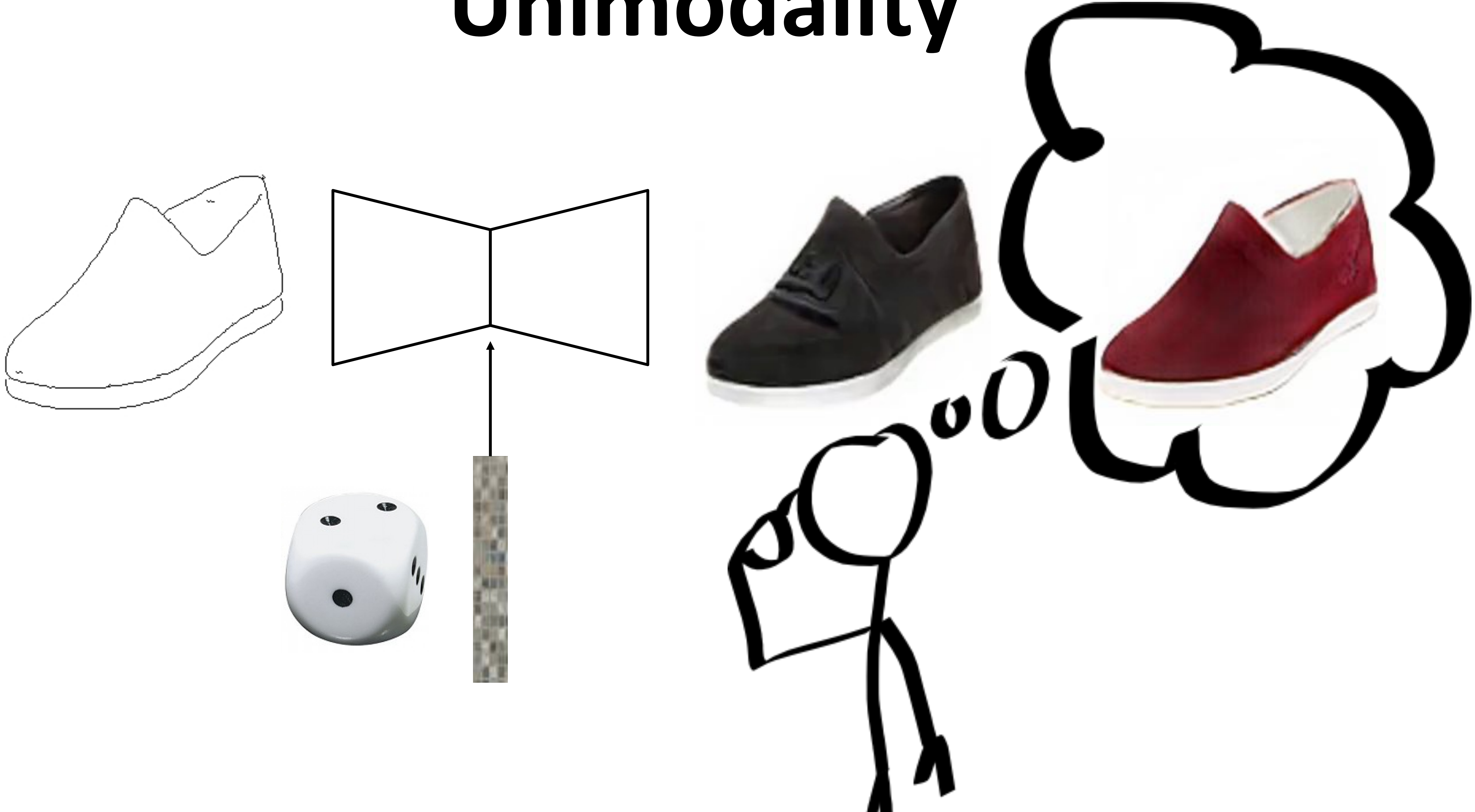


Disentangling the Latent Space

- 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**



Unimodality

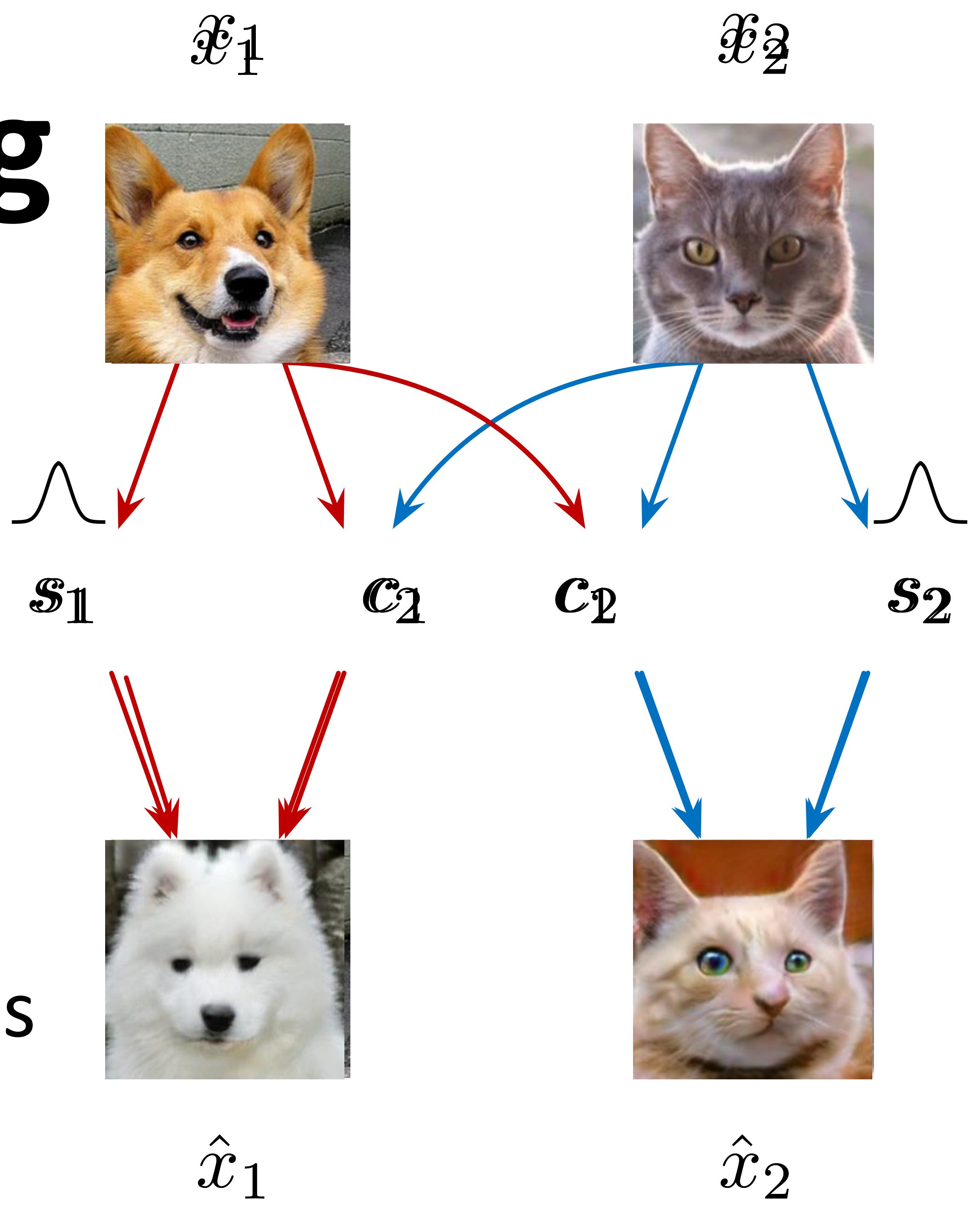


Towards Multimodality

...

Training

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

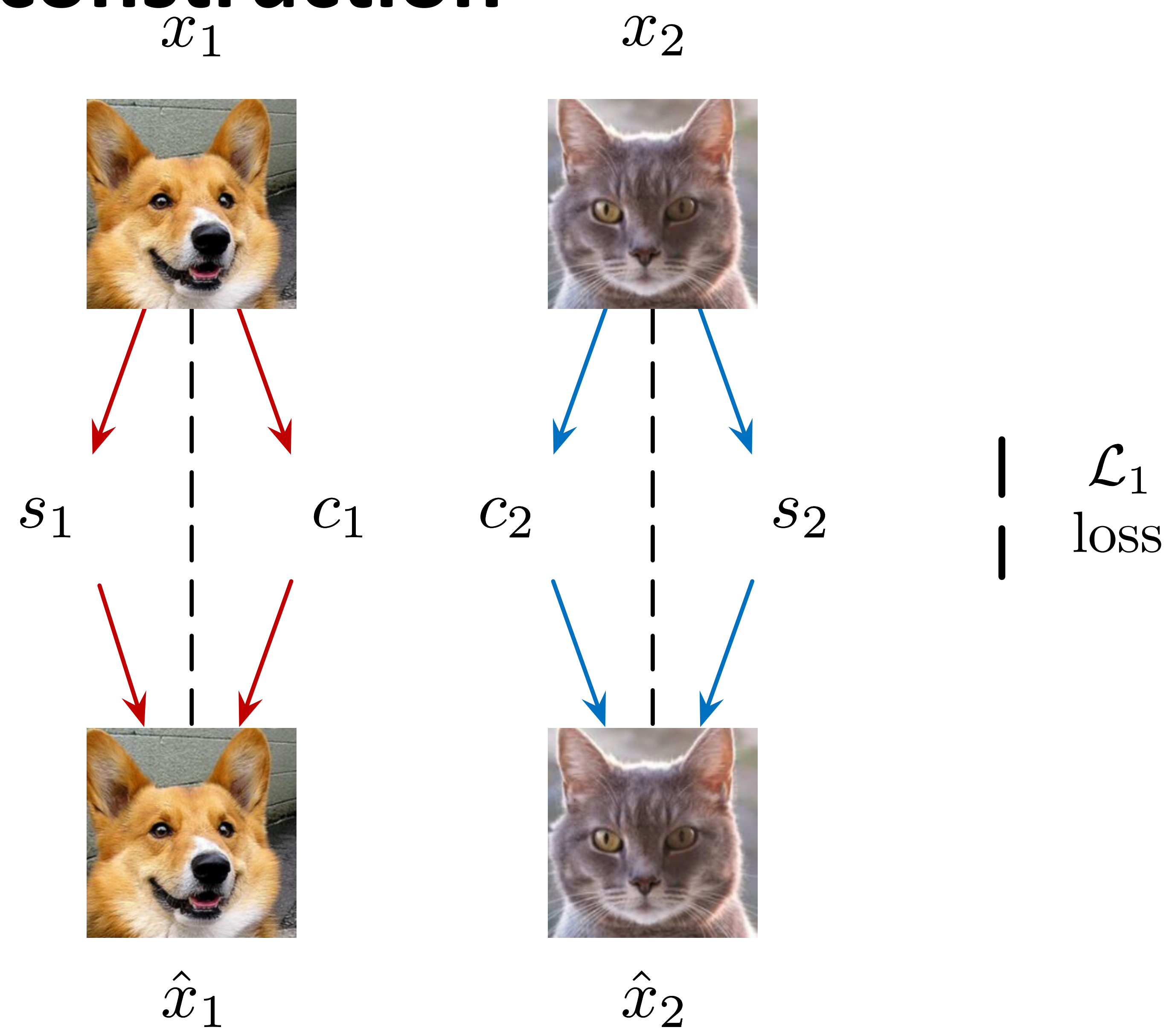


Within-domain reconstruction
Cross-domain translation

Bidirectional Reconstruction Loss: Image Reconstruction

Notations:

