

Vision & Perception 2019/20

Project Presentation

Francesco Starna

Sapienza University of Rome



Human Face Translation with GAN

Project

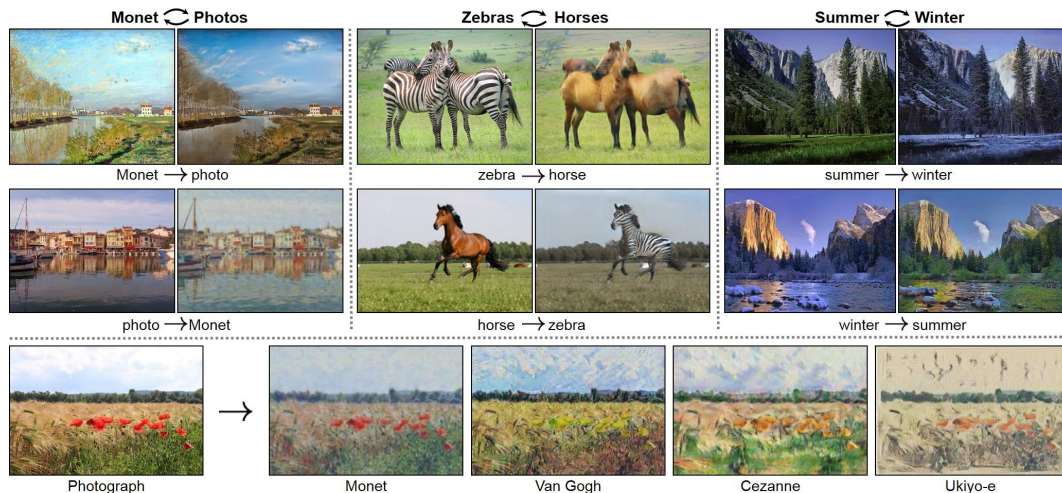
Project Goal

Image Translation using Generative Adversarial Networks

CycleGAN

(Jun-Yan Zhu et al.)

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



Generative Adversarial Networks

- Learns to generate new data with the same **statistic distribution** of the training set

- Supervised and Unsupervised learning methods

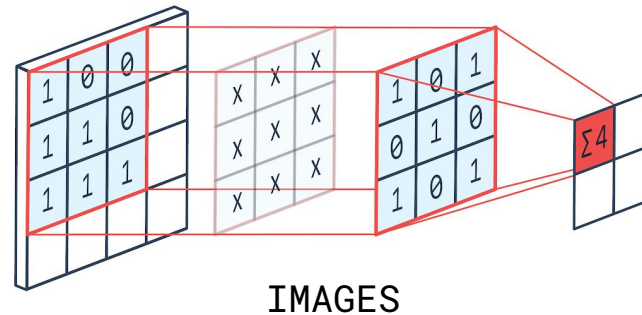
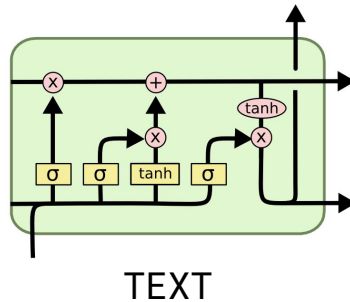
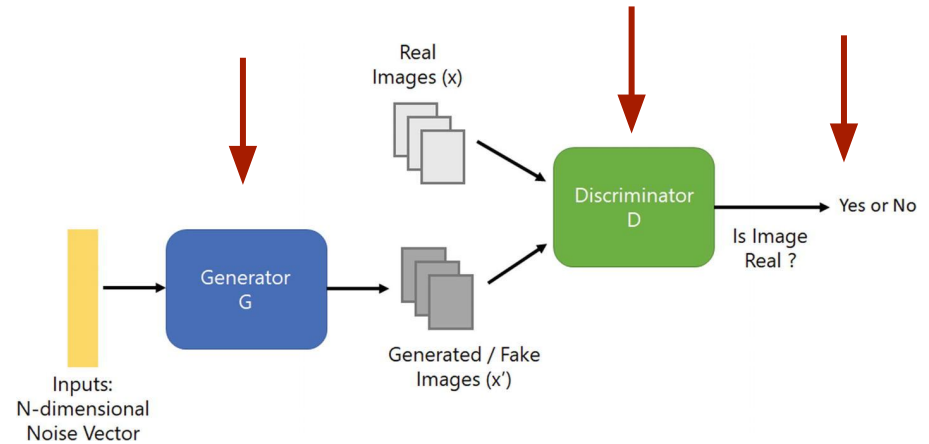
- Large domains application:

- Image Generation
- Image Translation
- Super Resolution
- Style Transfer
- Text to Image
- ... and many more



GAN Architecture

- Generator
- Discriminator
- Loss Function
- Model



CycleGAN

- Unsupervised Learning of mappings

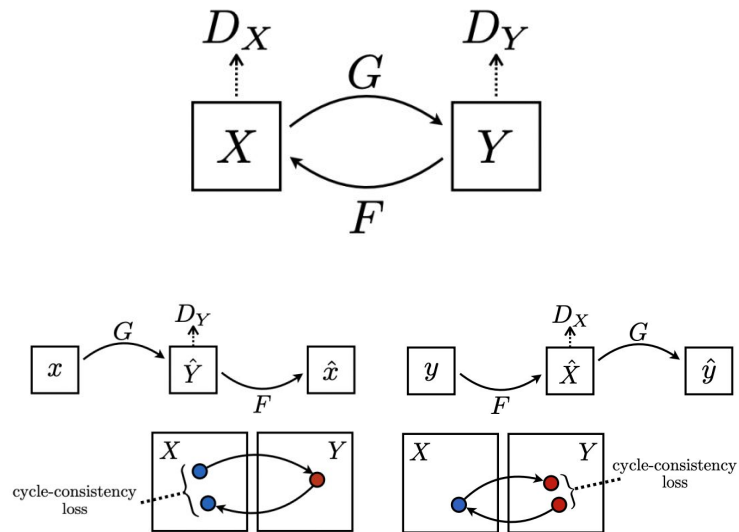
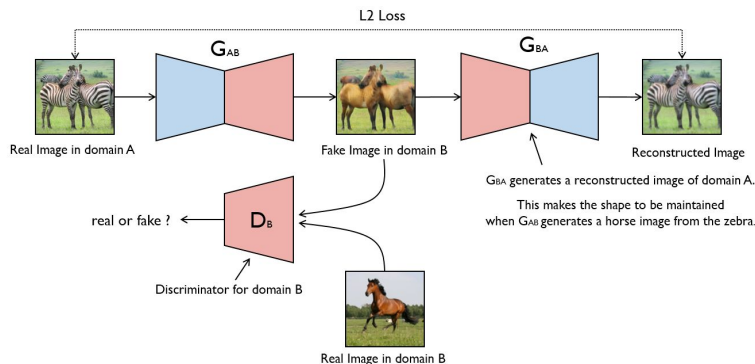
G: $X \rightarrow Y$ and inverse **F:** $Y \rightarrow X$

- **Cycle** consistency loss

- **Architecture**

2 Generator

2 Discriminator



Generator

ENCODER

- **Convolutional Layer**
- Highlights Extraction
- Downsample

TRANSFORMER

- **Residual Connection**
- Join Features
- Same dimension

DECODER

- **Transposed Convolutional Layer**
- Image Construction from low-level
- Upsample



Discriminator

- PatchGAN

(Isola et al.)

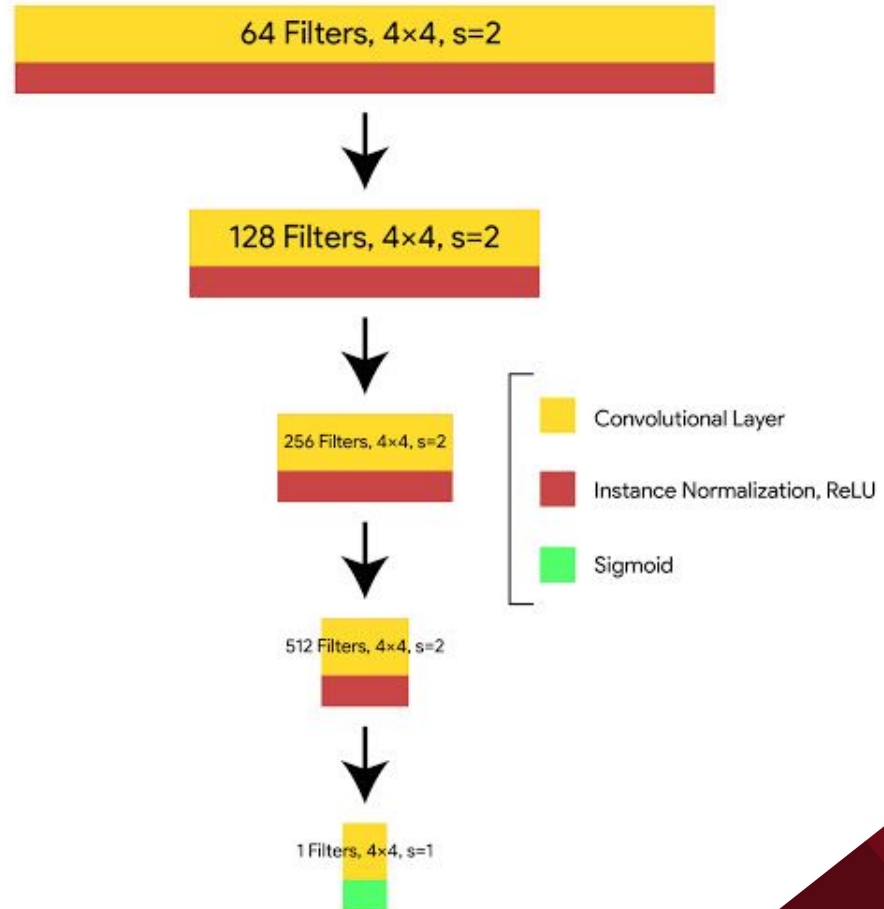
Image to-image translation
with conditional adversarial
networks.

