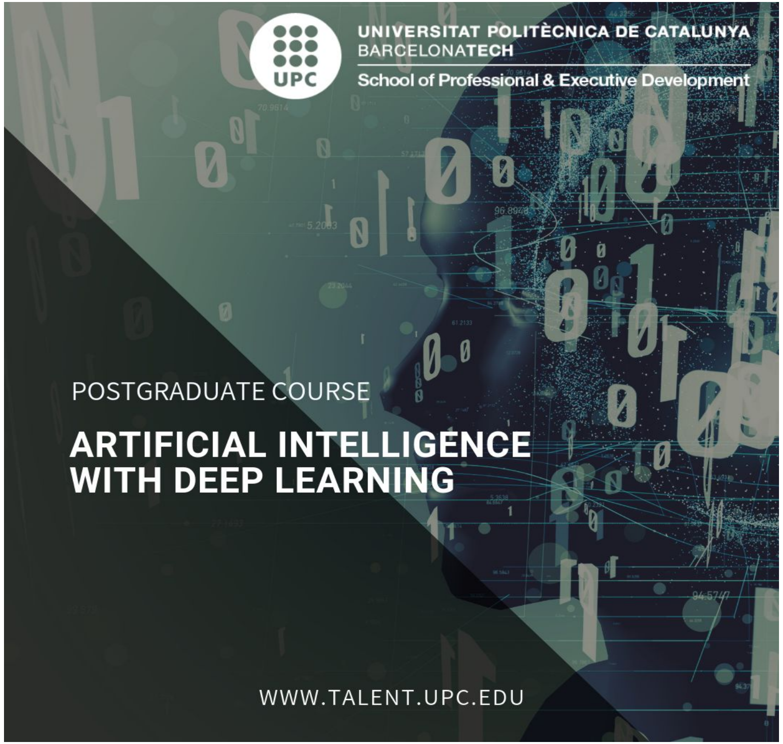


Image-to-Image Domain Translation: XGAN

github.com/rogeralmato/XGAN

Roger Almató
Claudio Curieses
Bernat Torres
Jorge Uribe

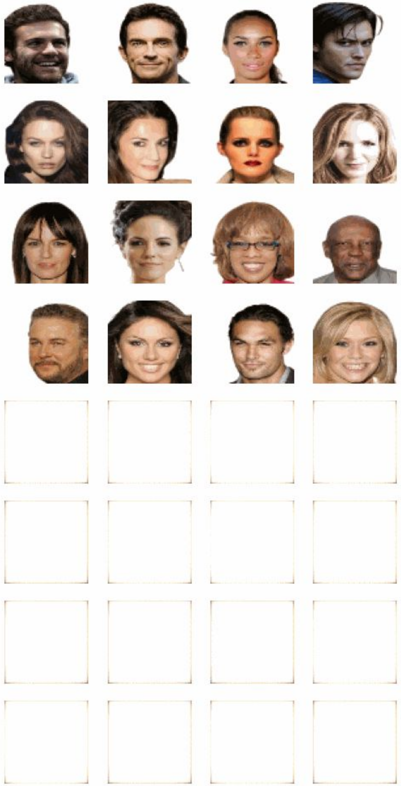


Image-to-Image Domain Translation: XGAN

Roger Almató
Claudio Curieses
Bernat Torres
Jorge Uribe

Contents

- ➔ **Proposal**
- ➔ Motivation
- ➔ Milestones
- ➔ Project Plan
- ➔ Datasets
- ➔ Results
- ➔ Conclusions

XGAN Paper

XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings

Amélie Royer[1 [0000−0002−8407−0705]], Konstantinos Bousmalis[2,6], Stephan Gouws[2], Fred Bertsch[3], Inbar Mosseri[4], Forrester Cole[4], and Kevin Murphy[5]

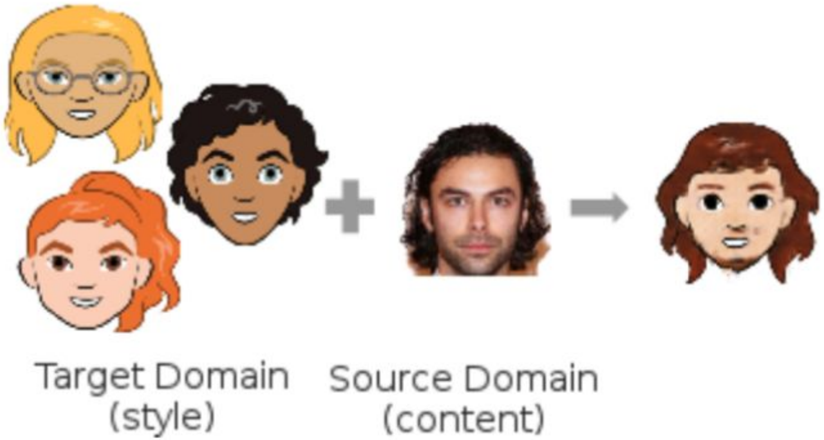
[1] IST Austria, 3400 Klosterneuburg, Austria
Work done while at Google Brain London, UK
aroyer@ist.ac.at
[2] Google Brain, London, UK
{konstantinos, sgouws}@google.com
[3] Google Brain, Mountain View, USA
[4] Google Research, Cambridge, USA
[5] Google Research, Mountain View, USA
[6] Currently at Deepmind, London, UK

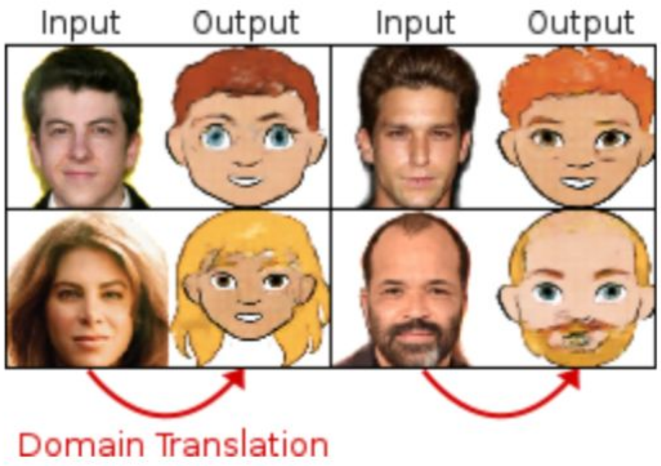
Abstract. Image translation refers to the task of mapping images from a visual domain to another. Given two unpaired collections of images, we aim to learn a mapping between the corpus-level style of each collection, while preserving semantic content shared across the two domains. We introduce XGAN, a dual adversarial auto-encoder, which captures a shared representation of the common domain semantic content in an unsupervised way, while jointly learning the domain-to-domain image translations in both directions. We exploit ideas from the domain adaptation literature and define a *semantic consistency loss* which encourages

https://arxiv.org/pdf/1711.05139.pdf

Image to Image translation (I2I)

“task of mapping images from a visual domain to another” Royer et al. (2018)





Source (human) - Target (cartoon)

Other applications

- Landscape pictures → Paintings
- Sketches → Images
- Text → Images

Contents

→ Proposal

→ **Motivation**

→ Milestones

→ Project Plan

→ Datasets

→ Results

→ Conclusions

Motivation

- Wide range of applications
- GANs are interesting for of us (prof. development)
- Easy-to-communicate and tangible results