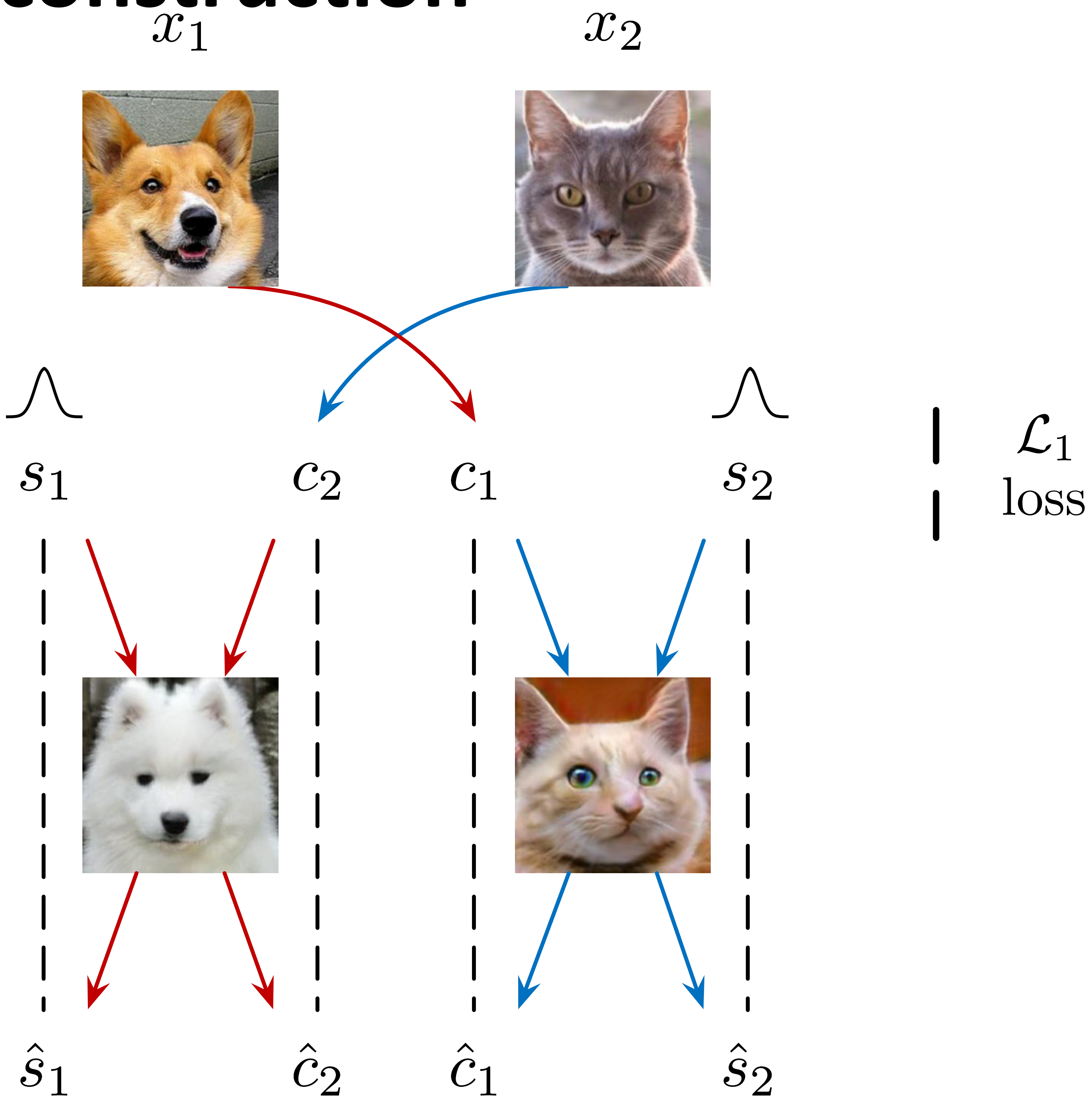
- x: images
- c: content
- s: style



Bidirectional Reconstruction Loss: Image Reconstruction

Notations:

