Generation of Deepfakes using Normalizing Flows

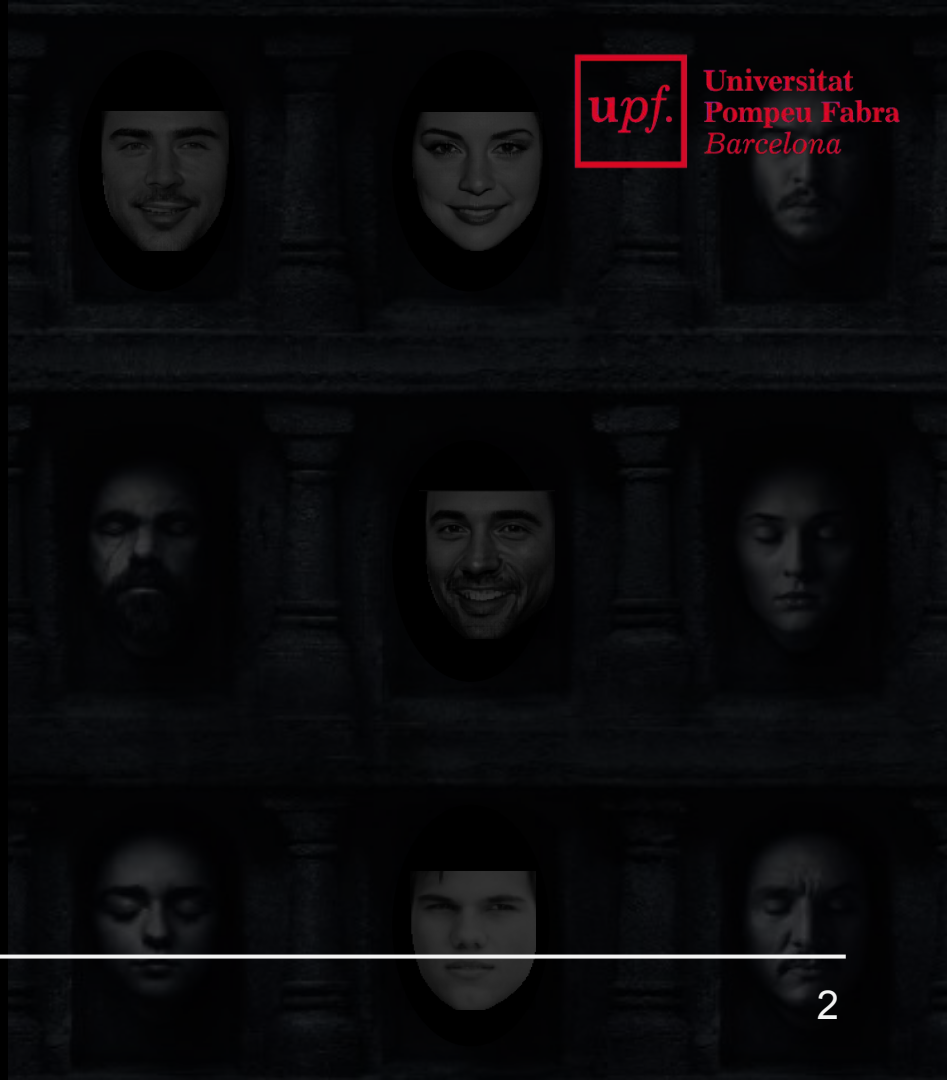
Andrea Valenzuela Ramírez

upf. Universitat Pompeu Fabra Barcelona

Telefonica

MIIS Master 2019-2020
2nd September 2020

INDEX



Universitat Pompeu Fabra Barcelona

INTRODUCTION

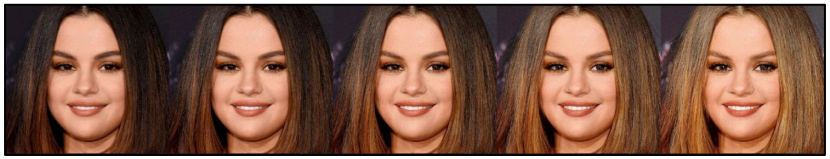
Deepfake content

Types of deepfake content:

- Face synthesis.

- Face swap.

- Facial attribute manipulation.



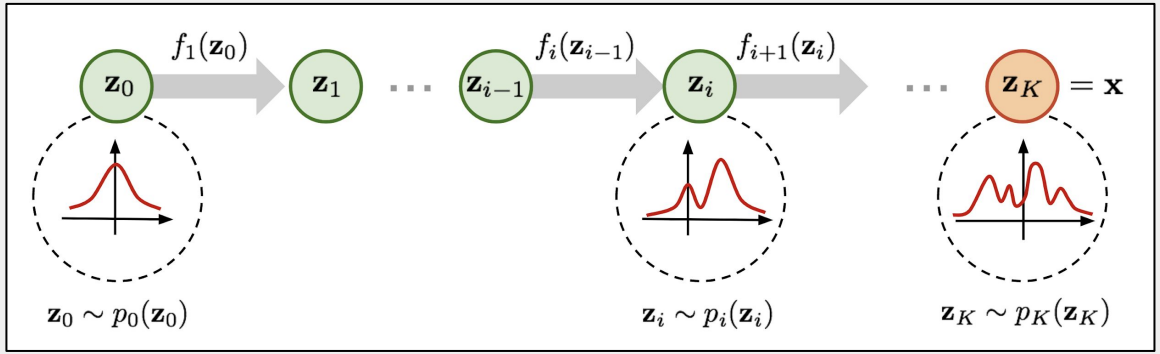
INTRODUCTION

Normalizing Flows and the Glow model

- The latent representation, **z**, of a given data point, **x**, can be expressed as:

$$\mathbf{z} = g(\mathbf{x}) = f^{-1}(\mathbf{x}).$$

- If instead of a simple function f_i one considers a composition of simple invertible transformations, one can initially model increasingly complex and even multi-modal distributions to generate **x**.



