Keep Smiling: Generative Adversarial Networks for Face Manipulation

Johannes Grimm   Emanuel Maus
Lecture: Object Recognition and Image Understanding, Prof. Dr. Björn Ommer

RUPRECHT-KARLS-
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Objectives

- Application which puts a smile on a photo, where the person is not smiling
- Dataloader including a cropper for the mouth (resource saving)
- Using a Generative Adversarial Network (GAN) for production of a smiling mouth
- Operating in feature space to switch between non-smile and smile state

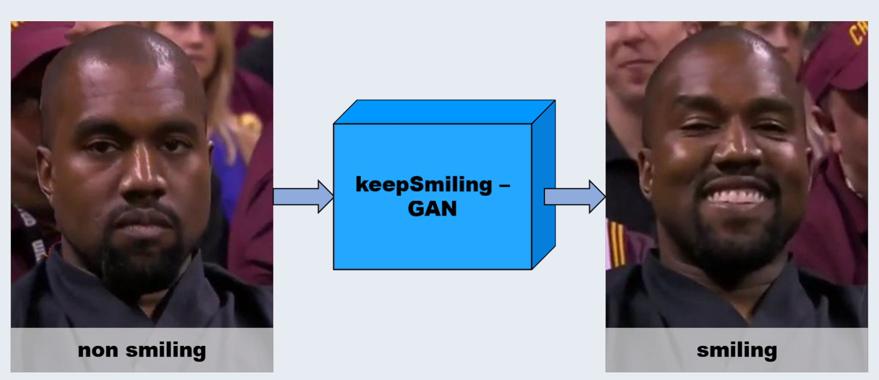

Figure 1: Goal of our project.[1]

Introduction: Dataloader

- Method for importing the dataset "Facial Landmark Detection by Deep Multi-task Learning"[2], which has a training and a testing dataset and provides 5 defined facial landmarks per picture (∼ 7500 training images, testing: smile ∼ 1000, non-smile ∼ 700)
- Preprocessing the images with cropping and resizing

Introduction: GAN

- Simple GAN[3]: consists of a Generator which learns to generate fake samples from a random noise vector and a discriminator, which distinguish between real and fake samples
- Complex GAN[4]: the generator starts from an image and the discriminator can also distinguish between a given label (smile, not smile)
- Splitting Generator: Generator can be split into decoder and encoder to access feature space

Method: Dataloader

The dataset delivers 5 facial landmark points (both eyes, nose and left and right mouth corners). With these information the mouth can be cropped using an ellipse to reduce resource power: The two mouth corners are the focus points of the ellipse (2,3) and the nose is the point on the ellipse (1), which is above its middle point (4). Figure 2 shows the described cropping procedure.

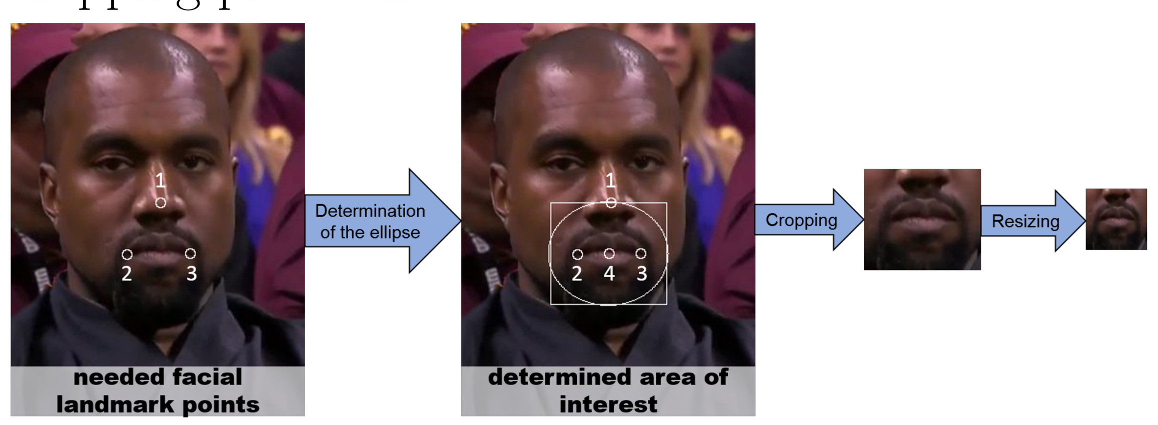

Figure 2: Cropping procedure of the Dataloader.[1]