- Layers

- Convolutional Layer
- Instance Normalization
- Leaky ReLU (0.01x if $x < 0$)

- Mapping

- 256x256 to NxN array
- Average to classify Real or Fake



Objective

- 2 Generators **G** and **F**
- 2 Discriminator **D_x** and **D_y**
- 1st Adversarial Loss

$$Loss_{advers} (G, D_y, X, Y) = \frac{1}{m} \sum (1 - D_y (G (x)))^2$$

- 2nd Adversarial Loss

$$Loss_{advers} (F, D_x, Y, X) = \frac{1}{m} \sum (1 - D_x (F (y)))^2$$

- **Cycle Consistency Loss**

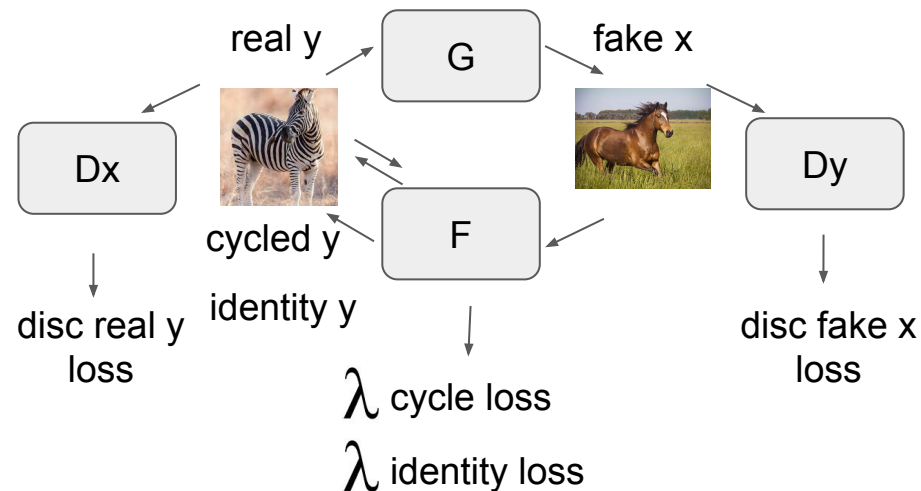
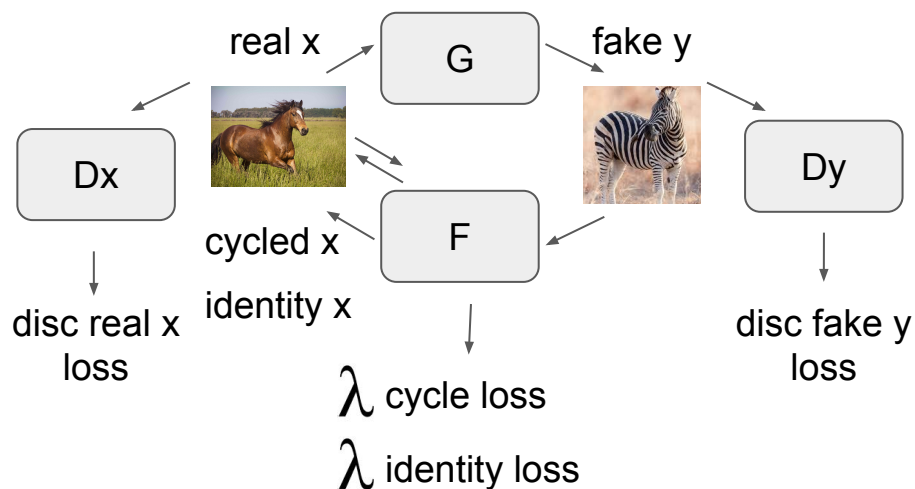
$$Loss_{cyc} (G, F, X, Y) = \frac{1}{m} [(F (G (x_i)) - x_i) + (G (F (y_i)) - y_i)]$$