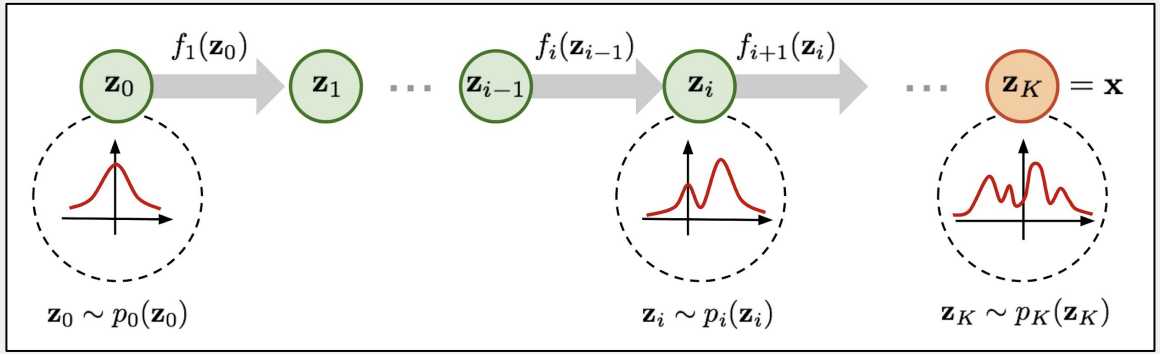
Bijective

INTRODUCTION

Normalizing Flows and the Glow model

- The (log-)probability density function of a given data point **x** can be expressed as:

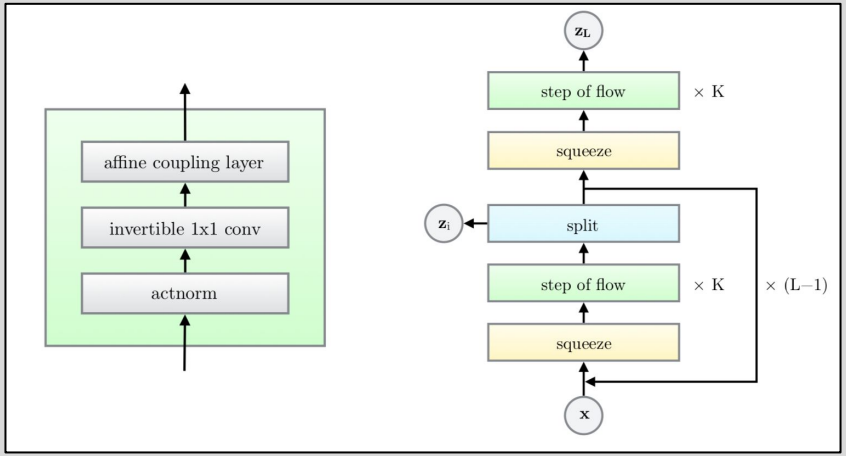
$$\log p(\mathbf{x}) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \log \left| \det \left(\frac{\partial f_k}{\partial \mathbf{z}_{k-1}} \right) \right|.$$



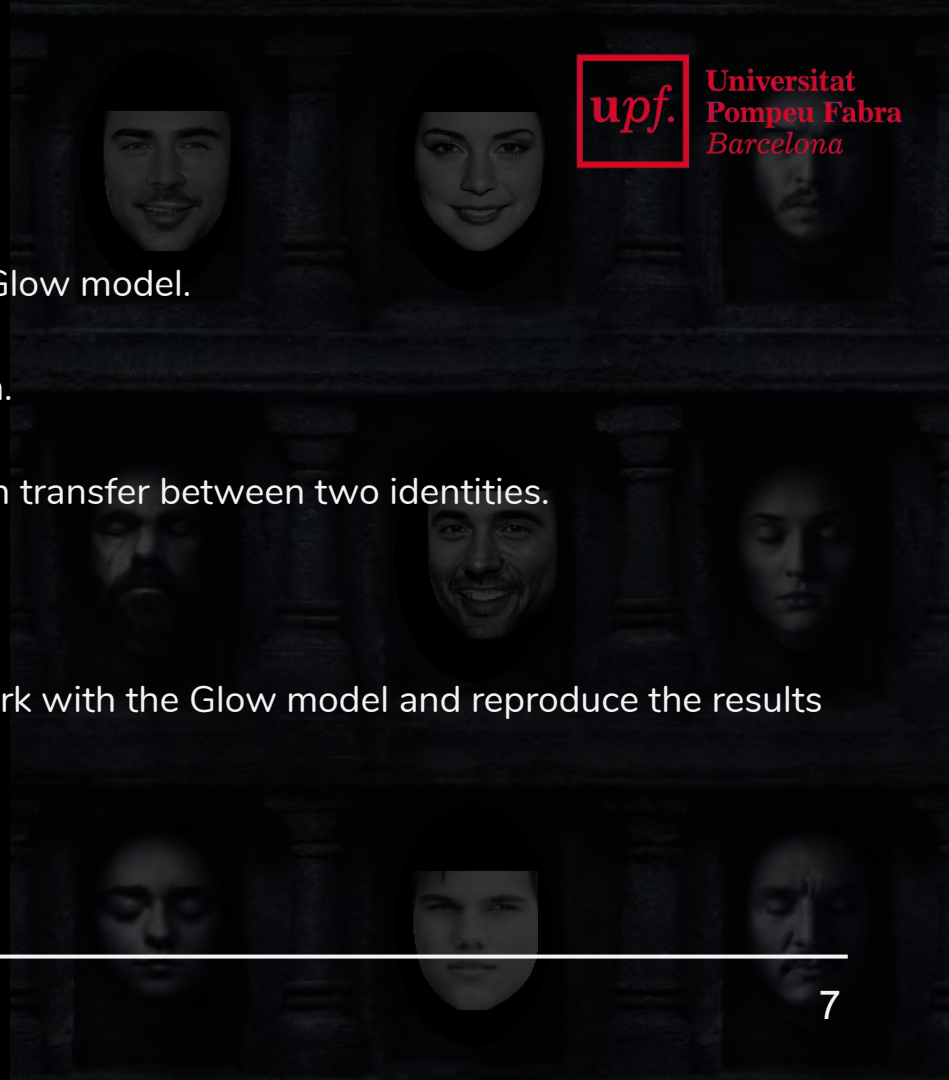
INTRODUCTION

Normalizing Flows and the Glow model

- *Glow* is a type of *flow-based generative model*.
- It is an extension of *NICE* and *RealNVP*.
- *Glow* improved *RealNVP* by adding a reversible 1x1 convolution and also by simplifying the overall architecture.
- The transformations used in the *Glow* model are:
  affine coupling layer, 1x1 convolution and actnorm layer.
- The three transformations are combined with a multiscale architecture.