Methods: GAN

1. Simple GAN: In general, a GAN consists of two neural networks, a generator, which generates new data instances and a discriminator, which evaluates this instances for authenticity. The exact steps are shown in figure 3. Both networks try to optimize different loss functions in dependence of each other.

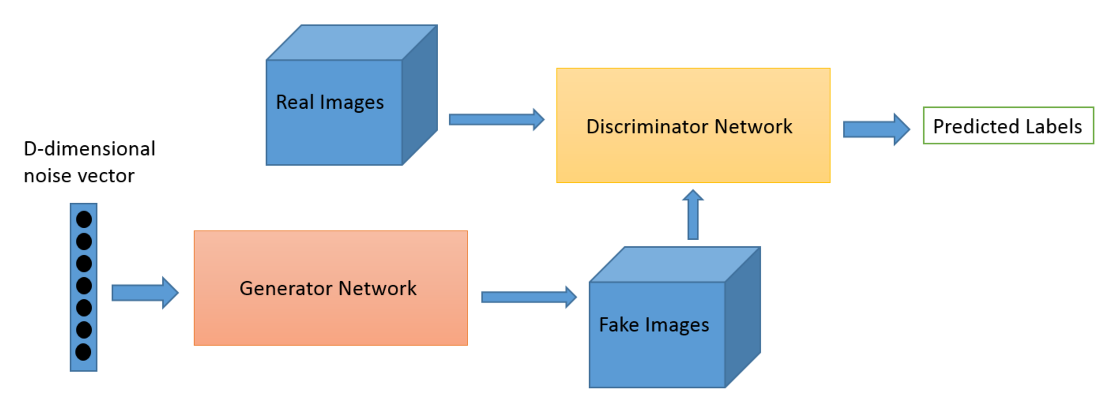

Figure 3: Overview of a simple GAN.[3]

2. Complex GAN: The Generator starts from a real image, then goes to the feature space and afterwards generates a resulting image depending on the target classification label (smile, not smile). This adds a further loss function (L1 norm between input image and generated image). Also the discriminator not only distinguishes between real and fake images but also between smiling and not smiling.
3. Split GAN: The Decoder (Generator) from the simple GAN is extended with a separate Encoder to allow mathematical operations in the feature space. There, the difference of the mean smile and mean non-smile vector represents the shift vector from non-smiling to smiling and vice versa. Besides this modification, the method of operation remains the same as for the Simple GAN.

Results

Method 1:

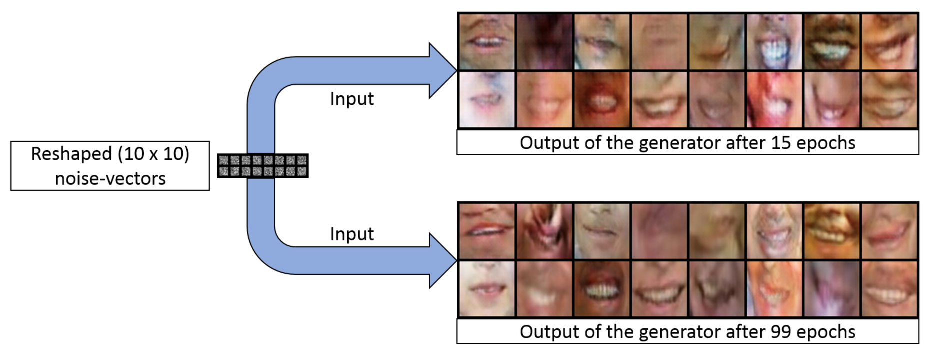

Figure 4: Output of the Generator after 15 and 99 epochs.

