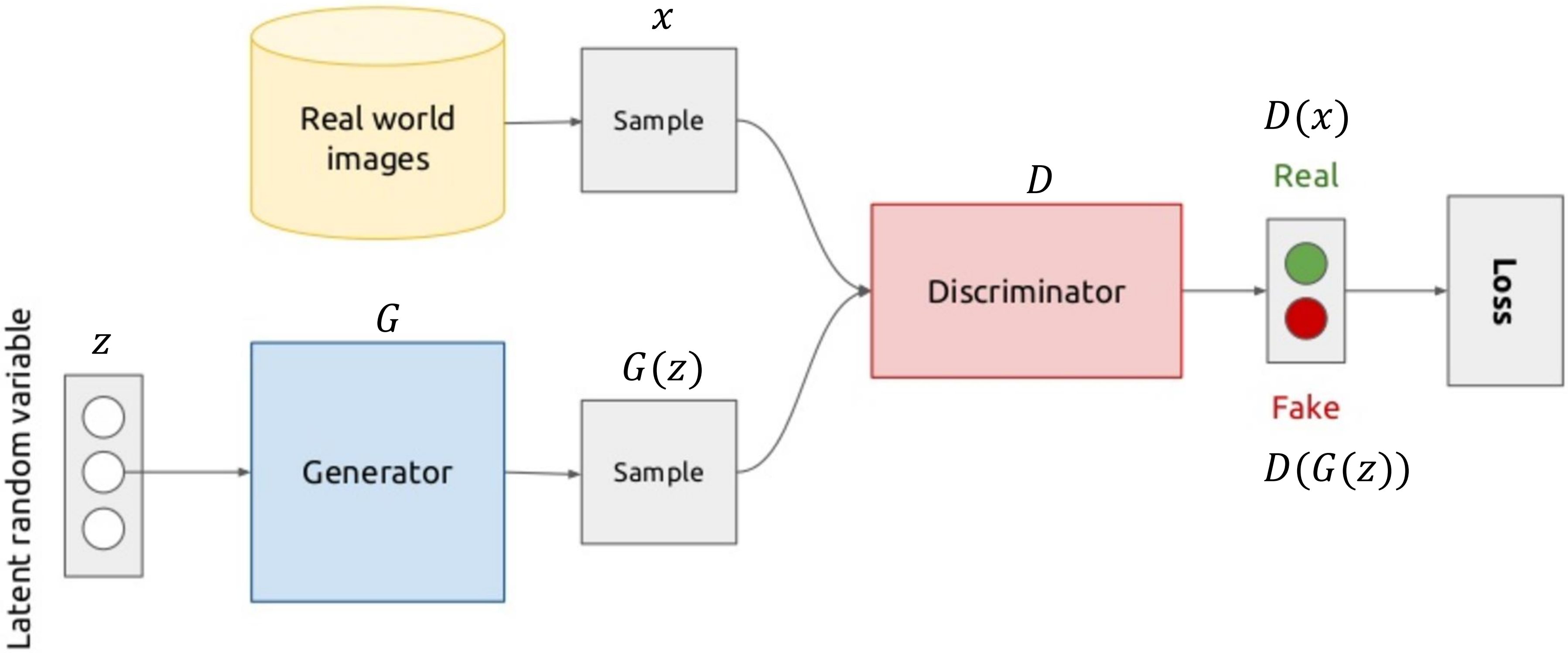
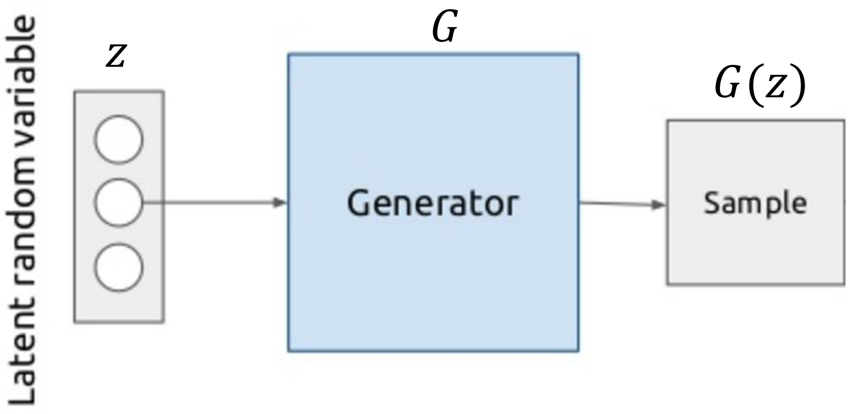
Conditional Generative Adversarial Networks (cGANs)

Generative Adversarial Networks (GANs)



Generative Adversarial Networks (GANs)

At test time: sample random variable -> obtain generated sample



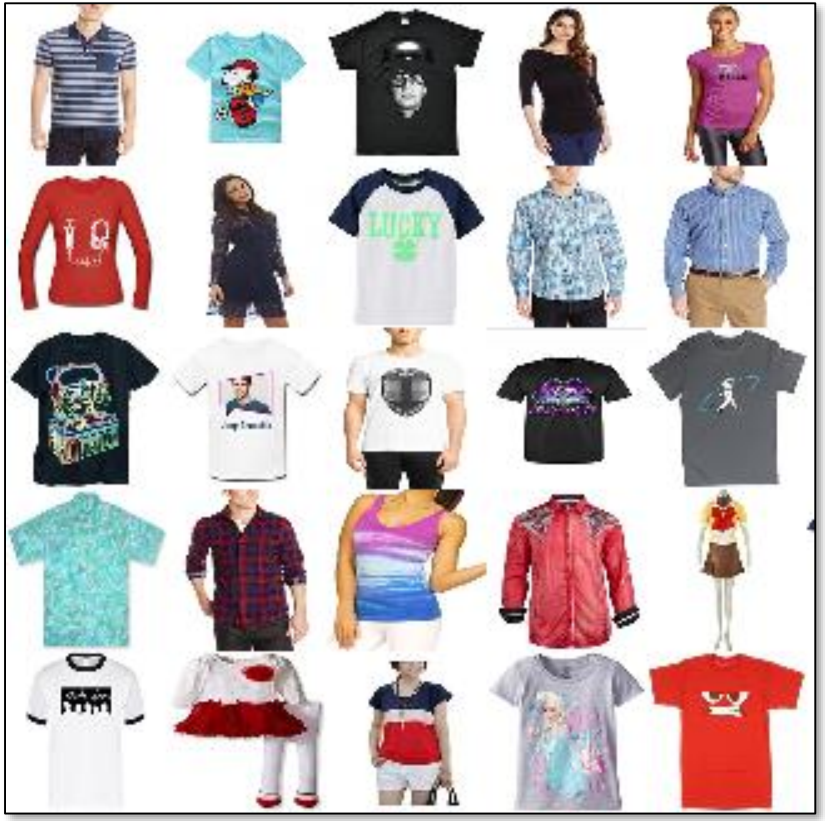
Conditional GANs (cGANs)

- Gain control of output

- Modeling (e.g., sketch-based modeling, etc.)
  - Add semantic meaning to latent space manifold

- Domain transfer
  - Labels on A -> transfer to B, train network on B, test on B
  - More later

GAN Manifold



Train Data

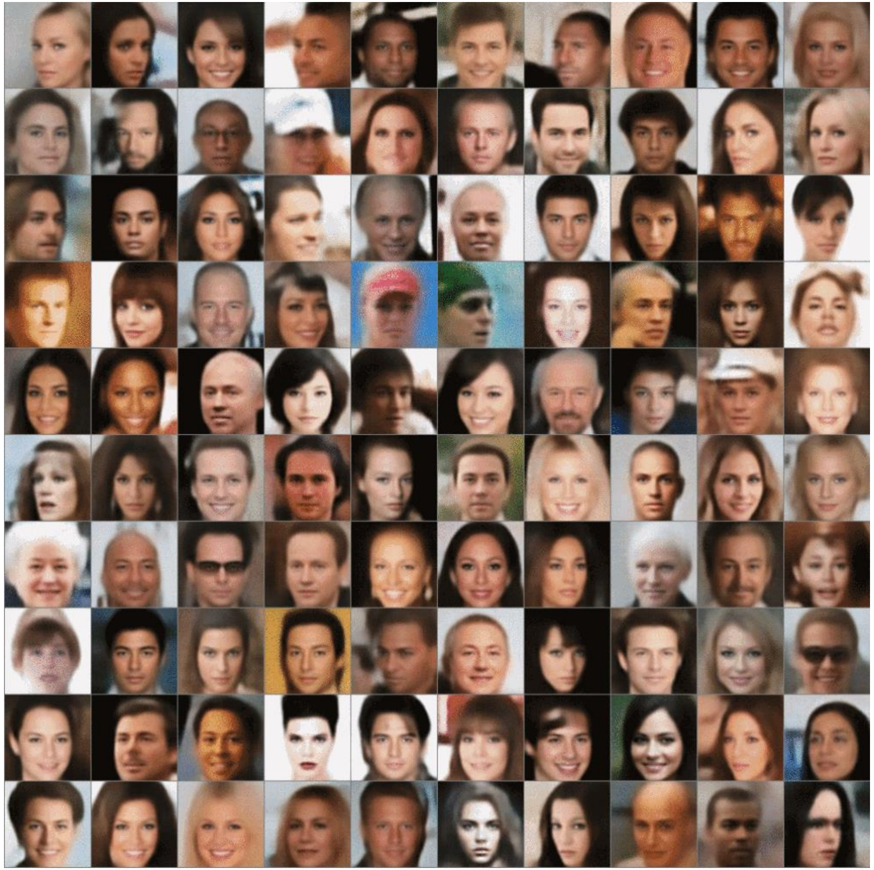
Sampled Data -> G(z)

[Radford et al. 15] Deep Convolutional Generative Adversarial Networks

GAN Manifold

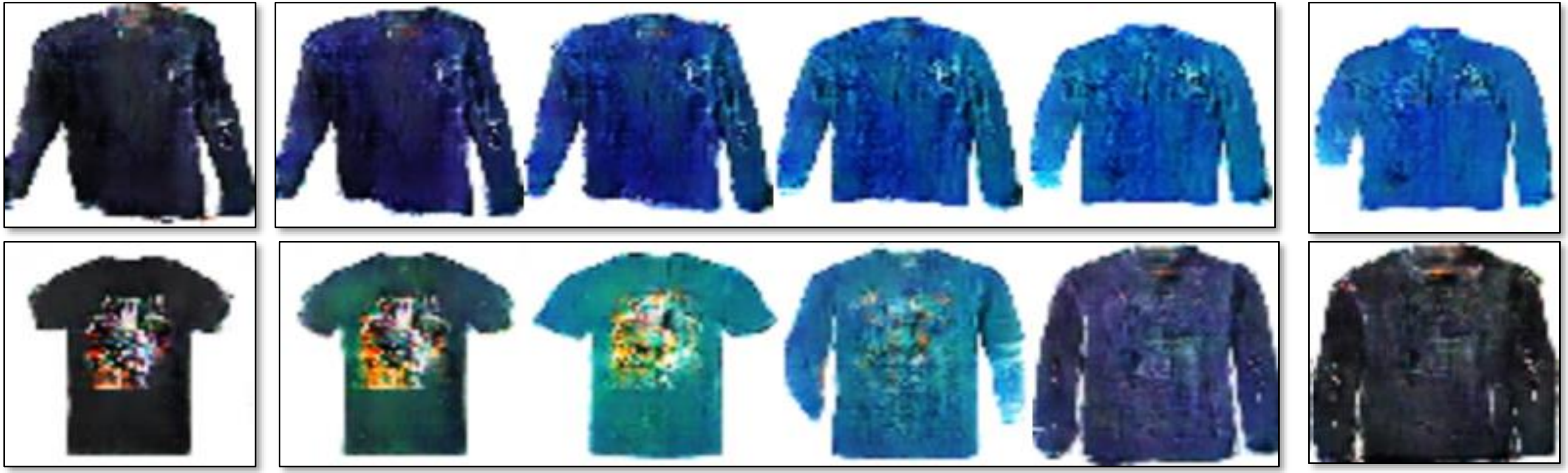
[Bojanowski et al. 17] Optimizing the Latent Space of Generative Networks

GAN Manifold



GAN Manifold

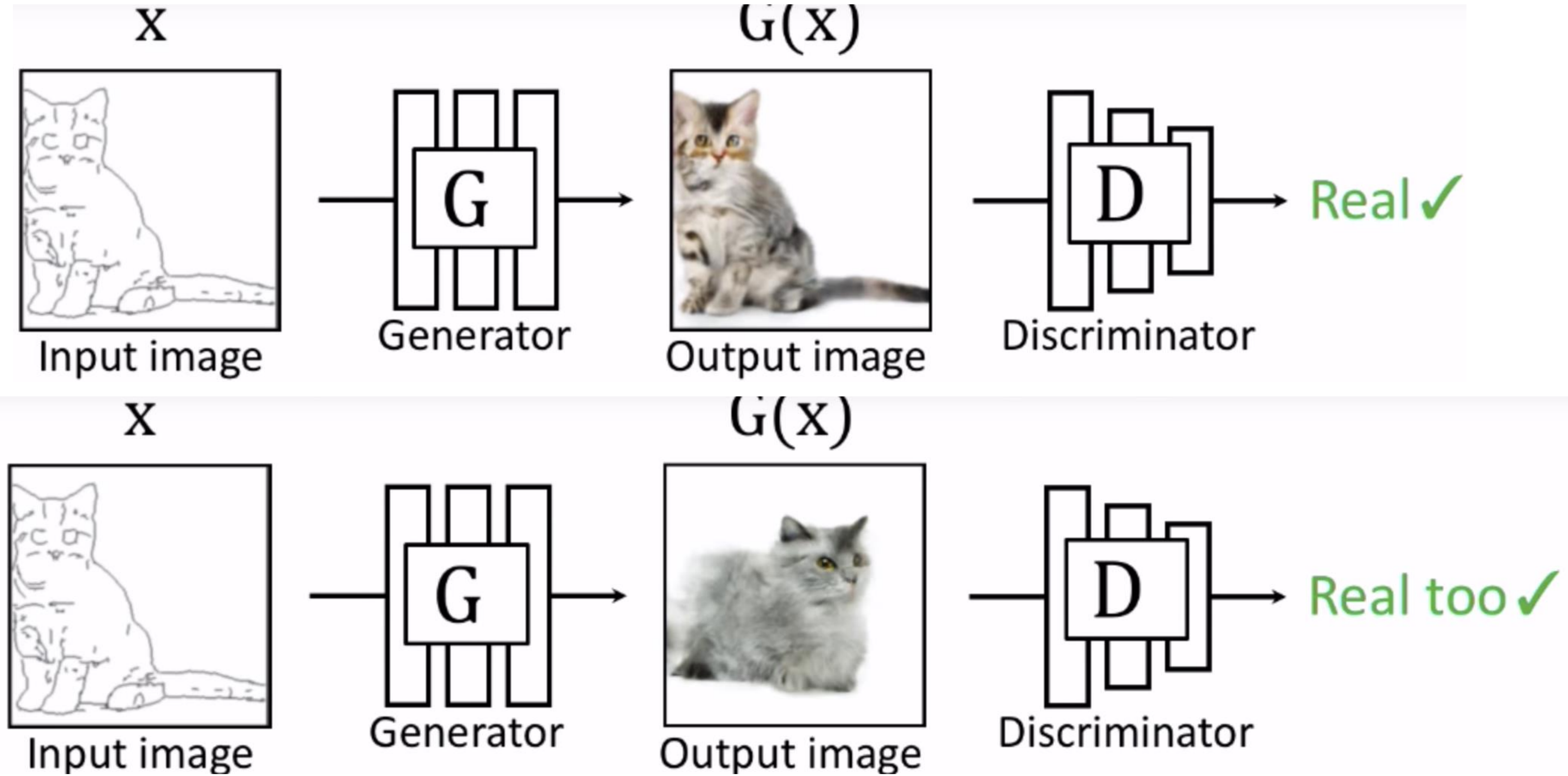
$G(z_0)$ Linear interpolation in z space: $G(z_0 + t \cdot (z_1 - z_0))$ $G(z_1)$