Contents

→ Proposal

→ Motivation

→ **Milestones**

→ Project Plan

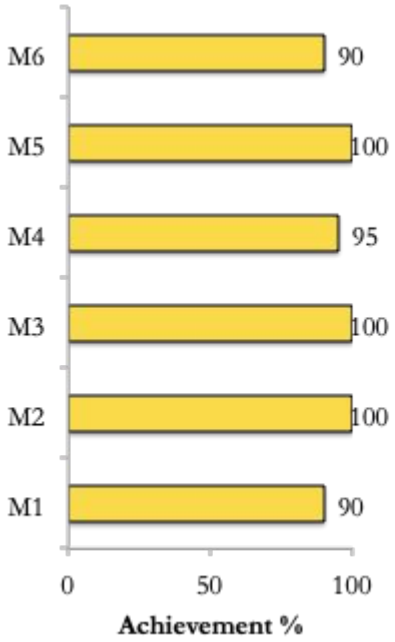
→ Datasets

→ Results

→ Conclusions

Milestones

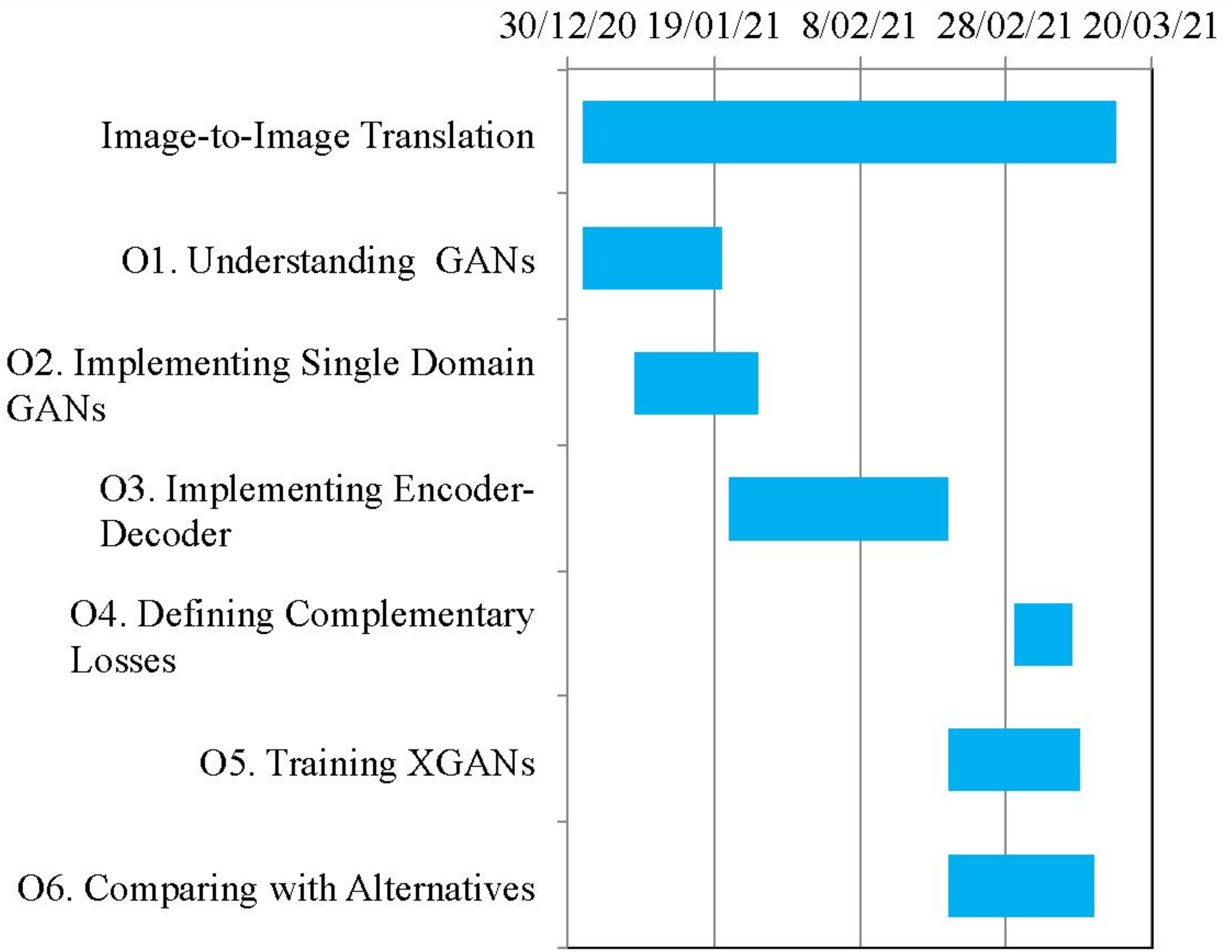
1. Understanding of GANs
2. Single-domain GANs- cartoon faces
3. Encoder-decoder- cross-domain generation
4. Definition of complementary losses for quality
5. Training XGAN
6. Comparison with alternatives



Contents

- → Proposal
- → Motivation
- → Milestones
- → **Project Plan**
- → Datasets
- → Results
- → Conclusions

Project Plan (Gann Chart)



Technology stack

- ➢ TensorFlow (latest, 2.4) with Keras
- ➢ At the beginning Google Colab, later plain Python
- ➢ Google Cloud VM

Contents

→ Proposal

→ Motivation

→ Milestones

→ Project Plan

→ **Datasets**

→ Results

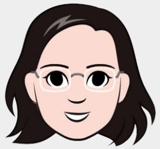
→ Conclusions

Dataset(s)

Dataset(s)

Same semantic

Cartoon faces



Human faces

Cartoon dataset

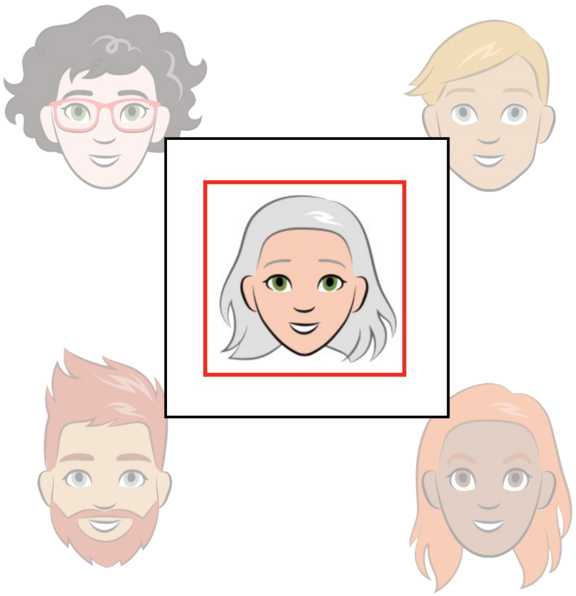
Each cartoon image is annotated with 16 different components indicating face components

Two datasets available:

- ➔ **10k** Small set
- ➔ **100k** Big set

Cartoon dataset

Image cropping to remove blank space



CelebA dataset

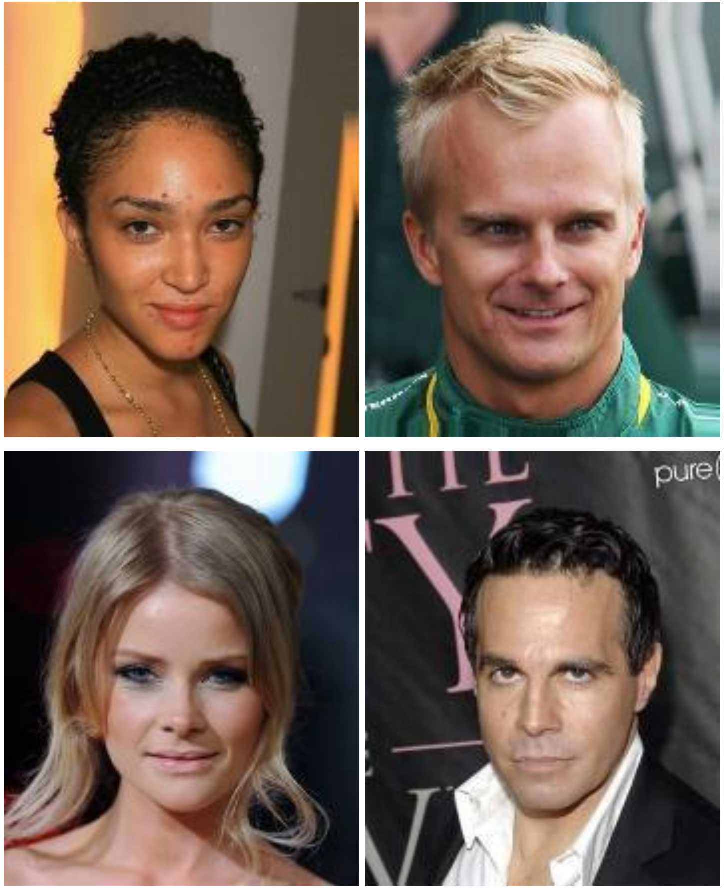
Face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations

CelebA dataset

All images are centered and cropped

CelebA dataset

Not background cleaned



CelebA dataset

We cleaned around 1k images




Contents

→ Proposal

→ Motivation

→ Milestones

→ Project Plan

→ Datasets

→ **Results**

→ Conclusions

Results

- ➔ **Single Domain GANs (Cartoon Domain)**
- ➔ Multiple Domain GANs (Human to Cartoon)
- ➔ Comparison with StarGAN-v2

Results: Single Domain GANs

We know how to generate numbers, now what..?

