Deep Adversarial Frameworks for Visually Interpretable Periocular Recognition

João Brito | Prof. Dr. Hugo Proença | November 2020

Goal

Given two images from the **periocular region**, determine whether they came from the **same subject** and **explain** why

Interpretable System

Non-interpretable answer: score 0.14

Binary answer: “Yes” / *“No”*

Visually interpretable answer: *e.g. image*

Currently available approaches

LIME (**L**ocal **I**nterpretable **M**odel-Agnostic **E**xplanation)

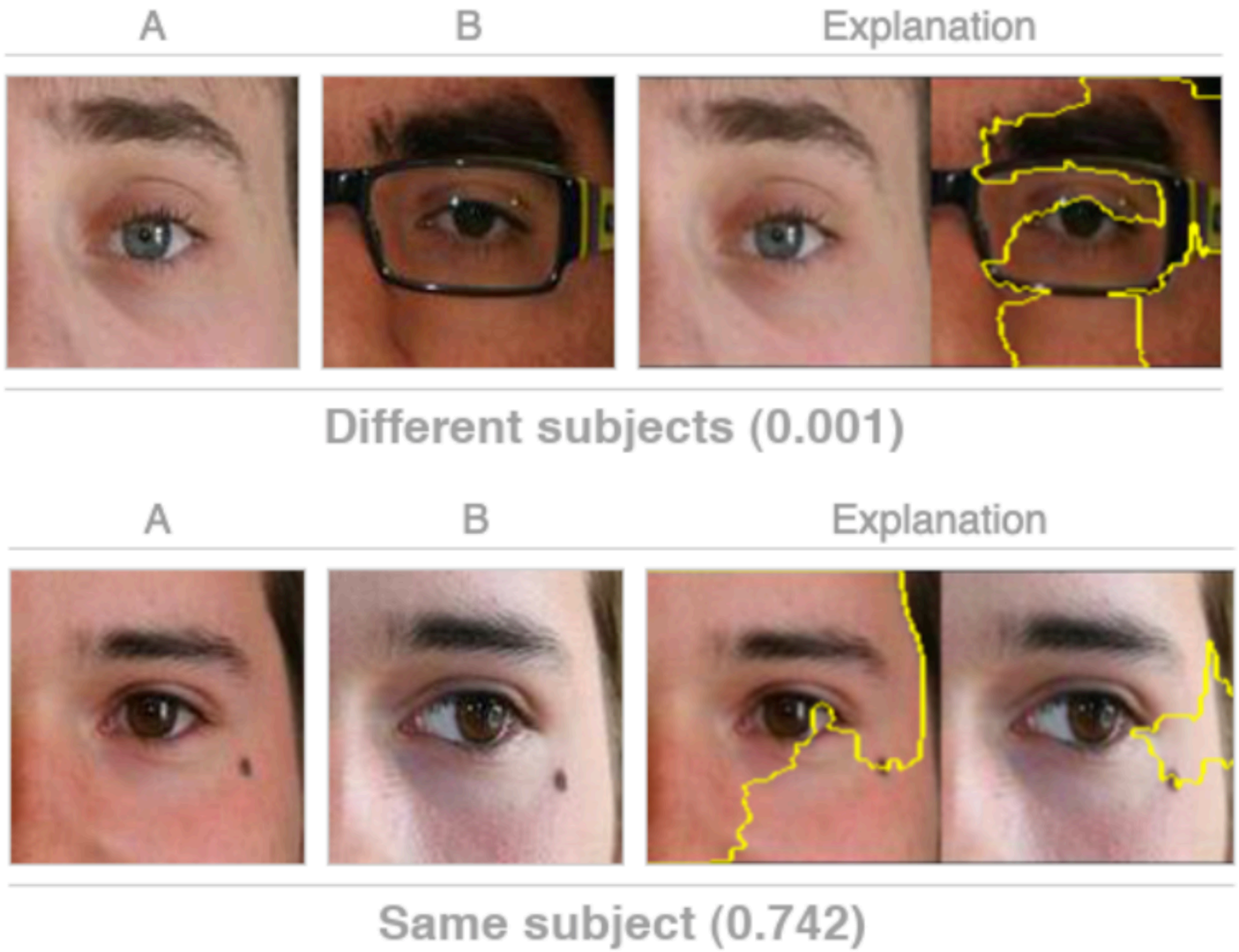
Simple Model

Relate the perturbations on the data and the changes in the original model's output

Black-box model's **output** on the "**altered**" data

$$explanation(x) = \arg\min_{g \in G} \underbrace{L(f, g, \pi_x)}_{\text{Loss}} + \underbrace{\Omega(g)}_{\text{Complexity}}$$

