A Method for Modifying Facial Features

Jing Bo Yang, Boning Zheng, Meixian Zhu

jingboy, b7zheng, mxzhu @stanford.edu

CS229 Final Project, Department of Computer Science, Stanford University

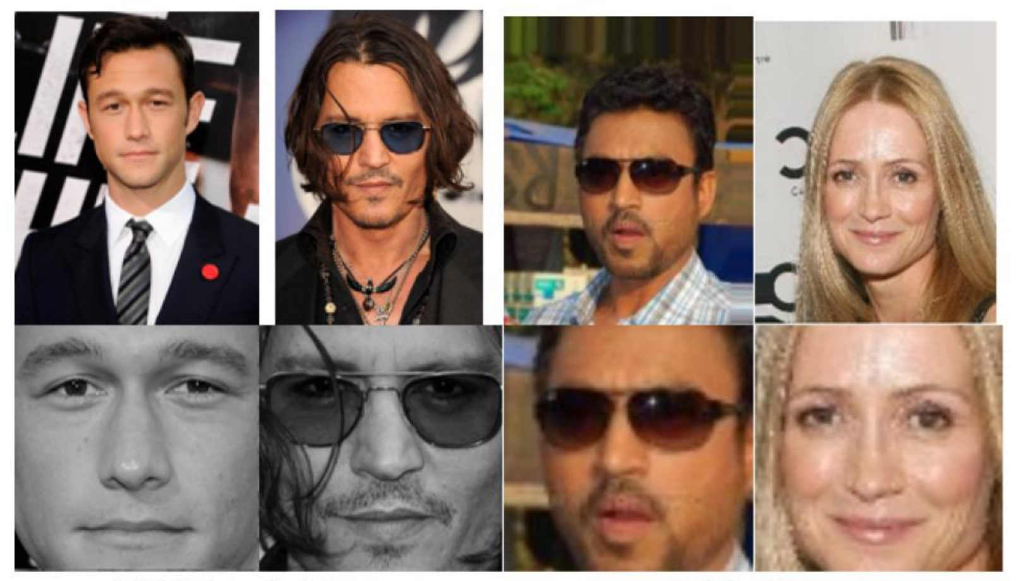
Stanford
Computer Science

Motivation

Facial recognition systems rely on original faces, but people's facial features, including beard and glasses, change frequently. A system capable of recovering the original human face or reconstructing disguise will be helpful for officers who need to manually verify ID photos and for assisting witnesses identify criminals with modified facial features.

Dataset

This project primarily uses two datasets: the dataset obtained by Wang and Kumar and the popular CelebA dataset.

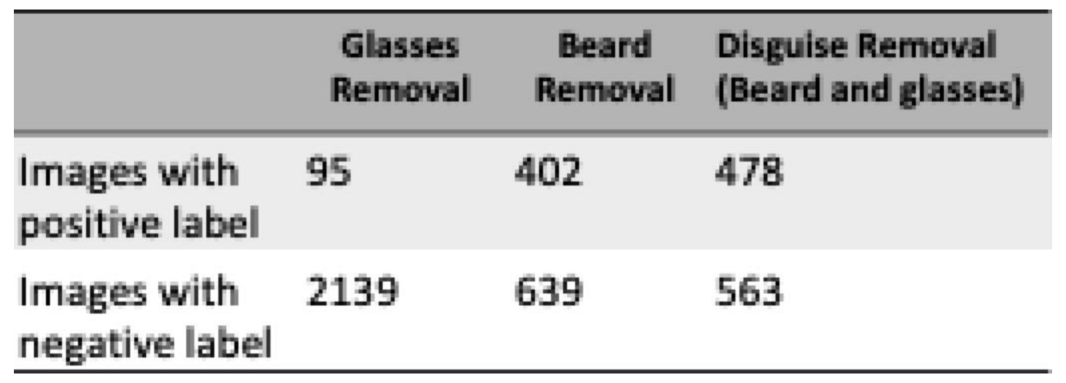


HK Polytechnic Dataset
Faces pre-processed by dataset provider

CelebA Dataset
Faces cropped using OpenCV

Fig 1. Two datasets used for this project.

Wang and Kumar's dataset consists of 2460 images of 410 different celebrities. The images are nicely aligned and cropped facial images pre-processed into gray scale along with multiple annotations.



	Glasses Removal	Beard Removal	Disguise Removal (Beard and glasses)
Images with positive label	95	402	478
Images with negative label	2139	639	563

The CelebA dataset is a large-scale face attributes dataset with more than 200K celebrity images. Approximately 10K images with appropriate beard/glasses tags were selected for this project. We used OpenCV to identify and then crop facial image to achieve similar samples as previous.