- Good output of the Generator after a few epochs
- Smiling can already be recognized

Method 2:

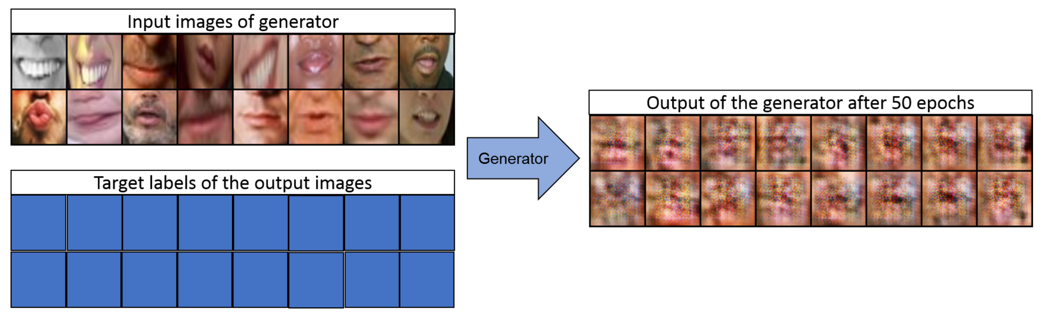

Figure 5: Input and output of the Generator after 50 epochs.

- Generated images still very noisy
  → Model complexity → more learning time
  → Dataset provides outliers (see figure 5)

Method 3:

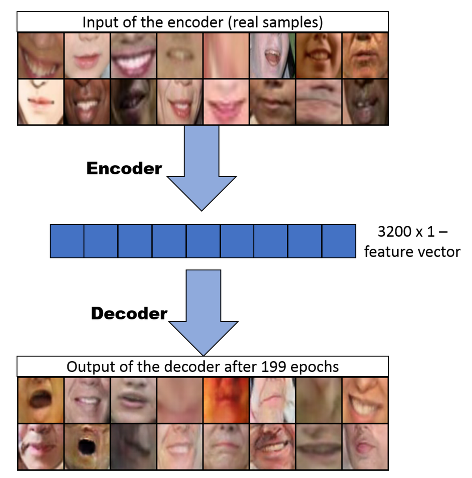

Figure 6: input and output of Split GAN.

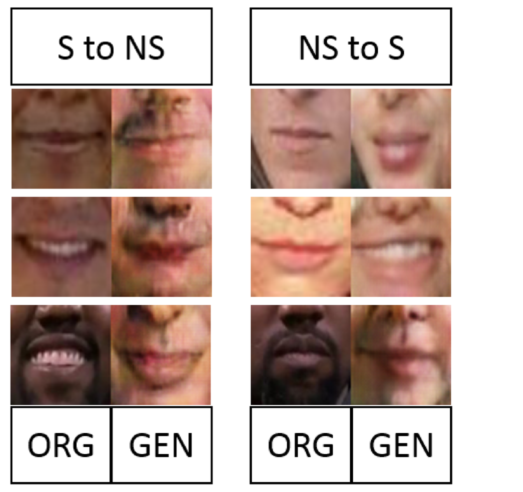

Figure 7: Original and reconstructed images of the Split GAN.[1]

- Figure 7 shows the reconstructed images to which the shift vector has been added or subtracted in feature space.

label	length(std)	#vectors in other conf. band
smile	36.7	48%
non-smile	36.2	50%

Table 1: Informations about the feature vectors.

Outlook

- Improve the Dataset
- More training time for the Complex GAN
- Better cropping method or taking the whole face into account
  ⇒ More computing power / time
- Reducing the feature space with more convolution layers

References

[1] Edward Roberts. Kayne West Image. https://www.mirror.co.uk.

[2] Zhang et al. Facial landmark detection by deep multi-task learning, in proceedings of european conference on computer vision.

[3] Xu et al. An empirical study on evaluation metrics of generative adversarial networks.

[4] Choi et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation.