Review

GAN-based Face Mask Removal using Facial Landmarks and Pixel Errors in Masked Region

published on 2022.

This paper propose to use generative adversarial networks (GAN) to complete the facial region that hidden by the mask . which is an important cue for communication facilitation, in human face image.

First of all the paper mentioned related works : Inpainting which is a technique used to repair scratches and holes in a part of an image , (Xie et al., 2012)by using a denoising auto encoder, and DeepFill (Yu et al., 2018; Yu et al., 2019), generative adversarial networks . On these methods when there is a huge hole concentrated in one place in the image, such as mask, they may fail.

So they propose to defined custom loss functions that focus on the errors of the feature point coordinates of the face and the pixels in the masked region .By using pix2pix as a baseline which is a GAN that learns the correspondence between a pair of images and given one image, generates the other corresponding image. Next they add two new custom terms to the loss function formula of the generator , mean of the squared error of the pixel values limited to the masked region, and the mean of the error of the feature point coordinates predicted from the face contour. Using dlib library to obtain the feature point coordinates.

The dataset was created by them using MaskTheFace to put the mask on the real faces images .1400 images from Flicker-Faces-HQ (FFHQ) Dataset ,588 from the Karolinska Directed Emotional Faces (KDEF) Dataset , 343 from UTKFace Dataset .

For quantitative metrics they use mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS) . after do training for 300 epochs, SSIM with **0.926** and LPIPS with **0.0459** give the best values . they also use human-rated quality score (HQS) gives **4.78±1.08** on the qualitative evaluation experiment , and accuracy **94.00±5.16**.

The advantage in ths paper that they designed facial expression identification with range of accuracy between (75-100)% for different expressions.

The disadvantage is the result may get worse if the given face image is wearing a mask different from the one used in the training dataset.