GANs for Generating Different Face Expressions

Helena Čeović Josipa Lipovac Antun Magdić Andrea Omićević Ante Žužul

***Abstract*—Since 2014 when they were first proposed by Ian Goodfellow, generative adversarial networks have been successfully used in various image generation tasks. In this project we use them to generate facial expressions corresponding to different emotions (Angry, Happy, Sad, Surprise). Developed GAN model takes an image of a person's face and a desired emotion as inputs. It outputs the image of the same person's face, but with the required emotion applied. A large part of generated images look like real photos of real people. Unfortunately, since GANs are still under research, a human eye can almost always distinguish real images from the fake.**

***Index Terms*—GAN, Facial Expressions, StarGAN**

I. Introduction

There have been several researches that tackled the problem of facial expression recognition. However, even though the problem of facial expression generation is more challenging than expected, it has been less investigated in the state-of-the-art. Being able to automatically animate the facial expression from a single image opens the door to many new exciting applications in different areas, including the movie industry, photography technologies, fashion, e-commerce business etc. As Generative Adversarial Networks (GANs) have become more successful and more prevalent, a big progress has been made in this task. The most successful architecture is StarGAN, which is able not only to synthesize novel expressions, but also to change other attributes of the face, such as age, hair color or gender. In this project, we will use StarGAN to generate facial expressions corresponding to different emotions. A GAN model in this project is developed to take an image of a person's face and a desired emotion as inputs. It outputs the image of that person's face with the required emotion applied, while personalized features of the face remain preserved.

II. Related Work

Generating a particular person's face with different facial expressions can be used in a variety of applications, including face recognition [1], [2], face verification [3], [4], emotion prediction, expression database generation, facial expression augmentation and entertainment.

Generative Adversarial Networks (GANs) are a powerful class of generative models based on game theory. A typical GAN optimization scheme consists of simultaneously training a generator network to produce realistic fake samples and a discriminator network to distinguish between real and fake data. This idea is embedded by the so-called adversarial loss [5].

DCGANs are generative convolutional networks based off of GANs. Based on work [6], DCGANs seem promising. Using them it is possible to reconstruct the original image with great accuracy. However, it is also shown that the network has not learned to modify the image.

According to the article [5], the GANimation (Anatomically Consistent Facial Animation) proposed additionally controlled generated expressions by Action Units labels, and allowed a continuous expression transformation. The authors introduced an attention-based generator to promote the robustness of their model for distracting backgrounds and illuminations.

Different approach was used in [7] to construct double encoder GAN. Double encoder GAN is used for facial expression synthesis to extract the latent vectors and conditional labels features of the real image.

Recent advances in GANs have shown impressive results for task of facial expression synthesis. The most successful architecture is StarGAN [8], that conditions GANs' generation process with images of a specific domain, namely a set of images of persons sharing the same expression [9]. Our work is based on this approach.

III. Proposed Solution

Proposed solution...

IV. experimental Results

Dataset: FER2013[1]. The data consists of 48×48 pixel grayscale images of faces. The faces have been automatically cropped so that the face is centered. Images are labeled with 7 different emotions: Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral. We used 4 of them to train our model due to lack of processing power: Angry (3395 images), Happy (7215 images), Sad (4830 images), Surprise (3171 images).

Training was performed using Adam optimizer [10] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. We train our model with learning rate 0.0001 for 400 epochs. Batch size is set to 8. Training execution time was roughly 1 day on NVIDIA Tesla T4 GPU.

Gradual improvement of desired facial expressions can be seen throughout epochs. After 100 epochs, there are only slight changes which are prevalent in the mouth area. Visible lines, which remind of moustache, appear around the mouth since it is where facial expressions most differ. By the 400th epoch, modifications are more obvious and the desired emotions more easily recognizable. Changes involve several facial areas, including eyebrows, eyes, etc.

Pronounced changes are visible on more expressive emotions such as happiness, anger and surprise while changes on sad faces are somewhat less apparent.

[1]`https://www.kaggle.com/msambare/fer2013?select=train`

V. Conclusion

In this work, we have presented how StarGAN model can be used to generate various facial expressions. Based on generative adversarial networks, generator competes with discriminator to achueve the best results. Noticeable differences are observed between the original set of images and the images generated for given emotions (Angry, Happy, Sad, Surprise). Even though the changes are visible, additional training of the model would improve the quality of results.

References

[1] Cao, Q., Shen, L., Xie, W., Parkhi, O.M., & Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 67–74). IEEE.

[2] Parkhi et al. 2015 - Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In BMVC (Vol. 1, p. 6).

[3] Sun et al. 2014; - Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification–verification. In Advances in neural information processing systems (pp. 1988–1996).

[4] Taigman et al. 2014 - Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701–1708).

[5] Pumarola, A., Agudo, A., Martinez, A.M. et al. GANimation: One-Shot Anatomically Consistent Facial Animation. Int J Comput Vis 128, 698–713 (2020).

[6] Suit, J. Generating Facial Expressions

[7] Chen, Mingyi & Li, Changchun & Li, Ke & Zhang, Han & He, Xuanji. (2018). Double Encoder Conditional GAN for Facial Expression Synthesis. 9286-9291. 10.23919/ChiCC.2018.8483579.

[8] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi- domain image-to-image translation. CVPR (2018)

[9] Pumarola, A., Agudo, A., Martinez, A.A., Sanfeliu, A., Moreno-Noguer, F.: Ganimation: Anatomically-aware facial animation from a single image. In: ECCV (2018)

[10] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.