Training

Optimizer: Adam (Kingma and Ba, 2017)

Epochs: 200 (100 fixed lr + 100 decay)



generator loss = total cycle loss + identity loss

discriminator loss = disc real loss + disc fake loss



Dataset

- Domain X
FLICKR FACE



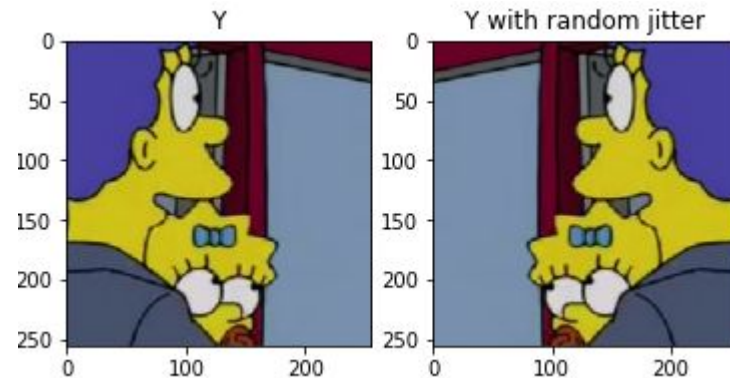
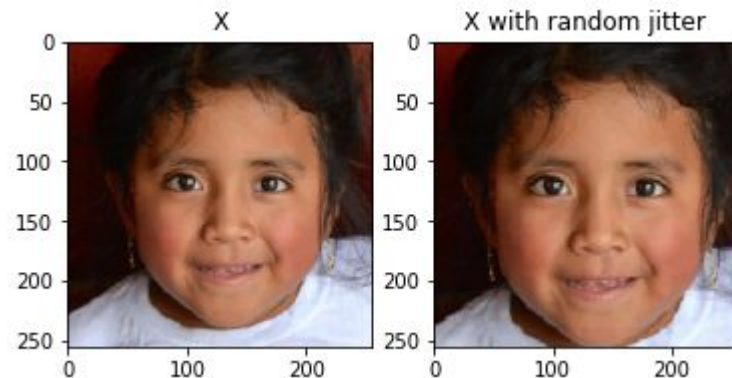
- Domain Y
SIMPSON FACE
ANIMAL FACE
BITMOJI FACE



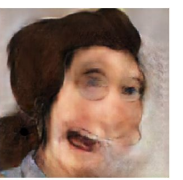
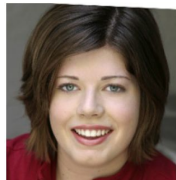
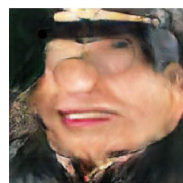
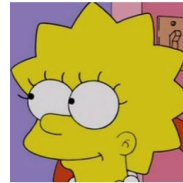
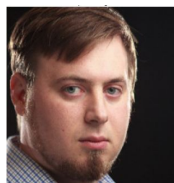
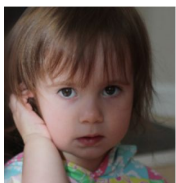
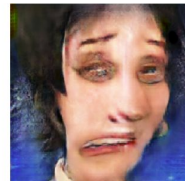
1000 Images Train
100 Images Test
Each

Preprocessing

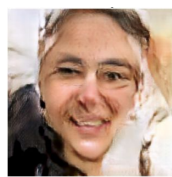
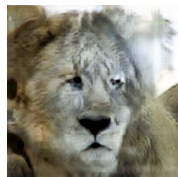
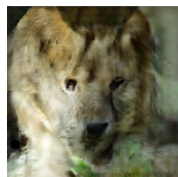
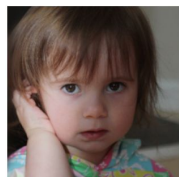
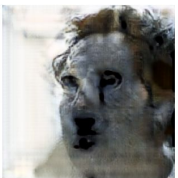
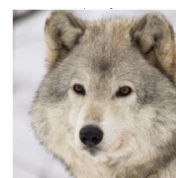
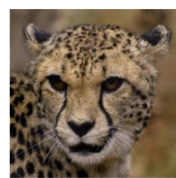
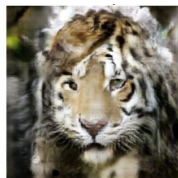
- Normalization
[0,255] to [-1,1]
- Data Augmentation
Random Jittering (resize, crop, flip)



