Vision & Perception 2019/20

Project Presentation

Francesco Starna

Sapienza University of Rome

SAPIENZA
UNIVERSITÀ DI ROMA

Human Face Translation with GAN

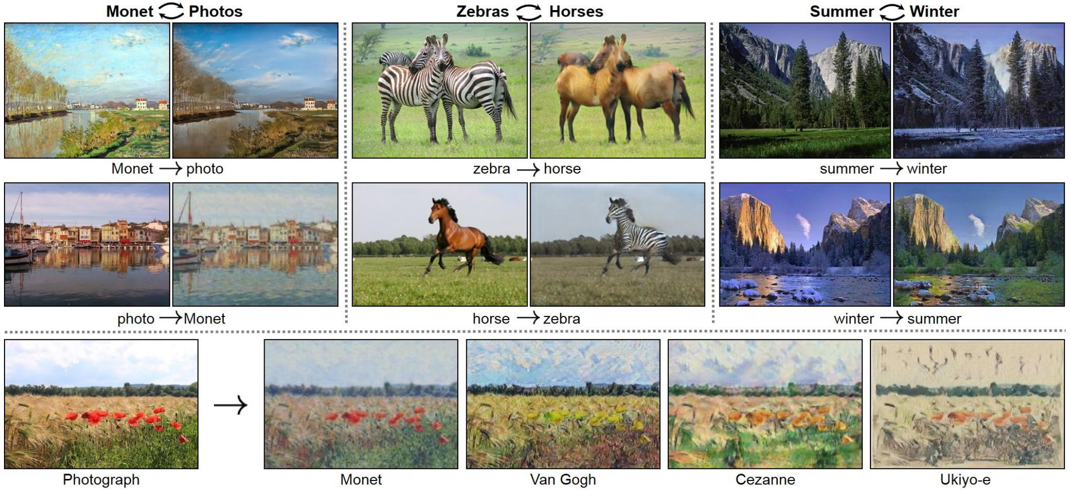
Project

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

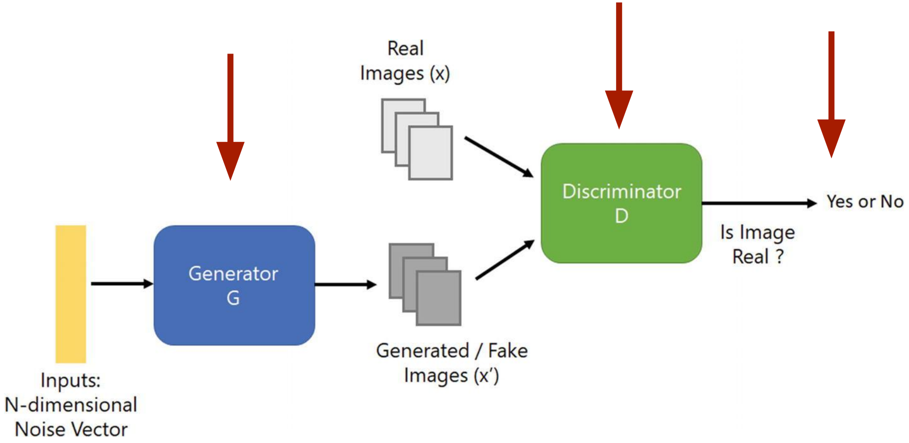
- Supervised and Unsupervised learning methods

- Large domains application:
  - -Image Generation
  - -Image Translation
  - -Super Resolution
  - -Style Transfer
  - -Text to Image
  - … and many more

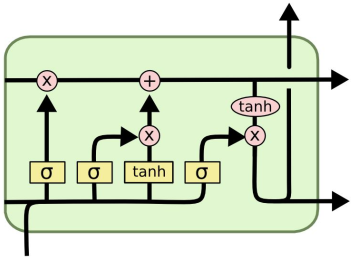
this bird is red with white and has a very short beak
horse → zebra

GAN Architecture

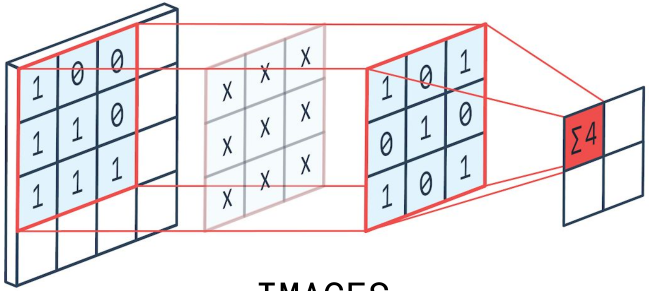
- Generator
- Discriminator
- Loss Function
- Model