OBJECTIVES



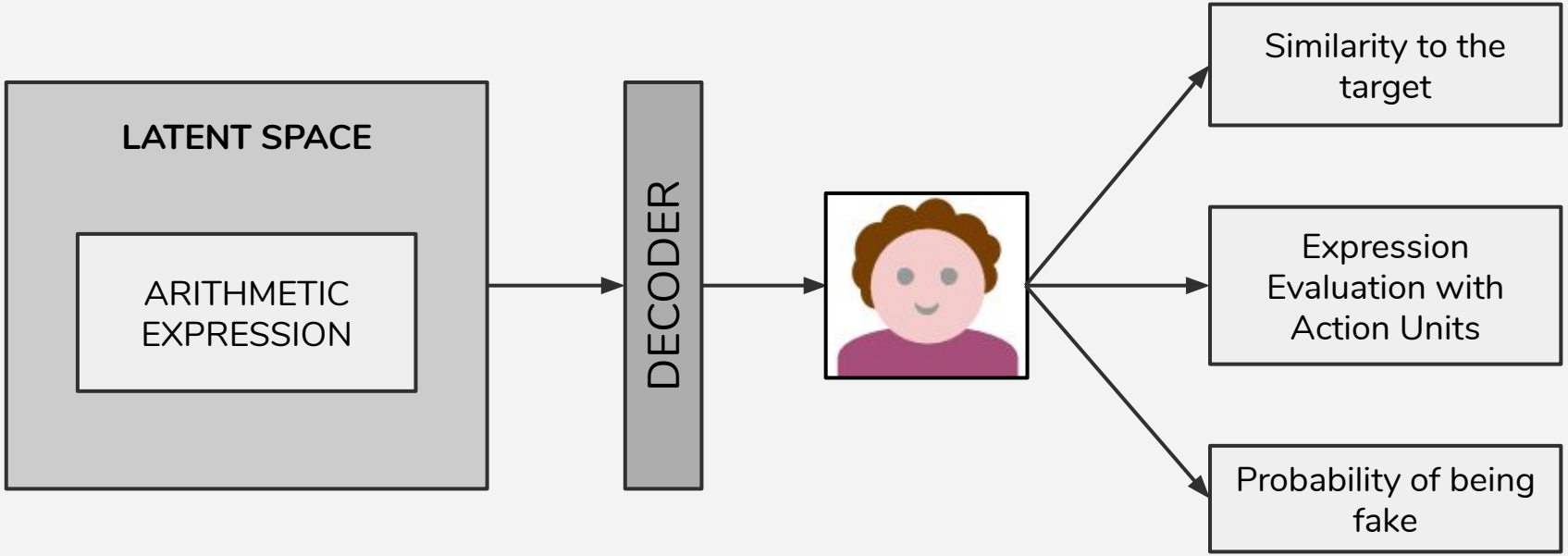
- Reproduce the state-of-the-art results of the Glow model.
- An extended analysis of attribute manipulation.
- Propose a new vector arithmetic for expression transfer between two identities.
  - Define a proper expression.
  - Evaluation of the obtained fake images.
- Release a public open-source repository to work with the Glow model and reproduce the results of this thesis.

METHODOLOGY

- The pre-trained model provided by OpenAI has a latent space of dimension *196,608* and *6* layers.
- The model is pre-trained in the CelebA dataset.
- Five different popular identities have been selected to work with during the project.
- Implementations in python based on the demonstration provided by OpenAI.

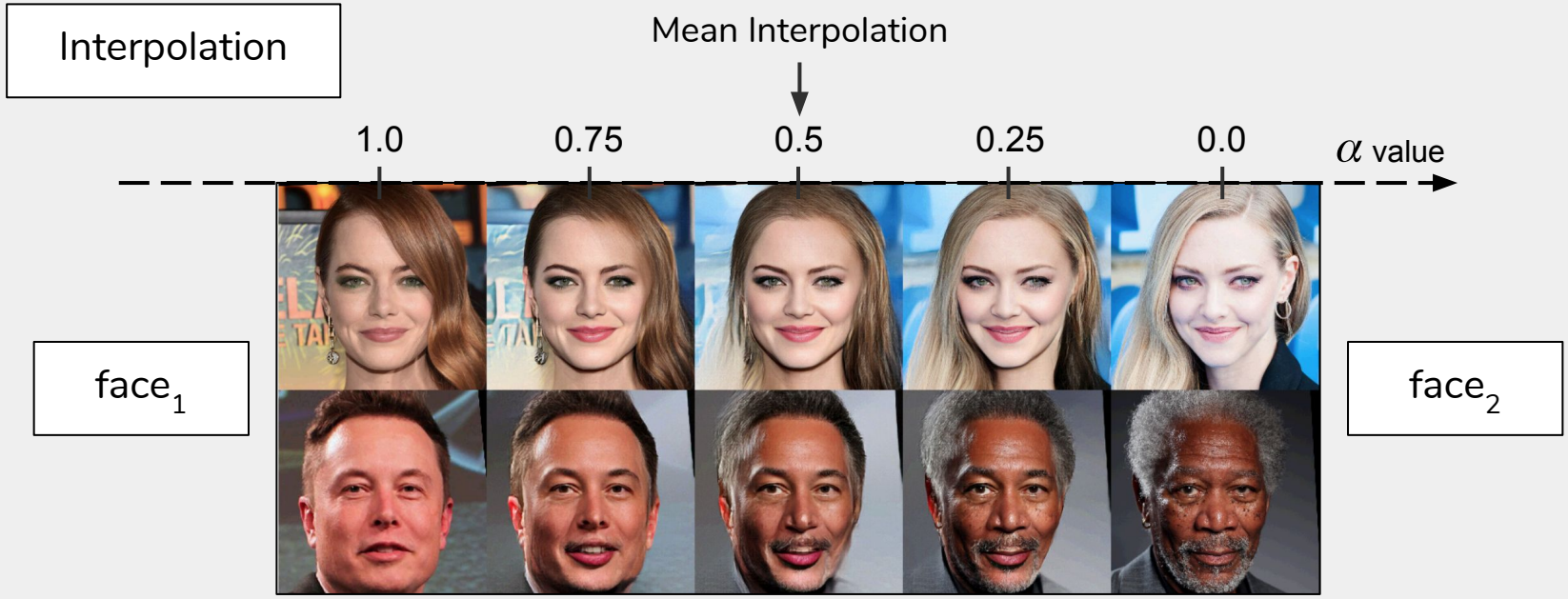
METHODOLOGY

- Generation and evaluation pipeline.



STATE-OF-THE-ART RESULTS

Universitat Pompeu Fabra Barcelona



$$\mathbf{z}_{\text{interpolated}} = \alpha \cdot \mathbf{z}_{\text{face1}} + (1 - \alpha) \cdot \mathbf{z}_{\text{face2}},$$

