

# Photo attribute editing using neural networks

Student: *Evgeny Zholkovskiy*  
Research Advisor: *Prof. Ivan Oseledets*

# Introduction

Photo editing is now widely used in entertainment, media, medicine and many other areas.

Automated image editing with desired properties can simplify routine work and make editing available for everyone.

# Overview

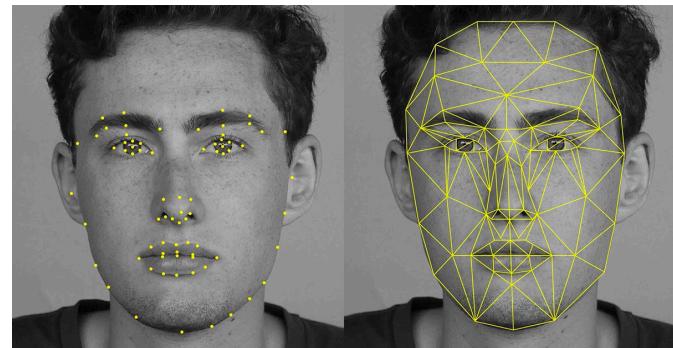
Generative networks



Segmentation



Facial landmarks



# Objectives

1. Obtain training dataset with examples of features we want to edit using the pipeline
2. Build a pipeline for editing photos.

# Dataset mining

First way:

1. Parse face photos from photo banks
2. Try to find face and key points
3. Crop mouth if found

+ Better classes after parsing,  
correct domain  
- Low resolution images

Second way:

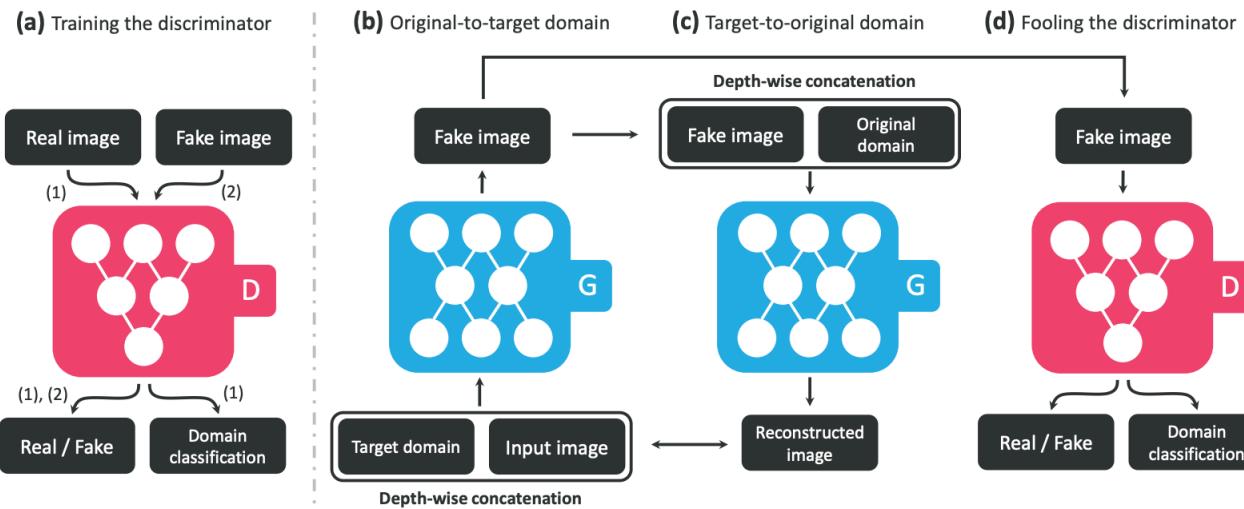
1. Parse teeth photo banks
  2. Try to face and key points
  3. Crop mouth if found
- + Noisy dataset, needs to be cleaned  
- High resolution images

# Dataset mining

1. Parse Bing Image (such queries as “braces smile”, “teeth smile”, etc)
2. Remove duplicates
3. Semi-automated data cleaning
4. Two classes with ~1k images each