$\min_G$



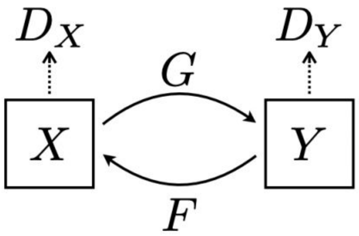
TEXT



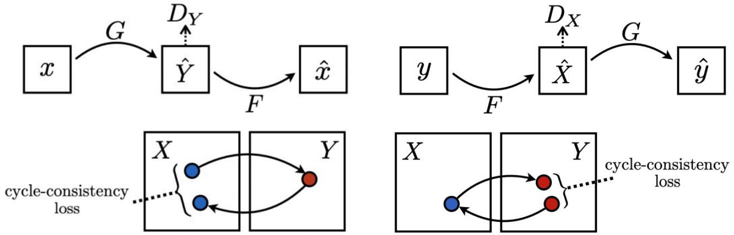
IMAGES

$(G(z)))]$

CycleGAN



- Unsupervised Learning of mappings
  G: X → Y and inverse **F: Y → X**

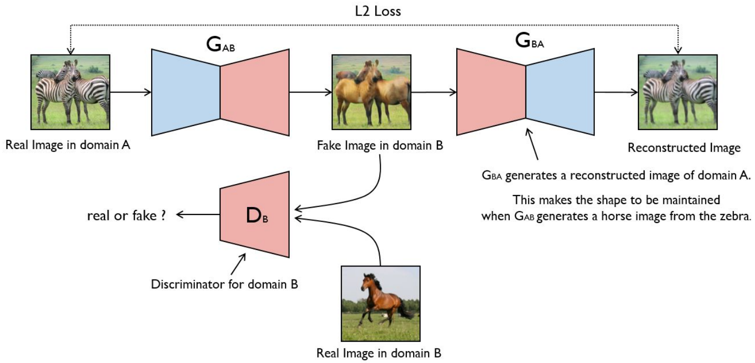


- **Cycle** consistency loss

- **Architecture**

  2 Generator

  2 Discriminator



G_{BA} generates a reconstructed image of domain A.

This makes the shape to be maintained when G_{AB} generates a horse image from the zebra.

Generator

ENCODER

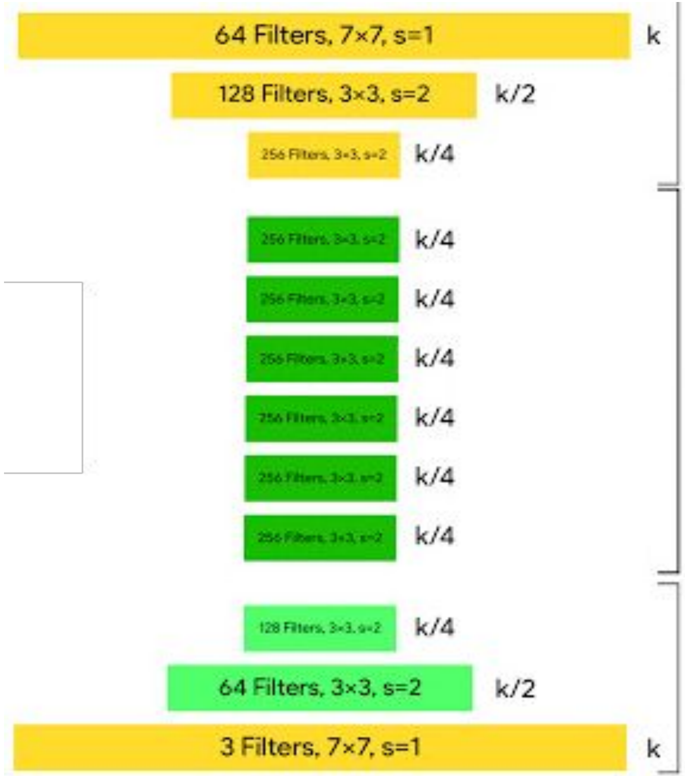
- **Convolutional Layer**
- Highlights Extraction
- Downsample