STATE-OF-THE-ART RESULTS

Universitat Pompeu Fabra Barcelona

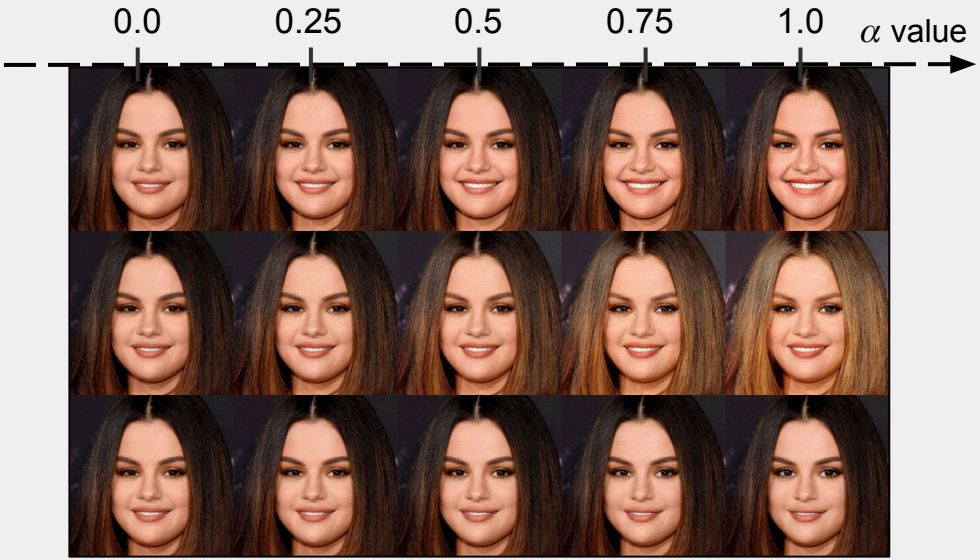
Attribute manipulation

Mustache

Smile

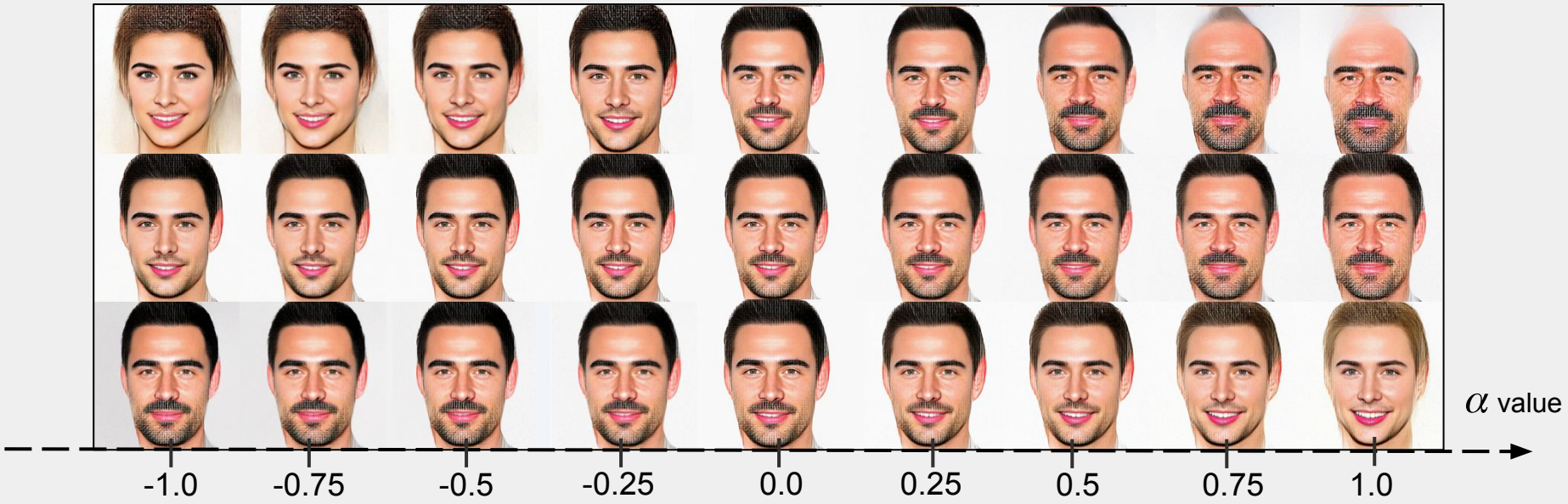
$$\mathbf{z}_{\text{manipulation vector}} = \mathbf{z}_{\text{positive}} - \mathbf{z}_{\text{negative}}$$

- Face attributes can be modified in the latent space.



$$\mathbf{z}_{\text{manipulated}} = \mathbf{z}_{\text{input}} + \alpha \cdot \mathbf{z}_{\text{manipulation vector}}$$

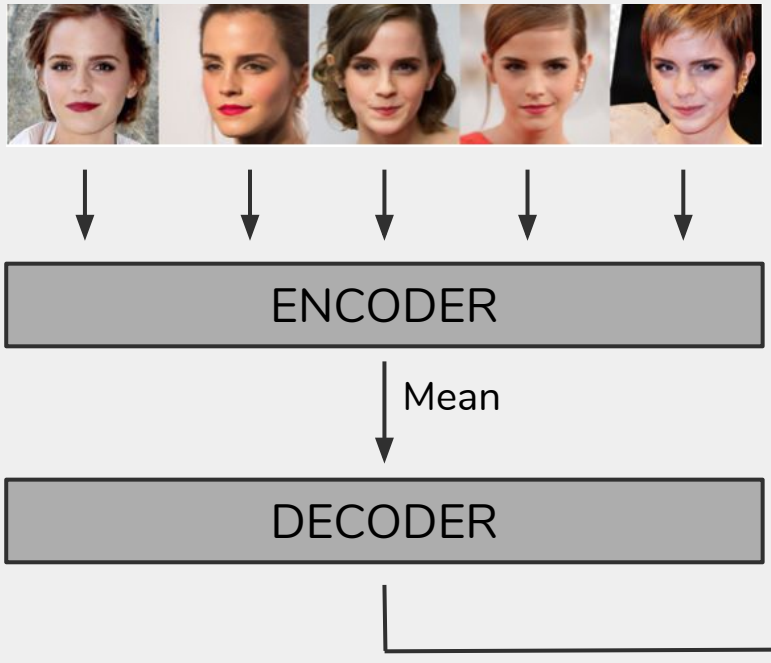
ATTRIBUTE MANIPULATION



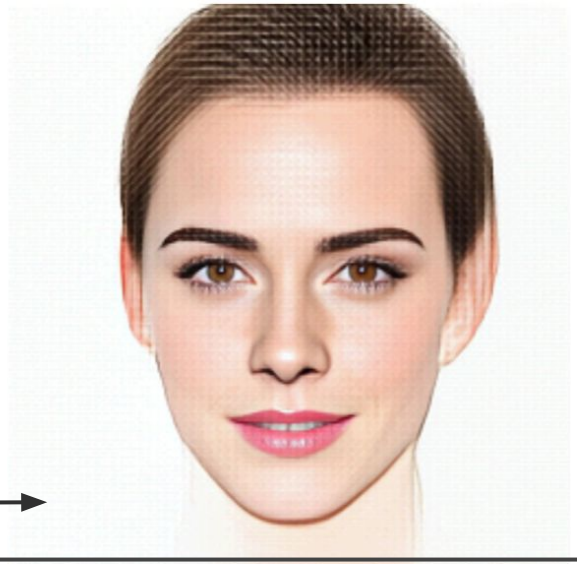
- Attribute manipulation influences other aspects of the original face apart from the selected attribute.