Methodology

For this task, we would like to learn mapping functions between two domains X and Y (original and disguised faces) given training samples $\{x_i\}_{i=1}^N$ and $\{y_j\}_{j=1}^M$.
The Cycle-GAN by Zhu et.al. [1] presents a method of learning the two mappings simultaneously using a forward GAN and backward GAN. The network architecture is presented below:

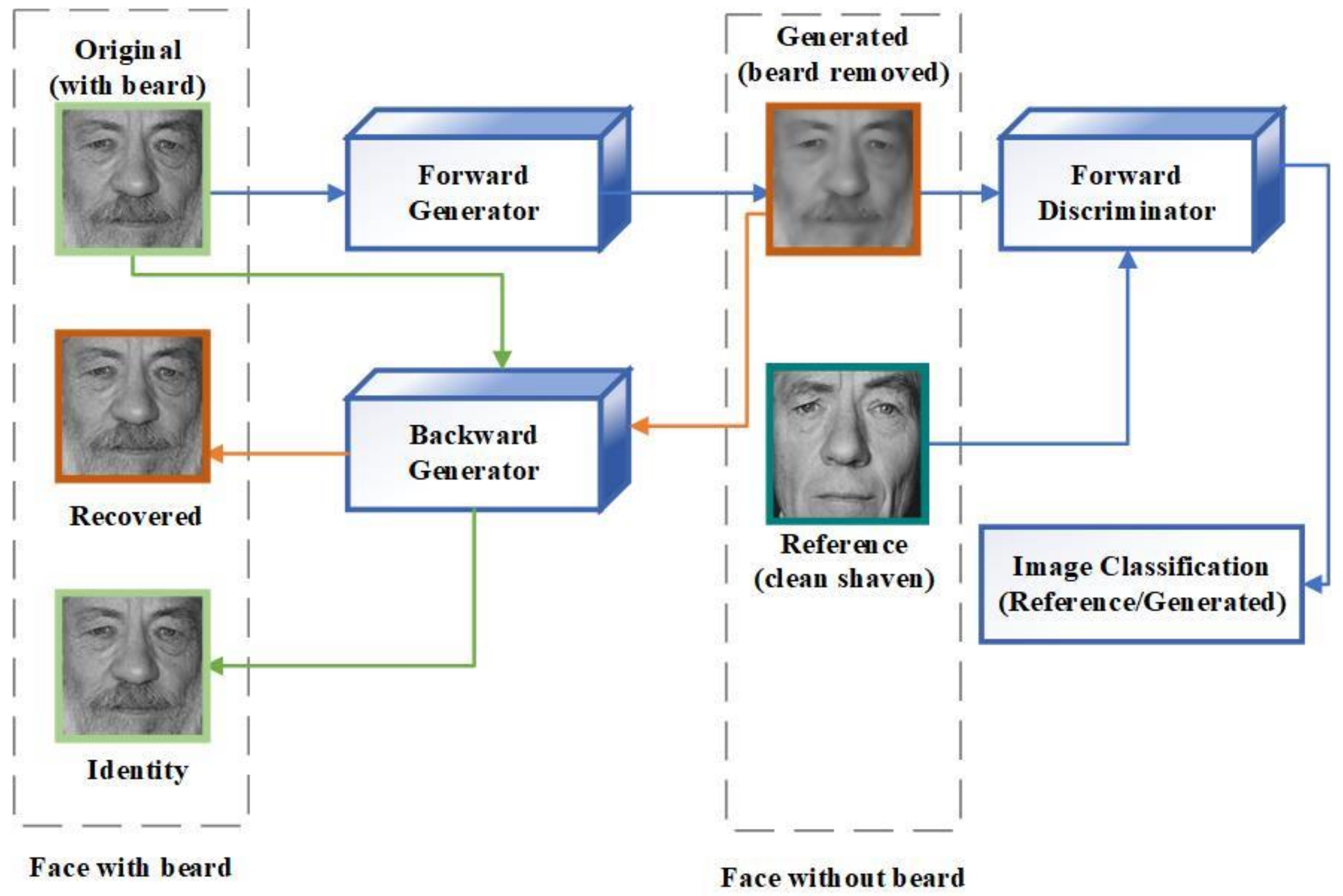


Fig 2. First half of CycleGAN. Forward generator constructs desired images while backward generator is trained for preserving the original image.