TRANSFORMER

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

DECODER

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



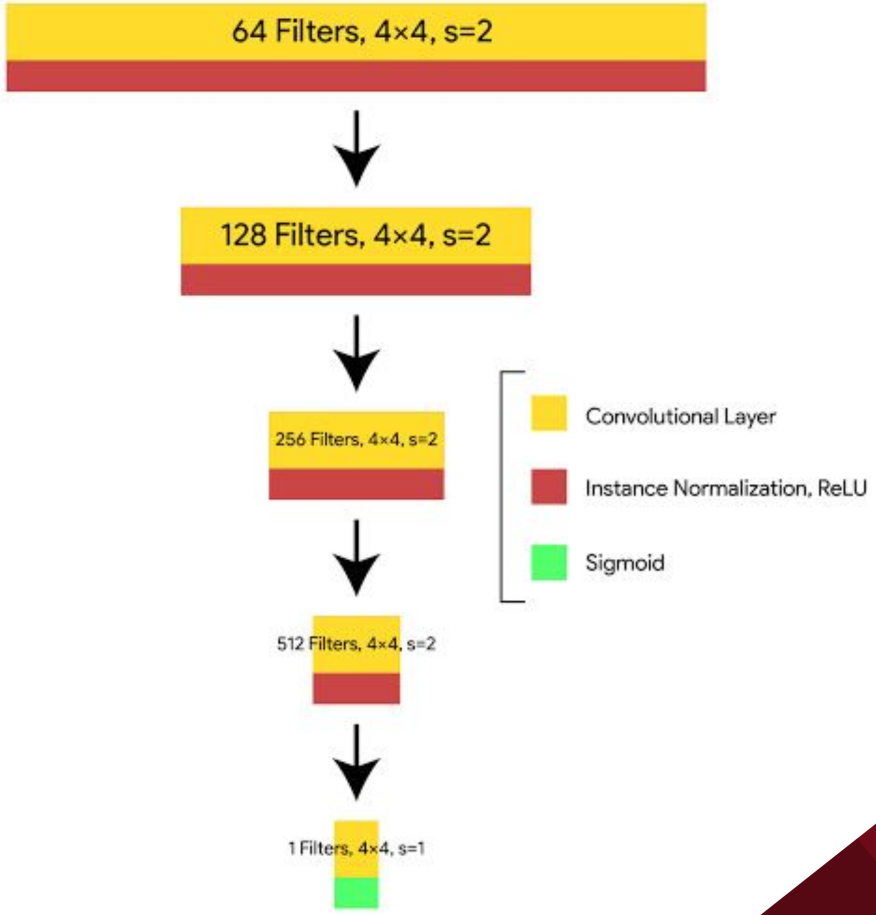
Discriminator

- 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



64 Filters, 4×4, s=2

128 Filters, 4×4, s=2

256 Filters, 4×4, s=2

512 Filters, 4×4, s=2

1 Filters, 4×4, s=1

🟨 Convolutional Layer

🟥 Instance Normalization, ReLU

🟩 Sigmoid

Objective

- 2 Generators **G** and **F**
- 2 Discriminator **Dx** and **Dy**
- 1st Adversarial Loss