# GAN Architecture



$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x, c), c')||_1]$$

$$\mathcal{L}_{cls}^r = \mathbb{E}_{x,c'}[-\log D_{cls}(c'|x)]$$

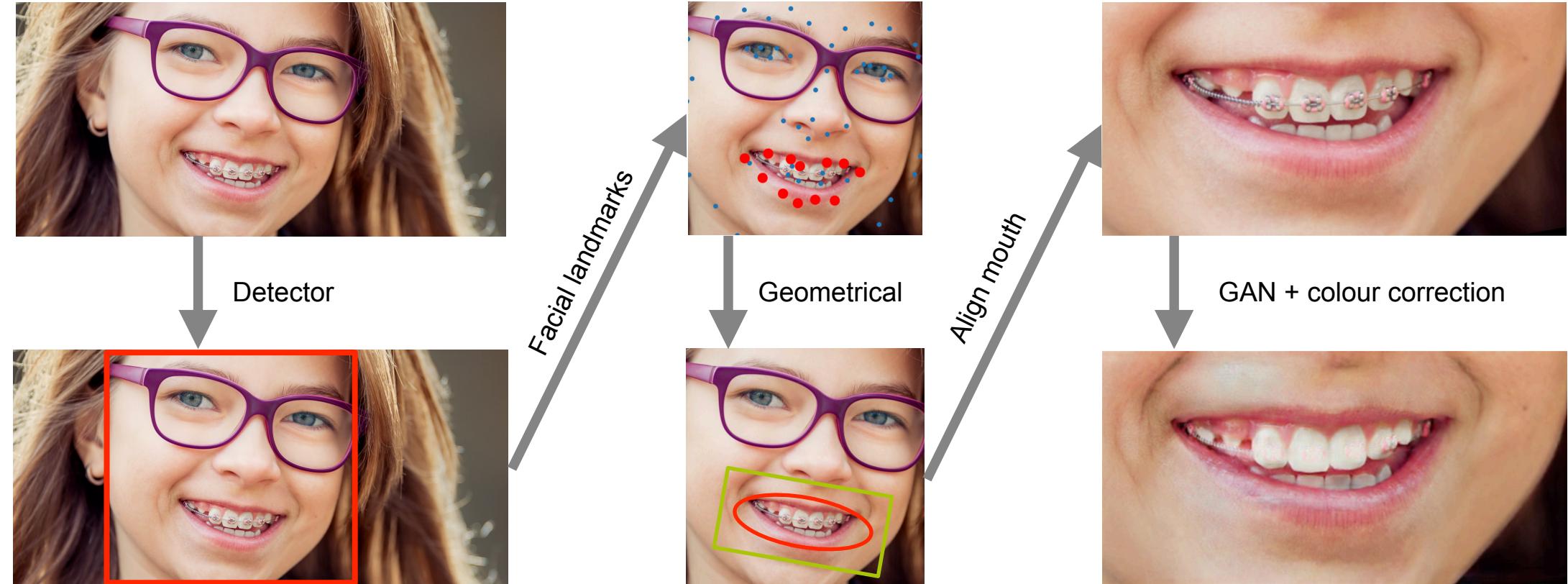
$$\begin{aligned} \mathcal{L}_{adv} &= \mathbb{E}_x[D_{src}(x)] - \mathbb{E}_{x,c}[D_{src}(G(x, c))] \\ &\quad - \lambda_{gp} \mathbb{E}_{\hat{x}}[(||\nabla_{\hat{x}} D_{src}(\hat{x})||_2 - 1)^2], \end{aligned}$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec}$$

# Processing pipeline

1. Face detection  
MTCNN, BCE + MSE  
FDDB dataset (3k images, bboxes)
2. Face landmarks detection  
Resnet-18, MSE,  
300w dataset (6k images, 68 keypoints)
3. Teeth align
4. Removing braces  
StarGAN based architecture, BCE + Adv.  
 $\lambda_{cls}=1$ ,  $\lambda_{gp}=10$ ,  $\lambda_{rec}=100$
5. Original photo editing

# Processing pipeline



# Results



# Results



# Results



# Results



# Discussion of results - problems

- Incomplete braces removal
- Blurry images
- Noise and artefacts
- Image tone change



# Novelty and Applications

1. A large new dataset was mined
2. The pipeline that was build allows to easily edit facial attributes on photos.

## Applications:

- Medicine and cosmetology
- Entertainment
- Photo editing and cameras

# Conclusions

1. Having a good dataset as much important as training a machine learning model. Obtaining appropriate dataset is a challenging task.
2. A pipeline of computer vision models for face detection, alignment and photo editing was built.

# Plans

1. Obtain more data
2. Perform experiments with training and architectures
3. Explore additional features and usages of the pipeline

# Thank you!