Local Facial Attribute Transfer through Inpainting

Ricard Durall[1,2,3], Franz-Josef Pfreundt[1] and Janis Keuper[1,4]

[1]Fraunhofer ITWM - [2]IWR, University of Heidelberg - [3]Fraunhofer Center Machine Learning -
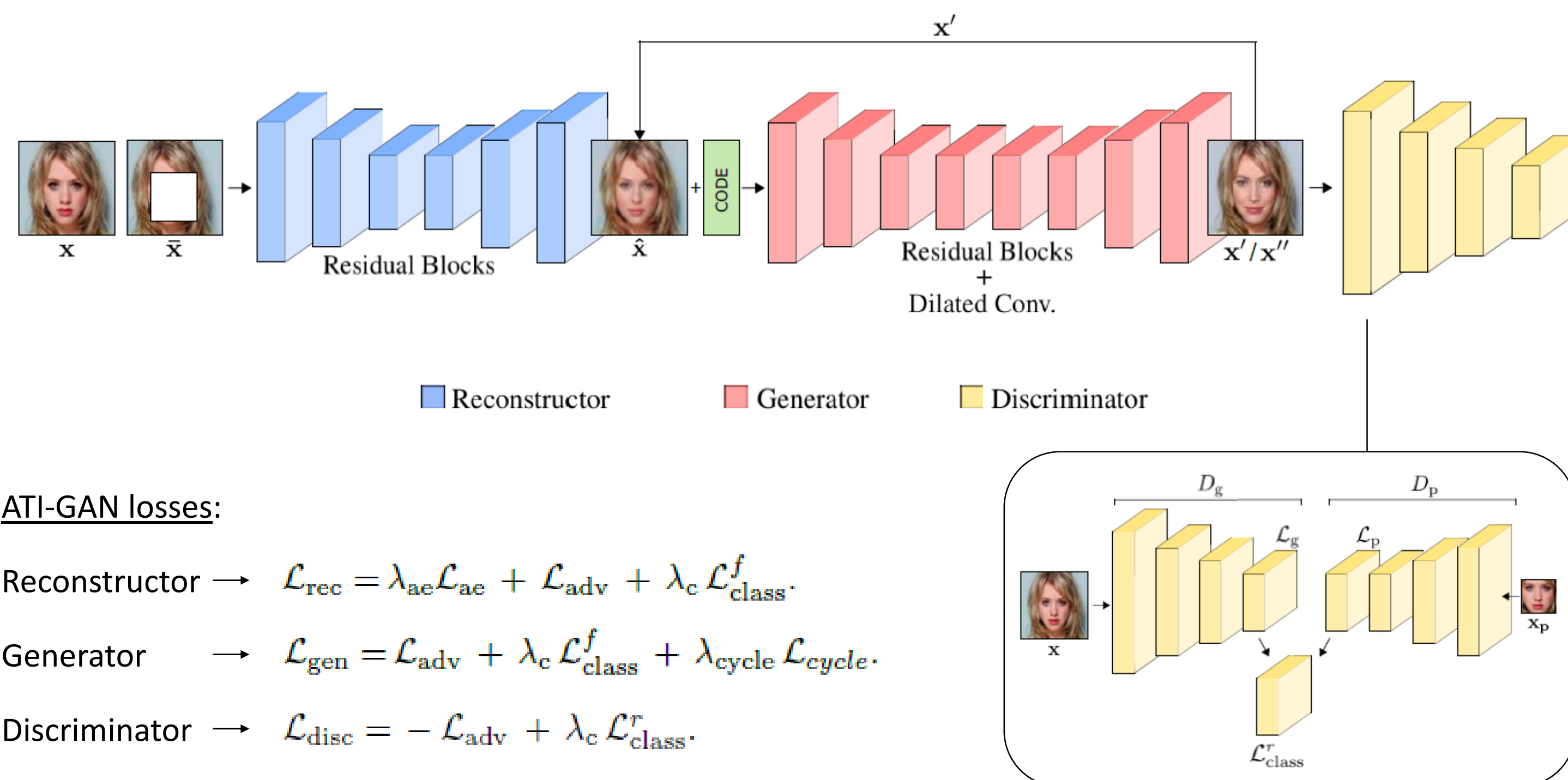[4]IMLA, Offenburg University

Abstract

The term "attribute transfer" refers to the tasks of altering images in such a way, that the semantic interpretation of a given input image is shifted towards an intended direction, which is quantified by semantic attributes.

