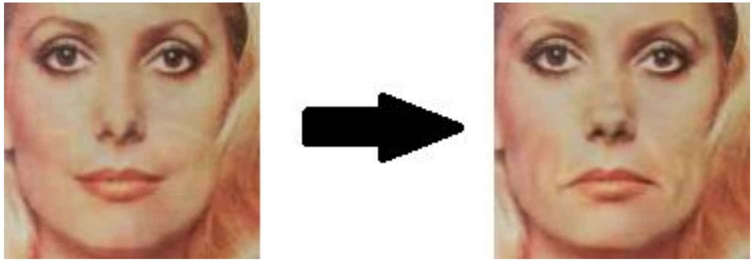
Computer Vision methods for GANimation

Jesse, Tommaso, Andrei-Daniel, Joel, Hilla

The Goal

- Generate a set of facial expressions given only one image of a face
- Conventionally: use GAN’s and generate images belonging to a given domain: e.g., people with a certain expression
- Presented model (introduced in the paper) tries not to be limited on the training expressions: done via a new GAN conditioning scheme

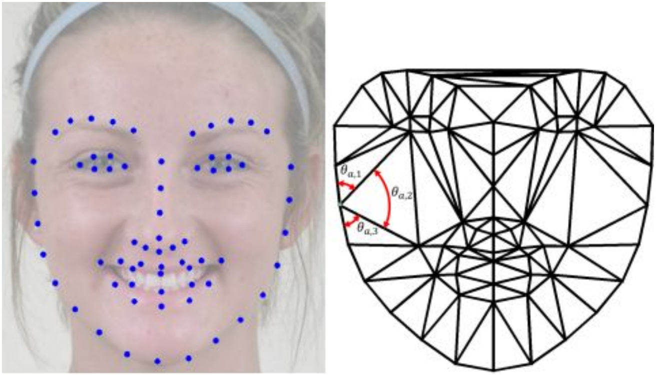


Problem Approach

- Emotion mapping via action units: given the original image and an action units target we want to create a new image that satisfies the target action units
- Image generation realized via a trained GAN
- Loss function composed of four components: image adversarial, attention, conditional expression, identity
- Action Units used to depict key components of facial expressions: e.g., Inner / Outer Brow Raiser / Lowerer, etc.
- By interpolation of AU’s the generated image is capable of depicting emotions that are not present in the training data set

Obtaining action unit values using feature detection

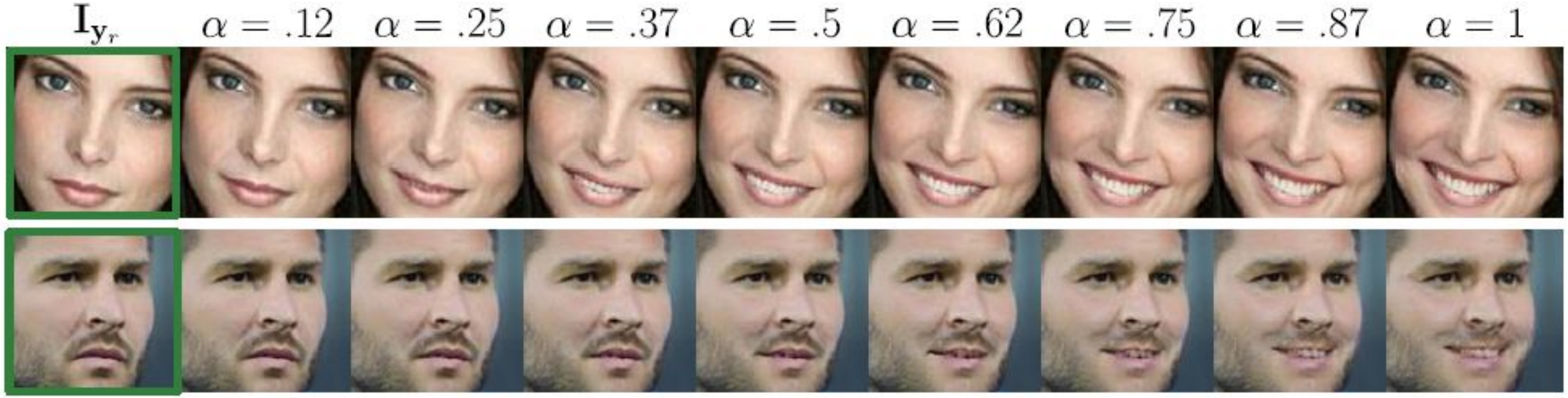
- Computer vision algorithms were used to obtain landmark points in faces (anatomical parts, corners, edges etc.)