PRE-EXPRESSION TRANSFER

Mean face of an identity



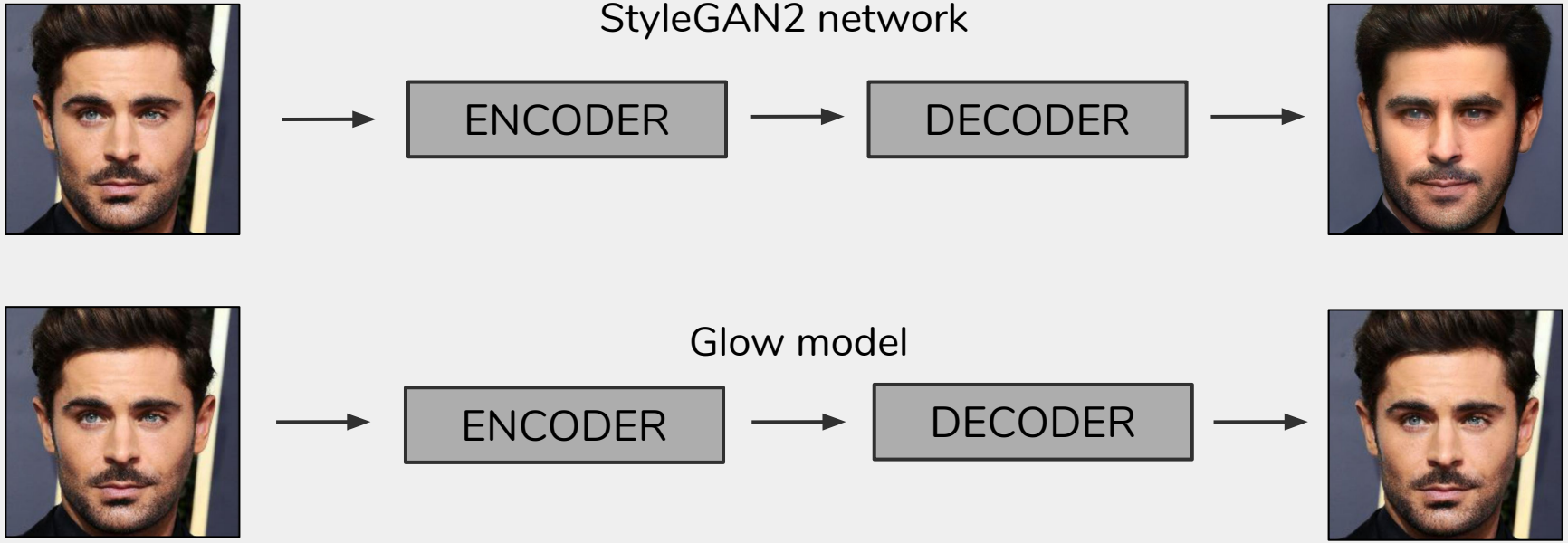
- A neutral face can be obtained by averaging several original images in the latent space.



PRE-EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Reconstruction of original images



PRE-EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Reconstruction of original images

- The Glow model performs a perfect reconstruction of original images that are not present on the training dataset.

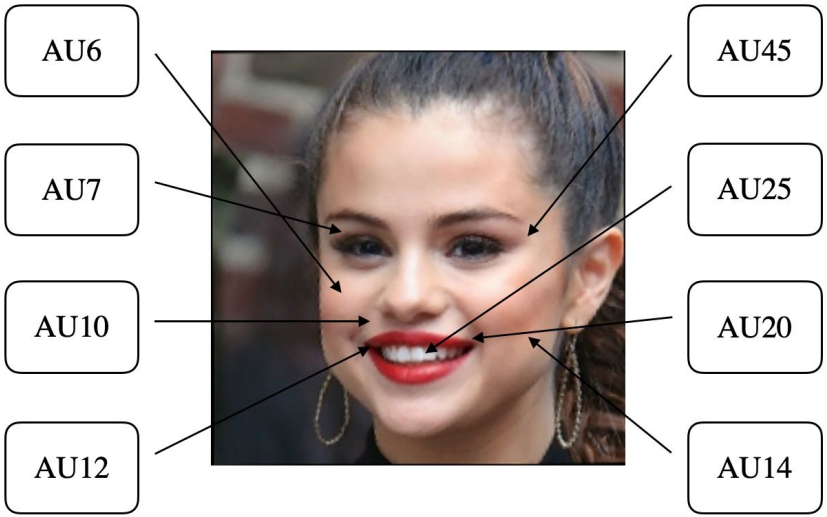
Glow model

ENCODER → DECODER

⚠ Check the reconstruction error in the Colab Notebook.

EXPRESSION TRANSFER

Source Expression



AU6: Check raiser.
AU7: Lip tightener.
AU10: Upper lip raiser.
AU12: Lip corner puller.
AU14: Dimpler.
AU20: Lip corner depressor.
AU25: Lipd part.
AU45: Blink.

- We are going to analyse the transfer of this original expression on the best cases of linear combination and variance.

EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

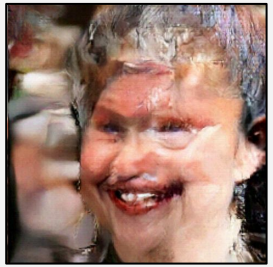
Best results on linear combination

$$\mathbf{z}_{\mathrm{OT}} = \alpha \cdot \mathbf{z}_{\mathrm{expression}} + \mathbf{z}_{\mathrm{MT}}$$

OT: Output Target.
MT: Mean of the Target.

OS: Original Source.
OM: Mean of the Source.

$$\mathbf{z}_{\mathrm{expression}} = \mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}$$



$$\mathbf{z}'_{\mathrm{expression}} = \frac{\mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}}{2}$$

EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Best results on linear combination

$$\mathbf{z}_{\mathrm{OT}} = \alpha \cdot \mathbf{z}_{\mathrm{expression}} + \mathbf{z}_{\mathrm{MT}}$$

OT: Output Target.
MT: Mean of the Target.