Currently available approaches



Problems

- Fails to realise the extent of some facial components
- Some explanations are harder to understand

Main ideas

> K-Neighbours on a big synthetic dataset

GAN

CNN

K-Neighbours

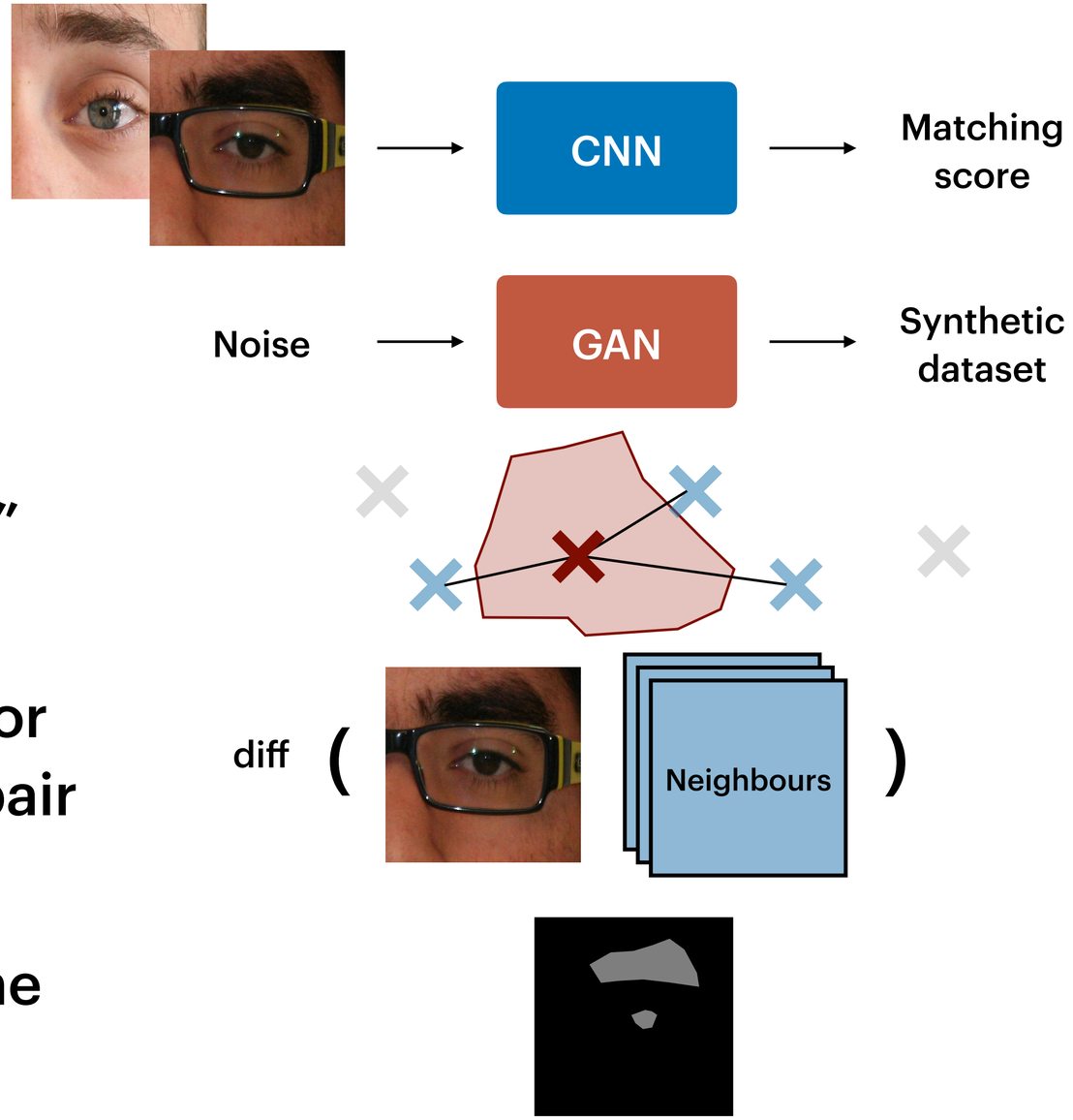
> Latent space directions and image manipulation