Results: Single Domain GANs

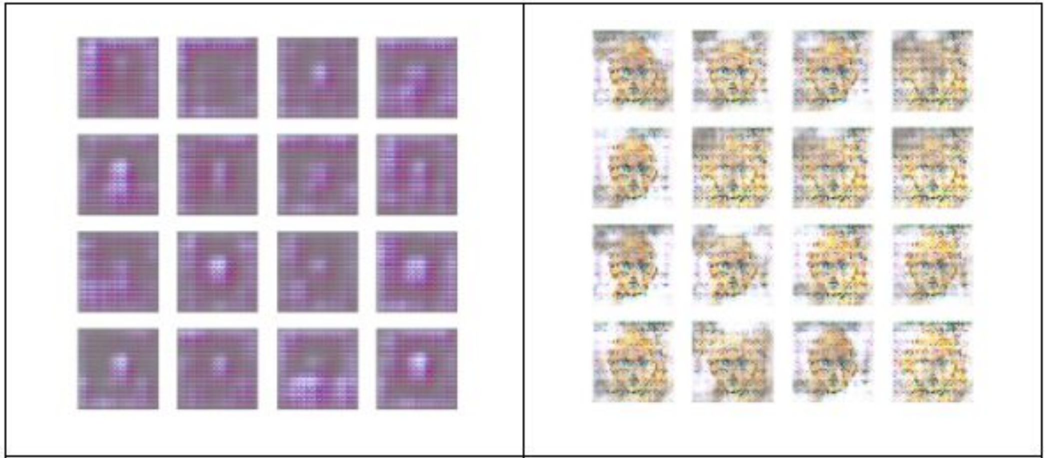
Goal: Generate Cartoon faces from a latent vector

Results: Single Domain GANs

- MNIST Lab **architecture** + extra conv layer + use of **RGB** channels
- **Binary Cross Entropy** for both generator and discriminator
- **Adam** as optimizer
- **Noise vector** of shape 100

Results: Single Domain GANs

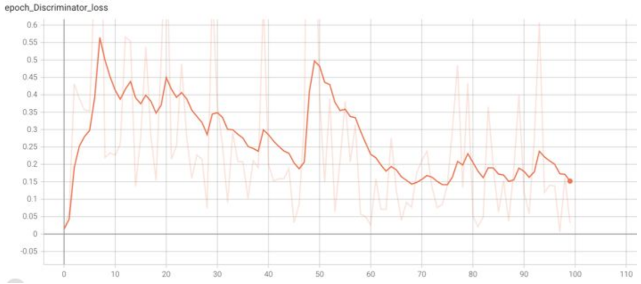
- Different learning rate values
- **Normalization** of data from -1 to 1



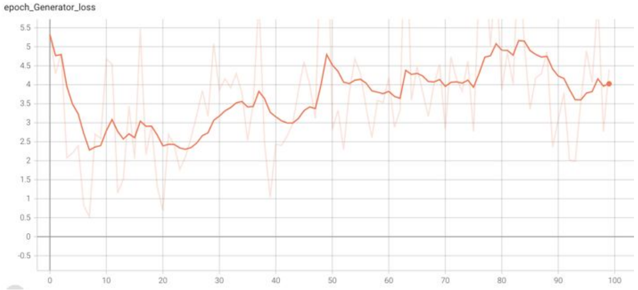
epoch 0 | epoch 10

Results: Single Domain GANs

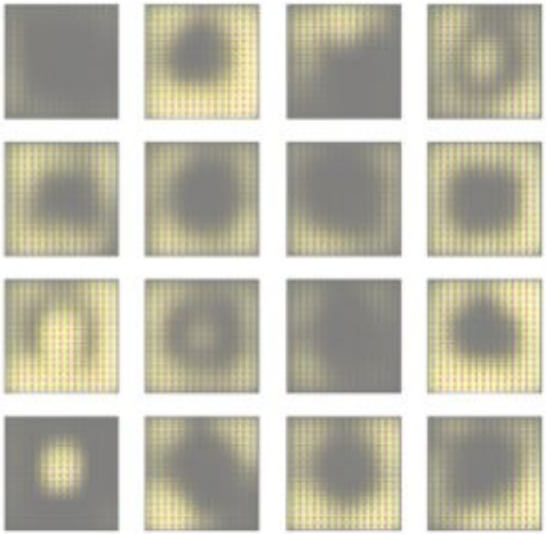
Discriminator Loss



Generator Loss



Epoch 0	**Epoch 100**



Results

- ➔ Single Domain GANs (Cartoon Domain)

- ➔ **Multiple Domain GANs (Human to Cartoon)**

- ➔ Comparison with StarGAN-v2

XGAN

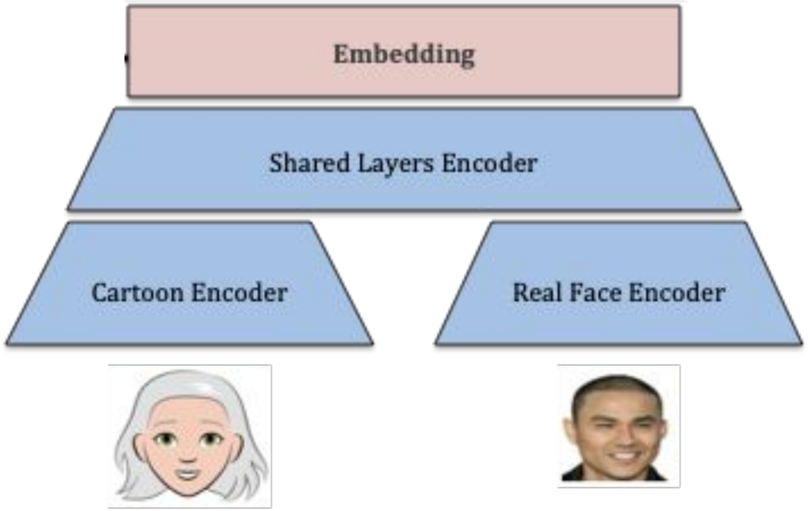
- Extension of GAN. Composed by 3 models:
  - Encoder - Decoder as a generator
  - Discriminator
  - Binary classifier

XGAN

➢ Encoder

Layer	Size
Inputs	64x64x3
conv1	32x32x32
conv2	16x16x64
(//) conv3	8x8x128
(//) conv4	4x4x256
(//) FC1	1x1x1024
(//) FC2	1x1x1024

