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.

1. Mask Segmentation

We made use of MediaPipe's FaceMesh [2] library to find facial landmarks over the face covered with the surgical mask and the reference photo. Facial landmarks are important to have an initial approximation of the region where to search the surgical mask and to warp the reference photo over the first one. To perform the segmentation of the mask we apply a k-means with k=3 over the polygon we created using face landmarks and pick the bigger region between the 3. The k has been choosen to be 3 as in the polygonal region we expect to find the mask, the background and the skin of the person's face. In the end, a binary image is created, with a 1 where the mask is present and 0 elsewhere, while in the original image, the mask area is filled with 0s.

2. Warping the reference photo

The objective of the reference photo is to guide the network to a more loyal reconstruction. As we allow the reference to have [avere un'angolazione diversa da quella frontale], we apply a thin-plate spline transformation to make it frontal [meh che traduzione brutta]. We use 30 specific landmarks as parameters as using more parameters lead to distortion given by the error in the landmarks detection and less lead to an imperfect warping. The same polygon region of Mask Segmentation is cut from the reference photo, the by applying the TPS is sticked to to main photo leading to a (partial) reconstruction.

3. 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.

3.1. 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 [1], rescaled to 256x256, during training. In other GAN inpainting architectures, the mask region to reconstruct is usually calcuated 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.

3.2. Architecture

Our architecture is higly inspired by Free Form Image Inpainting with Gated Convolution [4] and DeepGIN [3]. Like this two papers also us use a coarse to refine techniques. For this reason, our generator is composed of two different nets: a coarse net that fill the masked part with a rough image and a refine net that gets better the output image.

3.2.1 Coarse Net

3.2.2 Refine Net

References

[1] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[2] Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. Real-time facial surface geometry from monocular video on mobile gpus. *CoRR*, abs/1907.06724, 2019.

[3] 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.

[4] 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.