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

- 2nd Adversarial Loss

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

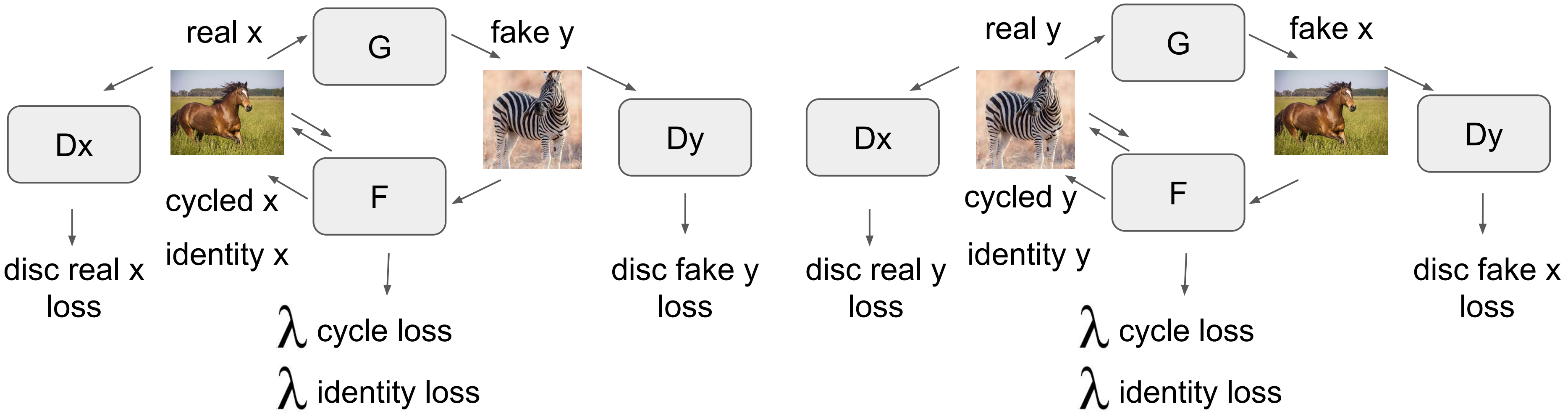
- **Cycle Consistency Loss**

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

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

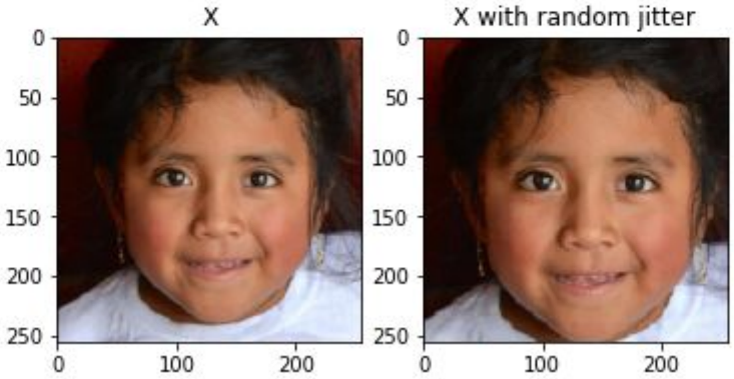
Preprocessing

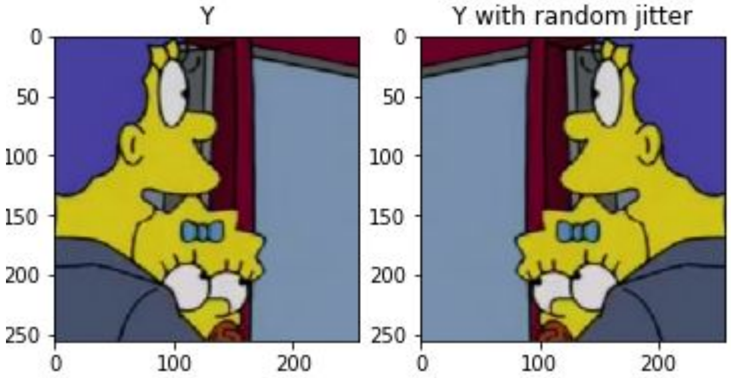
- Normalization

  [0,255] to [-1,1]

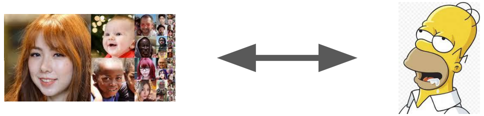
- Data Augmentation

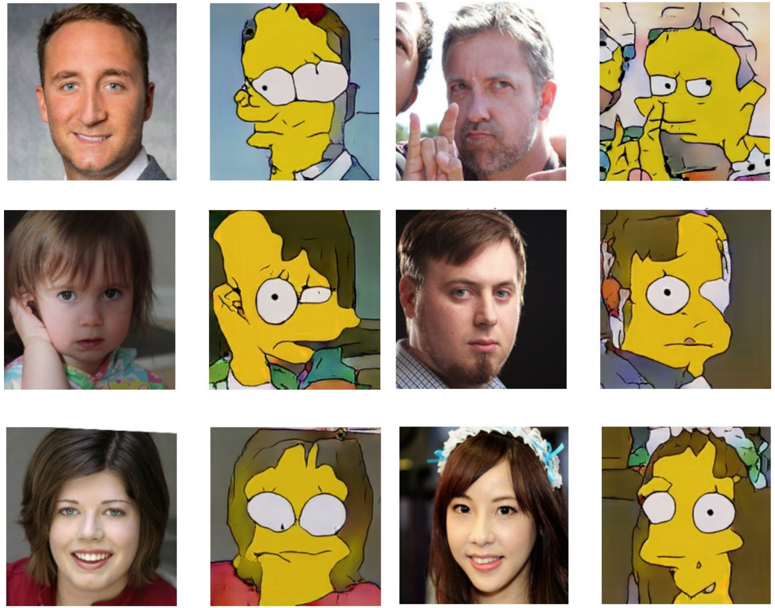
  Random Jittering (resize, crop, flip)

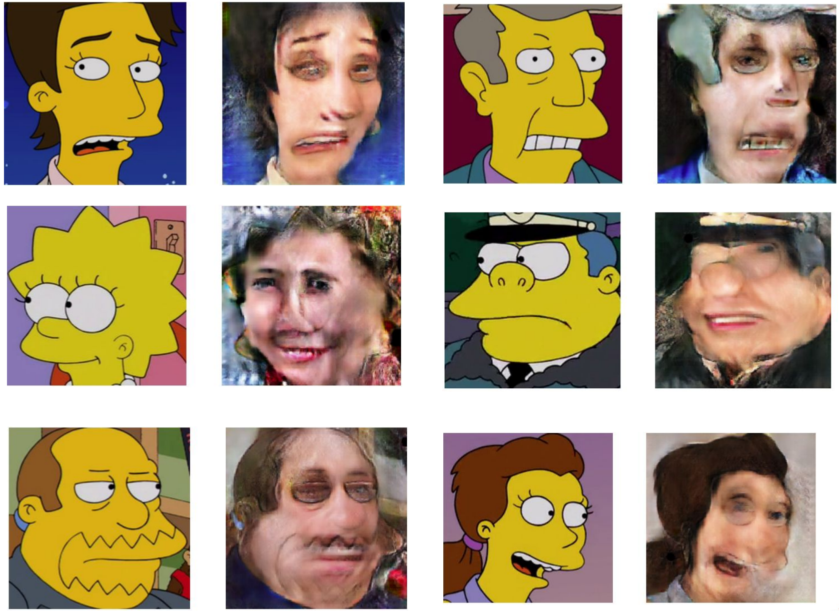




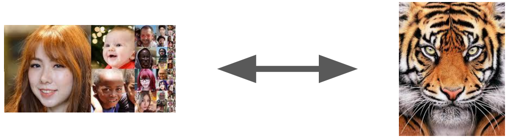
Results (Human & Simpson)

