UNIVERSITÀ DI PISA

Assignment N. 4:

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

ISPR Course A.Y. 2020/2021

Marco Petix

Marco Petix

Unpaired Image-to-Image Translation

Solving the problem of **Image Translation**

- $G : X \to Y$

Using **Unpaired training data**

- Paired data may not exist
- **Underlying relationship** between the domains **X** and **Y** (e.g. **context of a scene**)

Low Diversity Output and **Mode Collapse**

- Pairing up **x** and **y** in a **meaningful way**

Reverse Mapping and **Cycle-Consistency**

- $F : Y \to X$

Monet Paintings ↔ Photos

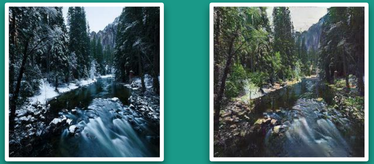
Style Transfer

Photo Monet Van Gogh Cezanne Ukiyo-e

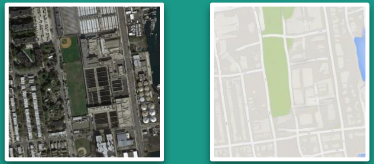
Object Transfiguration

Photo Enhancement

Season Transfer

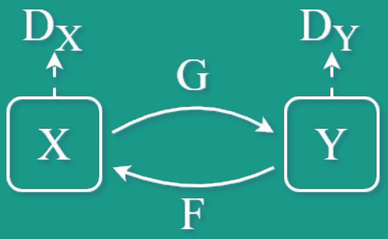


Aerial Photos ↔ Google Maps



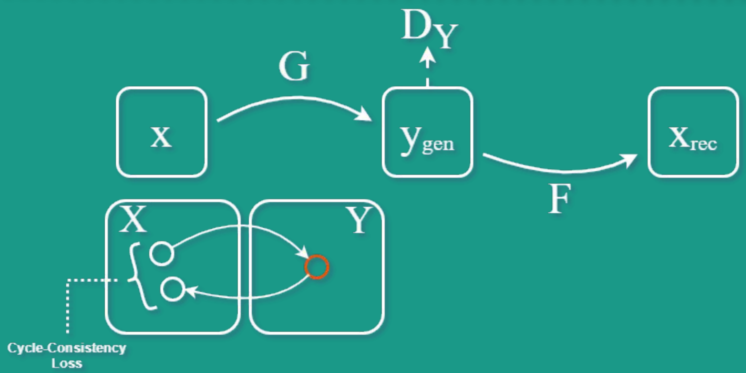
Marco Petix

Cycle - Consistent Adversarial Networks



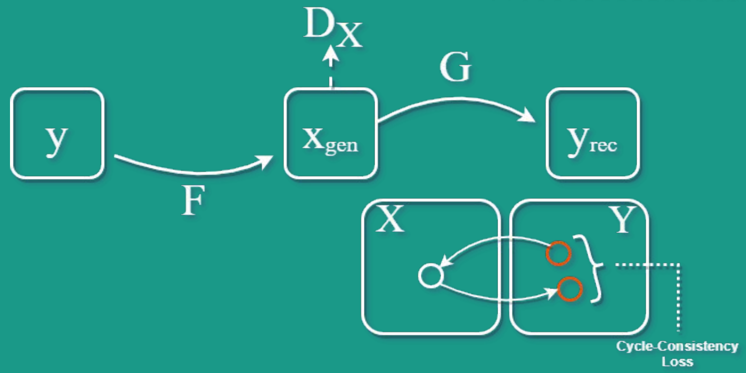
Applying **Adversarial Losses** to the **mapping functions G** and **F**

- $\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim P_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y \sim P_{data}(y)}[\log(1 - D_X(F(y)))]$
- $\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim P_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim P_{data}(x)}[\log(1 - D_Y(G(x)))]$



Avoiding **Mode Collapse** by enforcing **Cycle-Consistency**

- $\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim P_{data}(x)}[||F(G(x)) - x||_1] + \mathbb{E}_{y \sim P_{data}(y)}[||G(F(y)) - y||_1]$



Representing the **Full Objective**

- $\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F)$

Color Composition and **Identity Loss**

- $\mathcal{L}_{identity}(G, F) = \mathbb{E}_{x \sim P_{data}(x)}[||F(x) - x||_1] + \mathbb{E}_{y \sim P_{data}(y)}[||G(y) - y||_1]$

Marco Petix

Model Architecture

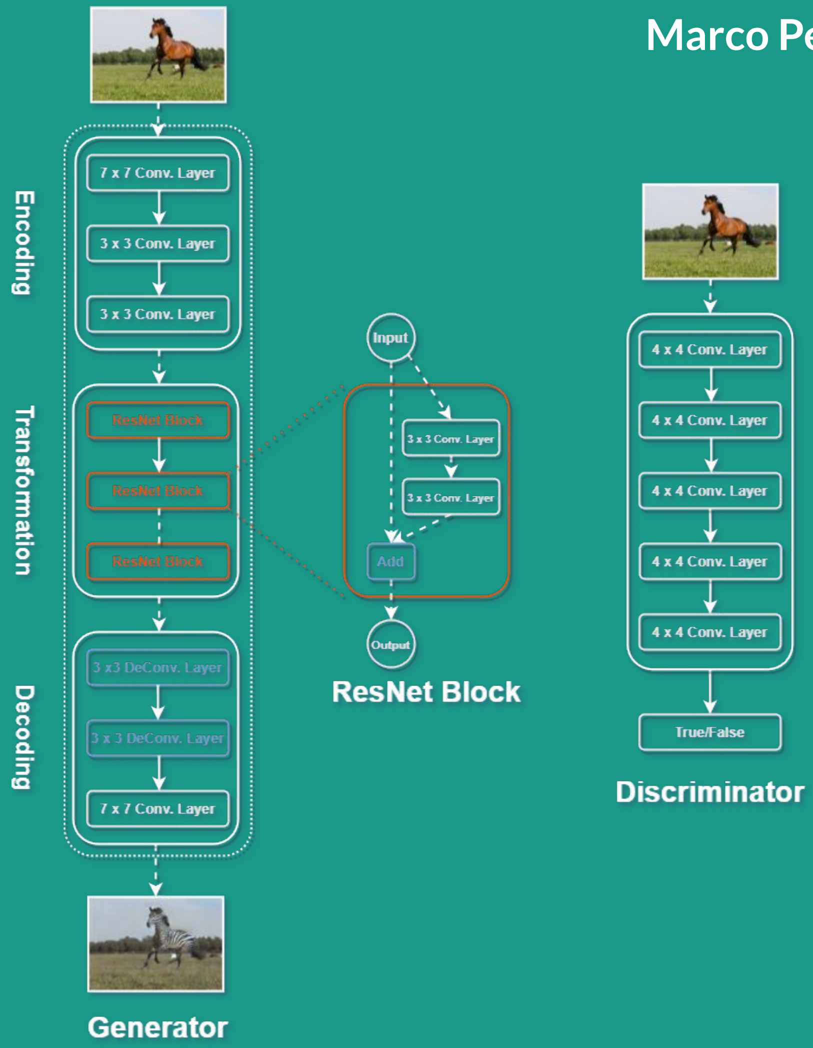
Generator

- Encoding
  - 3 Convolutional Layers (ReLU)
- Transformation
  - 6 ResNet Blocks for 128 × 128 images
  - 9 ResNet Blocks for 256 × 256 images
- Decoding
  - 2 De-Convolutional Layers (ReLu)
  - 1 Convolutional Layer (ReLU)