OS: Original Source.
OM: Mean of the Source.

$$\mathbf{z}_{\mathrm{expression}} = \mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}$$

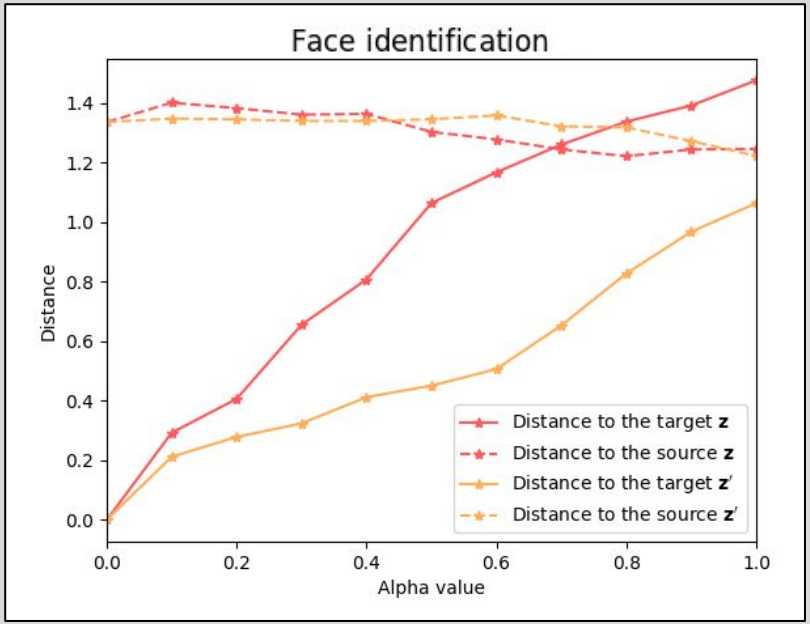
$$\mathbf{z}'_{\mathrm{expression}} = \frac{\mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}}{2}$$

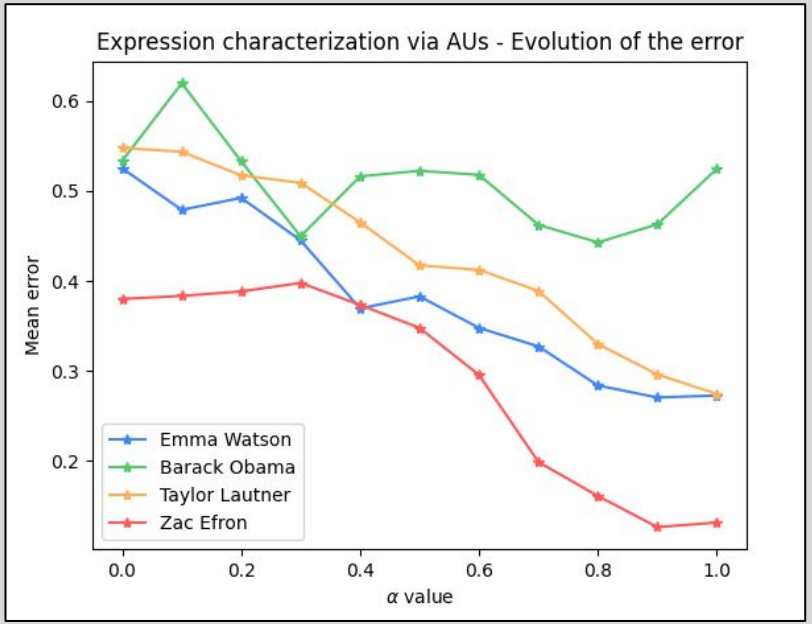
EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Analysis of the likeliness of the generated images to their target

- Which value of α shall we take?



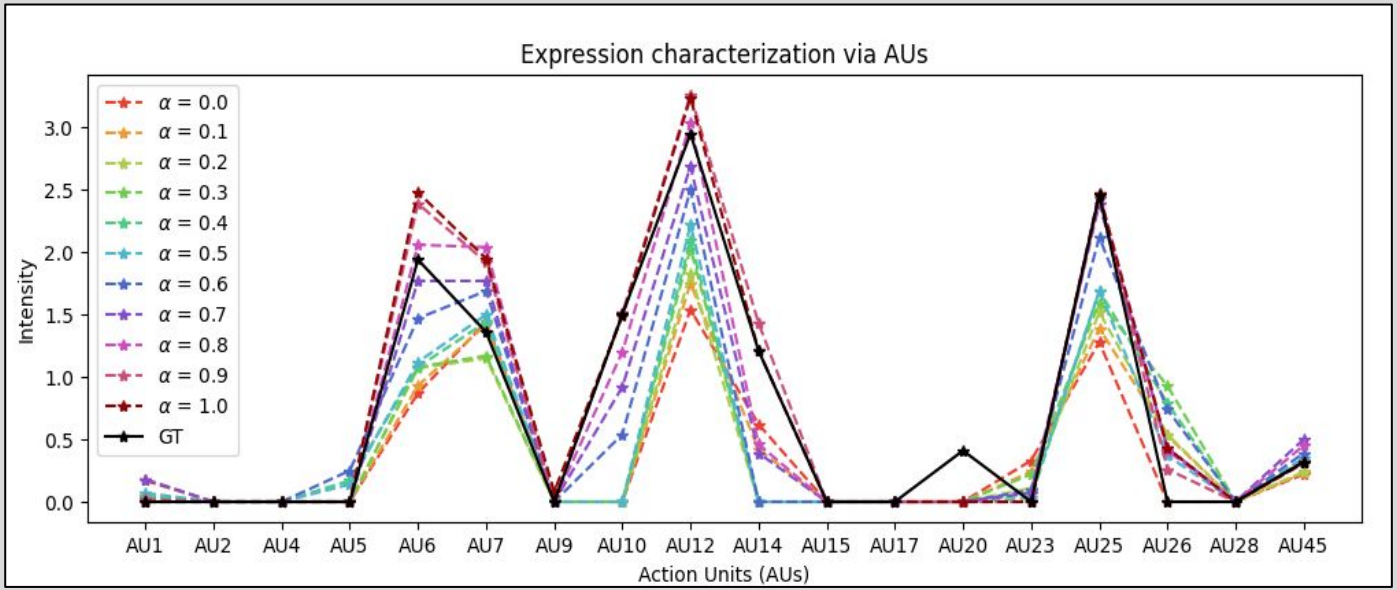


EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Expression analysis with Action Units (AUs)

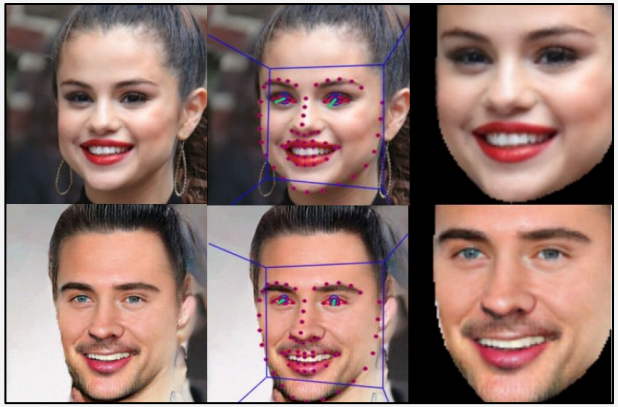
- Analysis of the generated image with smaller mean error.



EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

CONCRETE CASE



Distance Score = 0.97
Mean Error = 0.13
α = 0.9
Not Fake

GENERAL TREND

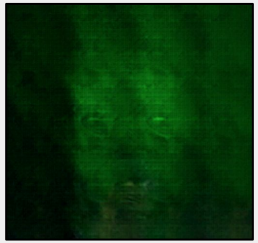
Distance Score = 0.59
Mean Error = 0.30
α^* = 0.57
12.50% of detected fakes

EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Best results on variance

Mean



Variance

$$\mathbf{z}_{\mathrm{OT}} = \frac{\frac{\mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}}{\mathbf{z}_{\mathrm{varS}} + \epsilon} + \mathbf{z}_{\mathrm{MT}}}{\mathbf{z}_{\mathrm{varT}} + \epsilon}$$

OT: Output Target.
MT: Mean of the Target.

OS: Original Source.
OM: Mean of the Source.

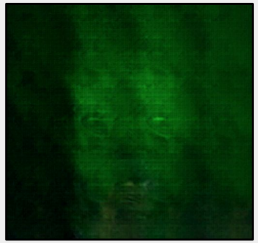
varT: Variance of the Target.
varS: Variance of the Source.

EXPRESSION TRANSFER

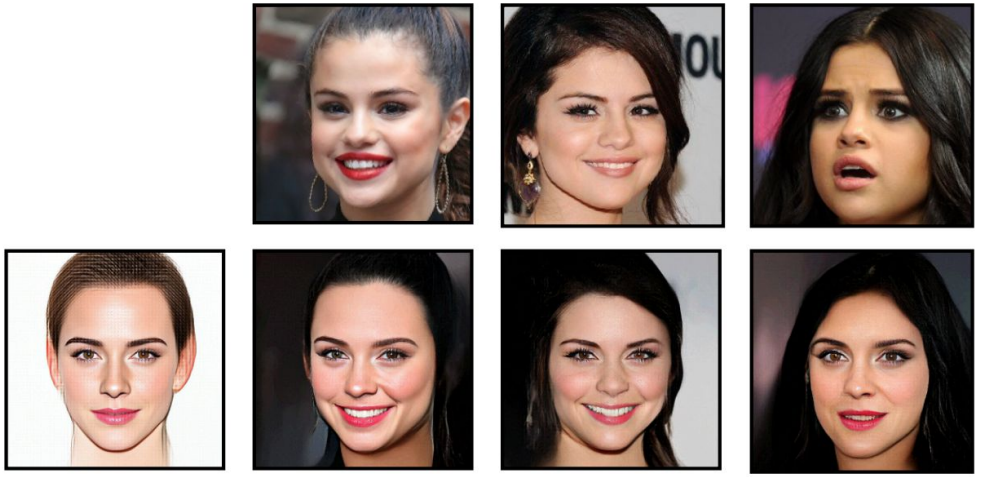
Universitat Pompeu Fabra Barcelona

Best results on variance

Mean



Variance

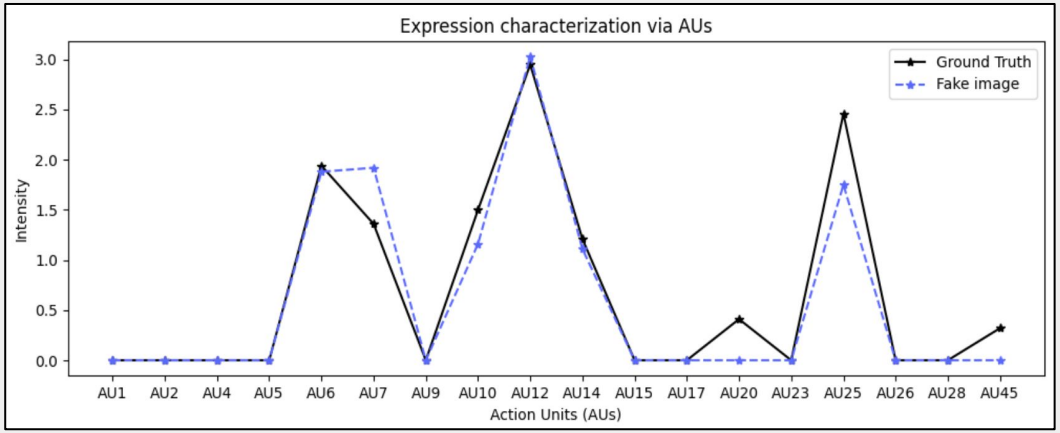


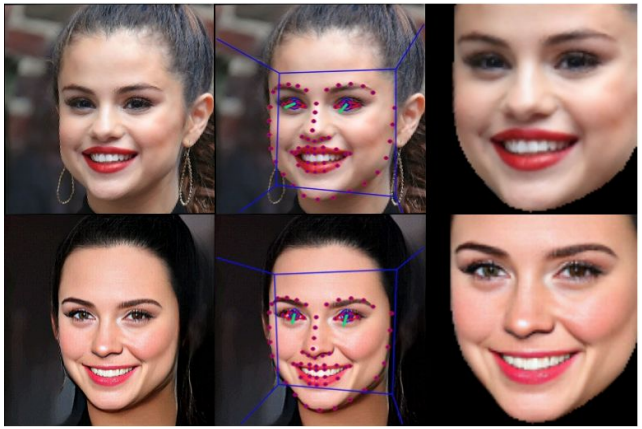
- The results obtained are of higher quality with respect to the linear combination.

EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

Best results on variance





GENERAL TREND

Distance Score = 0.92
Mean Error = 0.28
Any fake detected

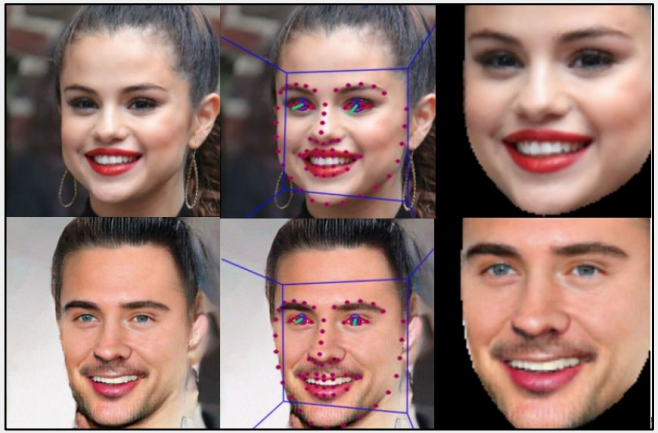
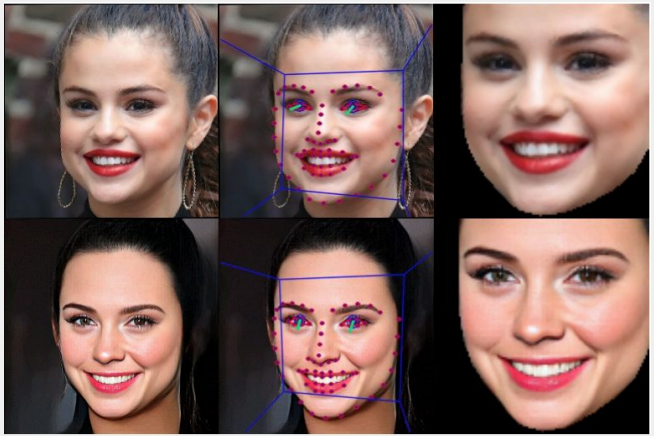
CONCRETE CASE