In this paper, we present a novel method called ATI-GAN (Attribute Transfer Inpainting Generative Adversarial Network) that modifies only parts of a face to achieve semantic changes (e.g. removing a mustache).
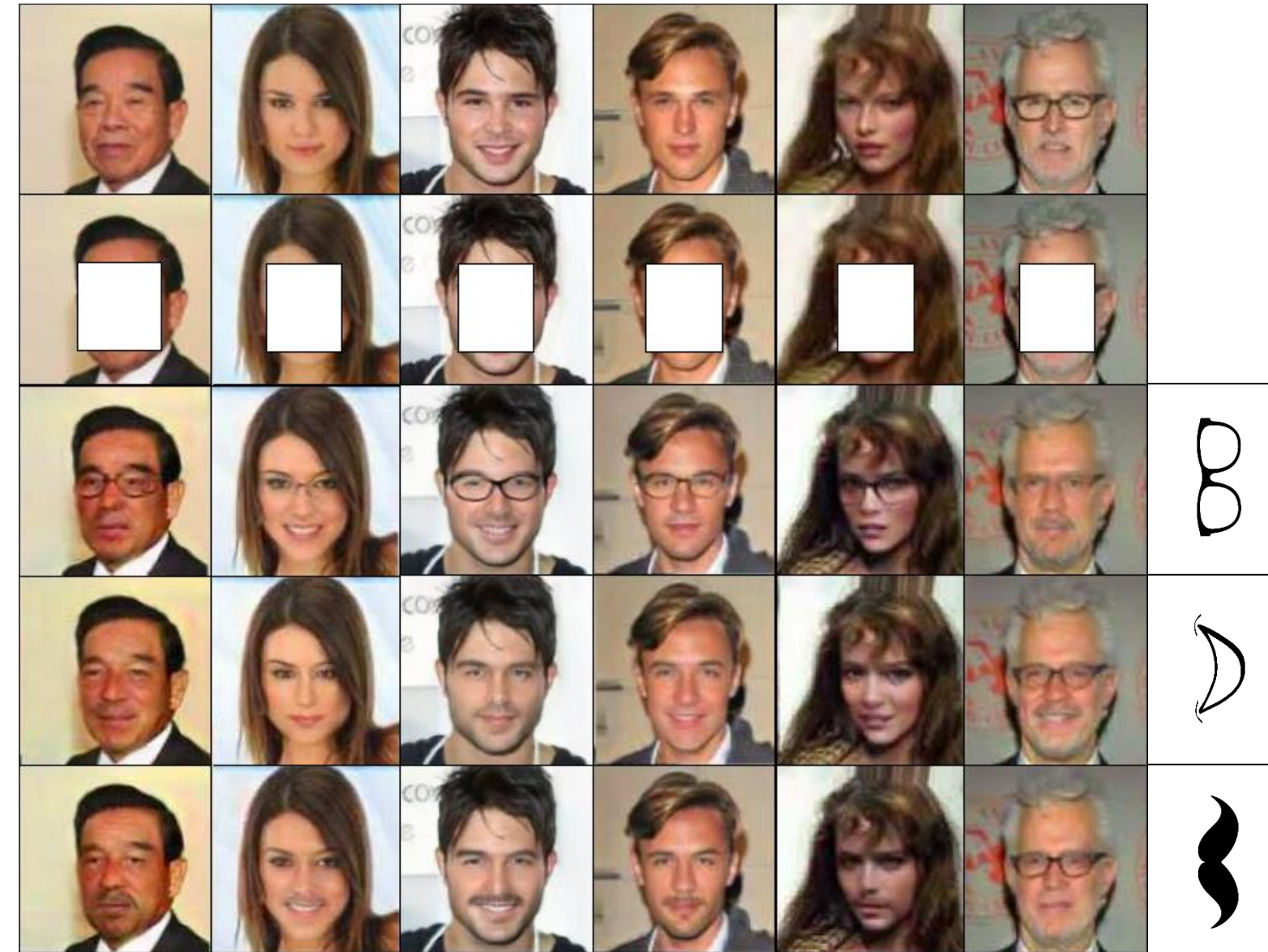
Architecture

The network architecture of our proposal is separated in an inpainting network (blue), generative network (red) and discriminative network (yellow). By combining each of these blocks in a sequential manner, the ensemble model is able to perform an end-to-end attribute transfer.