- Using the landmark points and shading changes of the images, each image’s expression is encoded as a vector using kernel methods.
- This was done for 1 000 000 images, which became the Emotionet database.

Action Units

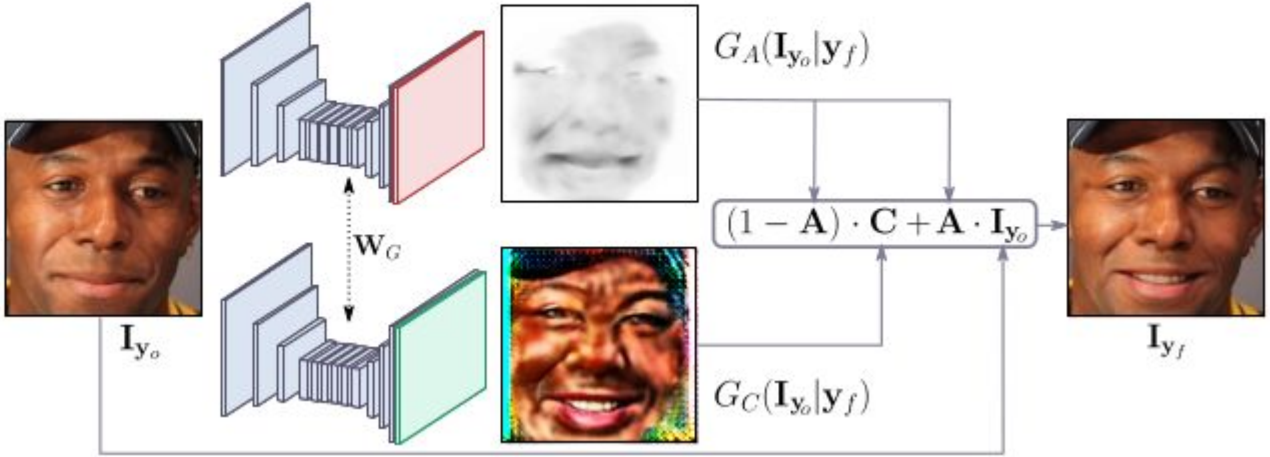
- Facial expressions are encoded as action unit vectors with length ~30
- Each action unit contains information about anatomical features caused by specific facial muscles.
- Action units can have activations between 0 and 1. This continuous mapping of facial expressions enables a large number of possible expressions by just using interpolation



Generative Adversarial Network

- A generator G is trained to realistically transform the original image to an image fulfilling the expression y (= encoded AU’s)
- A discriminator D is trained to evaluate the generated images in terms of their photo-realism and desired expression fulfillment

Network architecture



Network composed of:

- Generator network that produces two outputs:
  - Attention mask A
  - Color mask C
- Discriminator network

Use of convolution for feature detection

- Convolution (filtering) is used to process images small region at a time
- This can be done in order to modify an image, or to derive information with spatial awareness.
- Convolution layers used extensively in the image-to-image network
- Main purpose to detect features in input images
- Consecutive layers enable detecting larger, more complex features
- Multiple different convolution filters in a layer enables detecting many different features

Attention mask

- It is an output of the generator network together with the color mask
- It is used to allow the generator to focus only in the regions of the image that are responsible for the construction of the facial expression
- The rest of the elements of the image are just considered noise and they can left untouched, so they can just copied from the original image
    - Jewelry
    - Hats
    - Glasses
    - ...

How does it work in detail?

- The generator outputs two masks (attention and color mask)
- The **color mask** is an RGB color transformation over the entire image
- The **attention mask** defines a per pixel intensity specifying to which extend each pixel of the original image will contribute in the final rendered image
  - It relieves the color mask from having to accurately regress each pixel value
  - Only the pixels relevant to the expression change are carefully estimated, the rest are just noise.