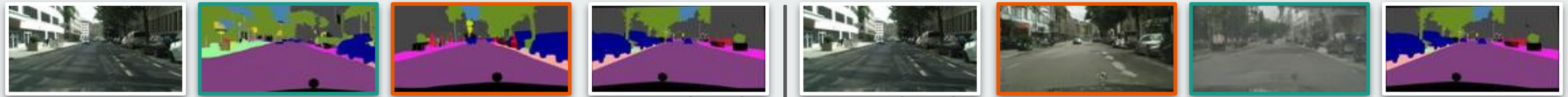
Discriminator

- 70 x 70 Patch-GAN
  - 4 Convolutional Layers (Leaky ReLU)
  - 1 Final Conv. Layer for the Decision

Also, Instance Normalization

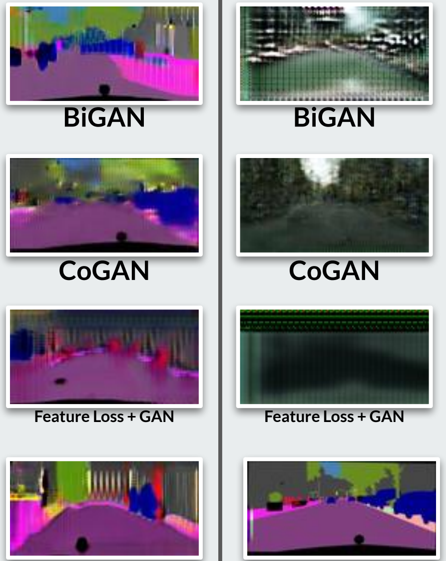


Empirical Results: Comparison on the Cityscapes dataset



Photos → Cityscapes labels

Loss	Per-pixel acc.	Per-class acc.	Class IOU
BiGAN	0.41	0.13	0.07
CoGAN	0.45	0.11	0.08
Feature Loss + GAN	0.50	0.10	0.06
SimGAN	0.47	0.12	0.07
CycleGAN	**0.58**	**0.22**	**0.16**
pix2pix	**0.85**	**0.40**	**0.32**



Cityscapes labels → Photos

Loss	Per-pixel acc.	Per-class acc.	Class IOU
BiGAN	0.19	0.06	0.02
CoGAN	0.40	0.10	0.06
Feature Loss + GAN	0.06	0.04	0.01
SimGAN	0.20	0.10	0.04
CycleGAN	**0.52**	**0.17**	**0.11**
pix2pix	**0.71**	**0.25**	**0.18**

Final Considerations

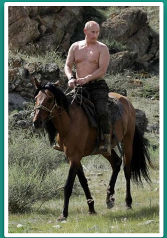
Pushing the boundaries of **Unsupervised Learning**

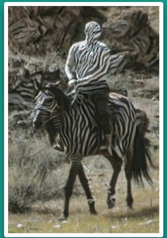
- **Compelling results** for tasks revolving around **color** and **texture changes**
- Dependent on the **distribution characteristics** of the training datasets
- Less performant on **shape changes**

CycleGAN's Descendants

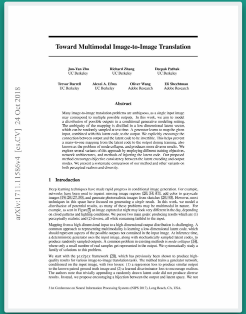
- **Toward Multimodal Image-to-Image Translation** (Zhu et Al., 2017)
- **CyCADA: Cycle-Consistent Adversarial Domain Adaptation** (Isola et Al., 2018)
- **Contrastive Learning for Unpaired Image-to-Image Translation** (Park et Al., 2020)

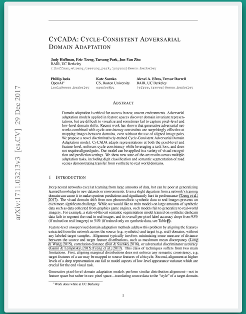
Man on Horse → Zebra Centaur

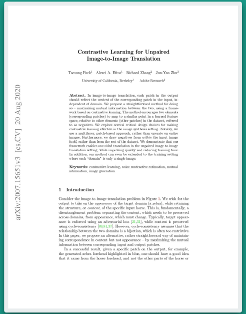




Cat → Dog?







Zhu et Al. 2017

Isola et Al. 2018

Park et Al. 2020