BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023

http://excel.fit.vutbr.cz

# Damaged Facial Image Reconstruction

Bc. Filip Pleško*

**Abstract**

Is it possible to use modern technology to fill in the missing or covered part of the face in an image? This work seeks to answer this question using a Generative Adversarial Neural Network (GAN). The work aims to investigate what approaches to face reconstruction exist and, based on the knowledge gained, propose a custom solution that could best accomplish the task. The paper further describes the different modifications to the Generator and compares how the different changes in the model architecture affect the output quality.

*filip.plesko@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

## 1. Introduction

Image reconstruction is a challenging task, in which it is necessary to be sure that the drawn missing part blends well with the known environment and at the same time that the resulting image is not blurred, which would make it clear at first glance that it is a generated image, which is not the sufficient quality of the reconstruction. This work focuses specifically on the reconstruction of the face image, which adds even more difficulty to this task as the face contains many unique key features. Suppose a key feature such as an eye, nose, or mouth is covered. In that case, generating this feature in the correct location is necessary, as even a slight deviation can cause the resulting face to be deformed and easily recognizable as the generated one.

Current existing solutions [1, 2, 3] try to solve this problem using different GAN architectures. This model contains two other models competing against each other, playing the adversarial game in which they are trying to beat each other and, by doing so, improve the quality of the output. Initially, the generated images are blurry, and it is easy for the Discriminator to detect them. Still, gradually the Generator improves and generates sharper images which are much more difficult to detect whether they are real or generated.

Training GAN models is a complex and time-consuming process of continuously improving both models. This means that any change in one model affects the training of the other and vice versa, making the training dynamic and often unstable. When training a GAN model, the goal is not to find a minimum as in classical models but to find an equilibrium point between the Generator and the Discriminator. There are no theoretical foundations for building a GAN model architecture or a proven training method to guarantee successful training without network collapse. Still, there are hints to increase the probability of success [4] that will be described in Section 2.

The output of this work could be used to enhance the performance of tasks such as face recognition, where another object can cover part of the input face, and the missing part can be reconstructed.

## 2. Best practices

As Section 1 mentions, face reconstruction using a GAN model is challenging. A considerable amount of attention must be spent on the correct configuration of the model architecture or the hyperparameters during training.

While no well-defined practices specify what the correct approach should look like, studies have been conducted [5] in the past years to investigate how changing certain parameters affects the stability of training and the achievement of desired results. For example, feature extraction using Convolutional Neural Network commonly ends with Dense layers. Those can be replaced with 2DConvolution layers. Or replacement of Pooling layers with strided convolutions can also improve output quality. Using Adam optimizer with a lower learning rate was proven to lower the risk of network collapse.

Just as the model architecture needs to be correctly defined to achieve the desired reconstruction quality, the correct loss functions for both nested models must be chosen.

All these verified practices have been tried in the proposed solution and compared to their effect on facial reconstruction as Section 3 describes.

## 3. Proposed solution

Based on the research, the chosen solution is the GAN model, whose Generator and Discriminator are two Convolutional Neural Networks. Firstly the base model was designed, and the Generator was composed of Encoder and Decoder. Encoder was inspired by VGG-19 architecture, and contained several 2DConvolution layers followed by MaxPooling layers and was finished with Dense layers. The Decoder, whose architecture can be seen in Figure 6, was designed to reverse Encoder feature extraction by combining 2DConvolutions and Transpose Convolution for upsampling, resulting in $256 \times 256 \times 3$ output reconstructed image.

### 3.1 Loss functions

Loss functions give the models feedback during training to let them know if adjusting the weights had the desired impact on the output. Several loss functions must be chosen correctly for the proposed solution to the reconstruction task, namely for the Generator and the Discriminator. An overview of the different loss functions is provided by Loss Functions of Generative Adversarial Networks (GANs): Opportunities and Challenges [6].

For the Generator to work correctly, a loss function must be chosen to measure how big an error the Generator made during the reconstruction. This function is called application loss or reconstruction loss. In addition to the reconstruction loss, the Generator also uses feedback from the Discriminator to tell it how easy it is to distinguish the generated images from the real ones. This helps the Generator produce more realistic and less blurry outputs.

In the case of the Discriminator, it is a Convolutional Neural Network that separates the inputs into two classes, real and generated. So it is a binary classifier that uses a Binary cross-entropy loss function, which we also call an adversarial loss in this case.

### 3.2 Model architecture

For successful training of GAN models, it is necessary to properly design their architecture so that the network does not experience instability or collapse during training. Radford & Metz [4] describe some valuable tricks to design such a network. Examples include using strided convolutions instead of pooling layers, removing fully connected layers when extracting low-level features of an image, and replacing them with convolutional layers but also using the ReLu activation function in the Generator and the LeakyReLu activation function in the Discriminator. Another significant change in the model architecture can be the addition of skip connections between the Encoder and Decoder parts of the Generator. Those changes are supposed to lead to more stable training, prevent the model from overfitting and improve the final quality of results. All those changes were implemented into the base model and tested. The results of those tests were reviewed, and those which increased the reconstruction quality were combined into the final model architecture that can be seen in Figure 5.

## 4. Results

This paper proposed GAN architecture that can reconstruct the damaged facial image. Several model architecture changes were implemented, tested, and compared. An example of the outputs of those different versions can be seen in Figure 2,3,4. The quality of generated image, measured by PSNR and SSIM metrics, is shown in Table 1 This paper also shows that reconstructed images can improve the performance of face recognition tasks, as shown in Figure 8.

## 5. Conclusion

This paper focuses on whether it is possible to reconstruct the covered part of the face in the image using modern technologies. Work done showed that it is possible to reconstruct the missing parts of the facial image by using GAN with special improvements to the Generator architecture. The most significant improvement in the quality of results was after using skip connections between Encoder and Decoder. It was also proven that face reconstruction significantly improves the results of face recognition tasks if part of the face is covered. The limitation of the work is that the input images must have a well-defined alignment for quality output.

## References


[1] Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang. Generative face completion. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5892–5900, 2017.

[2] Yanshun Zhao, Jinda Hu, and Xindong Zhang. Face restoration based on gans and nst. In *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, ICMAI 2020, page 198–203, New York, NY, USA, 2020. Association for Computing Machinery.

[3] Xin Hong, Pengfei Xiong, Renhe Ji, and Haoqiang Fan. Deep fusion network for image completion, 2019.

[4] Jason Brownlee. Tips for training stable generative adversarial networks, Sep 2019.

[5] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[6] Zhaoqing Pan, Weijie Yu, Bosi Wang, Haoran Xie, Victor S. Sheng, Jianjun Lei, and Sam Kwong. Loss functions of generative adversarial networks (gans): Opportunities and challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(4):500–522, 2020.