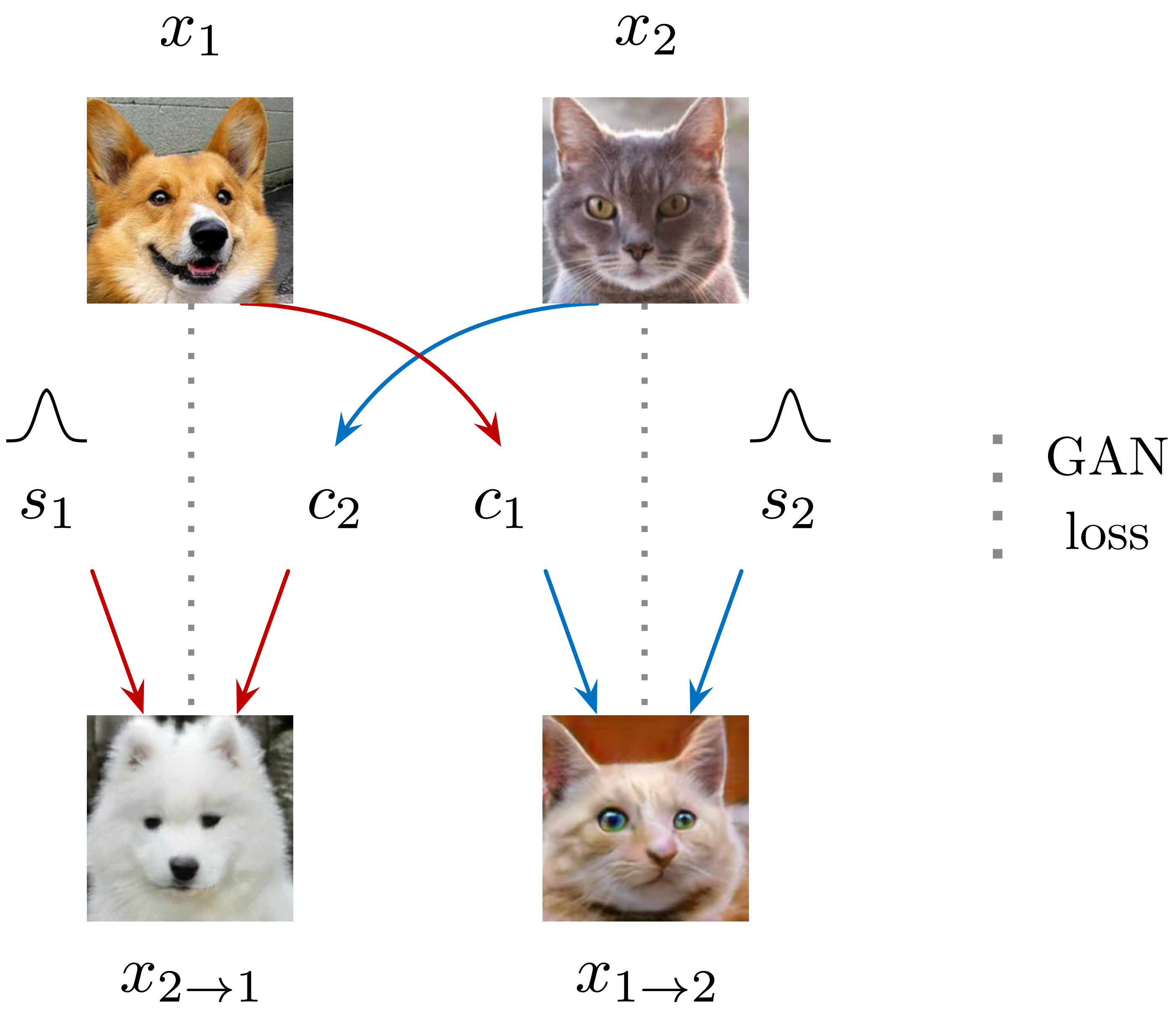
- x: images
- c: content
- s: style



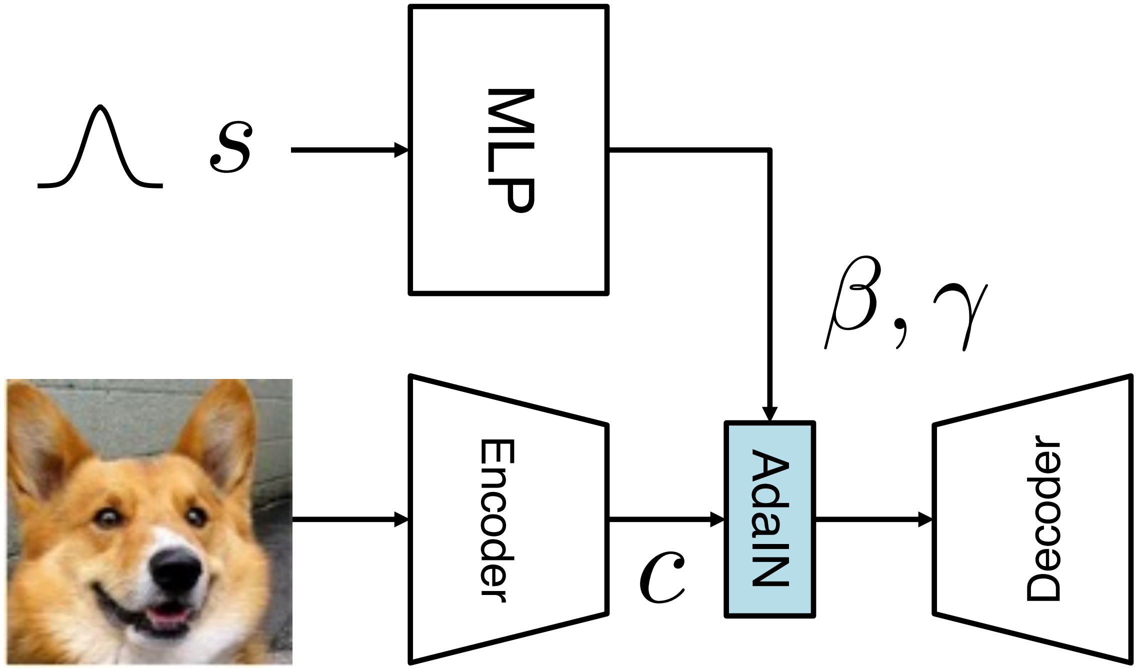
GAN Loss

Notations:

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



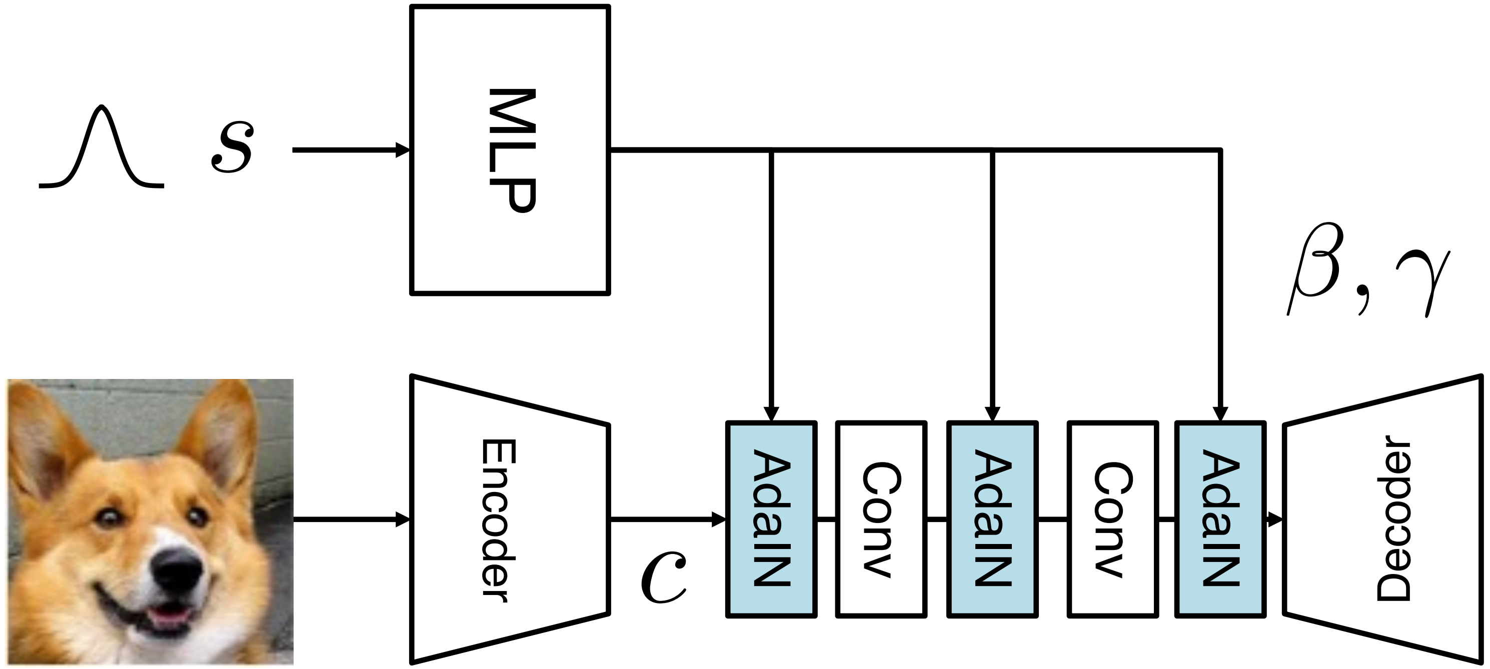
AdaIN in a Generative Network



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

AdaIN in a generative network