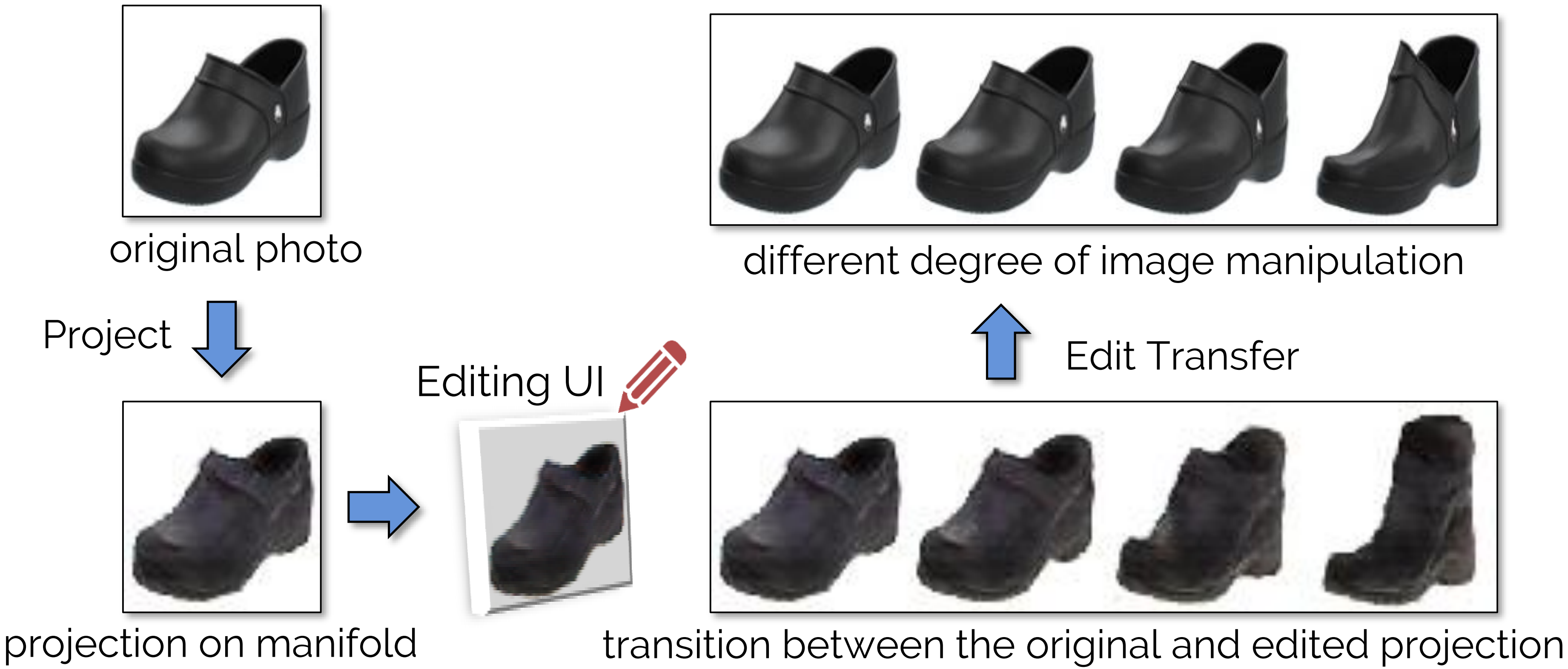
[Radford et al. 15] Deep Convolutional Generative Adversarial Networks

Conditional GANs (cGANs)

don't see input from same data distribution

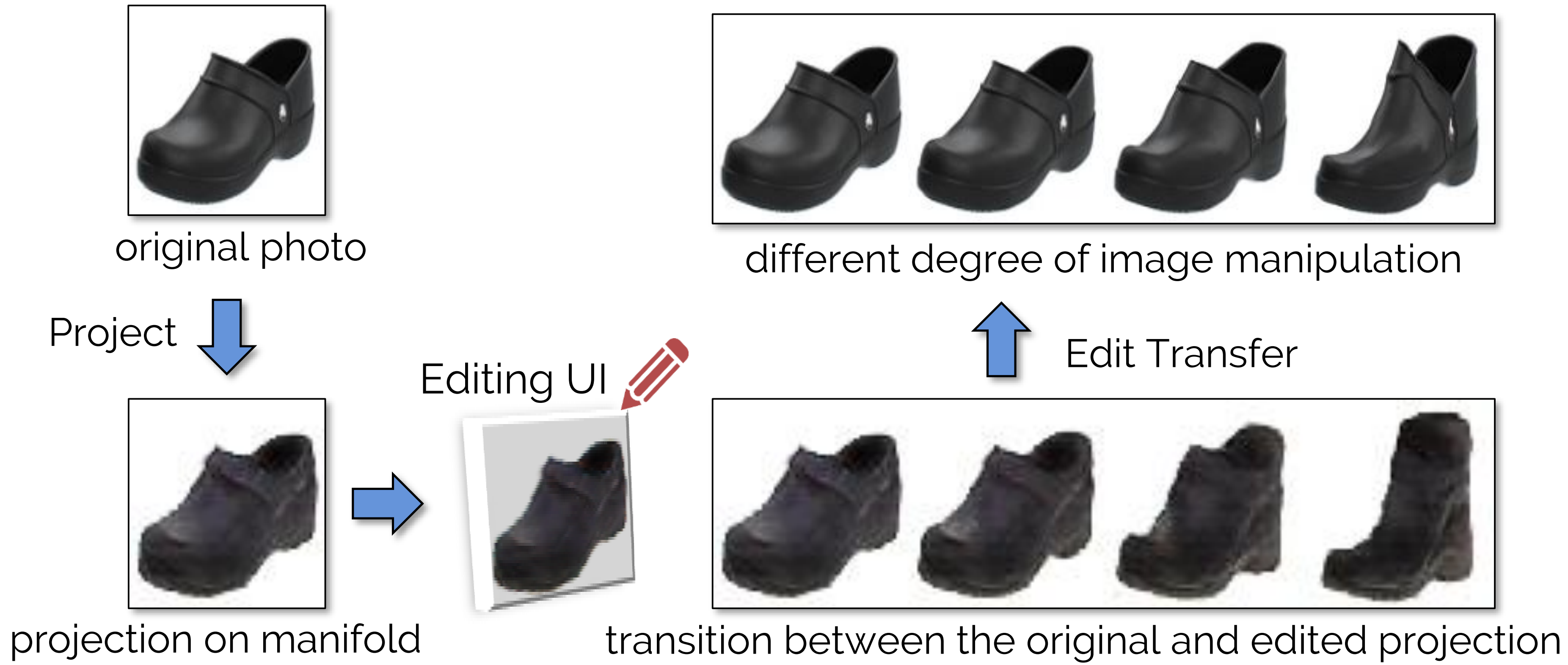


iGANs: Overview



[Zhu et al. 16] Generative Visual Manipulation on the Natural Image Manifold

iGANs: Overview



[Zhu et al. 16] Generative Visual Manipulation on the Natural Image Manifold

iGANs: Projecting an Image onto the Manifold

Input: real image x^R
Output: latent vector z

Optimization

$$z^* = \arg\min \mathcal{L}(G(z),\, x^R)$$

L_1, L_2

Reconstruction loss L

Generative model $G(z)$

not so realistic

iGANs: Projecting an Image onto the Manifold

Input: real image x^R
Output: latent vector z

Optimization

$$z^* = \arg\min \mathcal{L}(G(z), x^R)$$

Generate latent vector

Inverting Network $\mathrm{z} = P(x)$

$$\theta_P^* = \arg\min_{\theta_P} \sum_{x_n^R} \mathcal{L}(\underbrace{G(P(x^R;\theta_P))}_{\text{Auto-encoder}}, x^R)$$

Auto-encoder
with a fixed decoder G

iGANs: Projecting an Image onto the Manifold

Input: real image x^R
Output: latent vector z

Optimization

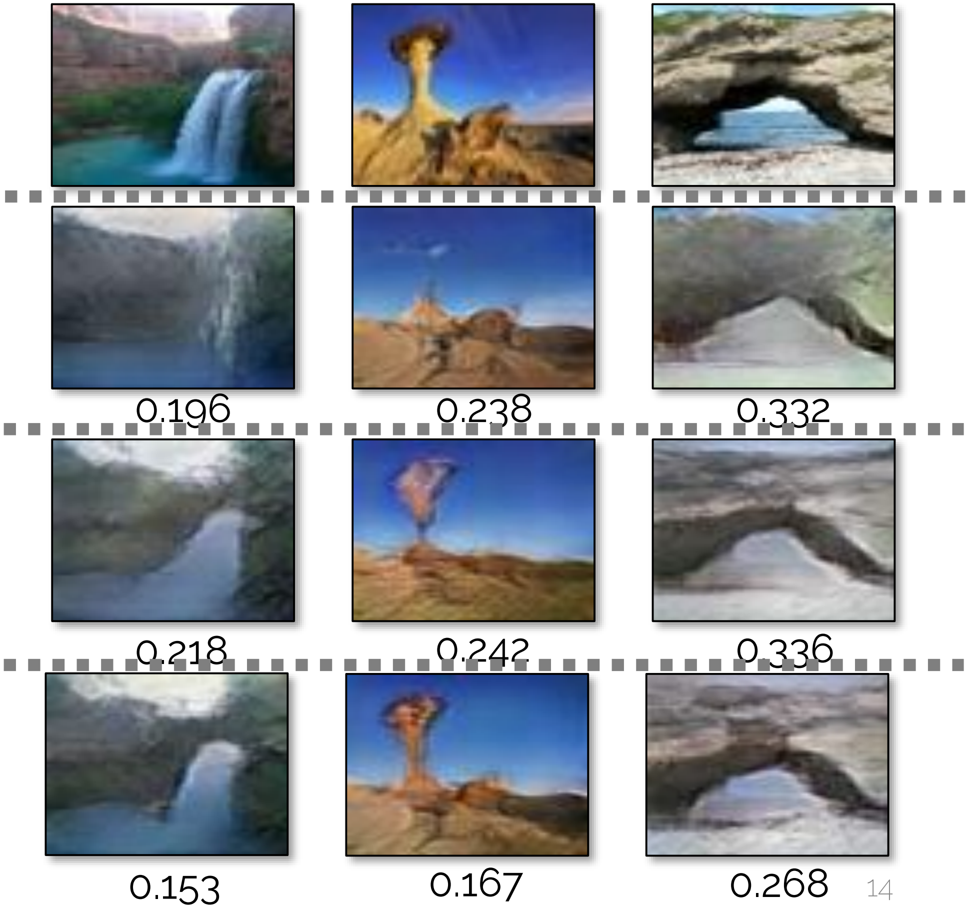
$$z^* = \arg\min \mathcal{L}(G(z),\, x^R)$$

Inverting Network $\mathrm{z} = P(x)$

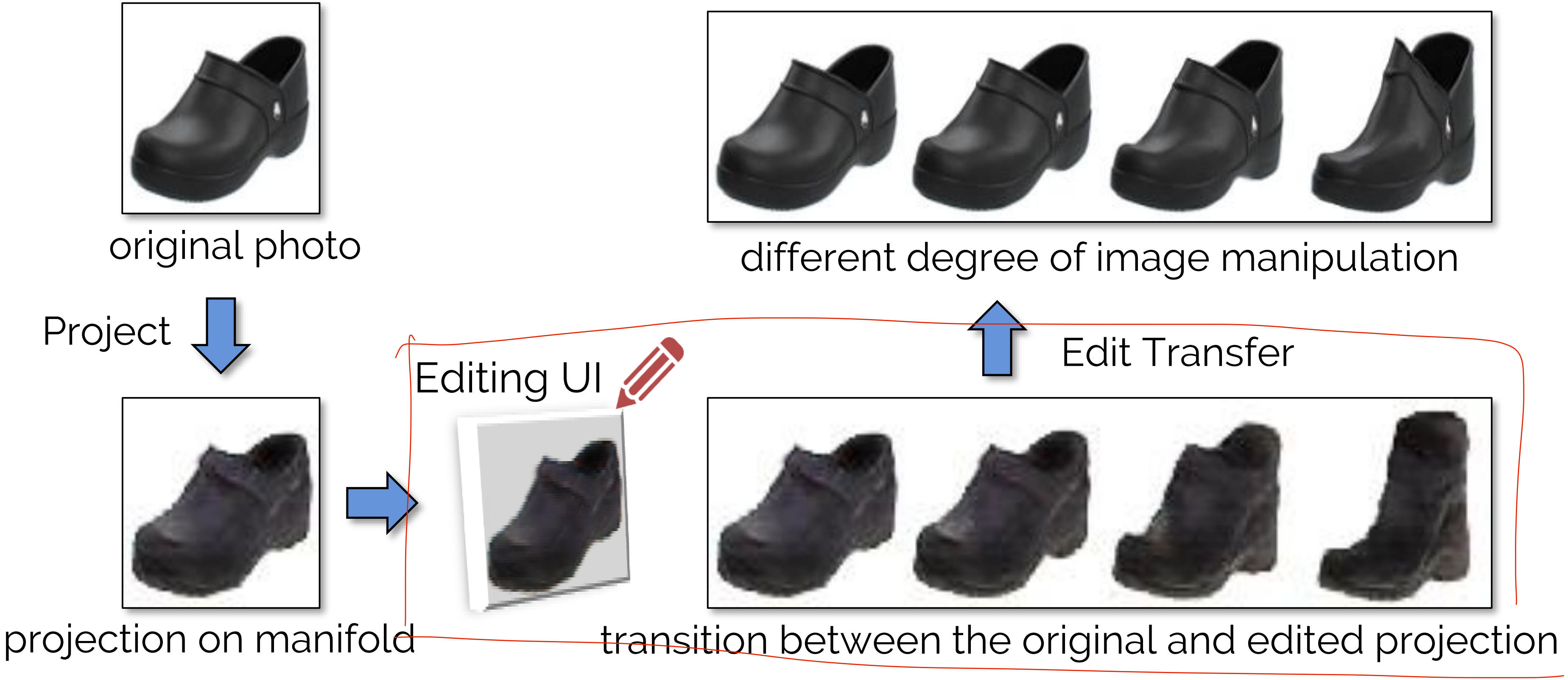
$$\theta_P^* = \arg\min_{\theta_P} \sum_{x_n^R} \mathcal{L}(\underbrace{G(P(x^R;\theta_P))}_{},\, x^R)$$

Hybrid Method

Use the network as initialization for the optimization problem



iGANs: Overview



iGANs: Manipulating the Latent Vector

constraint violation loss L_g

user guidance image

Objective: $$z^* = \arg\min_{z \in \mathbb{Z}} \Big\{ \underbrace{\sum_g \big(\mathcal{L}_g(G(z)), v_g\big)}_{\text{data term}} + \underbrace{\lambda_s \cdot \|z - z_0\|_2^2}_{\text{manifold smoothness}} \Big\}.$$

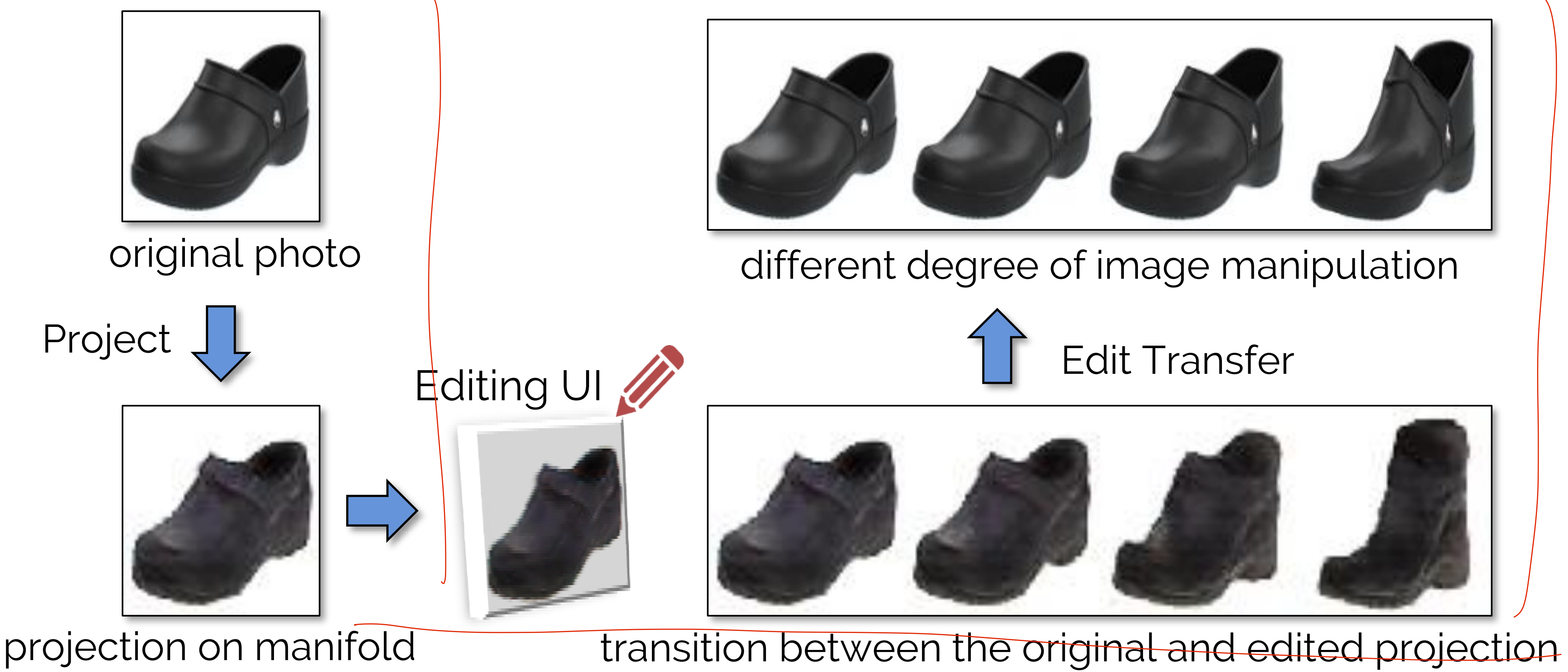
Guidance v_g

$G(z)$

z_0

[Zhu et al. 16] Generative Visual Manipulation on the Natural Image Manifold

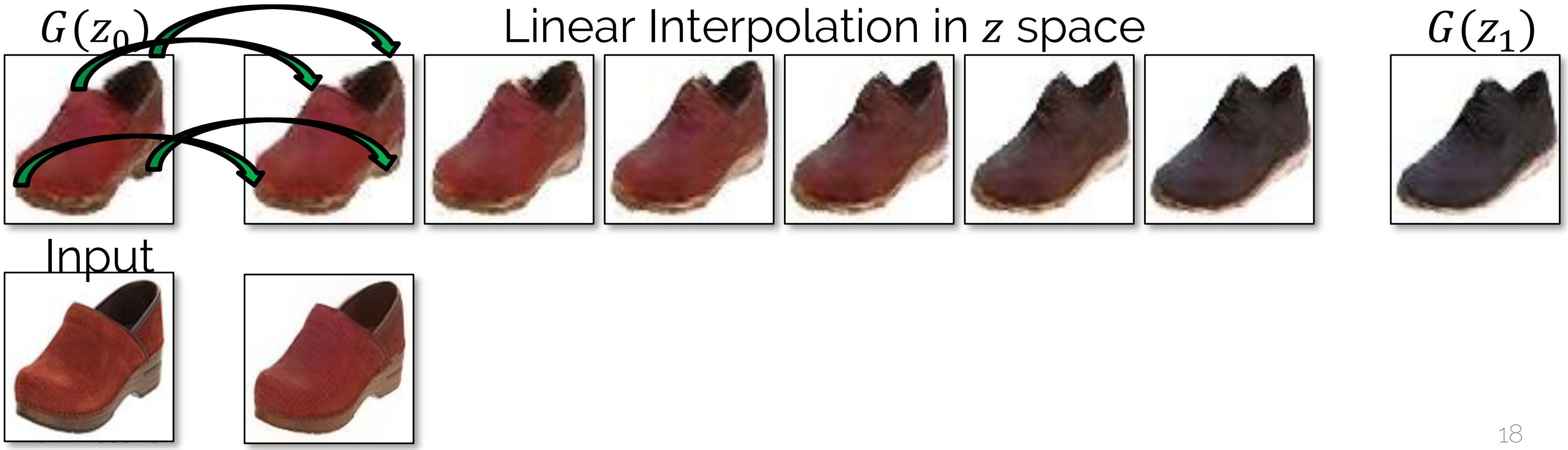
iGANs: Overview



iGANs: Edit Transfer

Motion (u, v)+ Color ($A_{3\times4}$): estimate per-pixel geometric and color variation

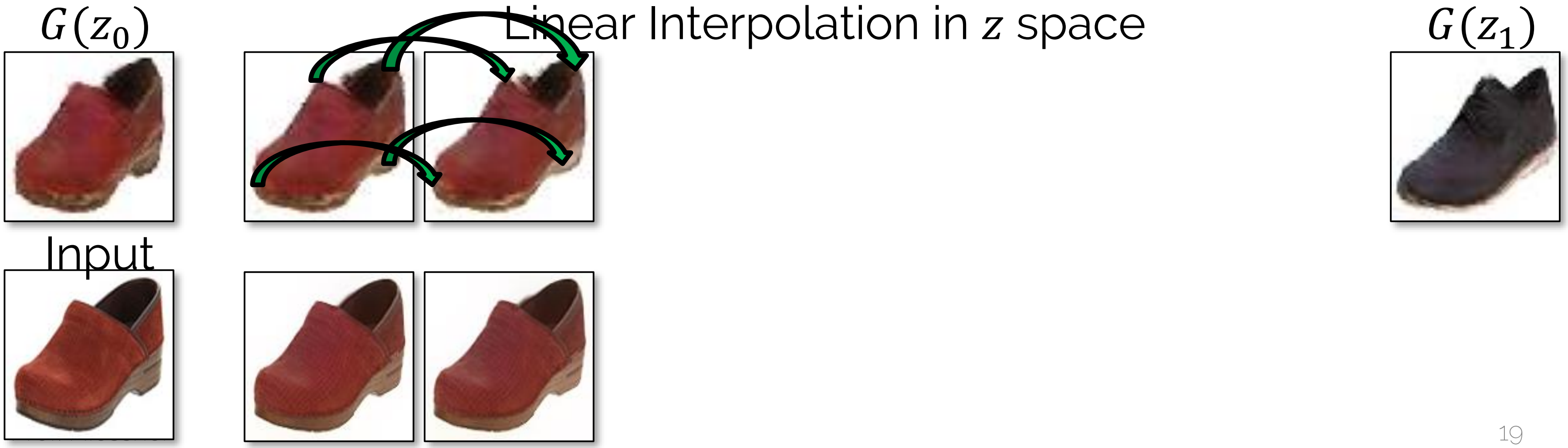
$$\iint \underbrace{\|I(x,y,t) - \boxed{A}\cdot I(x+\boxed{u}, y+\boxed{v}, t+1)\|^2}_{\text{data term}} + \underbrace{\sigma_s(\|\nabla u\|^2+\|\nabla v\|^2)}_{\text{spatial reg}} + \underbrace{\sigma_c\|\nabla A\|^2}_{\text{color reg}} dxdy$$



iGANs: Edit Transfer

Motion **(u, v)+** **Color** **($A_{3\times4}$):** estimate per-pixel geometric and color variation

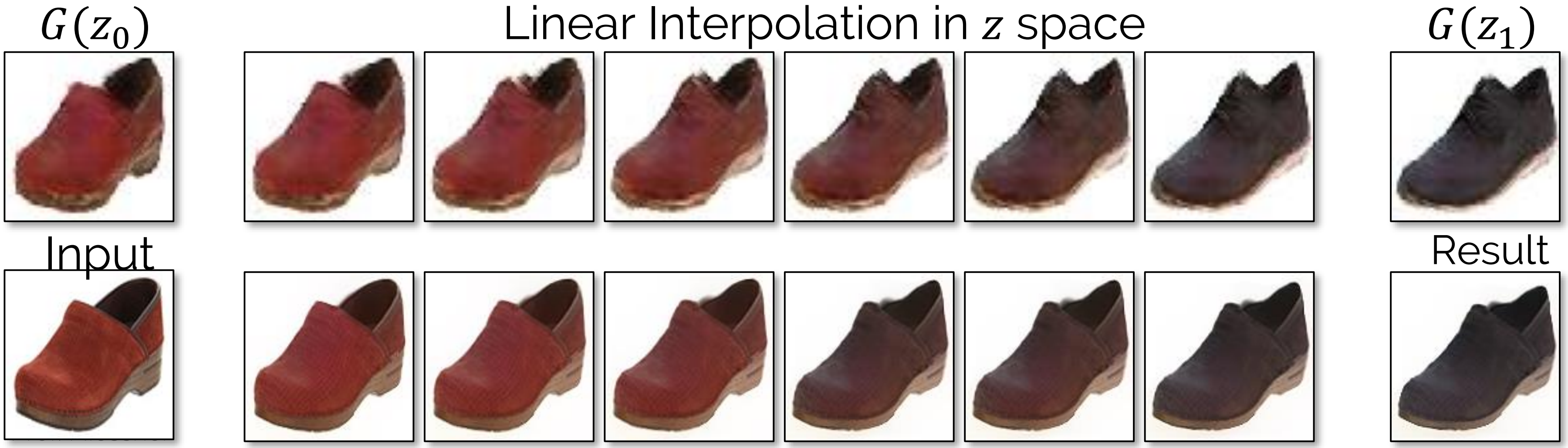
$$\iint \underbrace{\|I(x,y,t) - \boxed{A} \cdot I(x+\boxed{u}, y+\boxed{v}, t+1)\|^2}_{\text{data term}} + \underbrace{\sigma_s(\|\nabla u\|^2 + \|\nabla v\|^2)}_{\text{spatial reg}} + \underbrace{\sigma_c\|\nabla A\|^2}_{\text{color reg}} dxdy$$



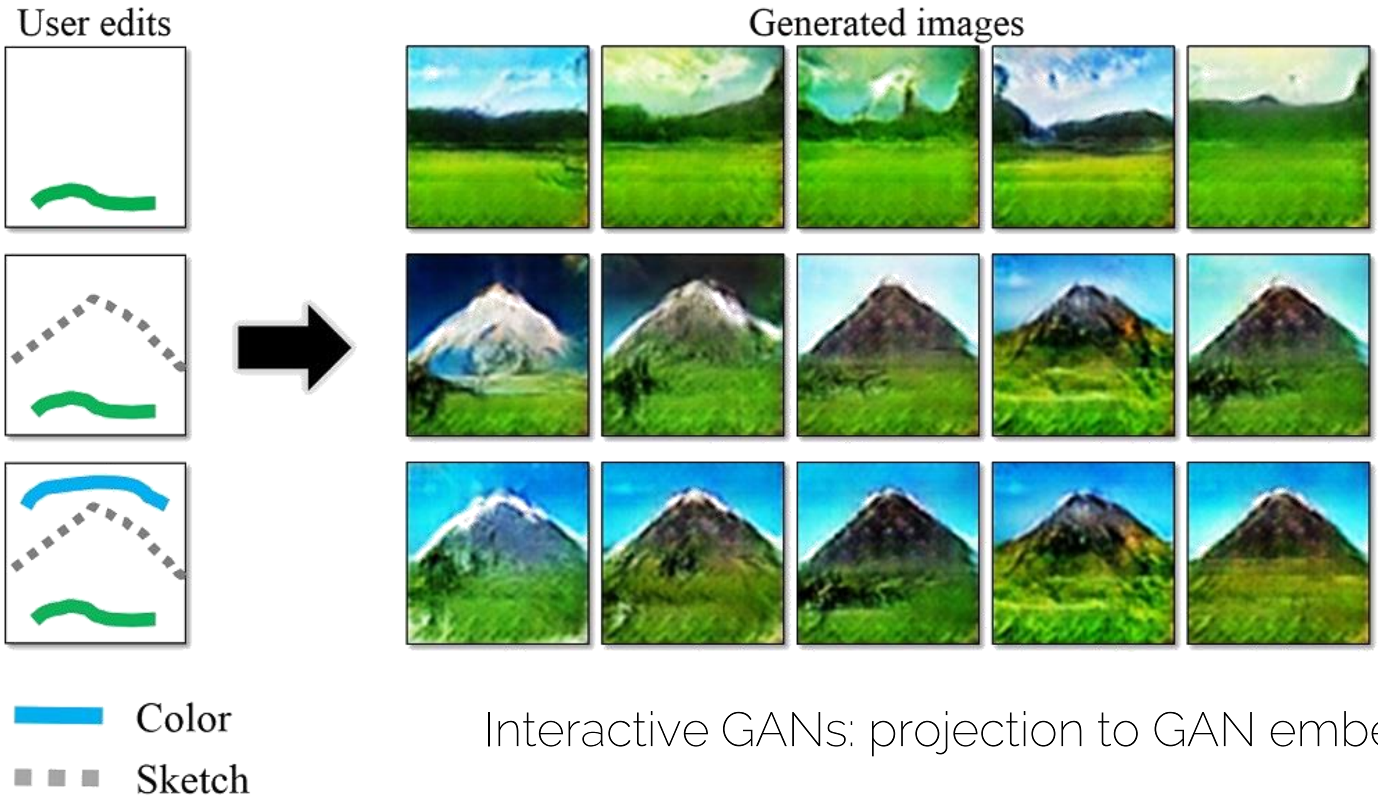
iGANs: Edit Transfer

Motion **(u, v)+** **Color** **($A_{3\times4}$):** estimate per-pixel geometric and color variation

$$\iint \underbrace{\|I(x,y,t) - \boxed{A} \cdot I(x+\boxed{u}, y+\boxed{v}, t+1)\|^2}_{\text{data term}} + \underbrace{\sigma_s(\|\nabla u\|^2 + \|\nabla v\|^2)}_{\text{spatial reg}} + \underbrace{\sigma_c\|\nabla A\|^2}_{\text{color reg}} dxdy$$



cGANs: Interactive GANs



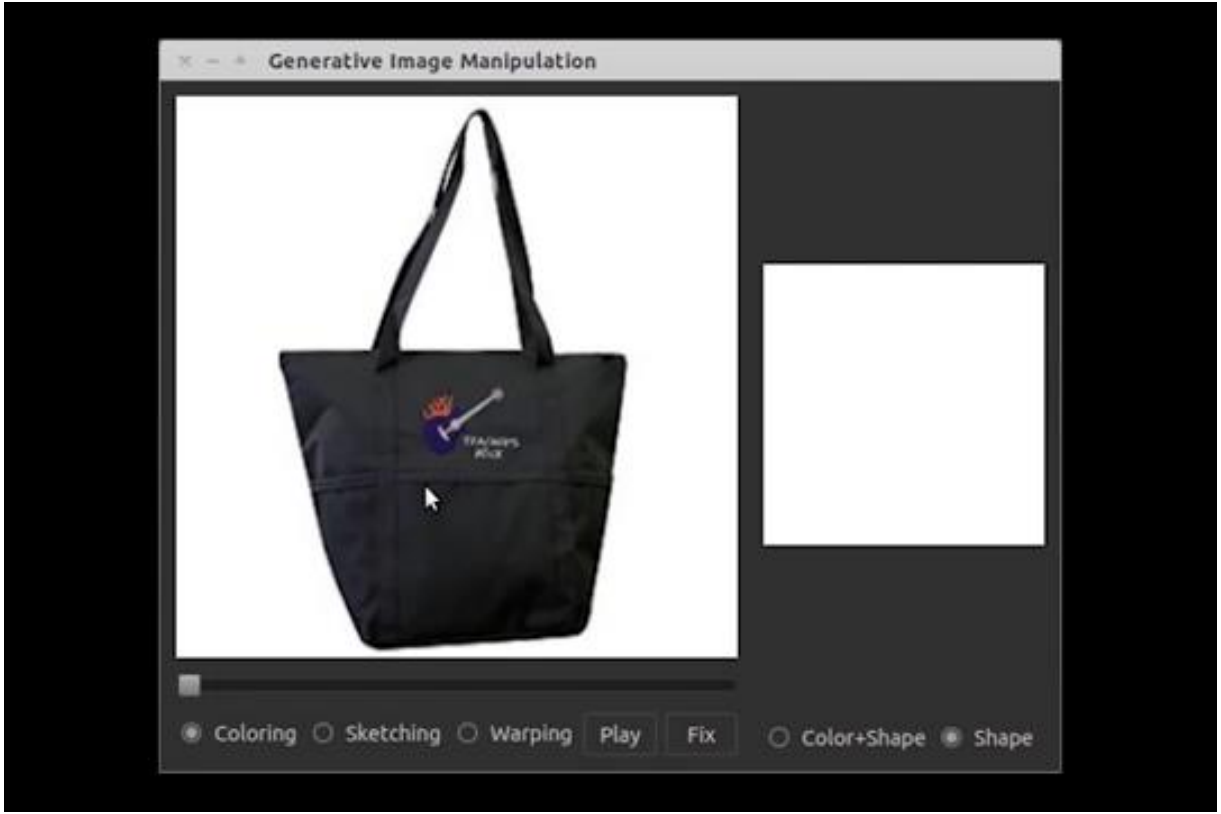
Interactive GANs: projection to GAN embedding

https://github.com/junyanz/iGAN [Zhu et al. 16.]

cGANs: Interactive GANs

https://github.com/junyanz/iGAN [Zhu et al. 16.]

cGANs: Interactive GANs



https://github.com/junyanz/iGAN [Zhu et al. 16.]