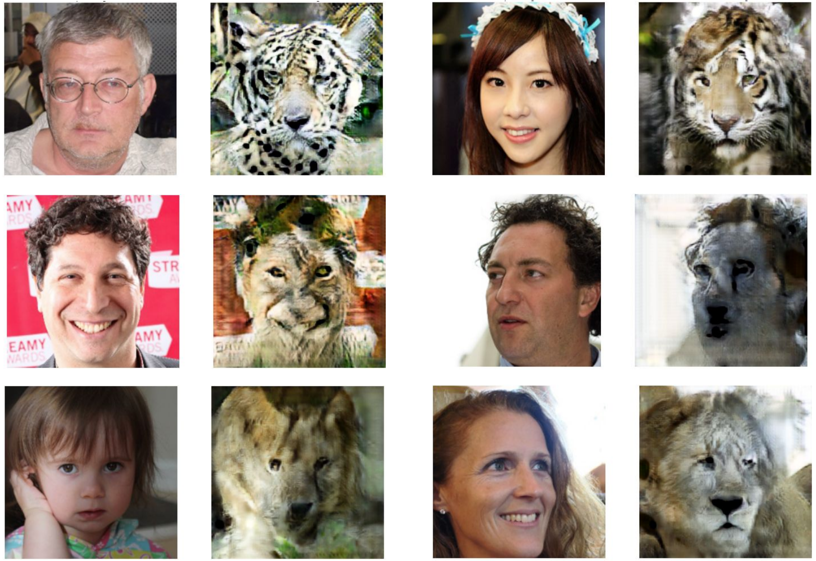


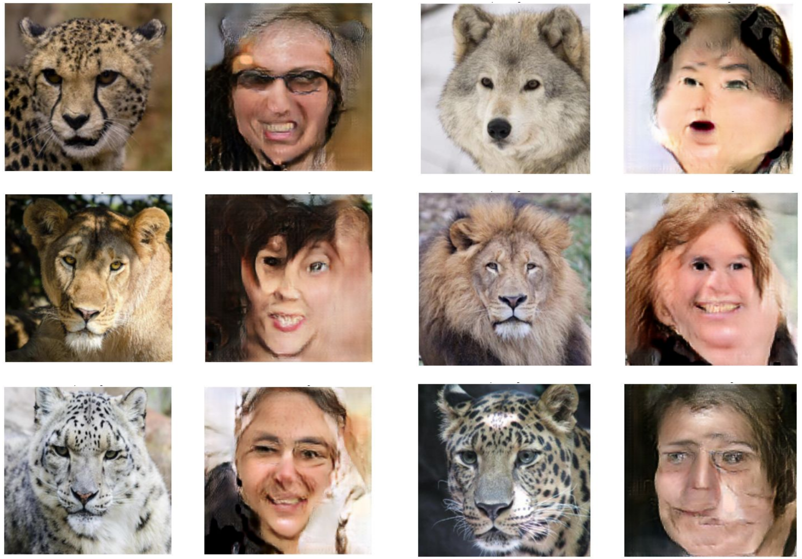




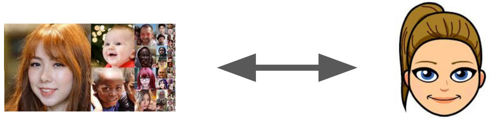
Results (Human & Animal)