ATI-GAN losses:

Reconstructor $\rightarrow$ $\mathcal{L}_{rec} = \lambda_{ae}\mathcal{L}_{ae} + \mathcal{L}_{adv} + \lambda_c \mathcal{L}_{class}^f$.

Generator $\rightarrow$ $\mathcal{L}_{gen} = \mathcal{L}_{adv} + \lambda_c \mathcal{L}_{class}^f + \lambda_{cycle} \mathcal{L}_{cycle}$.

Discriminator $\rightarrow$ $\mathcal{L}_{disc} = -\mathcal{L}_{adv} + \lambda_c \mathcal{L}_{class}^r$.

Contribution

The contribution of this paper is a novel attribute transfer approach that alters given natural images in such a way, that the output image meets the pre-defined visual attributes. To do so, our proposed architecture integrates an **inpainting block**. This block allows to focus all the attention into the **region of interest** (where the attributes are located) while keeping the background unmodified but still consistent with the changes. In particular, we take advantage of the fact that most facial attributes are induced by **local structures** (e.g. relative position between eyes and ears).
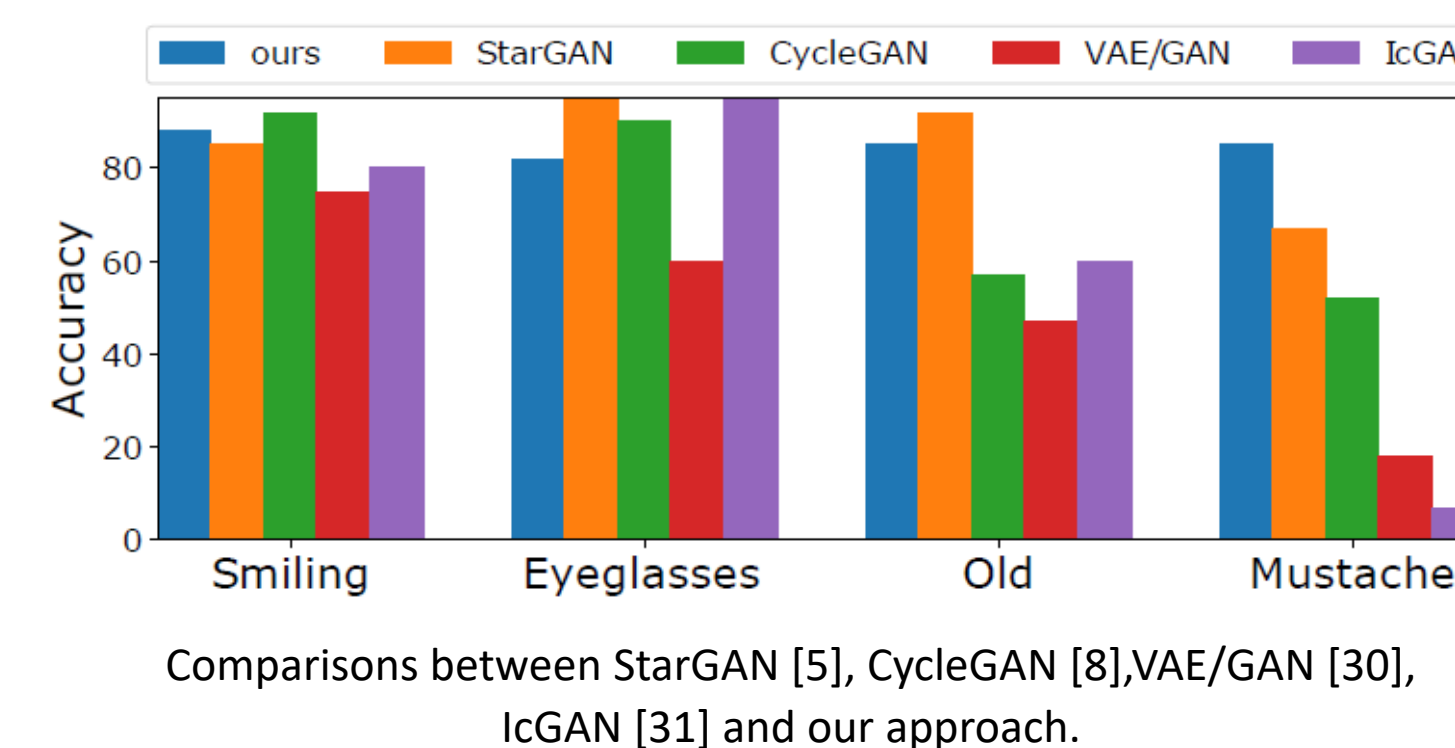
Qualitative Results