(a) Encoder

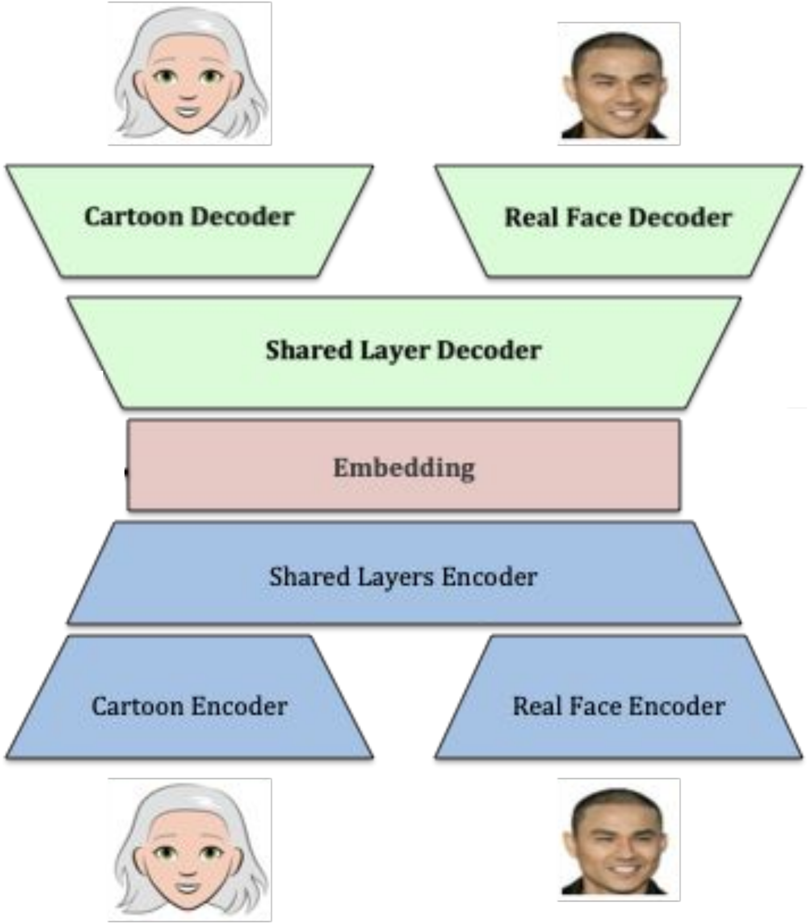


XGAN

➢ Encoder-Decoder

Layer	Size
Inputs	1x1x1024
(//) deconv1	4x4x512
(//) deconv2	8x8x256
deconv3	16x16x128
deconv4	32x32x64
deconv5	64x64x3

(b) Decoder

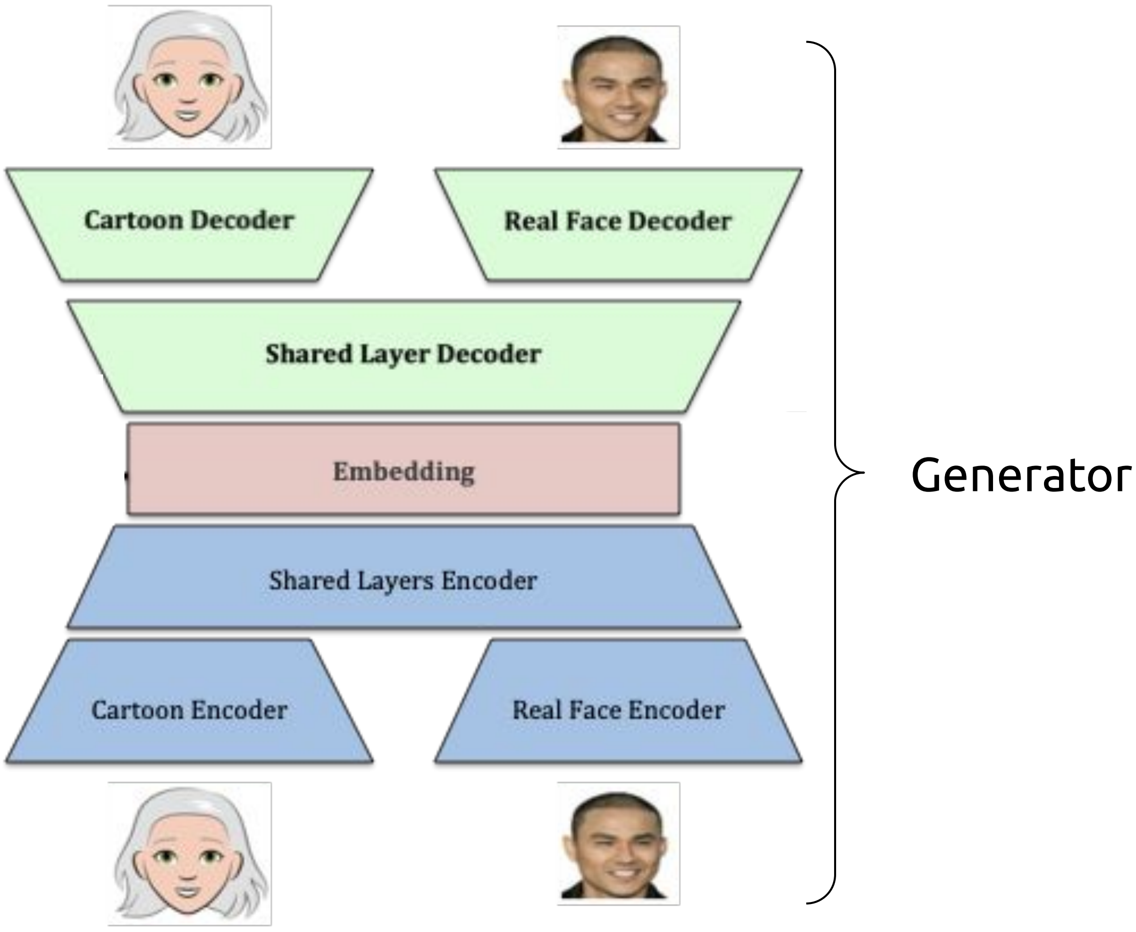


XGAN

➢ Encoder-Decoder

Layer	Size
Inputs	1x1x1024
(//) deconv1	4x4x512
(//) deconv2	8x8x256
deconv3	16x16x128
deconv4	32x32x64
deconv5	64x64x3