AdaIN in a Generative Network



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

AdaIN in a generative network

Sketches <-> Photo

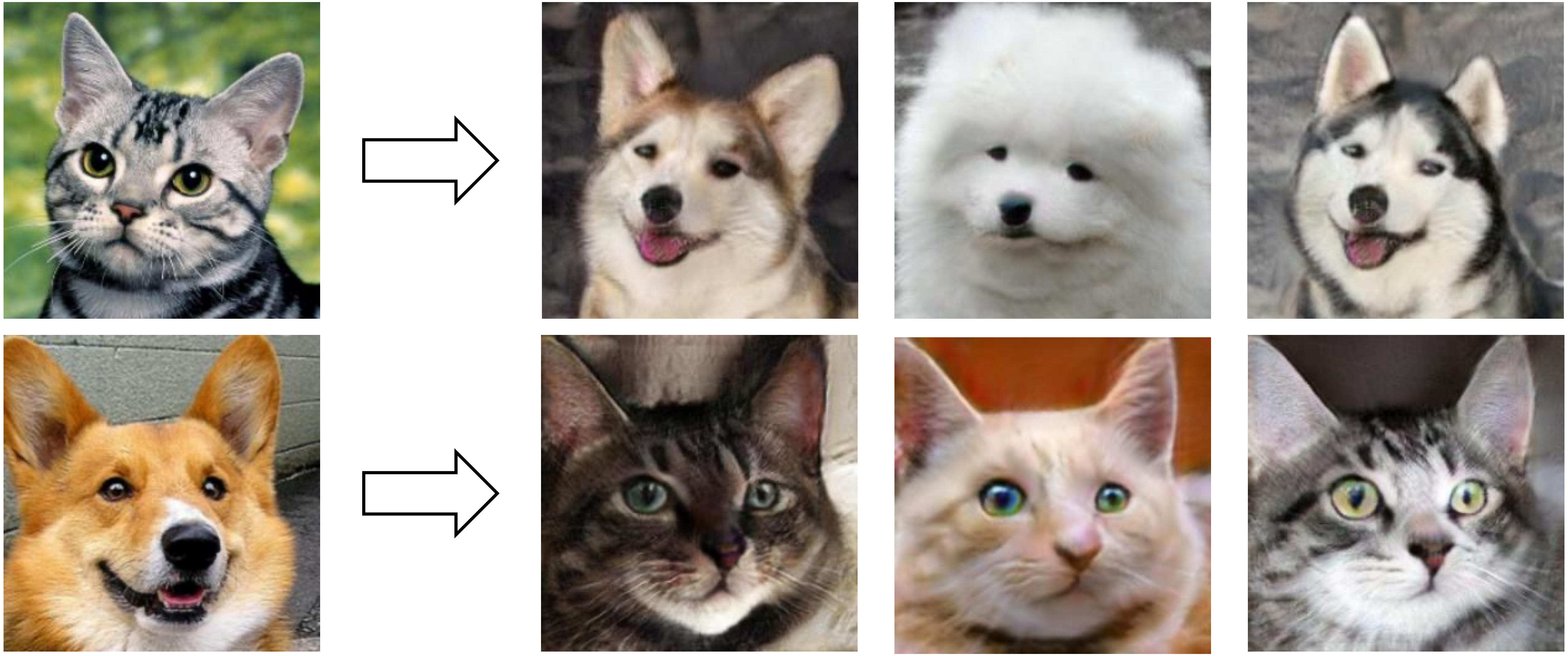
Input

Outputs

Cats ↔ Dogs

Input