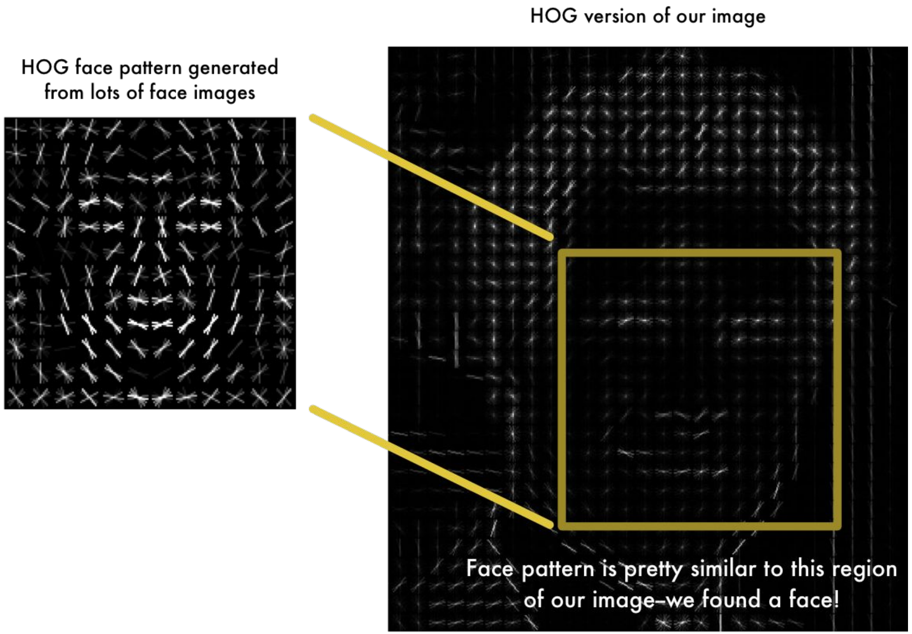
Example of attention mask

Advantages of the attention mask

- Allows applying the transformation only on the cropped face, and put it back onto the original image without producing any artifact
- Allows handle images in the wild with complex backgrounds and illumination conditions
- Results are sharper and more realistic because the network is focused on facial movements

Preprocessing: Face Detection using Histogram of Oriented Gradients

- Divide image into cells and calculate a histogram of gradient orientations for each cell
- Form feature vectors for each image, use those to train a classifier
- For GANimation preprocessing: use the classifier to detect the face and crop the bounding box



Adam Geitgey: https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78

Learning the Model

Image Adversarial loss:

- Discriminator tries to maximize the probability of correctly classifying real and generated images while the generator tries to fool the discriminator

Attention loss:

- We do not have ground truth for attention mask A or color mask C
- Regularize to prevent the masks from saturating and to ensure smooth spatial color transformation
- If the mask saturates the generator won’t produce any effect -> output image is a copy of the input image

Learning the Model

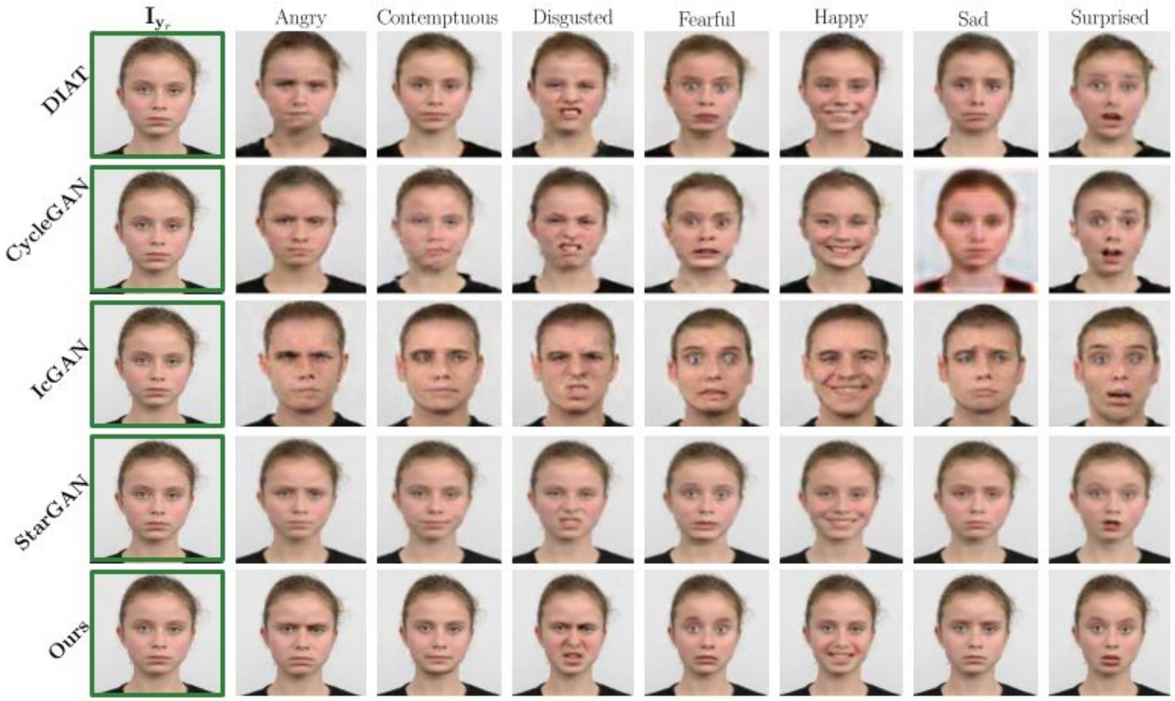
Conditional Expression loss:

- The discriminator also estimates the AUs activations in the generated image to satisfy the target expression

Identity loss:

- As we have no ground truth, there is no constraint to guarantee that the input and output images correspond to the same person
- Solution: use cycle consistency loss - penalize the difference between the original image and its reconstruction

Improvements compared to previous methods



How did everyone contribute?

Tafseer

- **The ghost of the team**

Jesse

- **Demo particularly**

Tommaso

- **Attention mask (its importance, advantages and examples)**
- **Selection of images for the presentation**

Andrei-Daniel

- **Goals, problem approach, identified computer vision concepts**

Joel

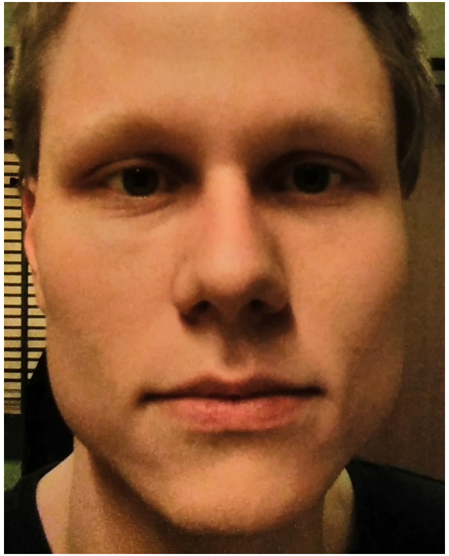
- **Wrote summaries, about feature detection, AUs, convolution**

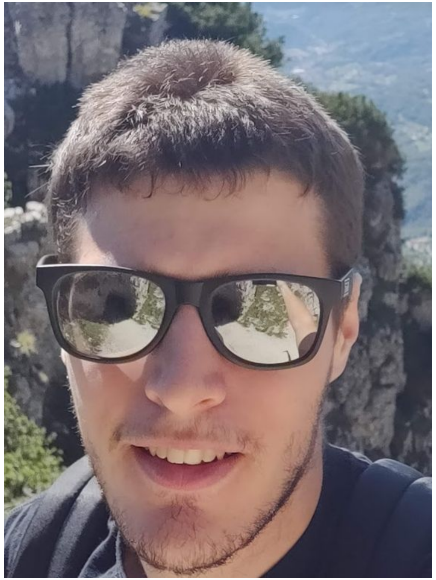
Hilla

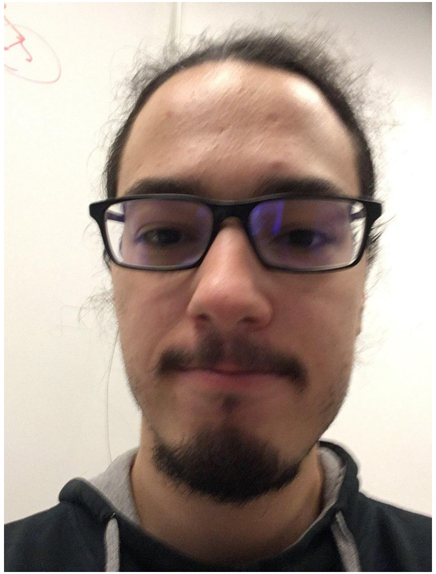
- **Image preprocessing, learning the model, GAN**

Everyone participated in the discussions and preparing of the presentation (not sure about the ghost though)