(b) Decoder

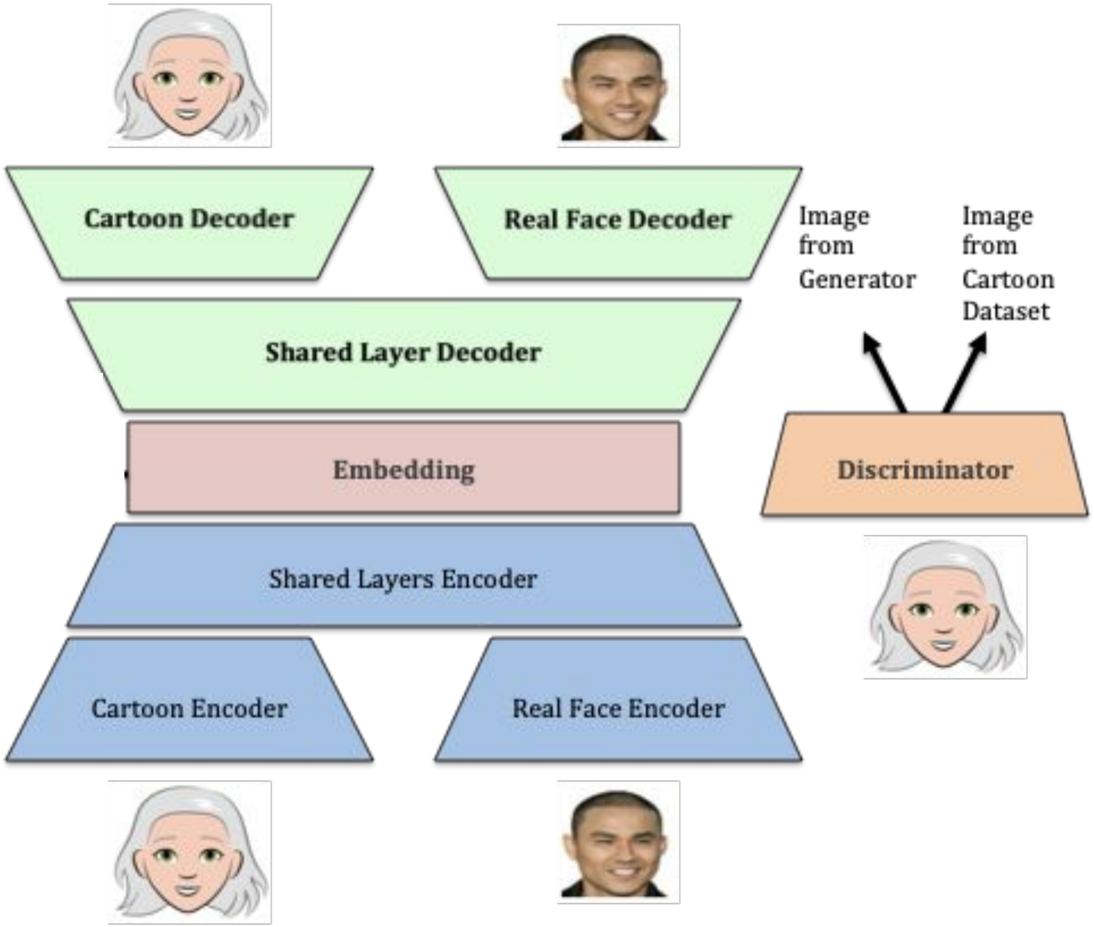


XGAN

➢ Discriminator

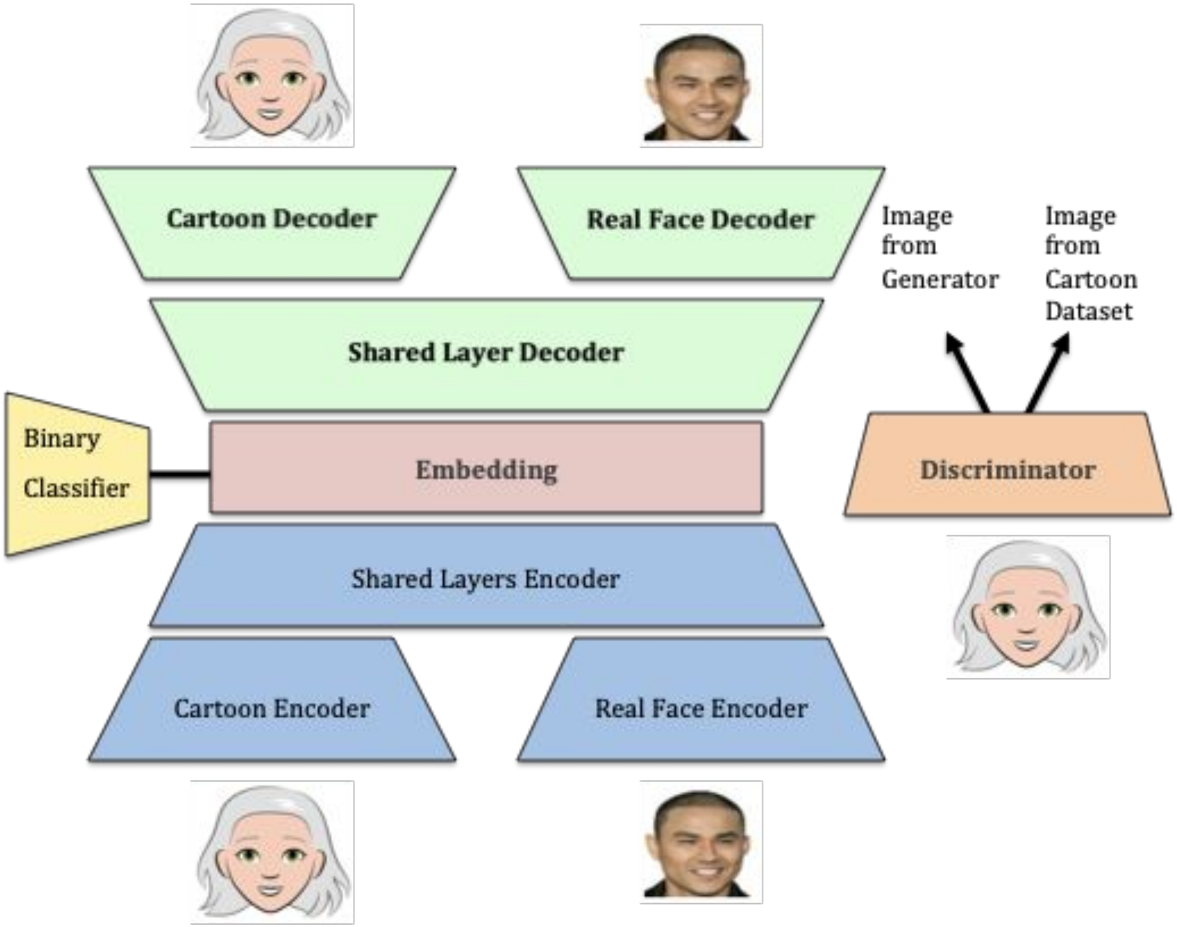
Layer	Size
Inputs	64x64x3
conv1	32x32x16
conv2	16x16x32
conv3	8x8x32
conv4	4x4x32
FC1	1x1x1

(c) Discriminator



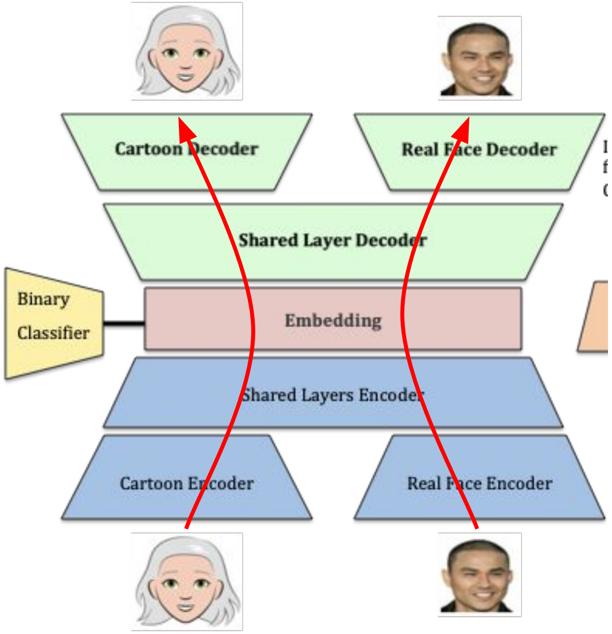
XGAN

- ➢ Binary classifier

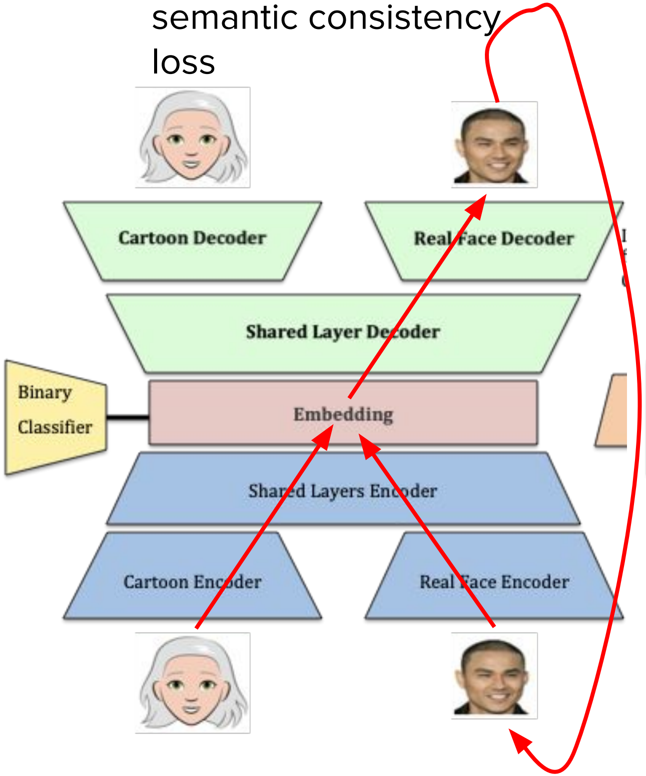


XGAN - Generator losses

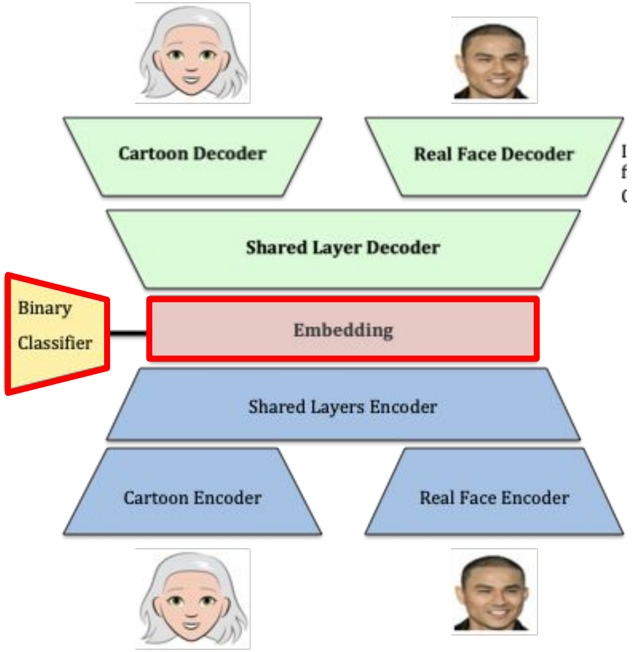
encoder-decoder loss



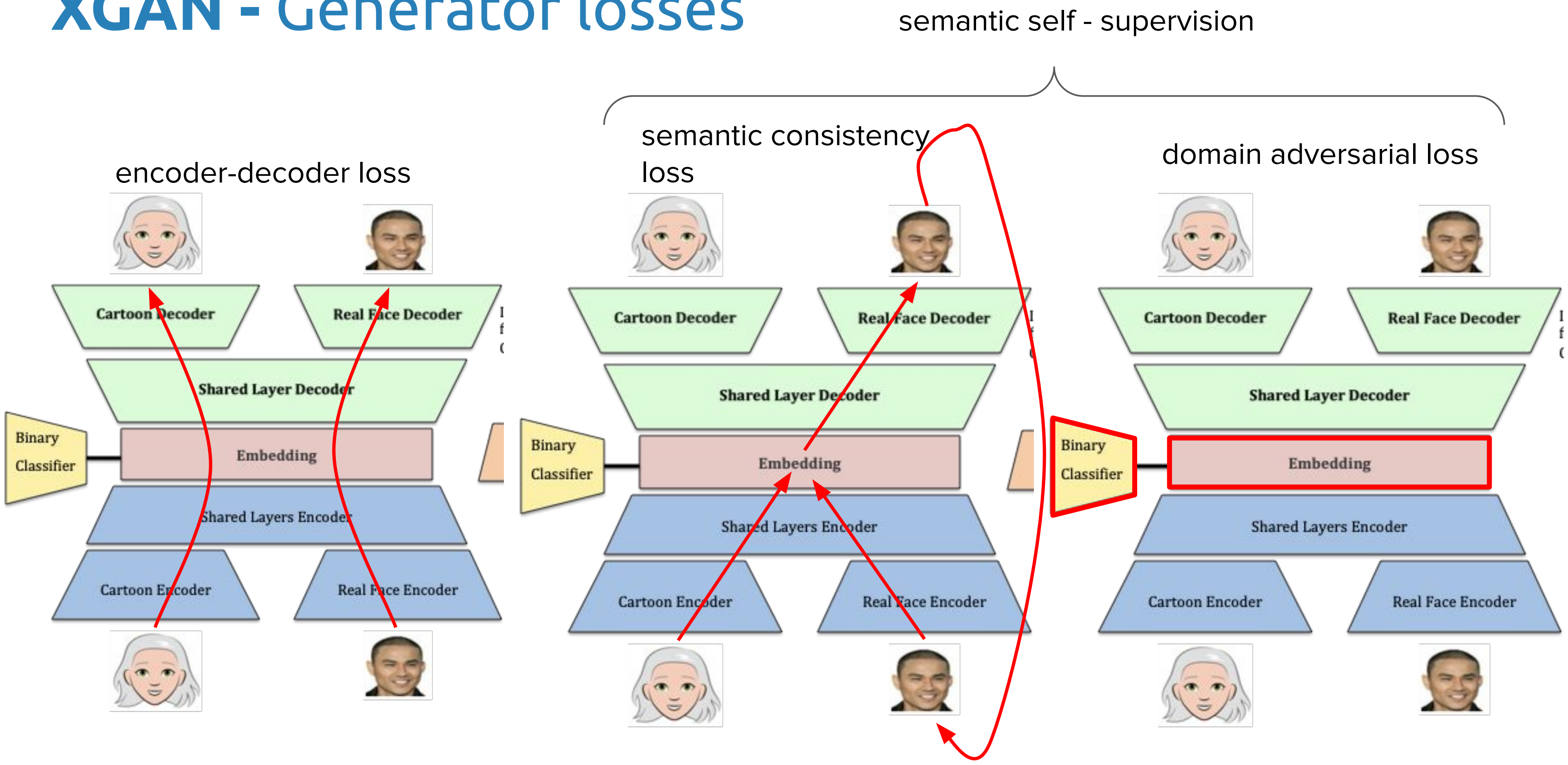
semantic consistency loss



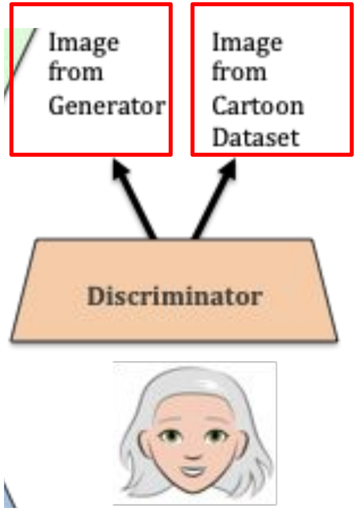
domain adversarial loss

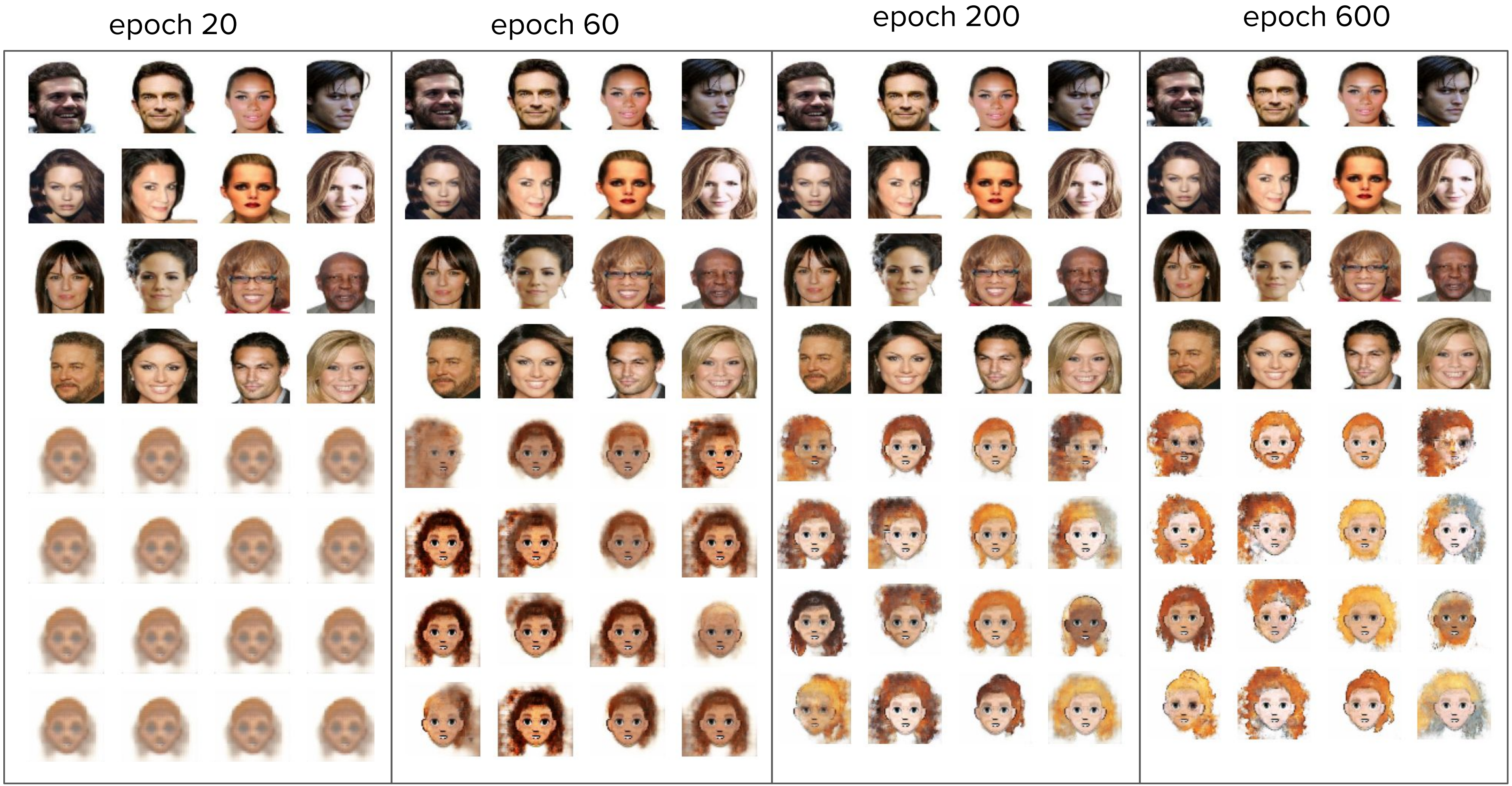


XGAN - Generator losses

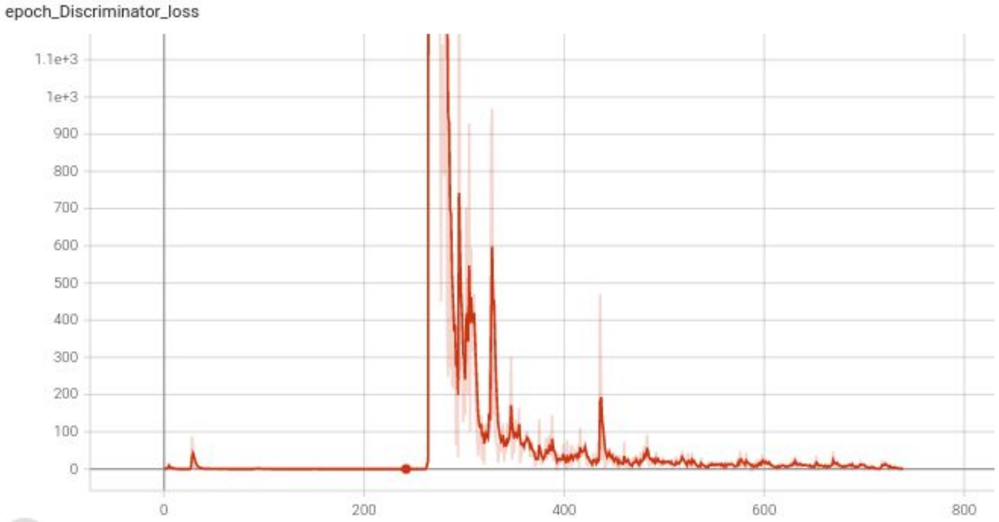


XGAN - Discriminator loss

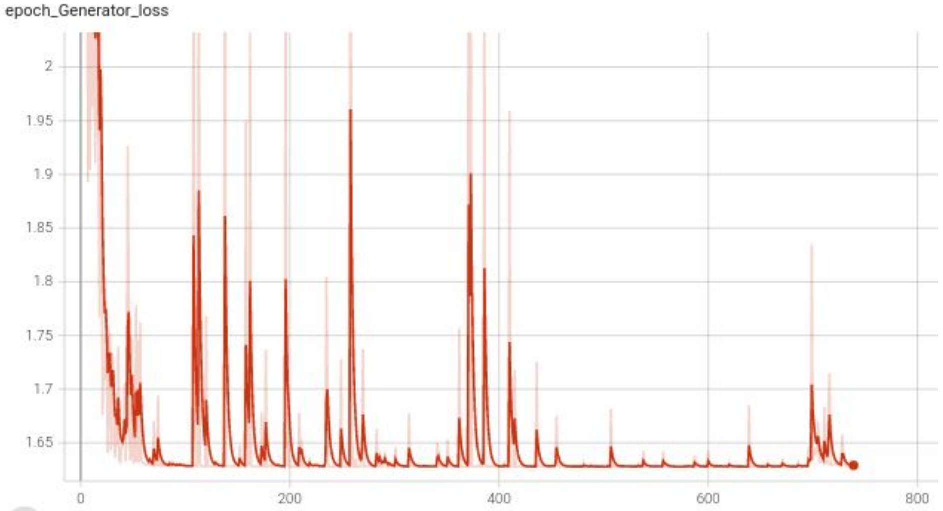



epoch 20
epoch 60
epoch 200
epoch 600

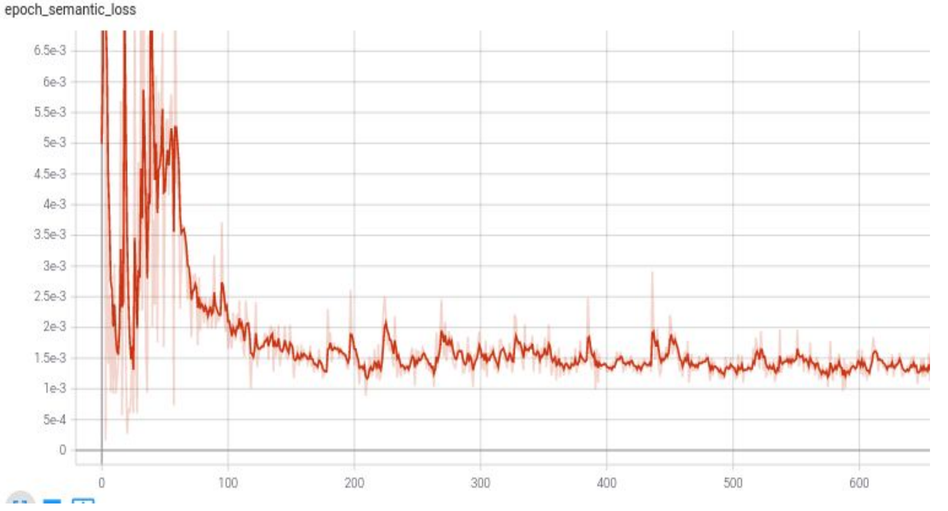
XGAN



epoch_Discriminator_loss
1.1e+3
1e+3
900
800
700
600
500
400
300
200
100
0
0
200
400
600
800



epoch_Generator_loss
2
1.95
1.9
1.85
1.8
1.75
1.7
1.65
0
200
400
600
800



epoch_semantic_loss
6.5e-3
6e-3
5.5e-3
5e-3
4.5e-3
4e-3
3.5e-3
3e-3
2.5e-3
2e-3
1.5e-3
1e-3
5e-4
0
0
100
200
300
400
500
600