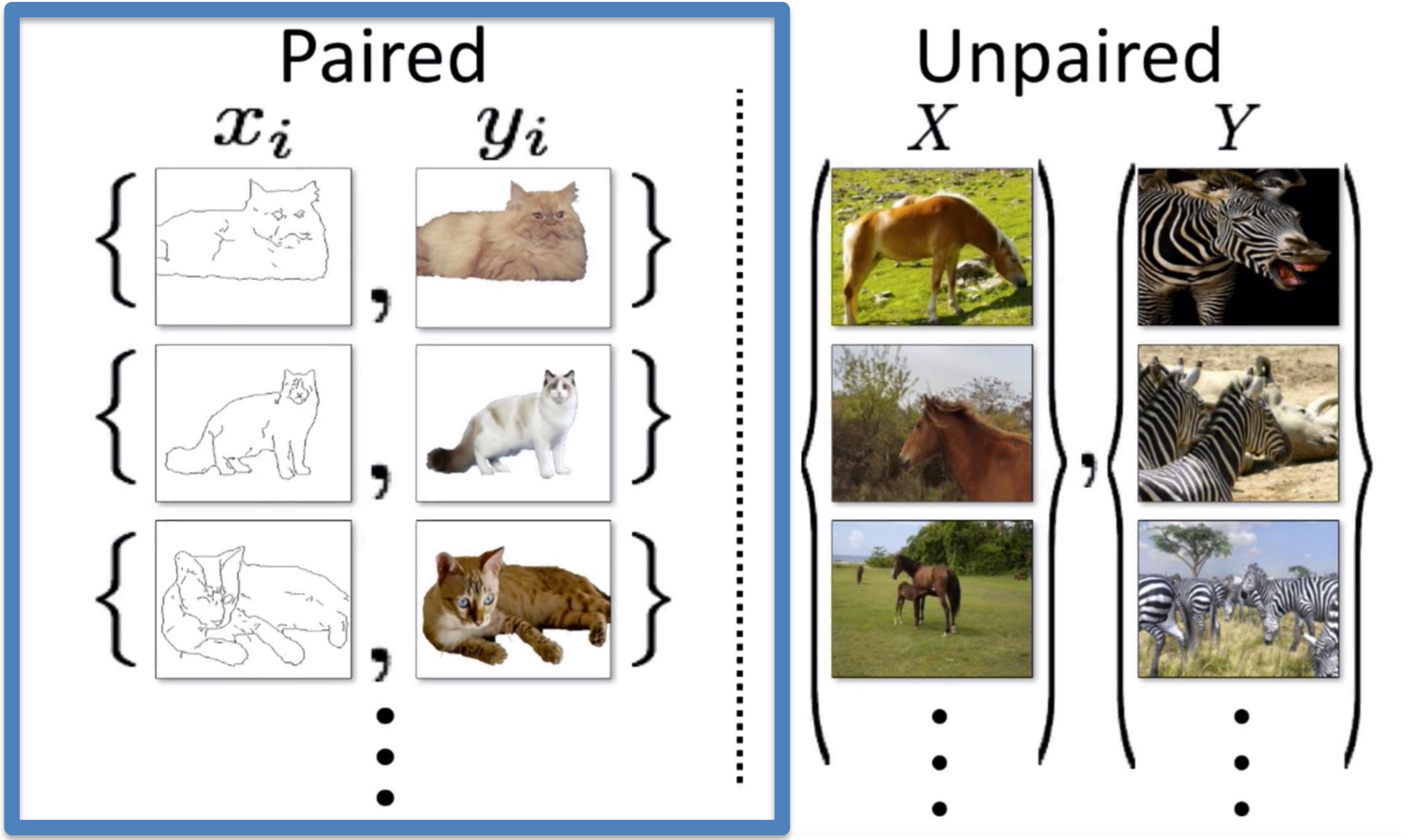
Mapping in Latent Space is Difficult!

- Semantics are missing
- In most cases, no labels available
- Ideally, need some unsupervised disentangled rep.

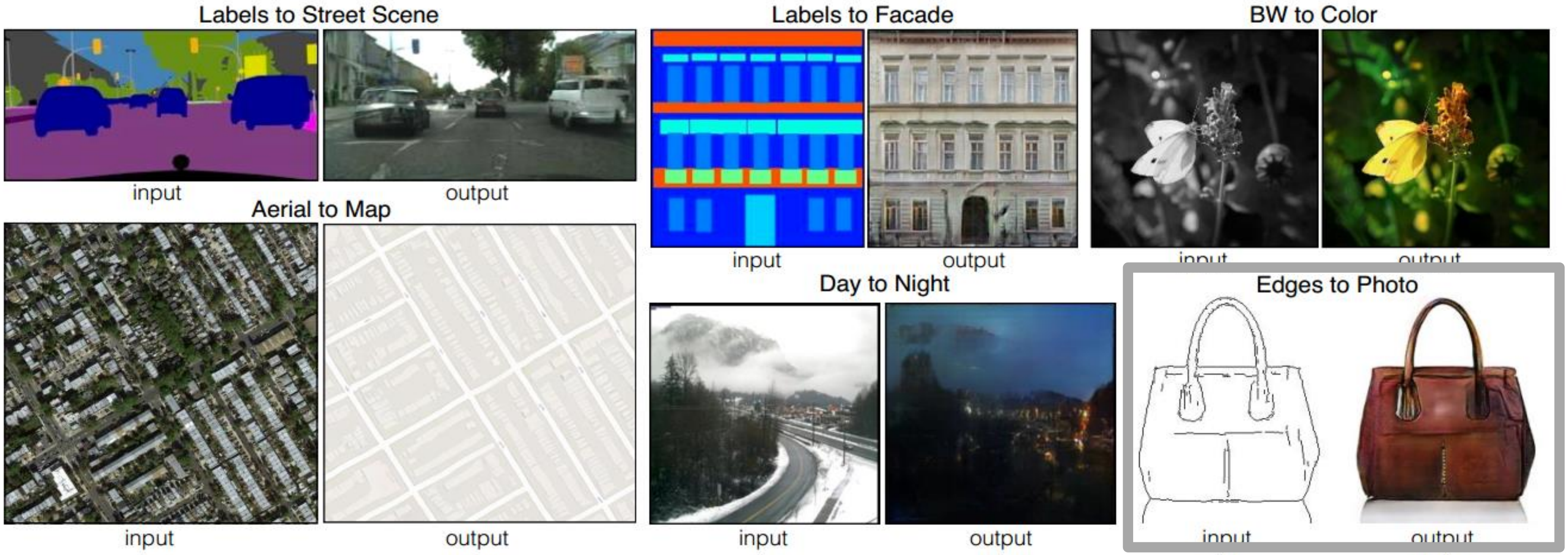
(a) Azimuth (pose) (b) Presence or absence of glasses

InfoGAN [Chen et al. 16]

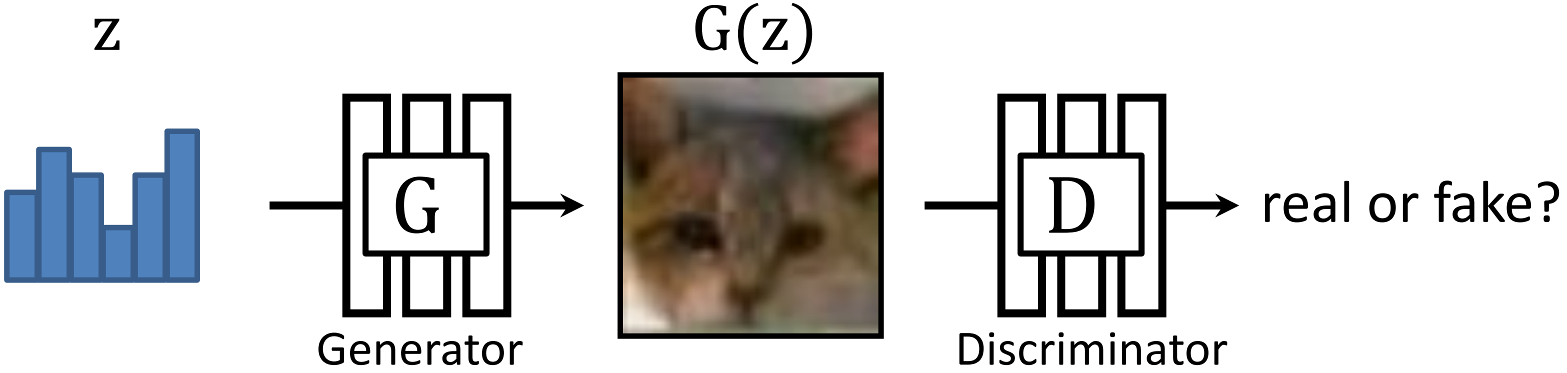
Paired vs Unpaired Setting



pix2pix: Image-to-Image Translation

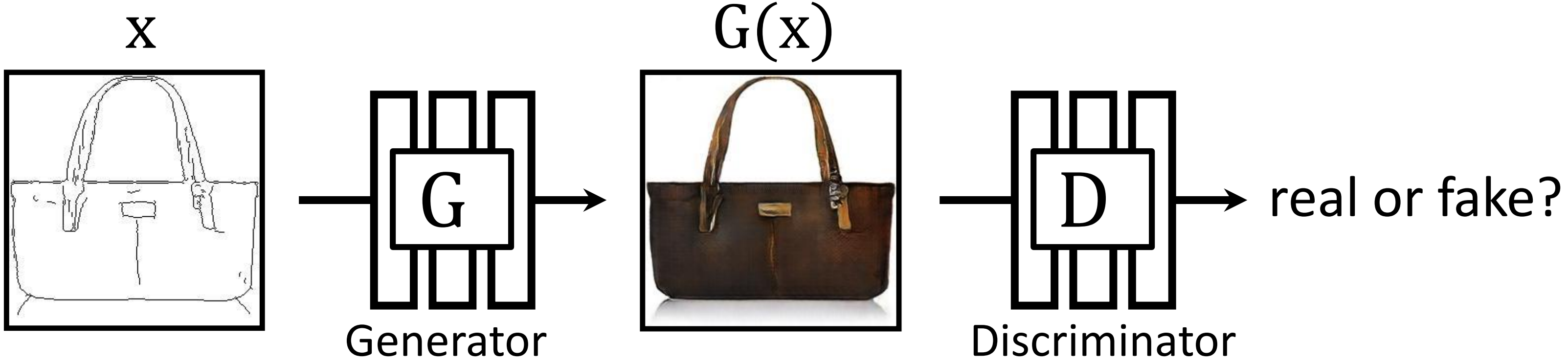


[Isola et al. 16] Image-to-Image Translation with Conditional Adversarial Networks



$$\min_G \max_D \mathbb{E}_{z,x}[\log D(G(z)) + \log(1 - D(x))]$$

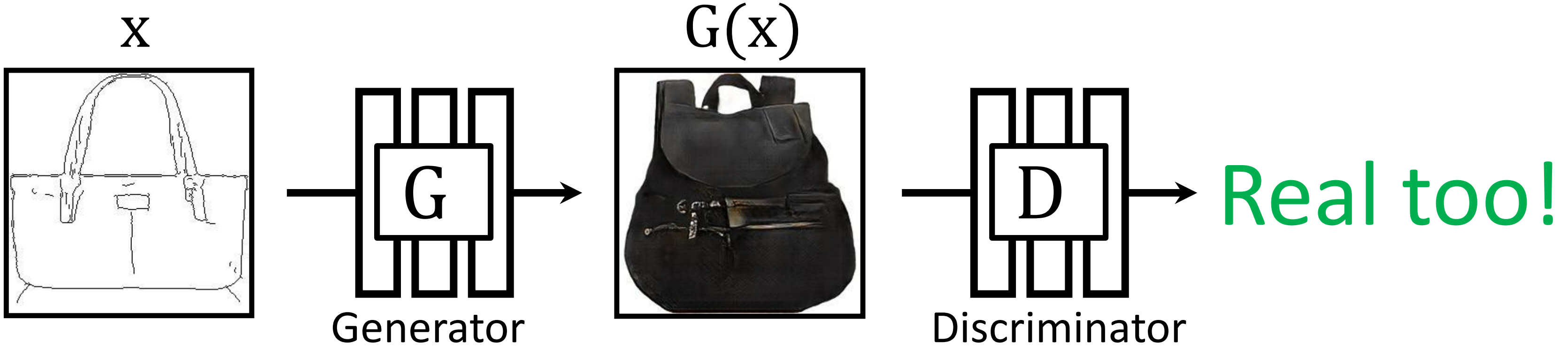
[Goodfellow et al. 2014] Generative Adversial Networks



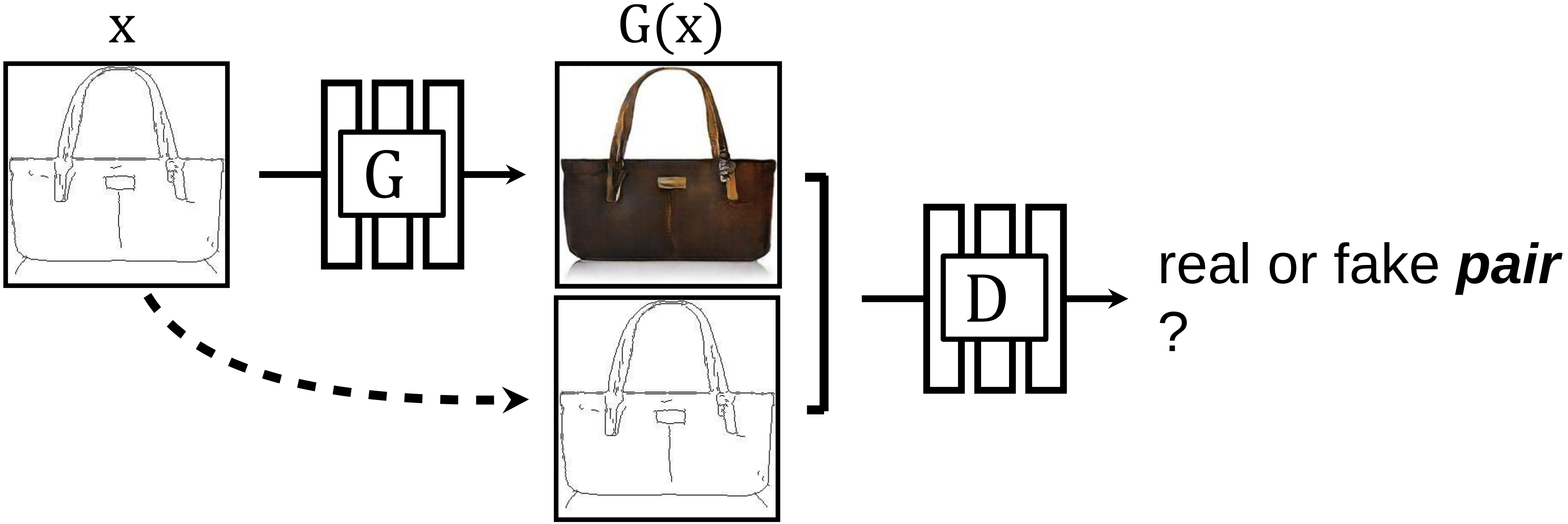
$$\min_G \max_D \mathbb{E}_{x,y}[\log D(G(x)) + \log(1 - D(y))]$$