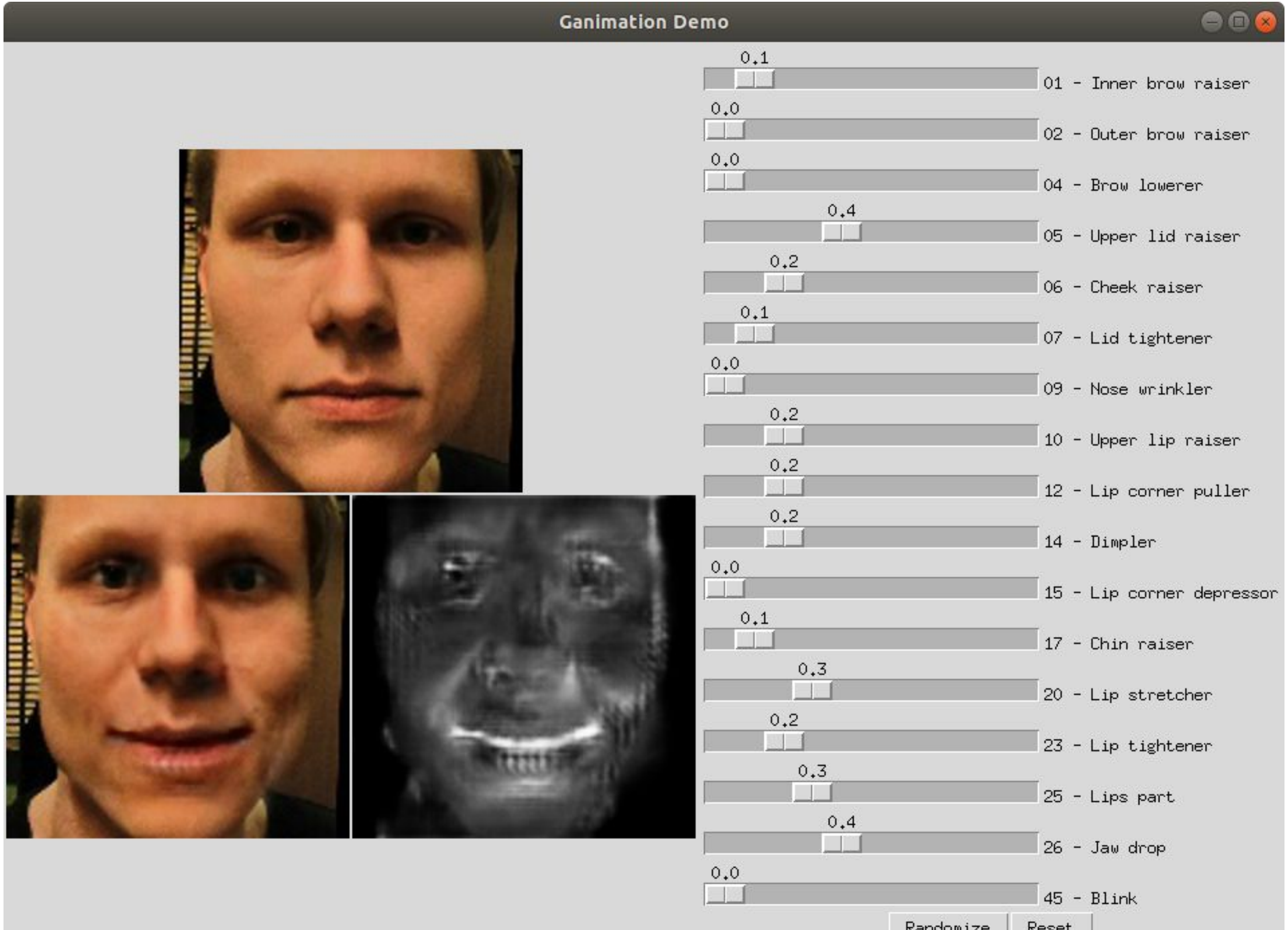
Team member faces for demo



