

Mask inpainting with a GAN network

Luca Lumetti

244577@studenti.unimore.it

Federico Silvestri

243938@studenti.unimore.it

Matteo Di Bartolomeo

241469@studenti.unimore.it

Abstract

Our project aims to remove a face mask over a person's face, by reconstructing the covered part of the face. To have a more precise reconstruction of the missing parts (mouth and nose) behind the mask, we plan to use a second photo of the same person without the mask as a reference during the facial reconstruction process. There are no constraints on the quality of the reference photo, for instance the face can be taken from a different point of view than the first one. To sum up, given as input an image containing a person's face partially covered by a medical mask and another photo of the same person without any occlusions, the output will be the first image with the mask-covered parts, mouth and nose, reconstructed. Future development could lead to generalizing the occlusion caused by the mask to any type of occlusion possible.

References

- [1] Chu-Tak Li, Wan-Chi Siu, Zhi-Song Liu, Li-Wen Wang, and Daniel Pak-Kong Lun. Deepgin: Deep generative inpainting network for extreme image inpainting. In *European Conference on Computer Vision*, pages 5–22. Springer, 2020.
- [2] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471–4480, 2019.

1. Image inpainting

Image inpainting (a.k.a. image completion) is the task to fill a missing region in an image by predicting the value of the missing pixels in order to have a realistic image which is semantically close to the original one.

1.1. Datasets

GAN networks are data-hungry and needs a lot of diverse training examples in order to generate quality images, for this reason we used the FFHQ 1024x1024 images, rescaled to 256x256, during training. In other GAN inpainting architectures, the mask region to reconstruct is usually calculated during the training in a randomized way. We do not have this randomization process, so for each image of FFHQ we precalculated the face region where the mask is weared using facial landmarks. For testing we used CelebA256.

1.2. Architecture

Our architecture is higly inspired by Free Form Image Inpainting with Gated Convolution [?] and DeepGIN [?].