problem : don't see pair

$$\min_G \max_D \mathbb{E}_{x,y}[\log D(G(x)) + \log(1 - D(y))]$$



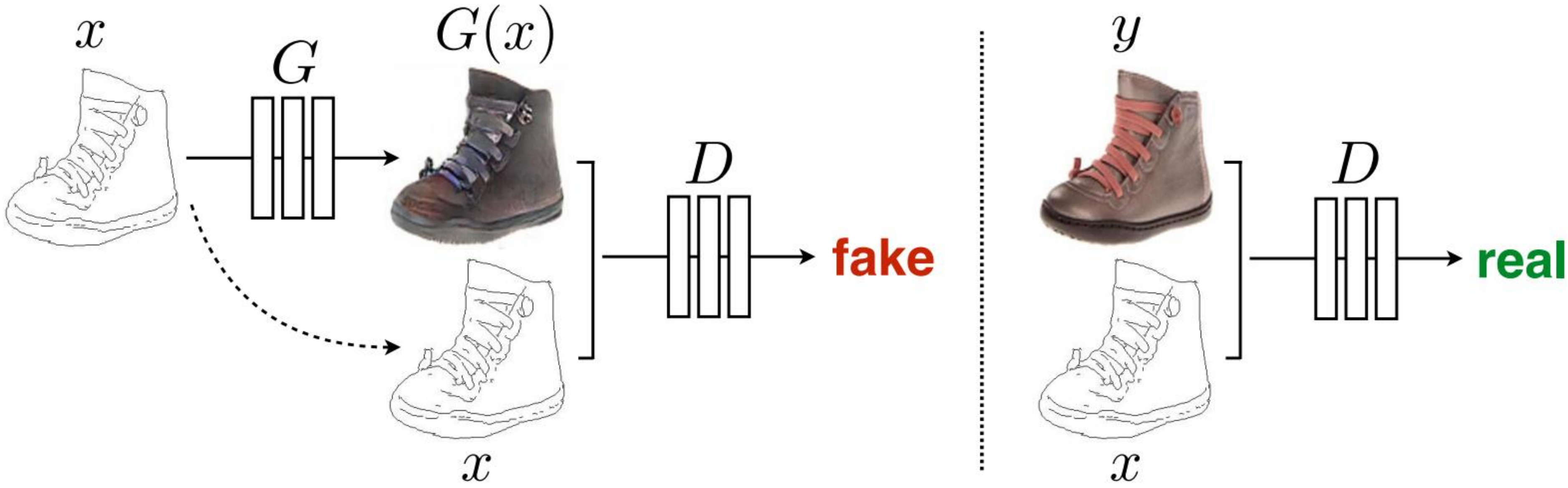
$$\min_G \max_D \mathbb{E}_{x,y}[\log D(G(x)) + \log(1 - D(y))]$$



$$\min_G \max_D \mathbb{E}_{x,y}[\log \underbrace{D(x, G(x))}_{\text{fake pair}} + \log(1 - \underbrace{D(x,y)}_{\text{real pair}})]$$

match joint distribution $p(G(x), y) \sim p(x, y)$

Pix2Pix



Pix2Pix: Paired Setting

- Great when we have ‘free’ training data
- Often called self-supervised
- Think about these settings ☺

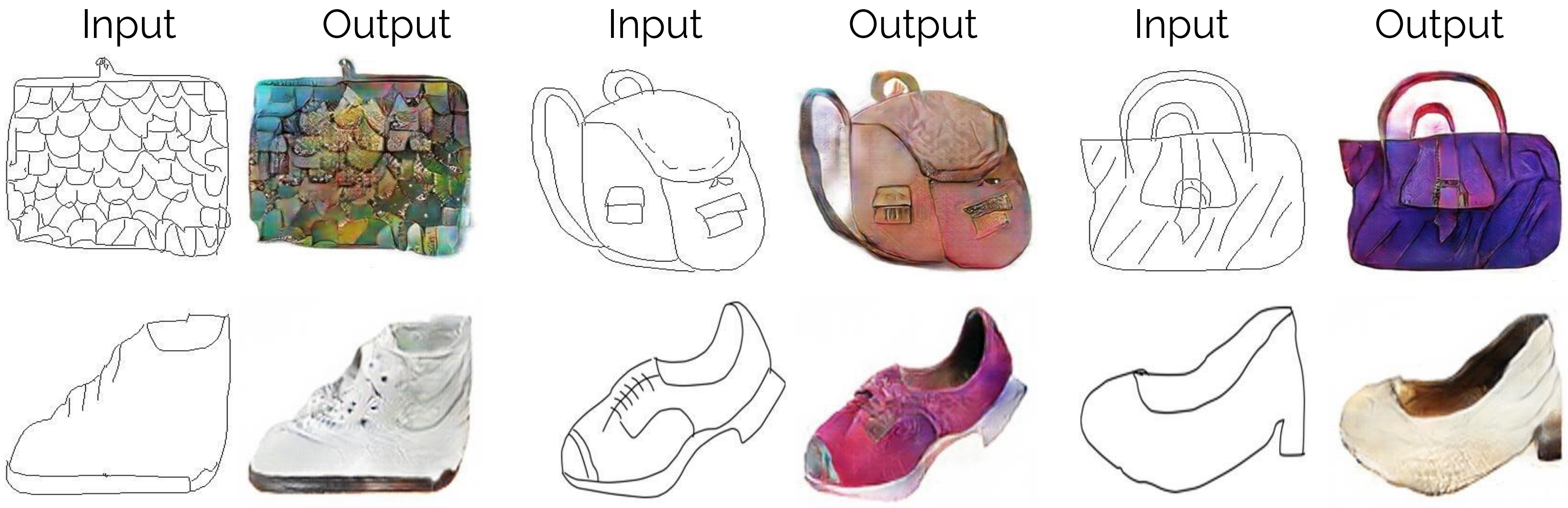
Pix2Pix - Examples

Edges → Images

[Xie & Tu 15] Holistically-Nested Edge Detection

Pix2Pix - Examples

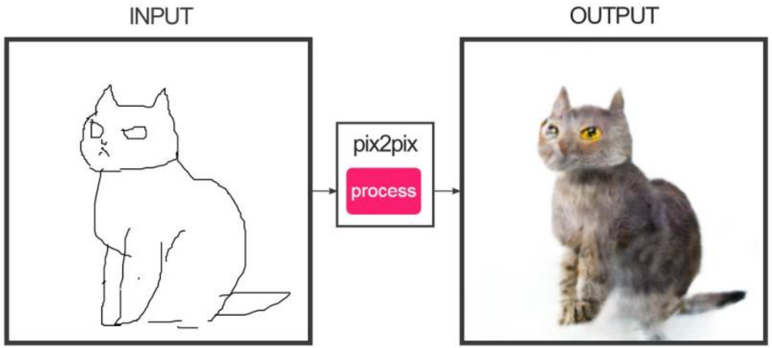
Sketches → Images

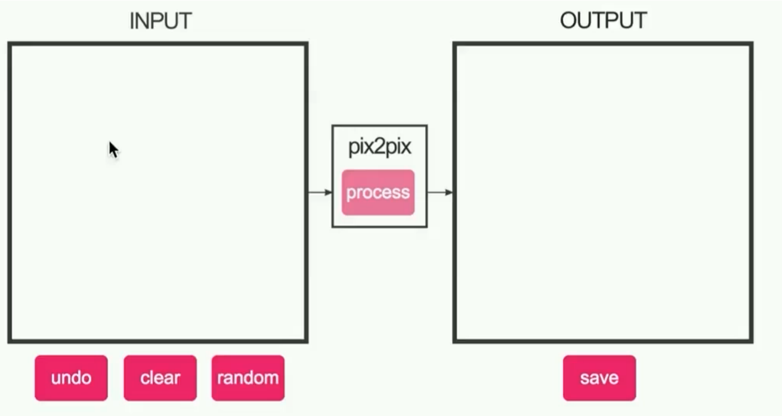


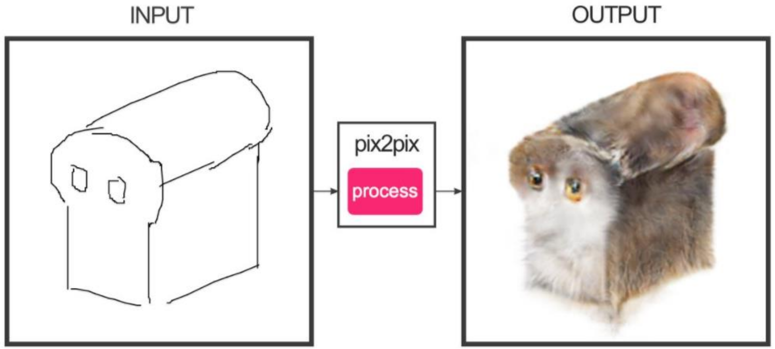
Trained on Edges → Images

How Do Humans Sketch Objects [Eitz, Hays, Alexa, 2012]

Pix2Pix - Examples







Vitaly Vidmirov @vvid

#edges2cats by Christopher Hesse https://affinelayer.com/pixsrv/

Pix2Pix - Examples

Input Output Groundtruth

Data from [maps.google.com]

Pix2Pix - Examples

$BW \rightarrow$ Color

Input	Output	Input	Output	Input	Output

Data from [Russakovsky et al. 2015 - ImageNet]

Ideas behind Pix2Pix

- $L = L_{GAN} + \lambda L_1$ (makes it more constraint)

- Unet / skip connections for preserving structure

- Noise only through dropout
  - cGANs tend to learn to ignore the random vector z
  - Still want probabilistic model

Ideas behind Pix2Pix

- L1 or L2 loss for low frequency details

- GAN discriminator for high frequency details

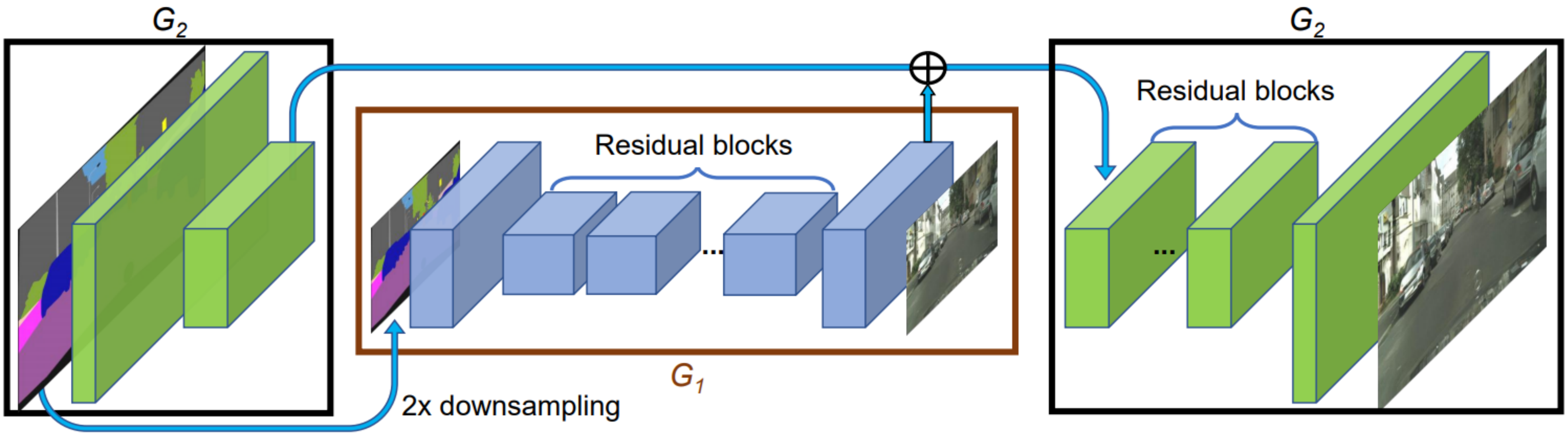
-> PatchGAN
  - GAN discriminator applied only to local patches
  - It's fully-convolutional; i.e., can run on arbitrary image sizes

Pix2PixHD

- Expand the pix2pix idea to multi-scale
- Coarse-to-fine generator + discriminator
- G's and D's are the same but since they operate on different resolutions, they have effectively a larger receptive field

[Wang et al. 18] pix2pixHD

Pix2PixHD



Pix2PixHD

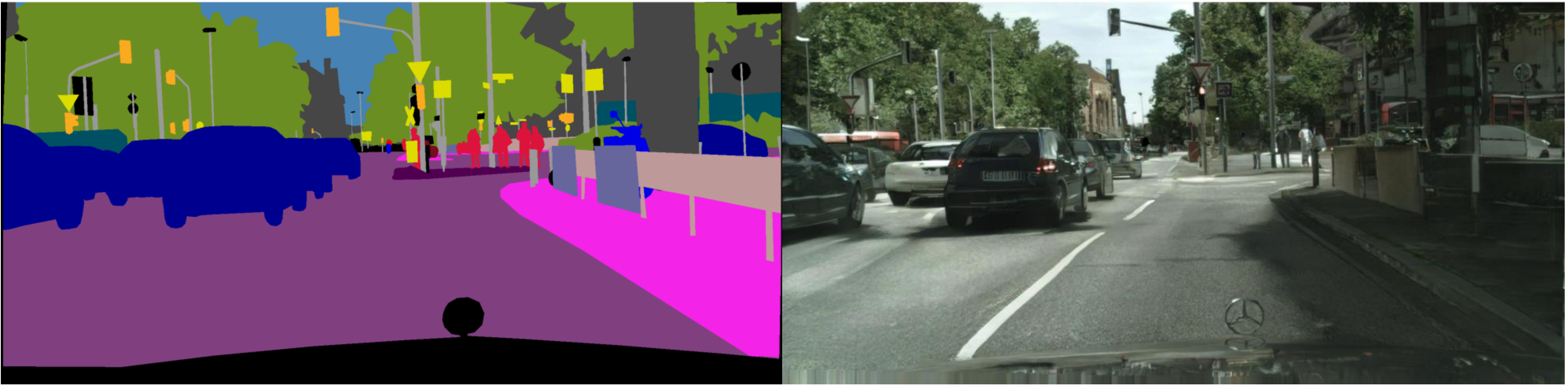
- Use of multi-scale discriminators

- $\min_{G} \max_{D_1, D_2, D_3} \sum_{k=1,2,3} L_{GAN}(G, D_k)$

- Can make various combinations of stacking discriminator and generator
  - E.g., have a single G and downsample generated and real images – or have intermediate real images (cf. ProGAN)

Pix2PixHD

Input labels Synthesized image



[Wang et al. 18] pix2pixHD

Pix2PixHD

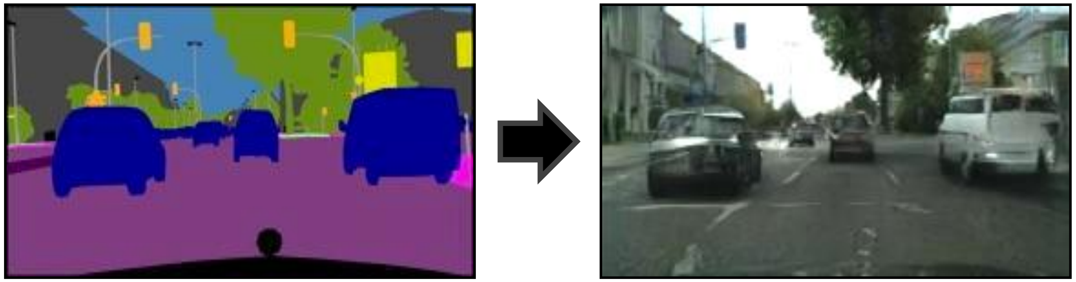
[Wang et al. 18] pix2pixHD

Pix2PixHD (Interactive Results)

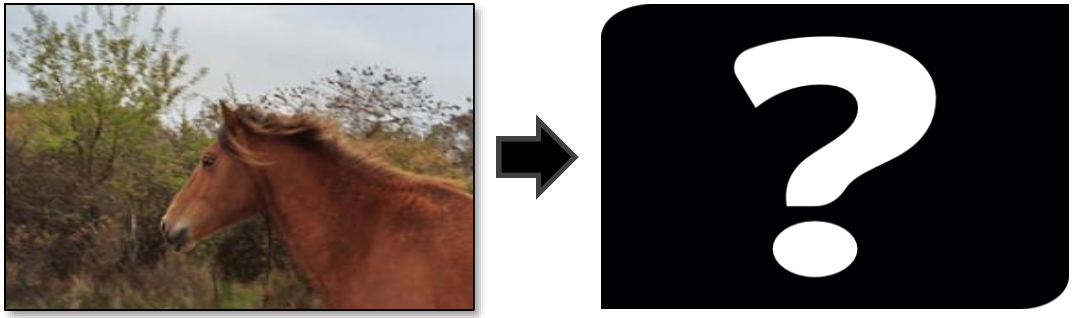
[Wang et al. 18] pix2pixHD

Paired

x_i y_i



Label ↔ photo: per-pixel labeling

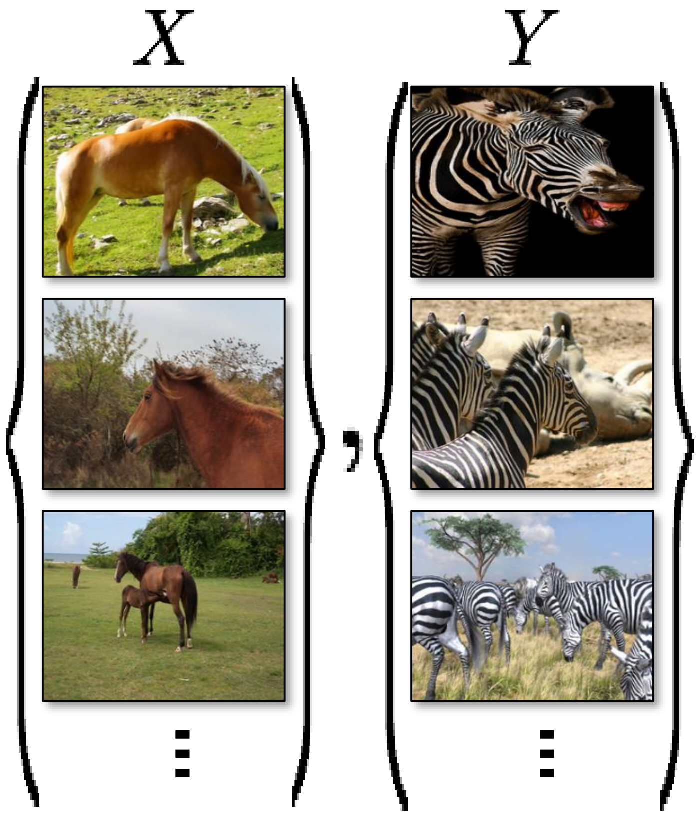


Horse ↔ zebra: how to get zebras?