Distance Score = 0.86
Mean Error = 0.14
Not Fake

EXPRESSION TRANSFER

Universitat Pompeu Fabra Barcelona

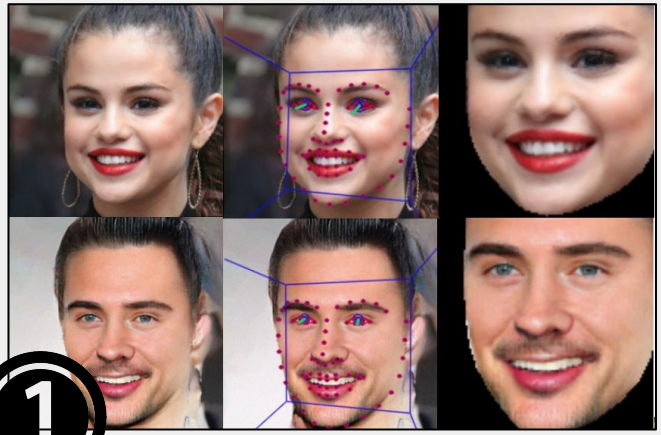
Discussion

Linear Combination	Variance
	
Distance Score = 0.59 Mean Error = 0.30 12.50% of detected fakes	Distance Score = 0.92 Mean Error = 0.28 Any fake detected

EXPRESSION TRANSFER

Discussion

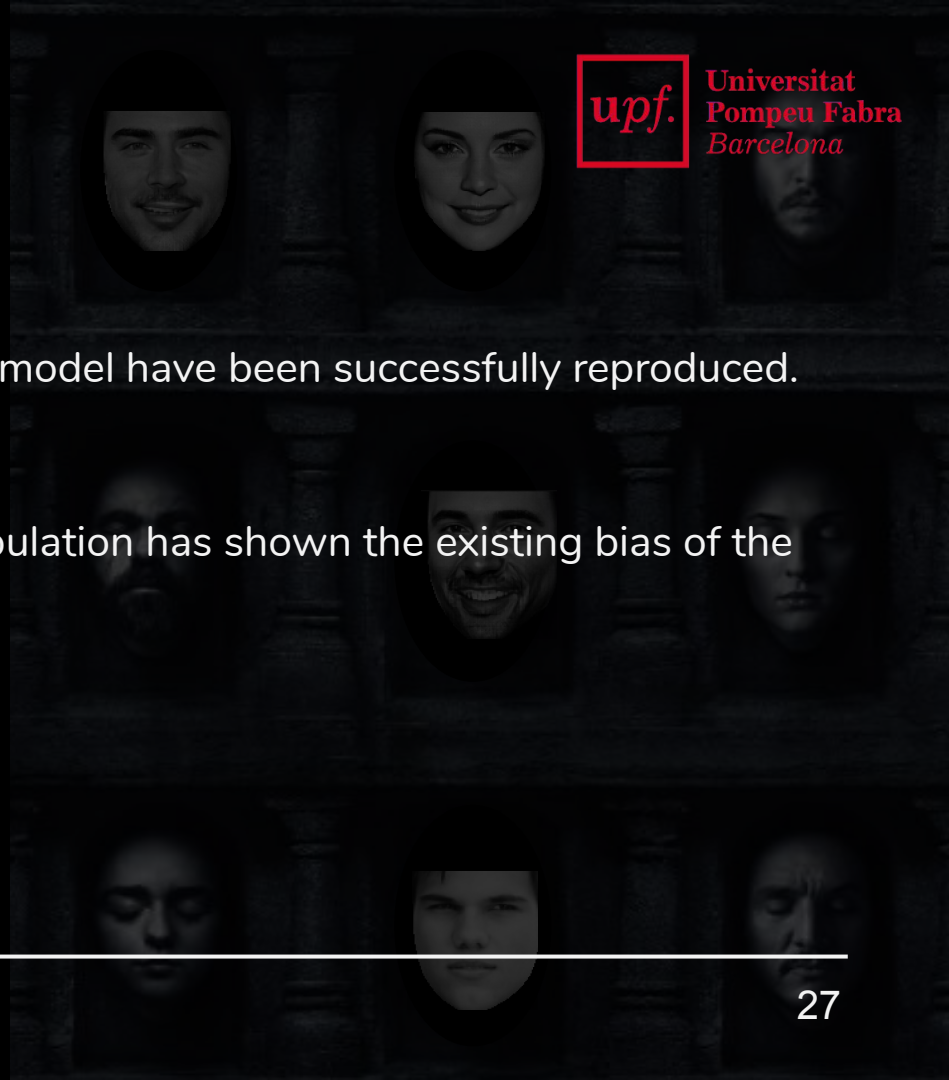
Linear Combination



Distance Score = 0.59
Mean Error = 0.30
12.50% of detected fakes

- Regarding the results, we prioritize a higher closeness to the target identity since the mean error is similar in both cases and the fake detection rate is not as high.

CONCLUSIONS



- The state-of-the-art results of the Glow model have been successfully reproduced.

- The extended analysis of attribute manipulation has shown the existing bias of the CelebA dataset.

CONCLUSIONS

Universitat Pompeu Fabra Barcelona

- Vector arithmetic for expression transfer between two identities:
  - The proposed methods rely on a successful reconstruction of images, the computation of the neutral face of the identities and their variance.
  - The best expression transfer has a mean error of 0.3 in the original 0-5 intensity scale. It can be performed by using the following linear combination:

$$\mathbf{z}_{\mathrm{OT}} = \alpha \cdot \mathbf{z}_{\mathrm{expression}} + \mathbf{z}_{\mathrm{MT}}$$

with the expression vector being:

$$\mathbf{z}'_{\mathrm{expression}} = \frac{\mathbf{z}_{\mathrm{OS}} - \mathbf{z}_{\mathrm{MS}}}{2}$$

CONCLUSIONS

- A *Google Colab Notebook* has been released to provide user-friendly tools to reproduce, verify and extend this present work.

- We have shown that normalizing flows are good candidate methods for generating deepfakes, possibly competing with the existing state-of-the-art.

Generation of Deepfakes using Normalizing Flows

"Inspiration exists, but it has to find you working."

Pablo Picasso

upf. Universitat Pompeu Fabra Barcelona

Telefónica