We train ATI-GAN on the CelebA dataset which consists of ~200K celebrity face images with variations in facial attributes.

- Evaluation of inpainting results:

Method	PSNR (dB)	SSIM
SIIWGAN [27]	19.20	0.920
SIIDGM [10]	19.40	0.907
CE [19]	21.30	0.923
GL [4]	23.19	0.936
GntInp[12]	23.80	0.940
GMCNN [28]	24.46	0.944
GL+LID[29]	25.56	**0.953**
ours	**31.80**	0.946

Comparisons in terms of PSNR and SSIM metrics
(Higher values are better).

- Evaluation of attribute transfer results:

Comparisons between StarGAN [5], CycleGAN [8],VAE/GAN [30],
IcGAN [31] and our approach.

*** Citations can be found in the paper.*

Quantitative Results

Smiling transformation.
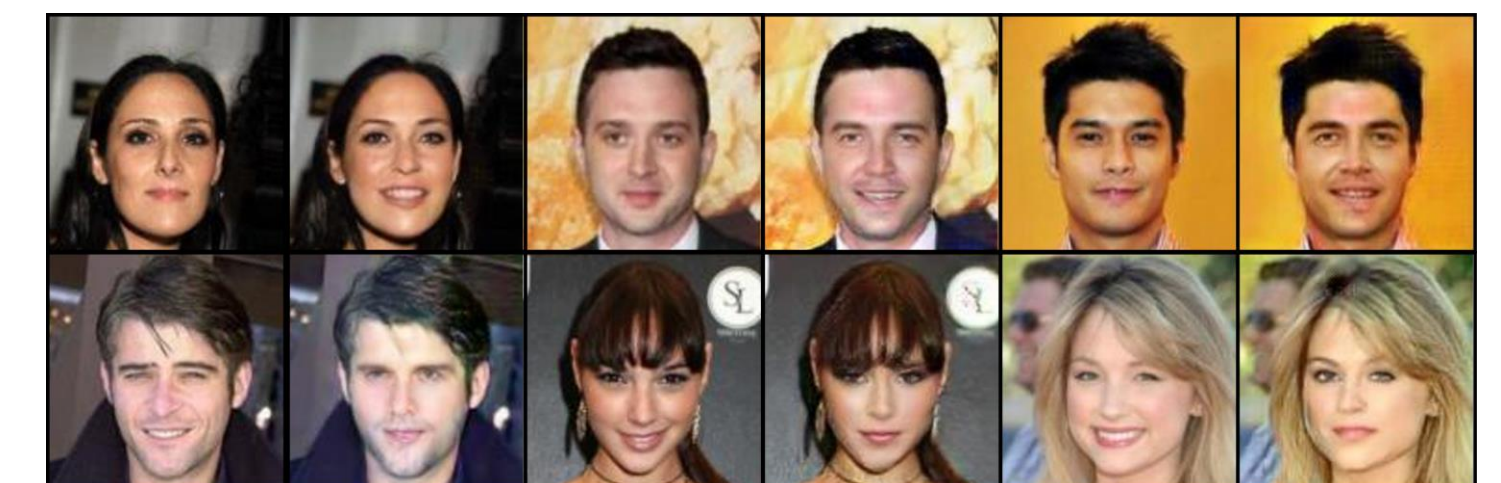
Age transformation.

Mustache transformation.

Eyeglasses transformation.