- Expensive to collect pairs.
- Impossible in many scenarios.

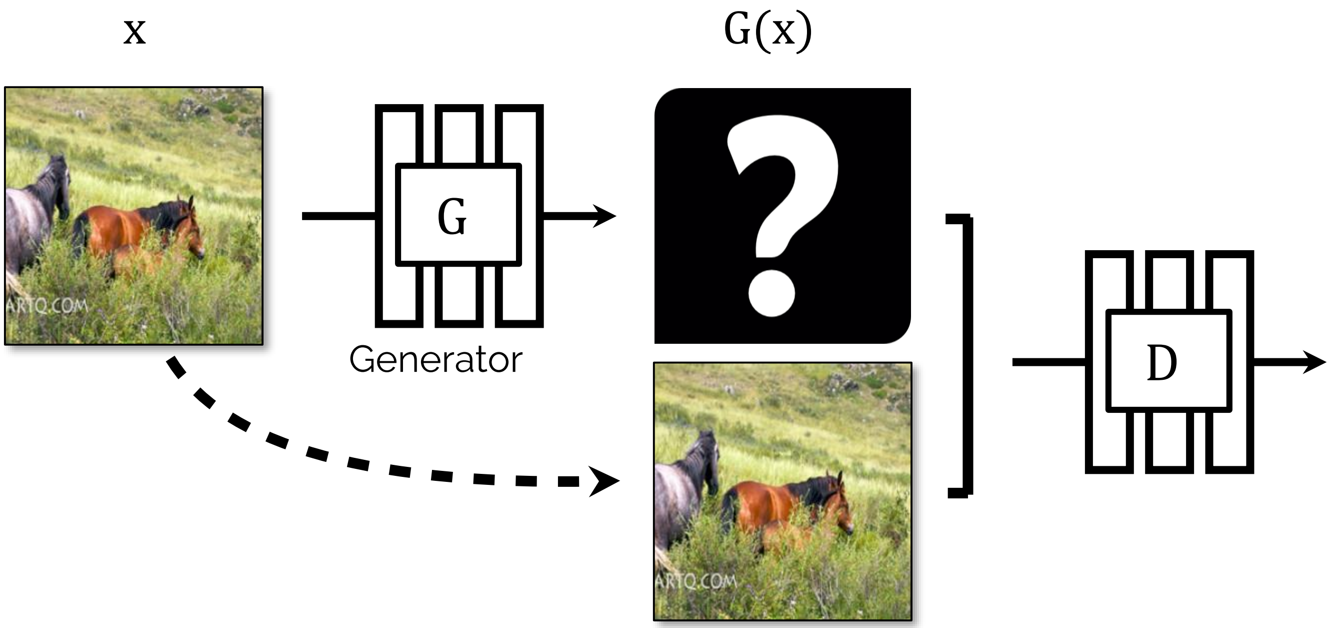
Paired

Unpaired



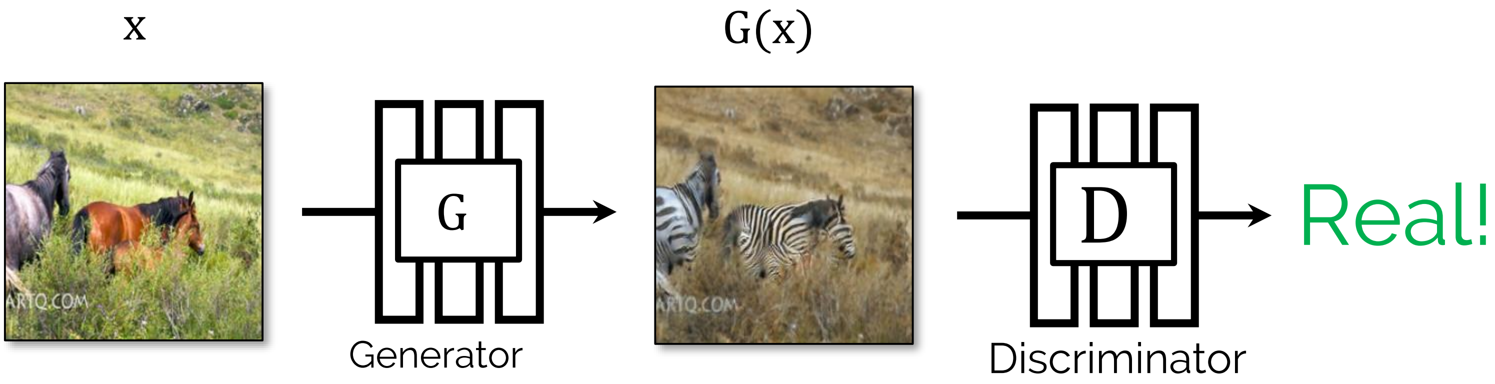
Cycle-Consistent Adversarial Networks

Cycle-Consistent Adversarial Networks

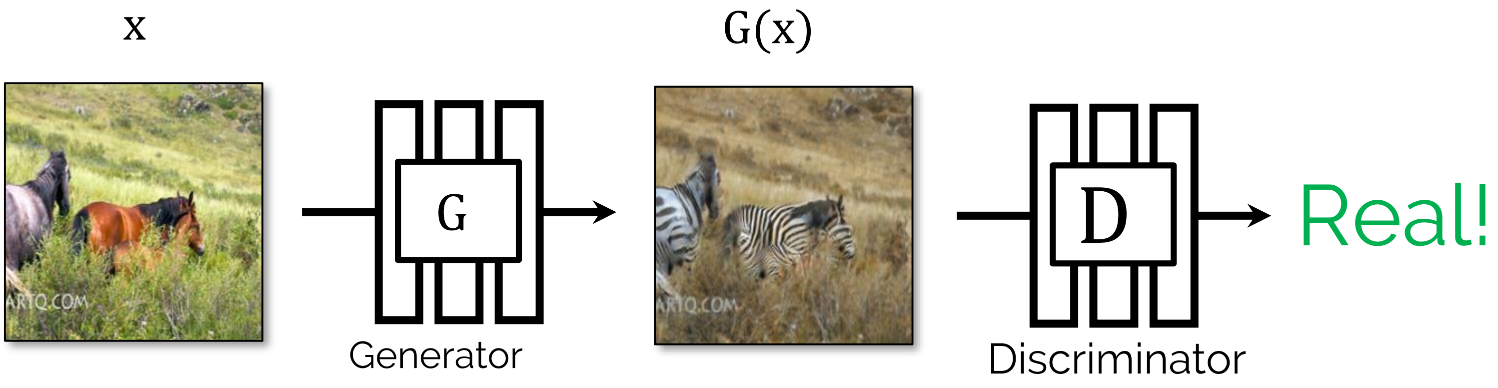


No input-output pairs!

Cycle-Consistent Adversarial Networks

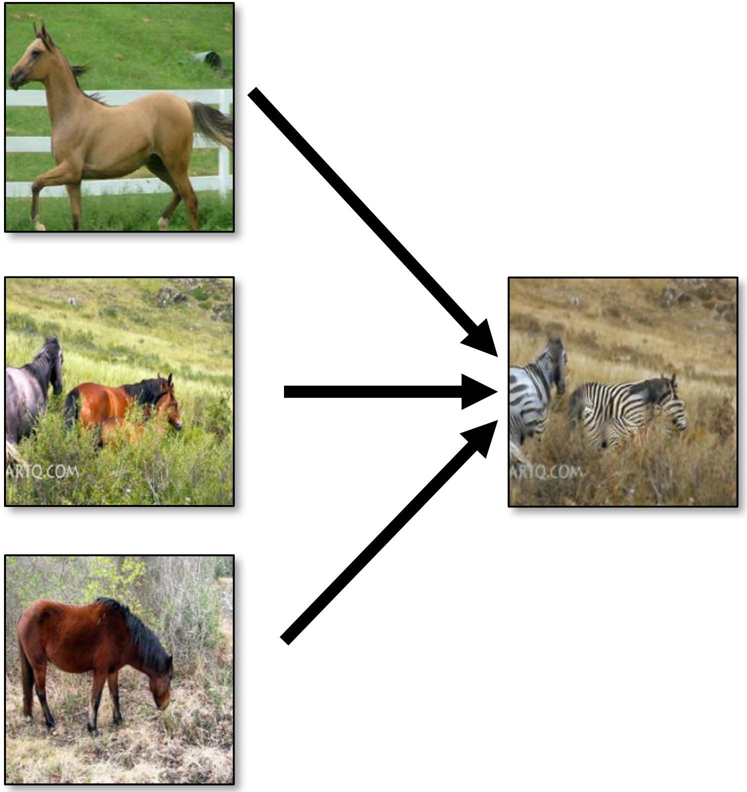


Cycle-Consistent Adversarial Networks



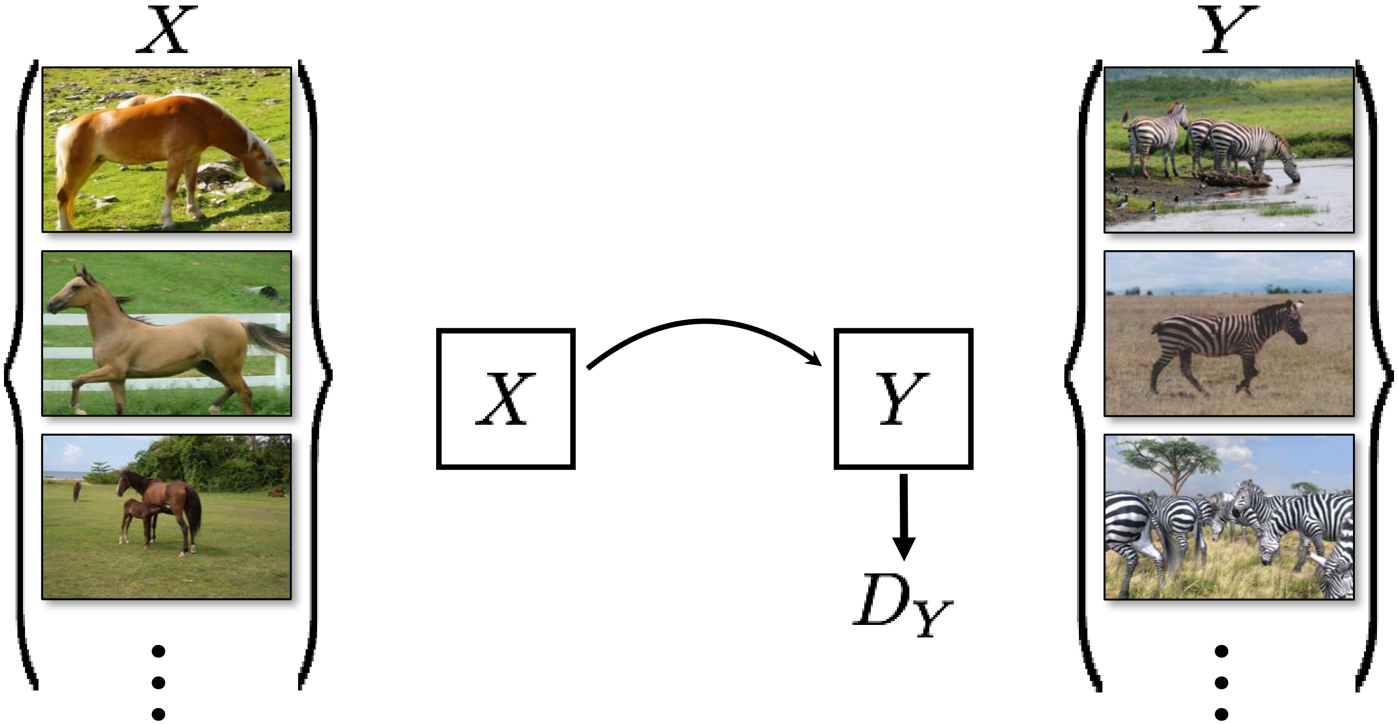
GANs doesn't force output to correspond to input

Cycle-Consistent Adversarial Networks



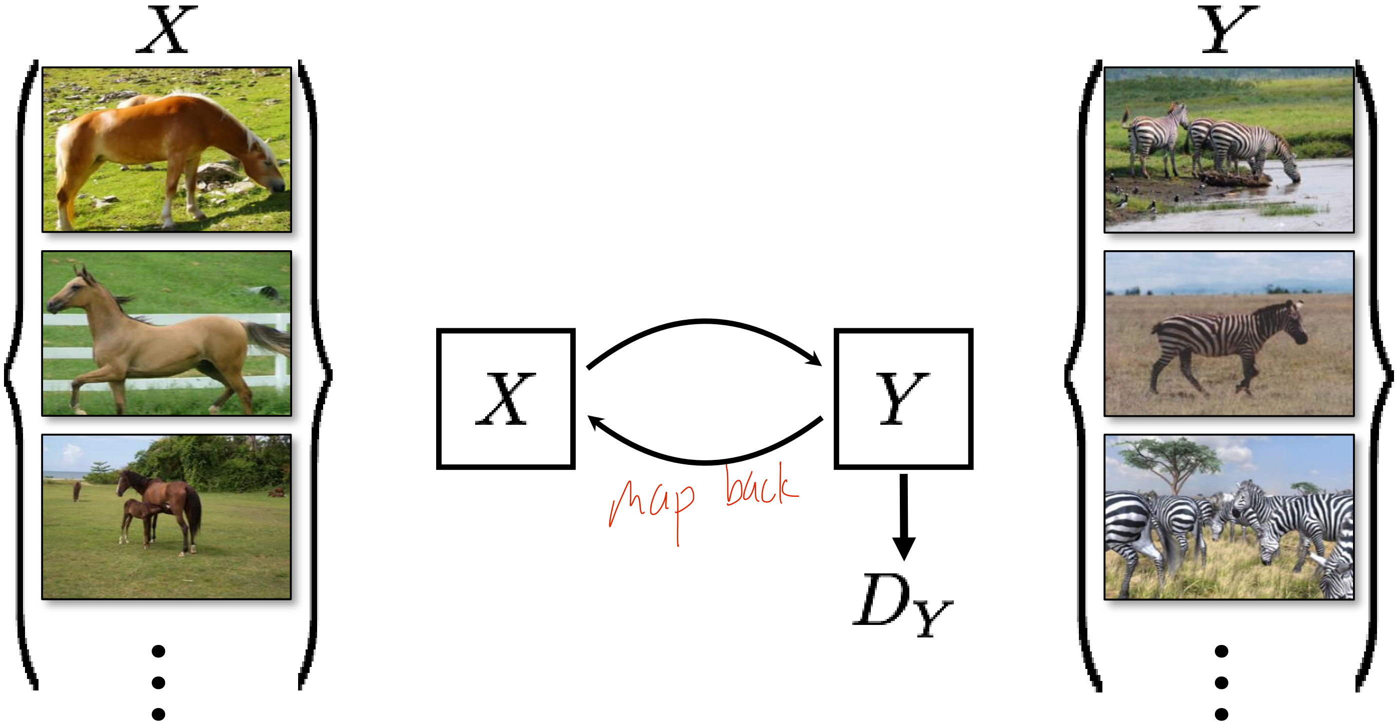
mode collapse!

Cycle-Consistent Adversarial Networks

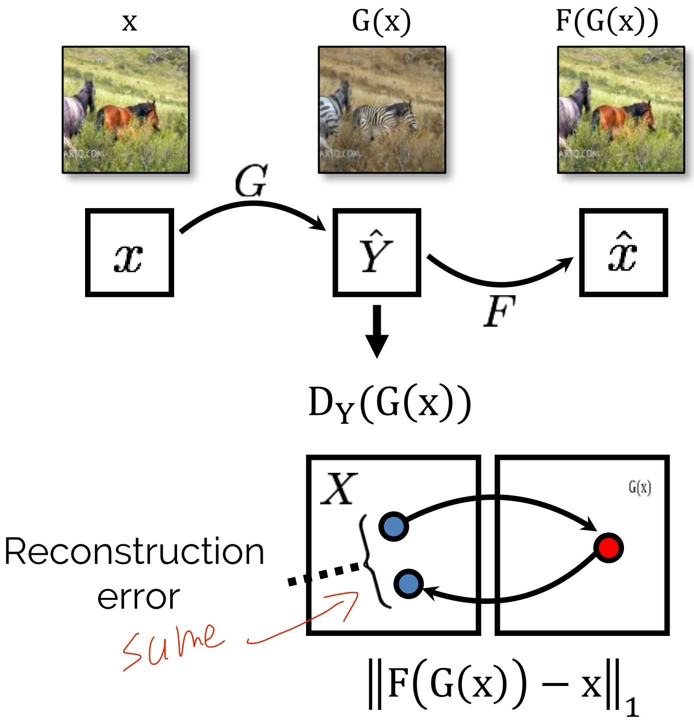


[Zhu et al. 17] Cycle-Consistent Adversarial Networks

Cycle-Consistent Adversarial Networks

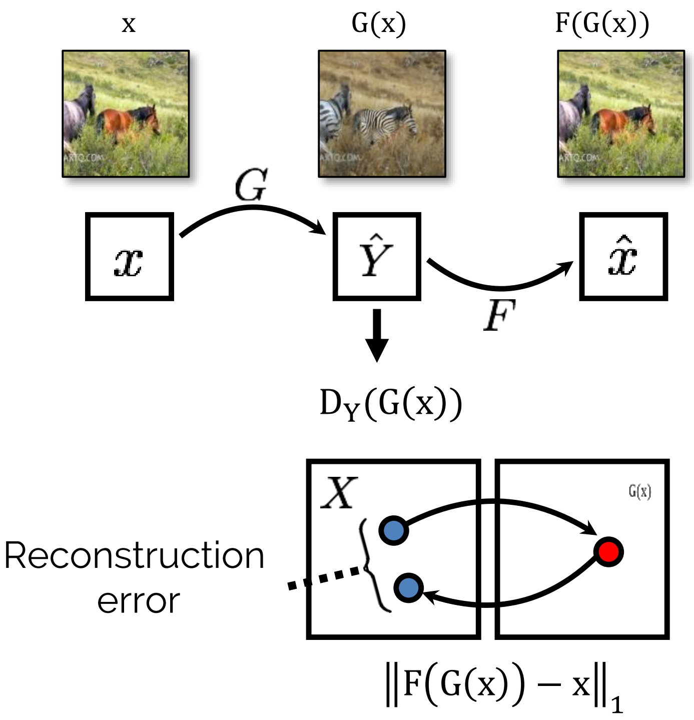


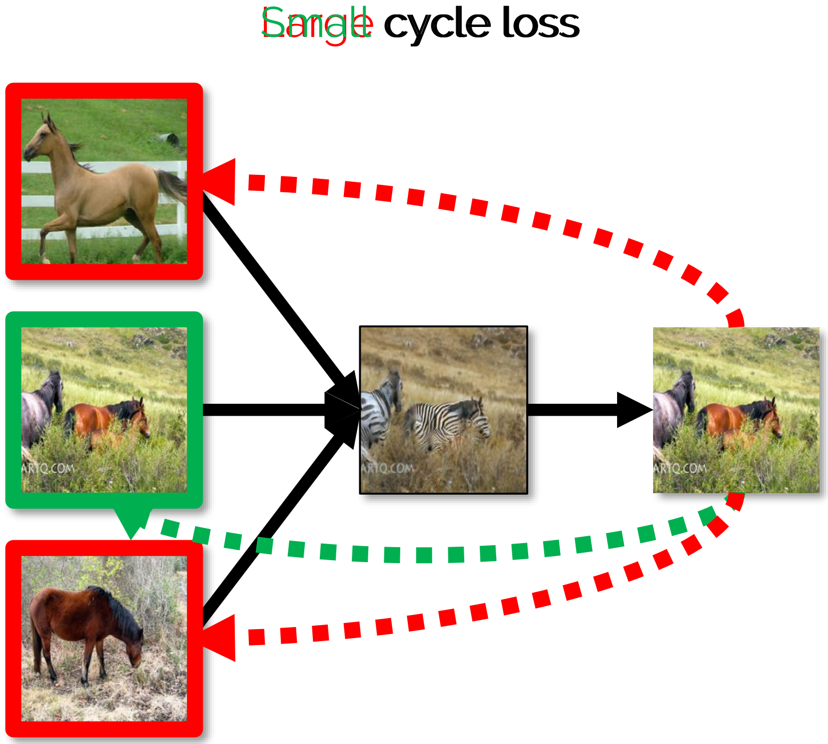
Cycle Consistency Loss



[Zhu et al. 17] Cycle-Consistent Adversarial Networks

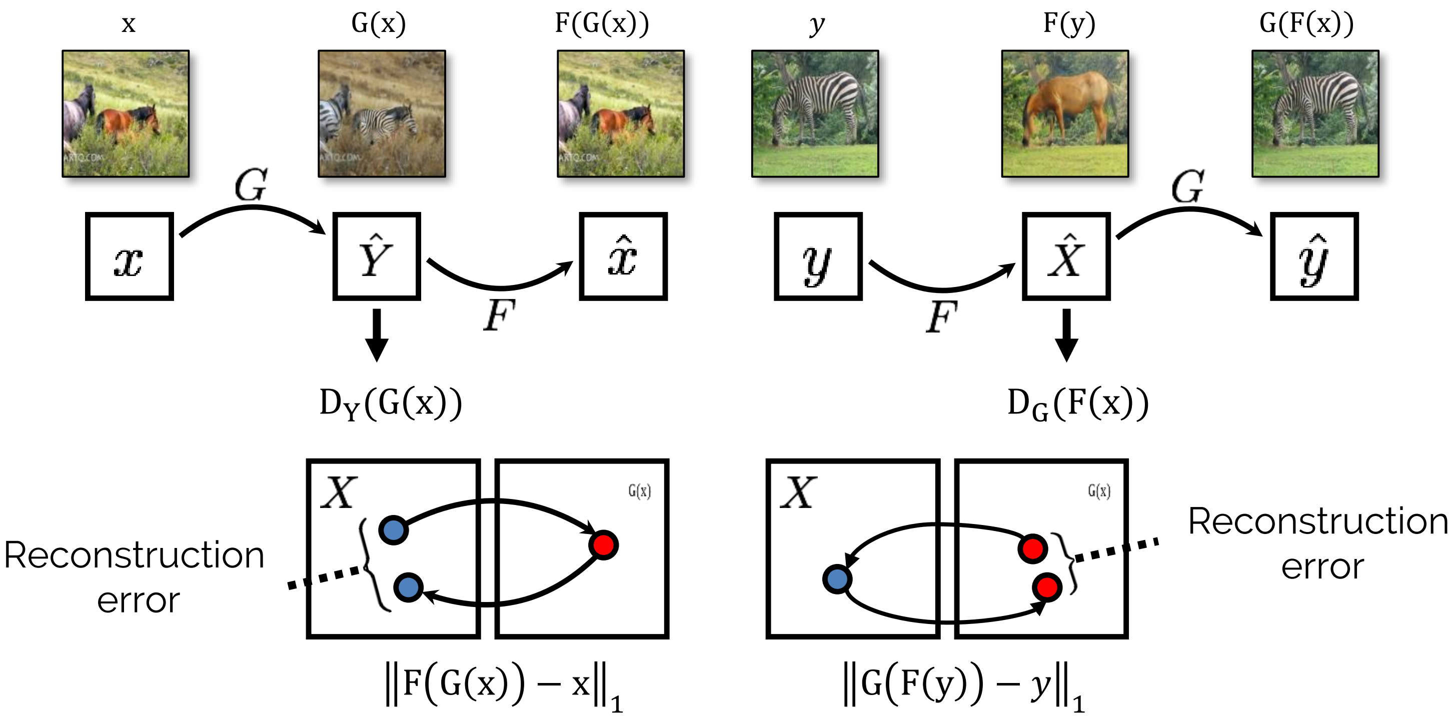
Cycle Consistency Loss





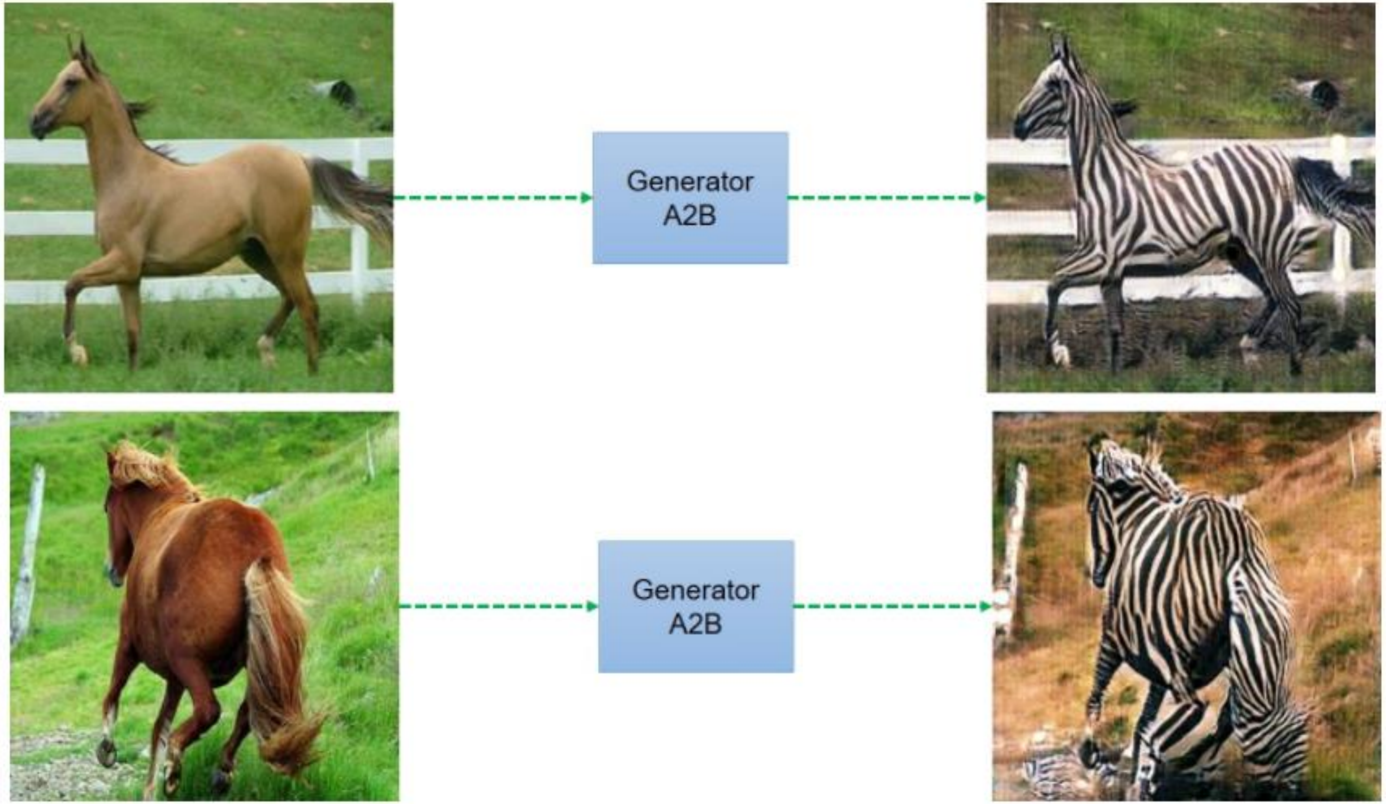
[Zhu et al. 17] Cycle-Consistent Adversarial Networks

Cycle Consistency Loss



[Zhu et al. 17] Cycle-Consistent Adversarial Networks

Cycle GAN - Overview



https://junyanz.github.io/CycleGAN/ [Zhu et al. 17.]

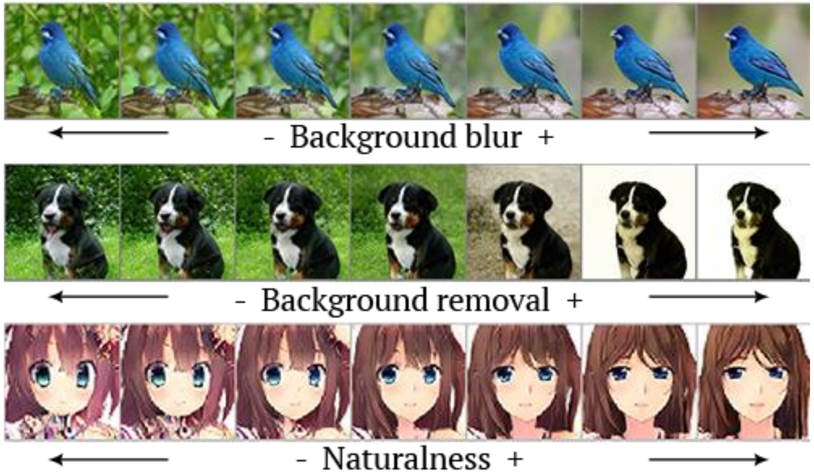
Monet's paintings → photos




Conditional GAN and Inversion Applications

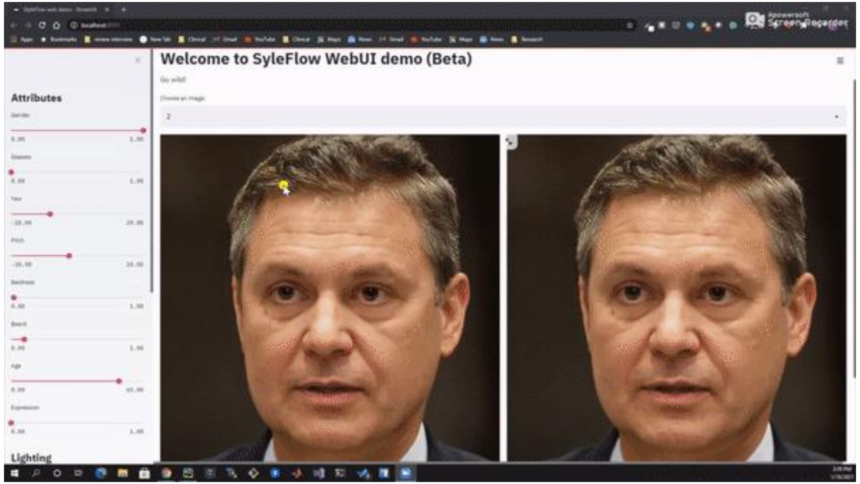
cGANs – Image Manipulation

Manipulate Properties



Voynov et al.: Unsupervised discovery of inter-
pretable directions in the GAN latent space

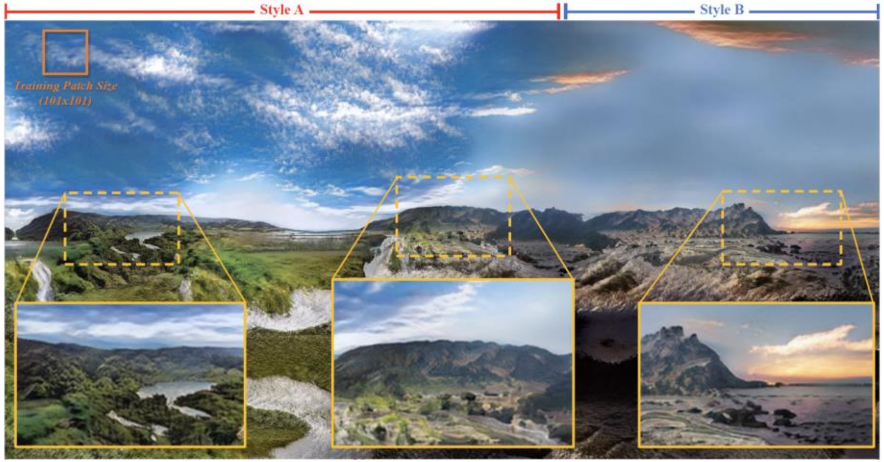
Manipulate Poses



Abdal et al.: StyleFlow

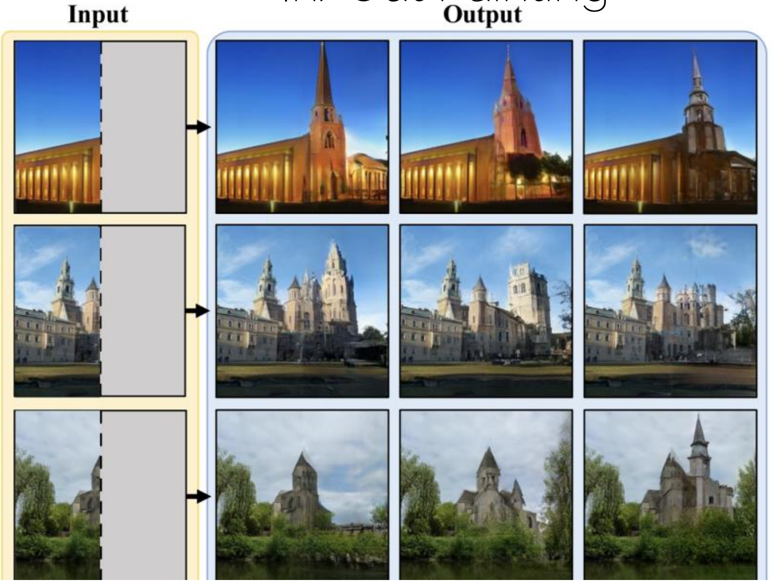
cGANs – Image Generation

Synthesis



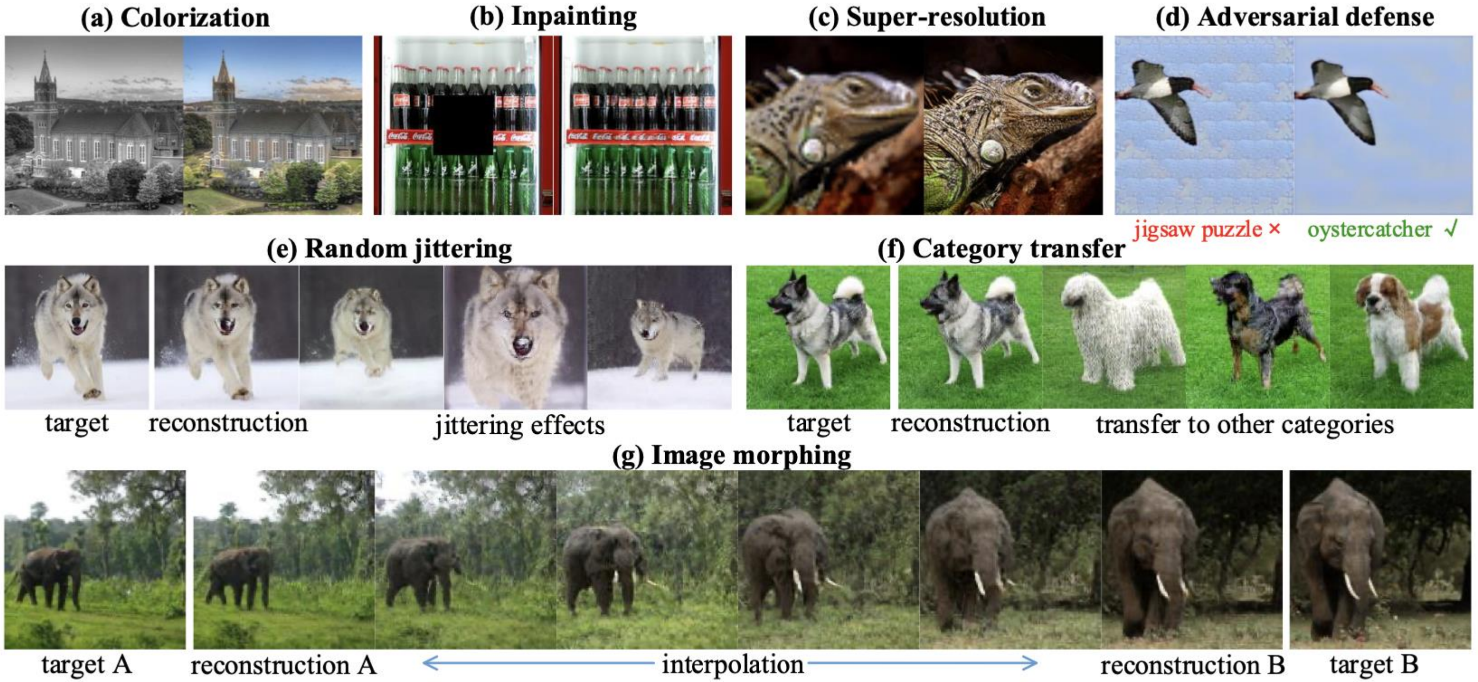
Lin et al.: InfinityGAN

In/Out Painting



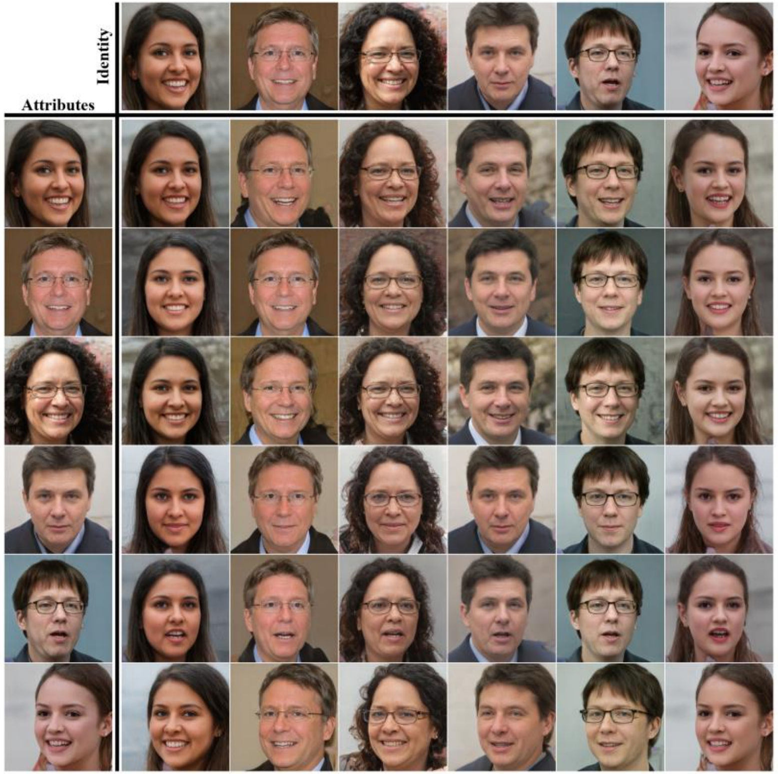
Cheng et al.: In&Out

cGANs – Image Restoration



Pan et al.: Exploiting Deep Generative Prior for Versatile Image Restoration and Manipulation

cGANs – Image Interpolation



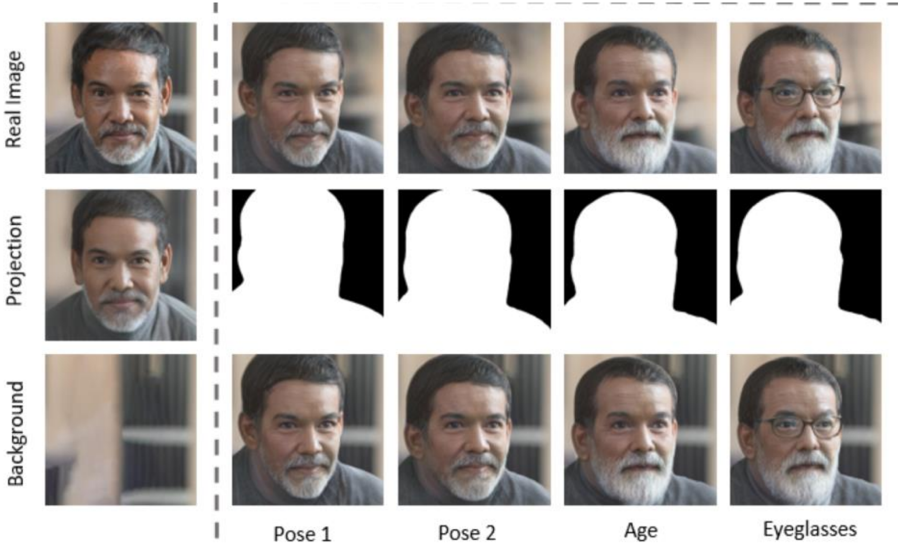
Nitzan et al.: Face identity disentanglement via latent space mapping

cGANs – Image Understanding

Few Shot Segmentation

Tritrong et al.: Repurposing GANs for One-shot Semantic Part Segmentation

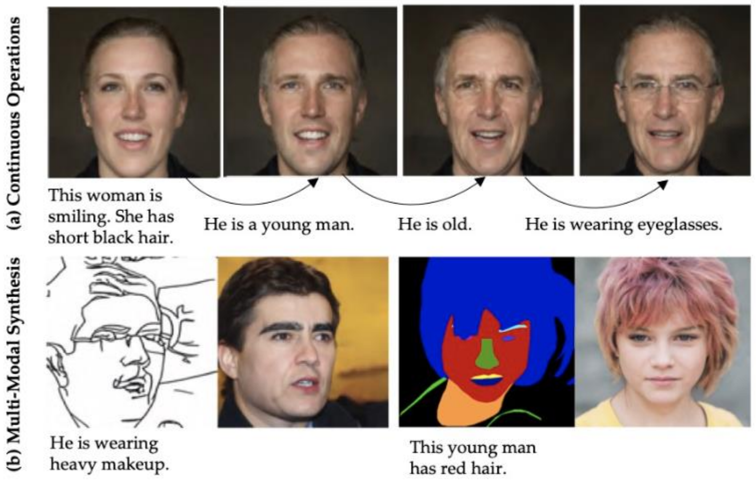
Alpha Matting



Abdal et al.: Labels4Free

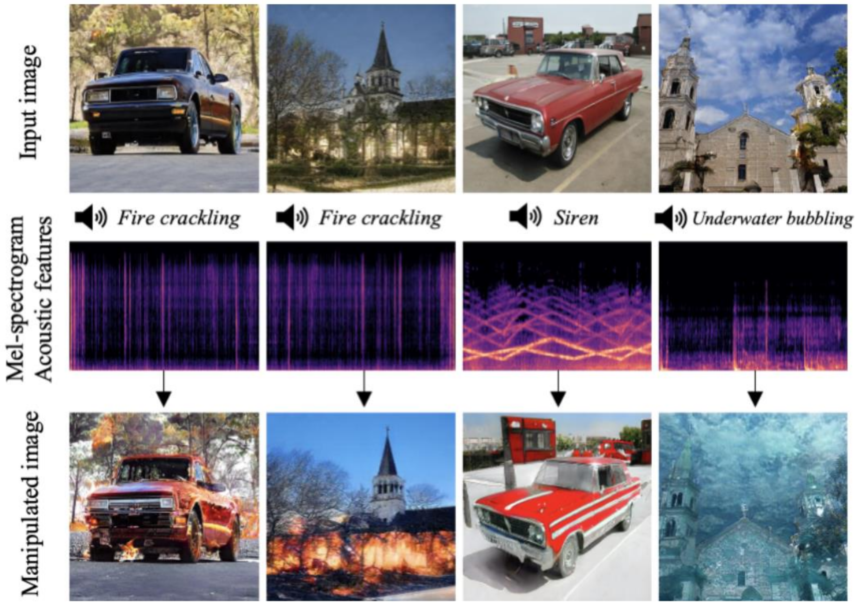
cGANs – Multimodal Manipulation

Text-to-Image



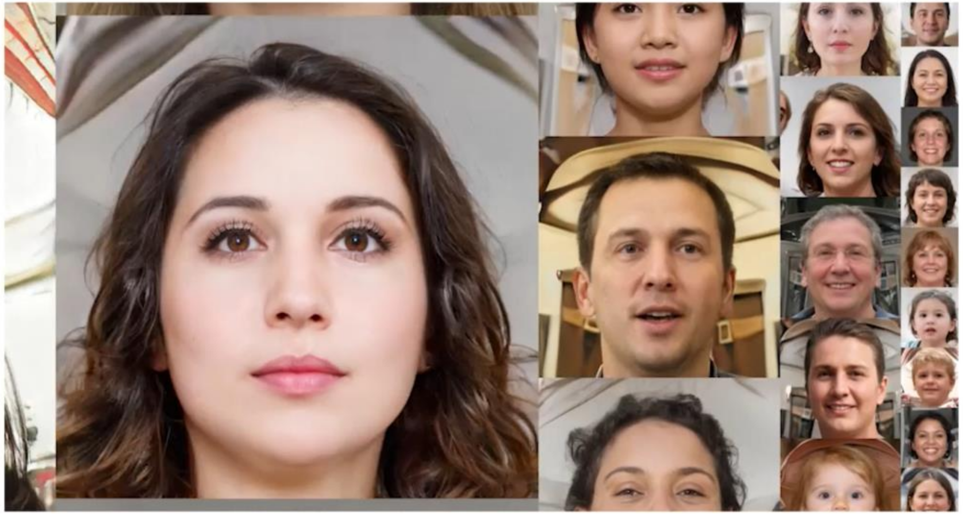
Xia et al.: TediGAN

Sound Manipulation



Lee et al.: Sound-guided semantic image manipulation, interactive results: https://kuai-lab.github.io/cvpr2022sound/

GANs for 3D Aware Synthesis

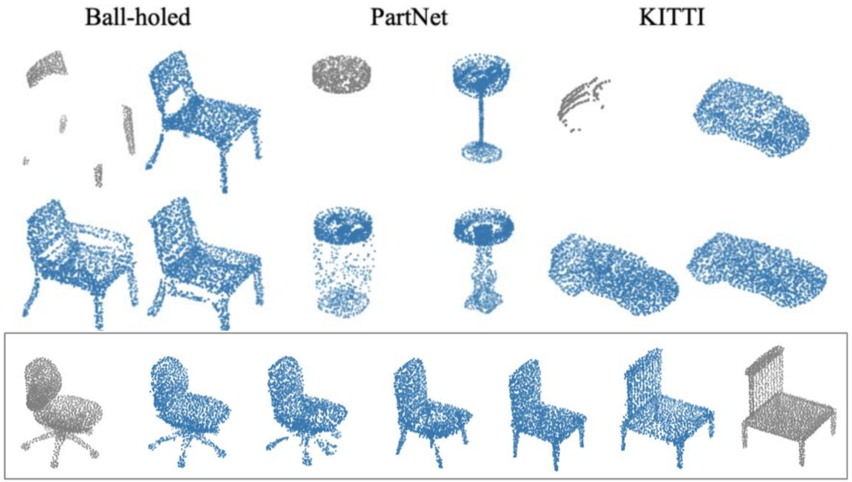


Gu et al.: StyleNerf

Chan et al.: Pi-GAN

GANs for 3D Aware Synthesis

G trained with 3D



Zhang et al.: Unsupervised 3D Shape Completion through GAN Inversion

G trained with 2D

Chan et al.: EG3D

Reading Homework

- [Zhu et al. 2016] Generative Visual Manipulation on the Natural Image Manifold
  - https://arxiv.org/abs/1609.03552

- [Isola, et al. 2017] Image-to-image translation with conditional adversarial networks
  - https://phillipi.github.io/pix2pix/

- [Zhu et al. 2017] Unpaired image-to-image translation using cycle-consistent adversarial networks
  - https://arxiv.org/abs/1703.10593

Thanks for watching!