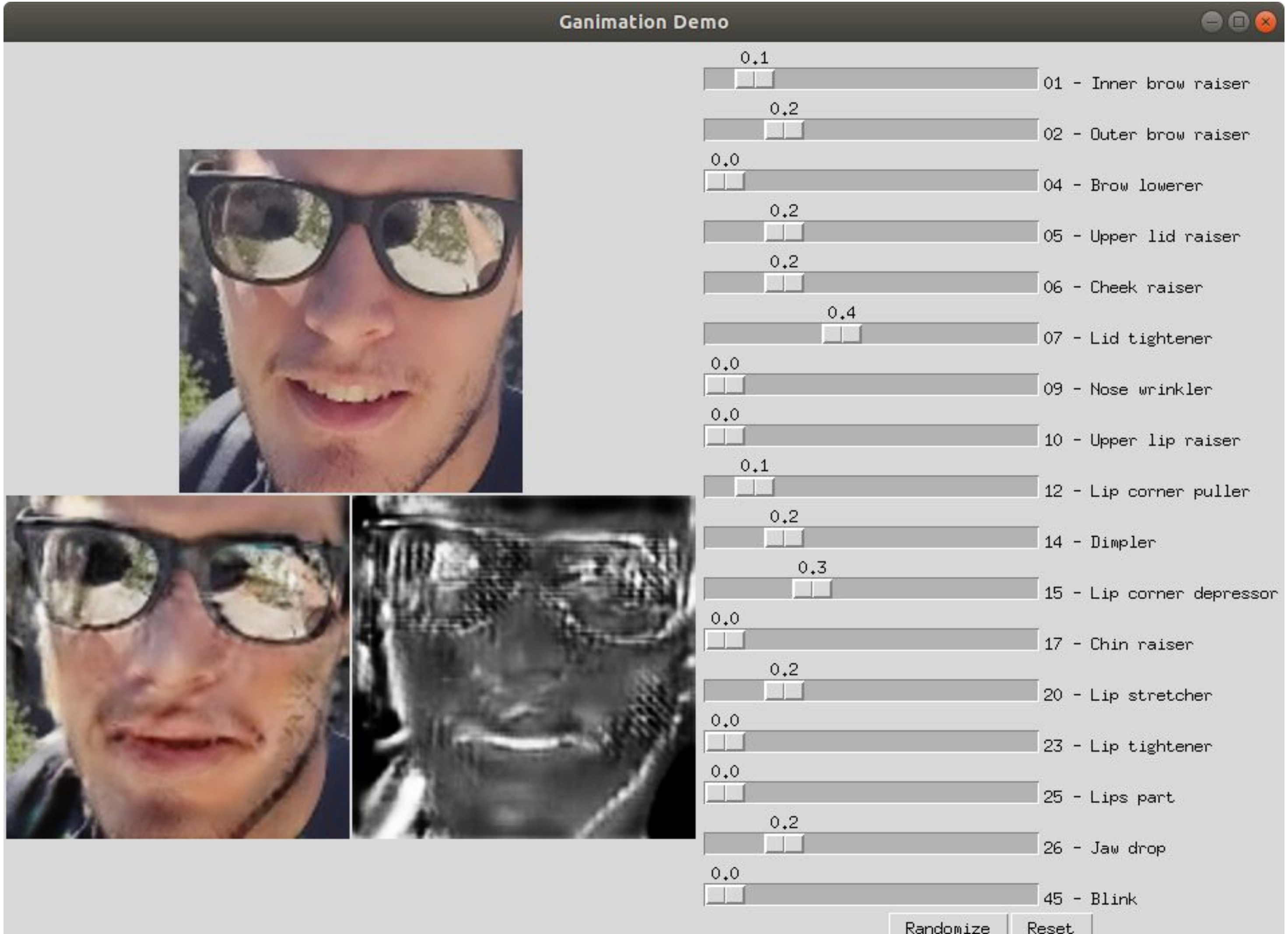


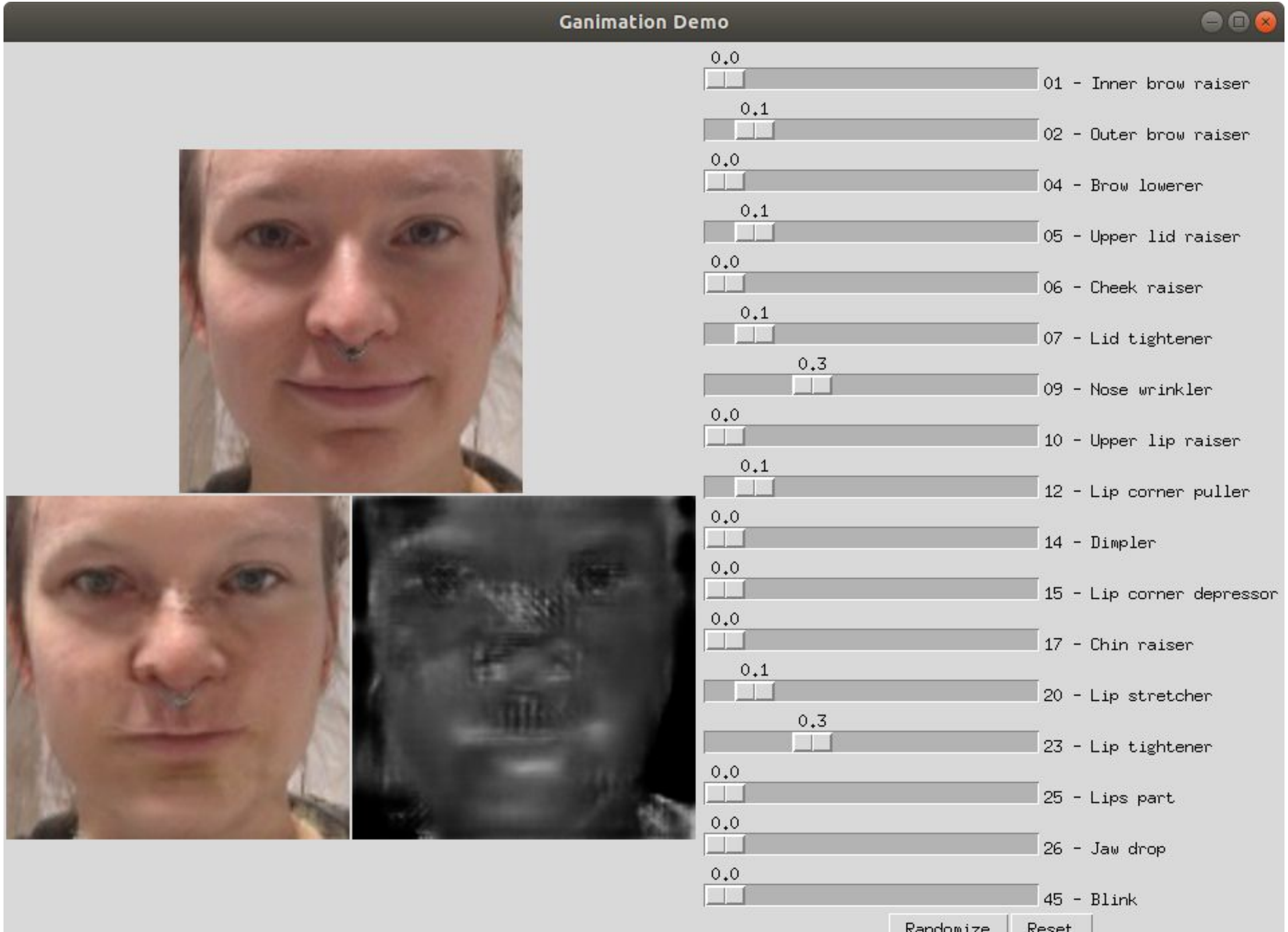


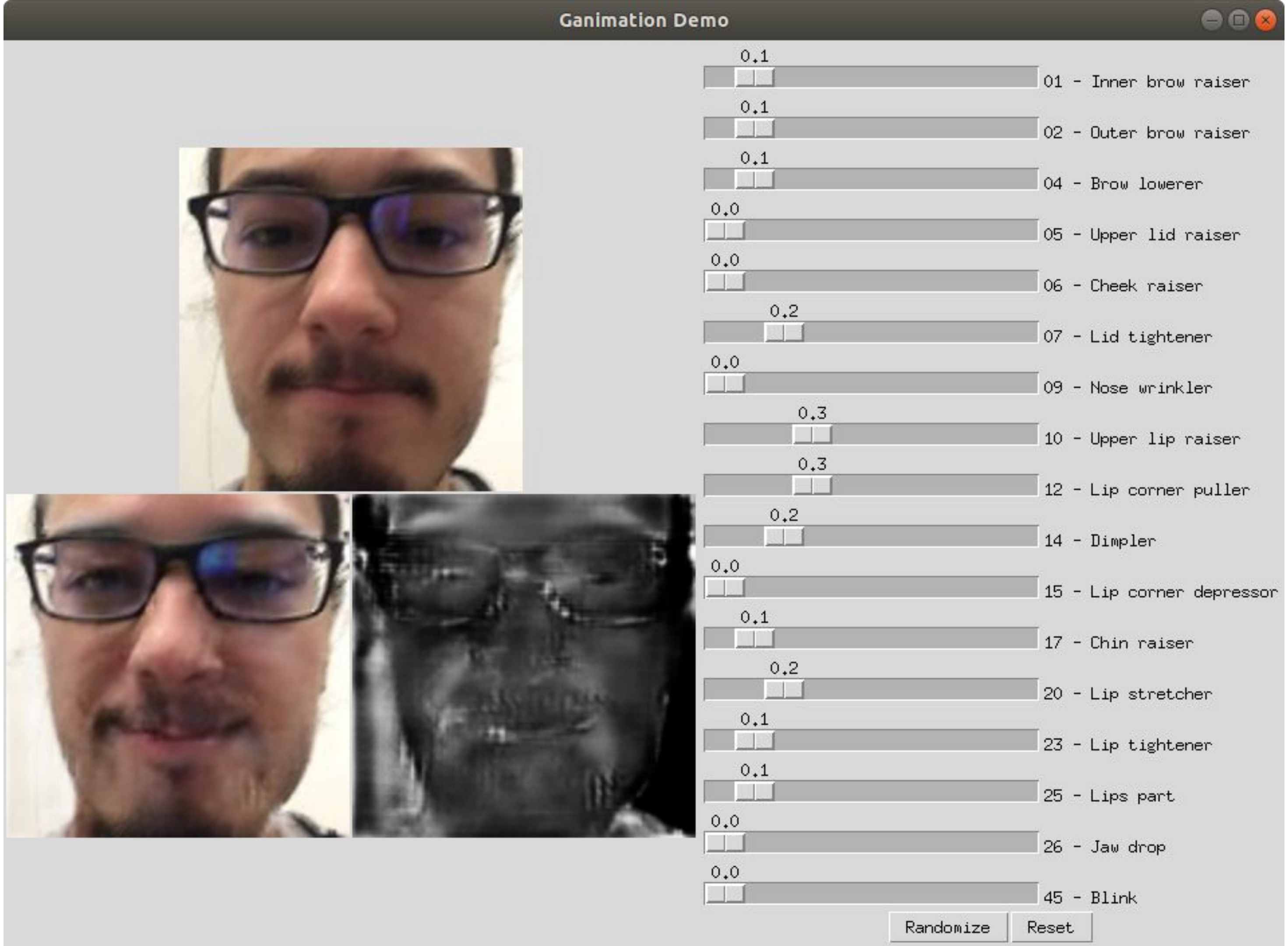
https://drive.google.com/file/d/1OWAx_Yb5S6_y3_fg88M77b_RS_SLkcM0/view?u