GAN

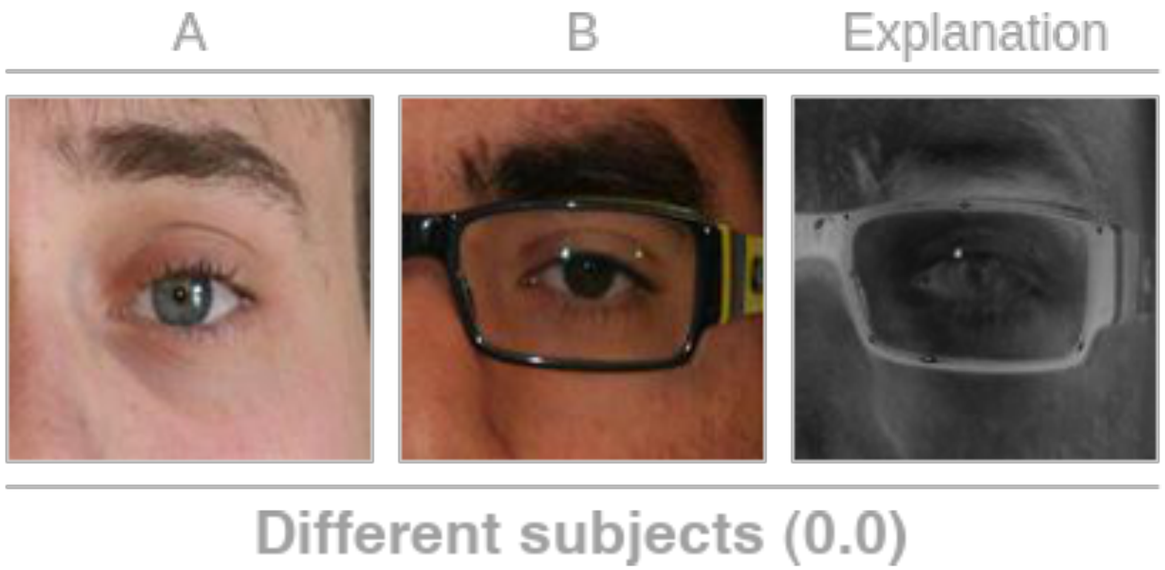
Latent space

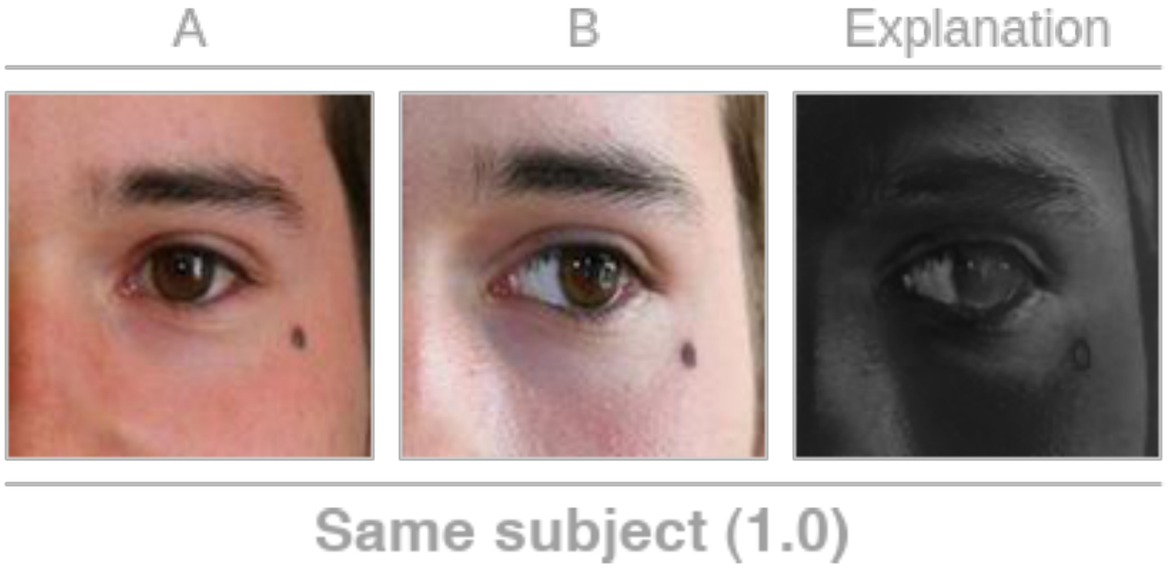
K-Neighbours on a big synthetic dataset

1. Train a regular **CNN** (e.g. DenseNet 161) to distinguish between **G**enuine and **I**mpostor pairs
2. Train a **GAN** (e.g. StyleGAN2) to generate **G**enuine and **I**mpostor pairs
3. Given the generated dataset, find the “**K**” **closest pairs** on “the other side”
4. Compute the **pixel difference** (e.g. MSE or SSIM) between the neighbours and the pair we want to explain
5. The areas with the most difference are the **interpretable reasons**



K-Neighbours on a big synthetic dataset

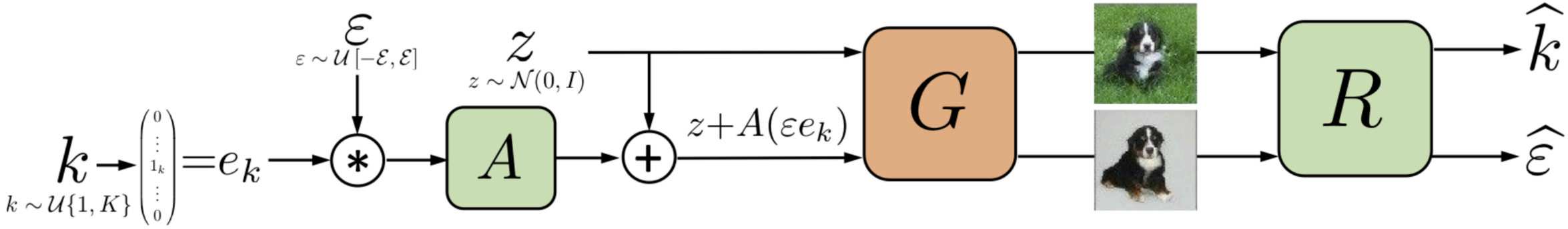




Problems

- Eyeglasses appear as difference factor
- Relies on generating an (ideally) infinite synthetic dataset
- Metric learning on latent codes

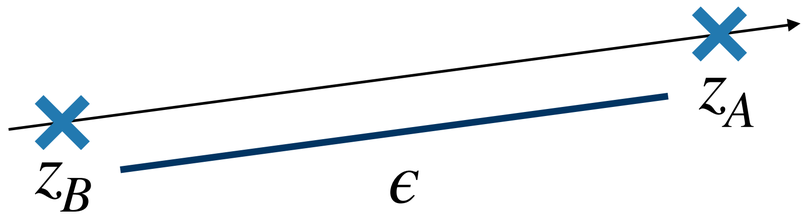
Latent space directions and image manipulation



in *"Unsupervised Discovery of Interpretable Directions in the GAN Latent Space"*

1 Train this method to find **usable directions** (e.g. change the color of the iris, skin, etc..)

2 Find the "ϵ" values for each **interpretable directions** (i.e. how far or close the two images are along that direction)



Latent space directions and image manipulation

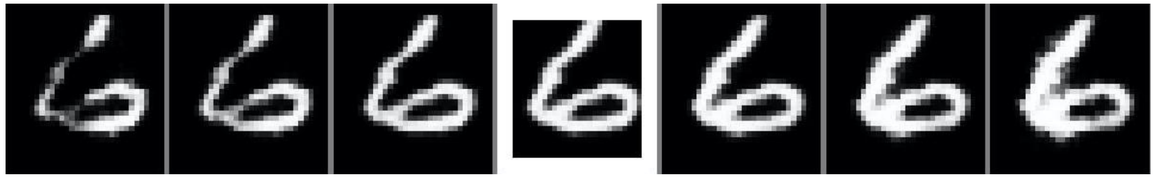
> Autoencoder (latent dim = 4096)

thickness

K = 5

thickness

K = 10



thickness

K = 64

advanced thickness

K = 128

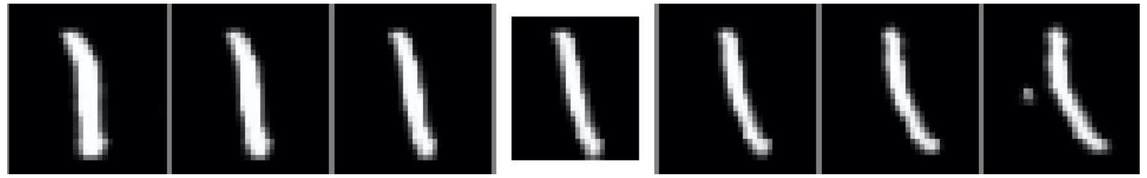
> Spectral Norm GAN (latent dim = 128)

thickness and angle

K = 5

angle

K = 10

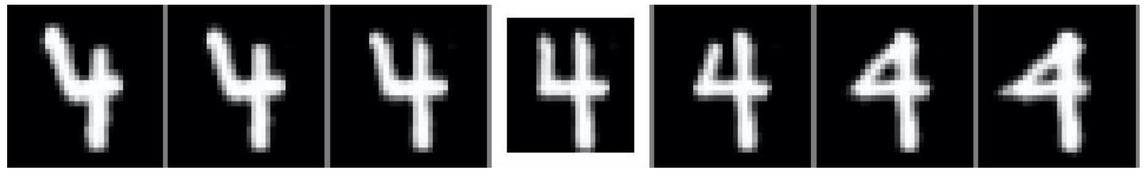


shape

K = 64

style

K = 128



Questions?

Thank you for your attention!