Results

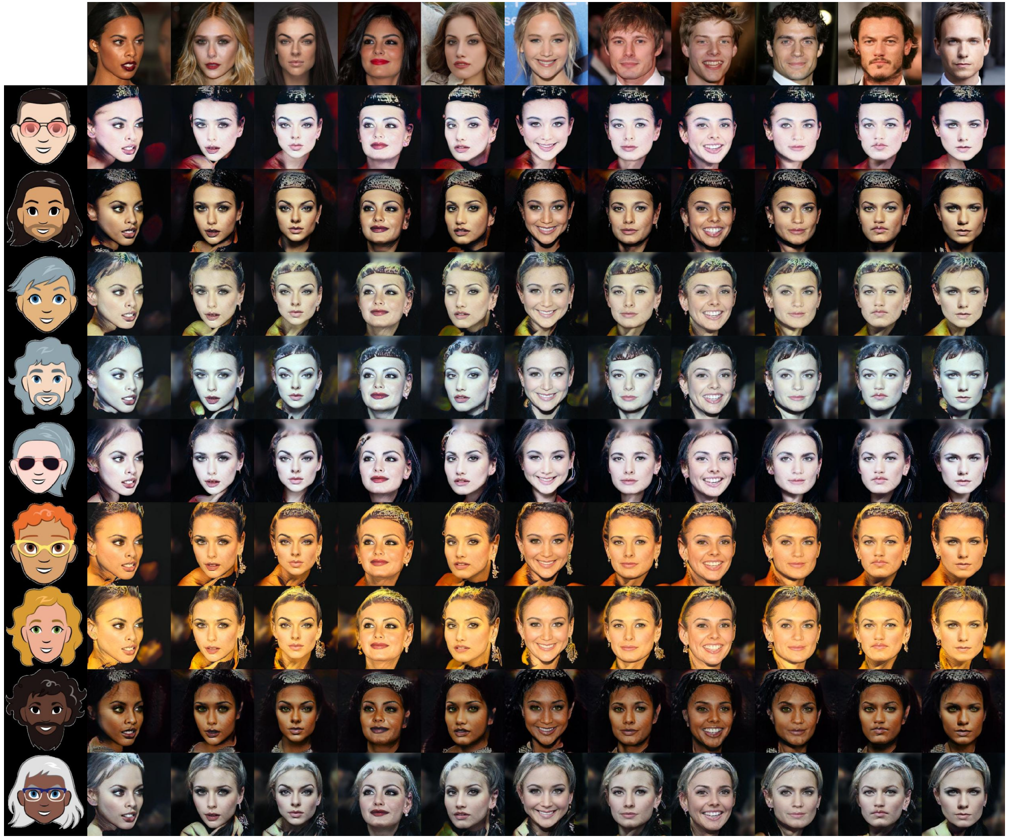
- → Single Domain GANs (Cartoon Domain)
- → Multiple Domain GANs (Human to Cartoon)
- → **Comparison with StarGAN-v2**

Results: Comparison with StarGAN-V2

Source

StarGAN-v2

Generation of multiple outputs, for a given "source", that reflect different "styles" in a "reference" domain



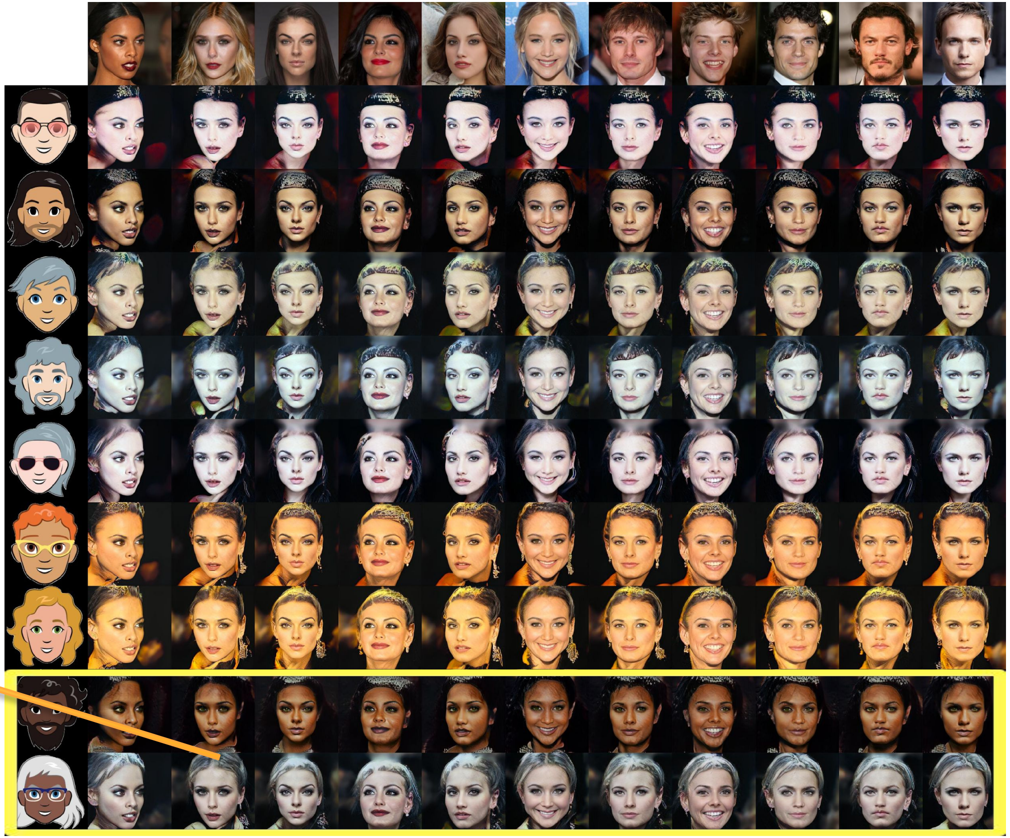
Reference

Results: Comparison with StarGAN-V2

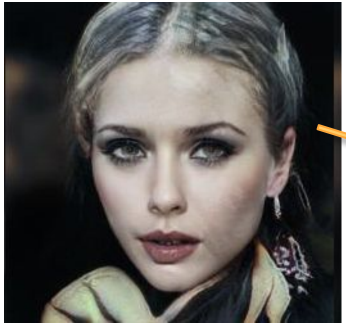
Source

StarGAN-v2

Cartoon Faces as a useful device to explore the inside machinery of I2I models.



Reference



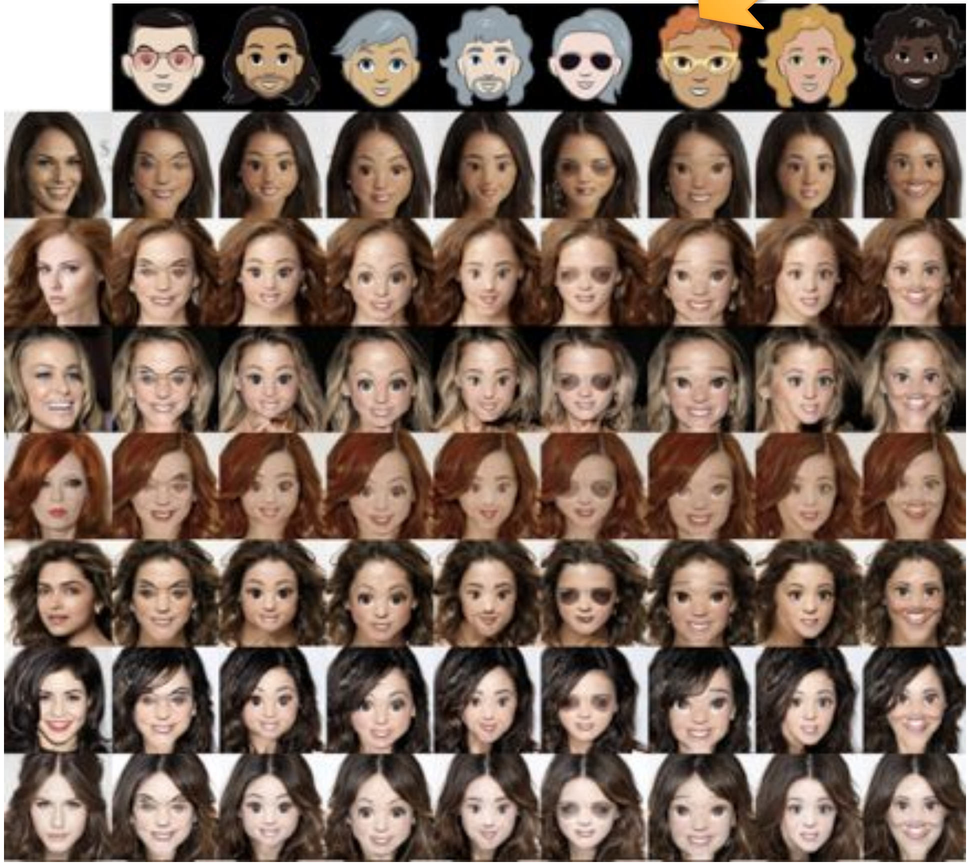
Results: Comparison with StarGAN-V2

Source

StarGAN-v2

Changing the source this time.

Reference



Contents

→ Proposal

→ Motivation

→ Milestones

→ Project Plan

→ Datasets

→ Results

→ **Conclusions**

Conclusions

✔ Reflections on Data-Preprocessing

✔ From single-domain to multiple-domain GANs

✔ Cartoon-Faces as useful devices to understand what is occurring in I2I tasks