Outputs

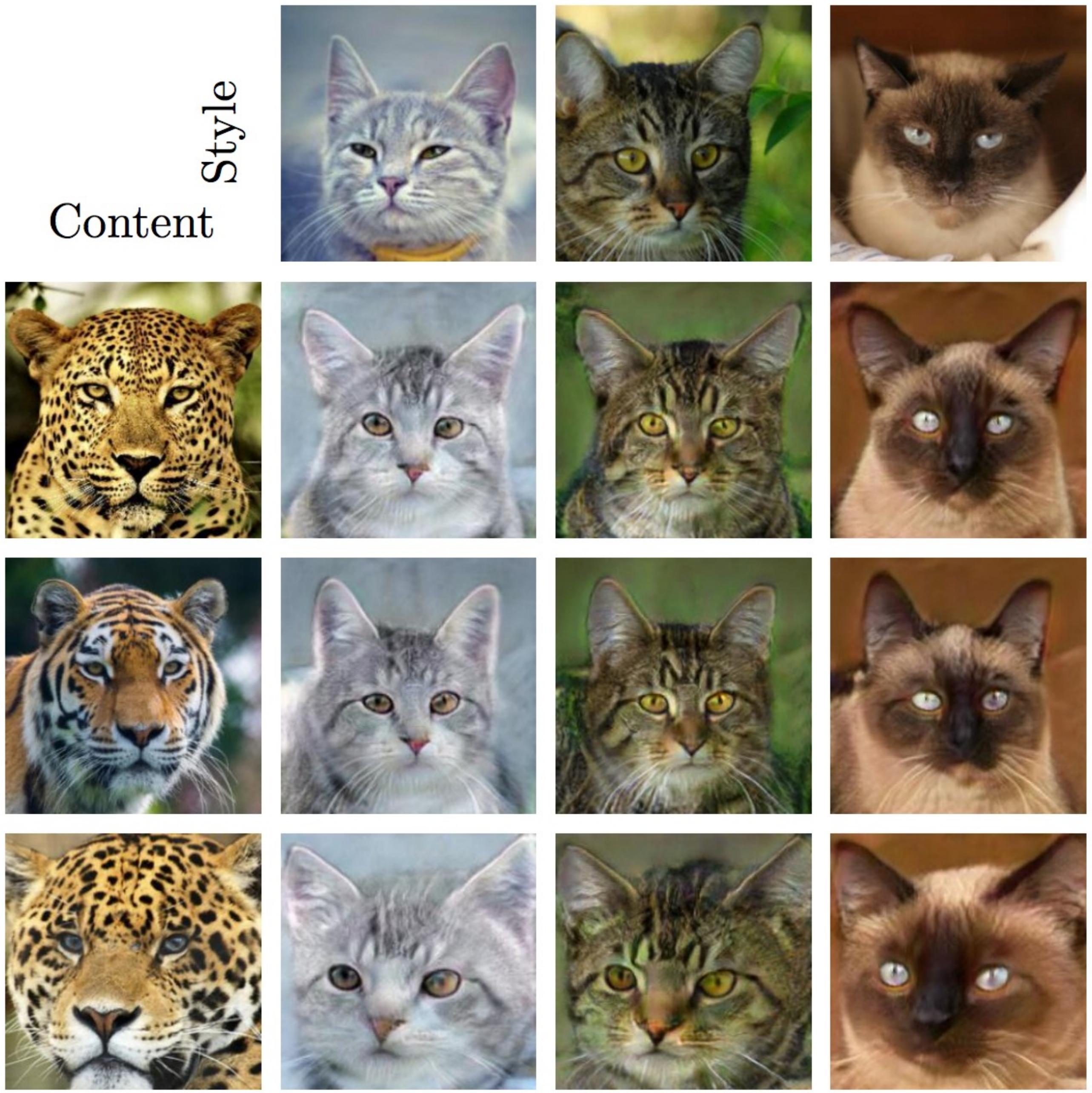


Synthetic ↔ Real

Input

Outputs

Example-guided Translation



Example-guided Translation

Thank You!

16-726, Spring 2023

https://learning-image-synthesis.github.io/