Ganimation Demo
0.1
01 - Inner brow raiser
0.0
02 - Outer brow raiser
0.0
04 - Brow lowerer
0.4
05 - Upper lid raiser
0.2
06 - Cheek raiser
0.1
07 - Lid tightener
0.0
09 - Nose wrinkler
0.2
10 - Upper lip raiser
0.2
12 - Lip corner puller
0.2
14 - Dimpler
0.0
15 - Lip corner depressor
0.1
17 - Chin raiser
0.3
20 - Lip stretcher
0.2
23 - Lip tightener
0.3
25 - Lips part
0.4
26 - Jaw drop
0.0
45 - Blink
Randomize
Reset


Ganimation Demo
0.1
01 - Inner brow raiser
0.2
02 - Outer brow raiser
0.0
04 - Brow lowerer
0.2
05 - Upper lid raiser
0.2
06 - Cheek raiser
0.4
07 - Lid tightener
0.0
09 - Nose wrinkler
0.0
10 - Upper lip raiser
0.1
12 - Lip corner puller
0.2
14 - Dimpler
0.3
15 - Lip corner depressor
0.0
17 - Chin raiser
0.2
20 - Lip stretcher
0.0
23 - Lip tightener
0.0
25 - Lips part
0.2
26 - Jaw drop
0.0
45 - Blink
Randomize
Reset


Ganimation Demo
0.0
01 - Inner brow raiser
0.1
02 - Outer brow raiser
0.0
04 - Brow lowerer
0.1
05 - Upper lid raiser
0.0
06 - Cheek raiser
0.1
07 - Lid tightener
0.3
09 - Nose wrinkler
0.0
10 - Upper lip raiser
0.1
12 - Lip corner puller
0.0
14 - Dimpler
0.0
15 - Lip corner depressor
0.0
17 - Chin raiser
0.1
20 - Lip stretcher
0.3
23 - Lip tightener
0.0
25 - Lips part
0.0
26 - Jaw drop
0.0
45 - Blink
Randomize
Reset


Ganimation Demo
0.1
01 - Inner brow raiser
0.1
02 - Outer brow raiser
0.1
04 - Brow lowerer
0.0
05 - Upper lid raiser
0.0
06 - Cheek raiser
0.2
07 - Lid tightener
0.0
09 - Nose wrinkler
0.3
10 - Upper lip raiser
0.3
12 - Lip corner puller
0.2
14 - Dimpler
0.0
15 - Lip corner depressor
0.1
17 - Chin raiser
0.2
20 - Lip stretcher
0.1
23 - Lip tightener
0.1
25 - Lips part
0.0
26 - Jaw drop
0.0
45 - Blink
Randomize
Reset

Thank you!