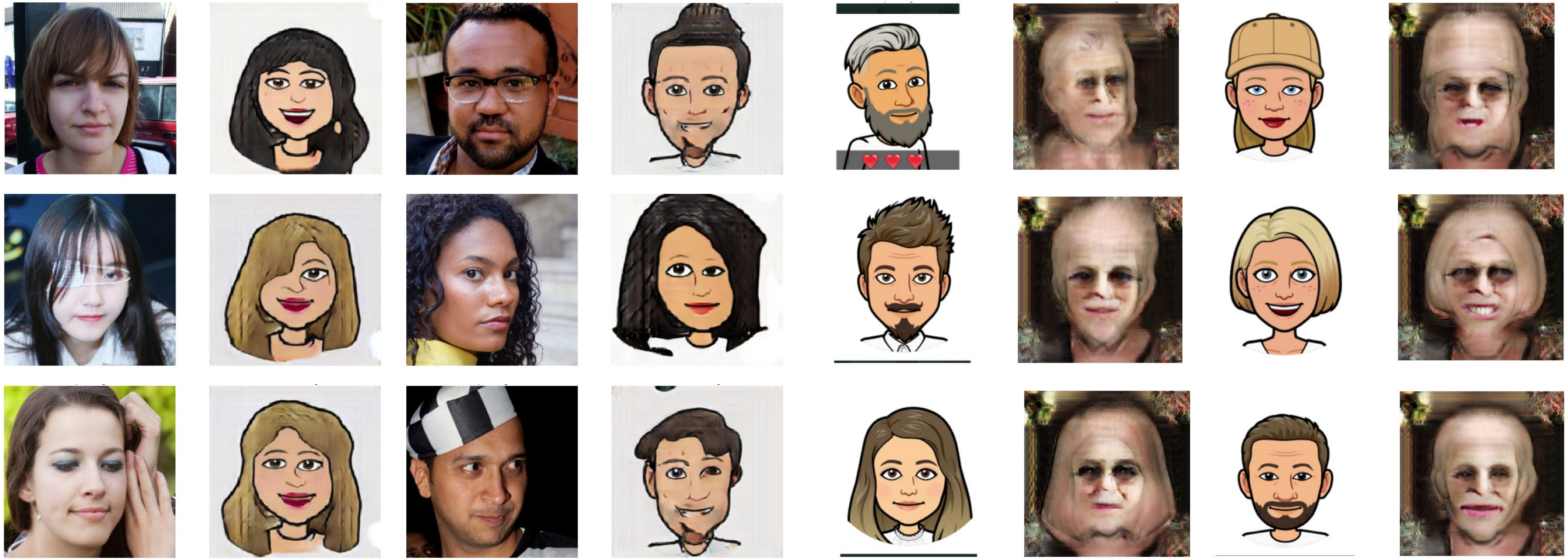




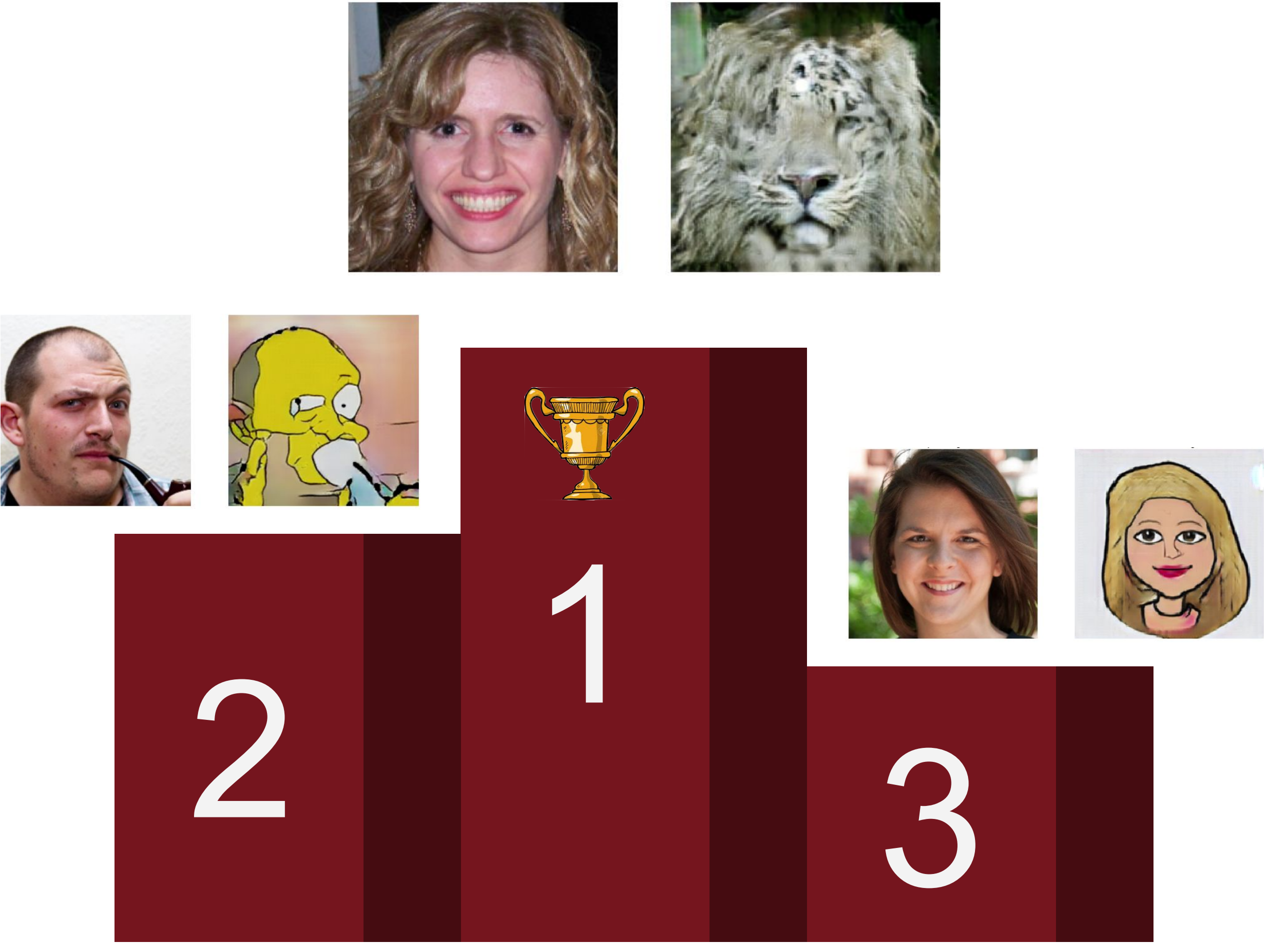


Results (Human & Bitmoji)





Results



Comments

- Data Augmentation > more training samples

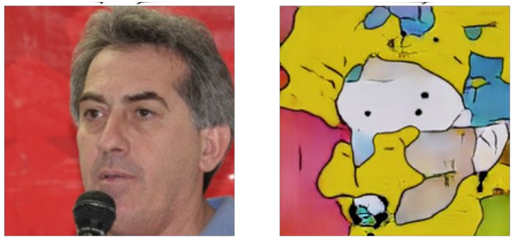
- Large Geometric Shifts are not Successful

- 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