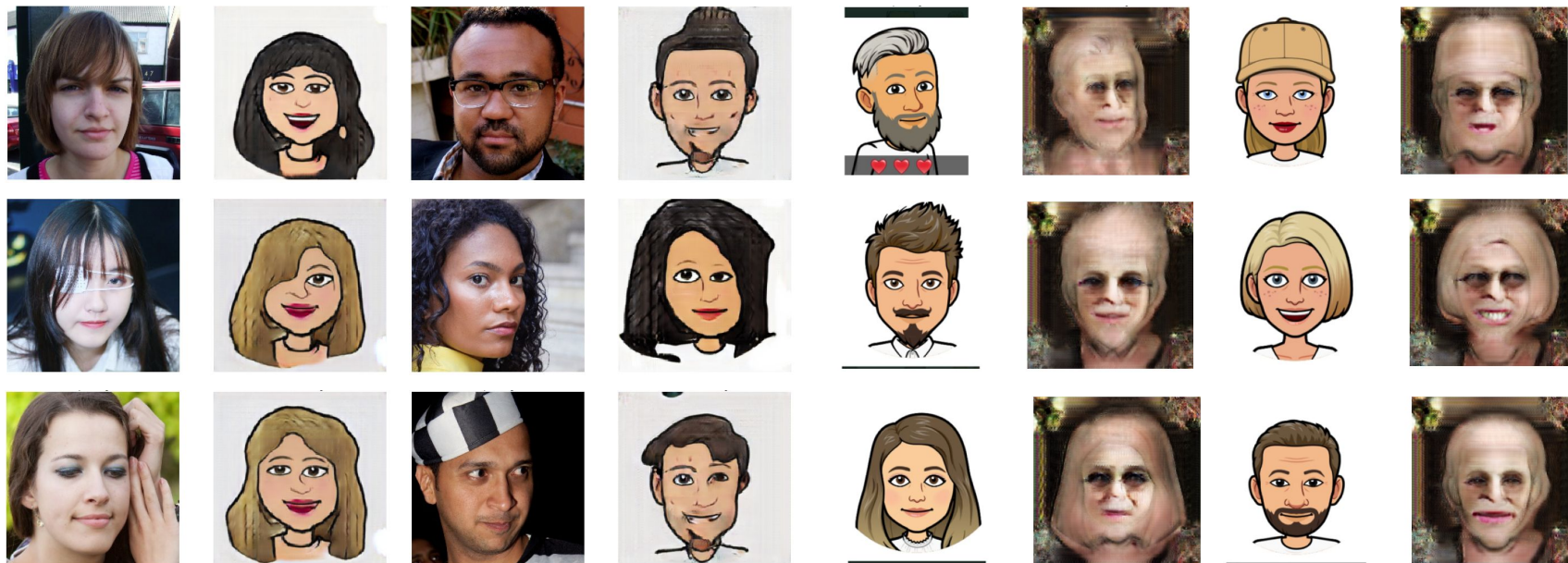
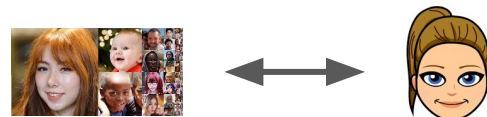
Results (Human & Simpson)



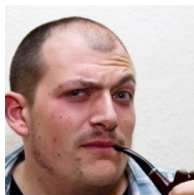
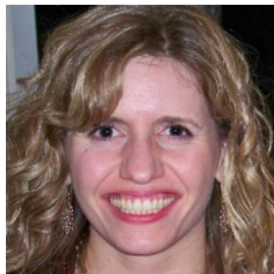
Results (Human & Animal)



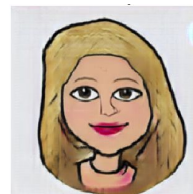
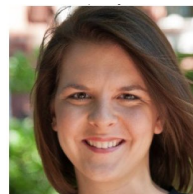
Results (Human & Bitmoji)



Results



1



2

3



Comments

- Data Augmentation > more training samples
- Large Geometric Shifts are not Successful
- Visual Inspection is better than more epochs
 - Simpson ~ 100 epochs
 - Animal ~ 150 epochs
 - Bitmoji ~ 120 epochs
- CycleGAN can be improved
Reduce oscillation by feeding the discriminator with a history of n generated images rather than last ones

Generator G result after 200 epochs training



Thanks