The forward generator G maps disguised faces to original faces, whereas the backward generator F maps original faces back to disguised faces.
We apply adversarial loss functions to both GAN's:

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = {} & \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_Y(y)] \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))],\end{aligned}$$

In addition to the adversarial loss functions, we have an additional cycle-consistency loss to preserve the individual identities through the generation process:

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = {} & \mathbb{E}_{x\sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_2] \\ & + \mathbb{E}_{y\sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_2]\end{aligned}$$

Such that our full objective would be:

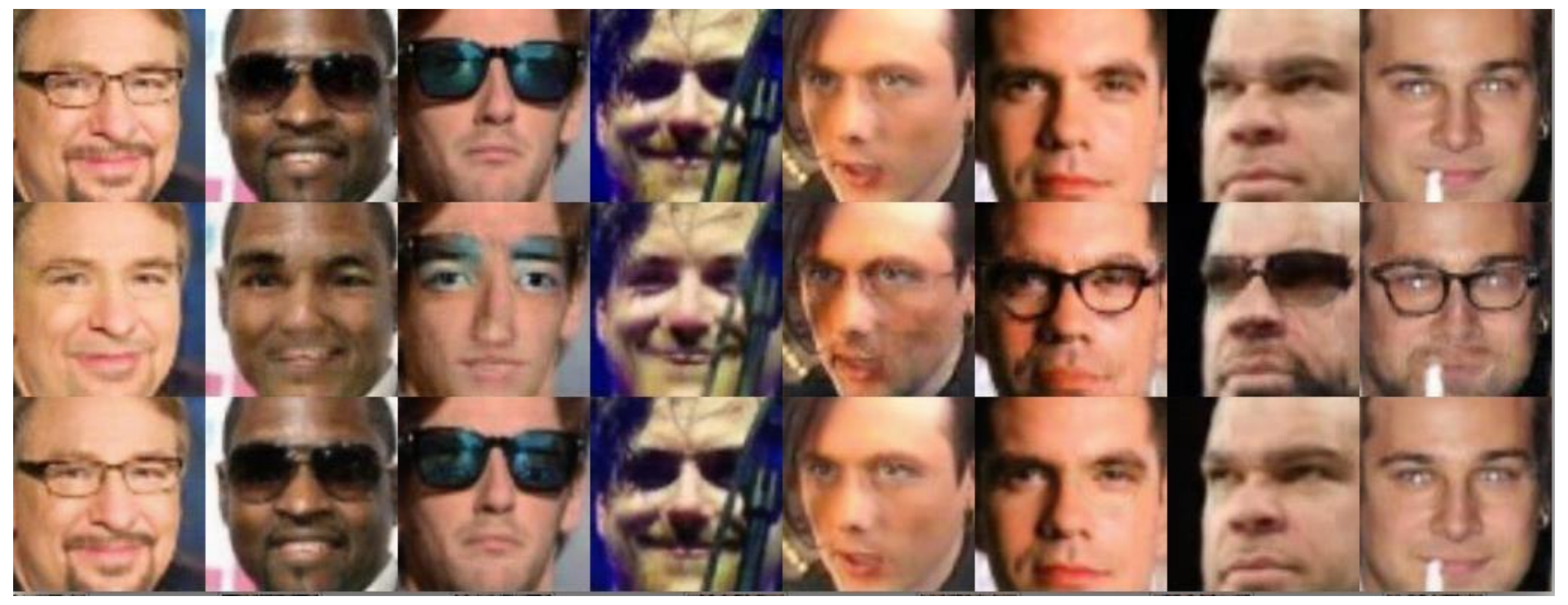
$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = {} & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ & + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ & + \lambda\mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

where λ is a hyperparameter that controls the relative importance of the two objective losses.

Results and Discussion

Left: beard and glasses removal

Right: beard and glasses reconstruction



Bottom row: Reconstructed images

Fig 3. Images generated using CycleGAN

Images generated using a basic GAN by Goodfellow [3] performs poorly in terms of preserving non-relevant facial features. Generated images are also blurred, potentially due to complex residual structure. In contrast, reconstruction losses and identity losses encourages CycleGAN to preserve features not affected by our manipulation. Decreasing combined GAN losses when individual component losses have plateaued could mean the network is refining image details. In addition, notice that reconstruction quality is higher for single feature difference training (higher quality beard/mustache for beard-only versus glasses and beard).
Added glasses depend on "type" of human. Old celebrities tend to get classic looking glasses whereas young celebrities often get sunglasses. We also want to point out that addition of features seems to be easier, likely because it hides details, compared to detail "creation" task of glasses and beard removal. We could attribute these details to the network making a choice based on characteristics of the training population.

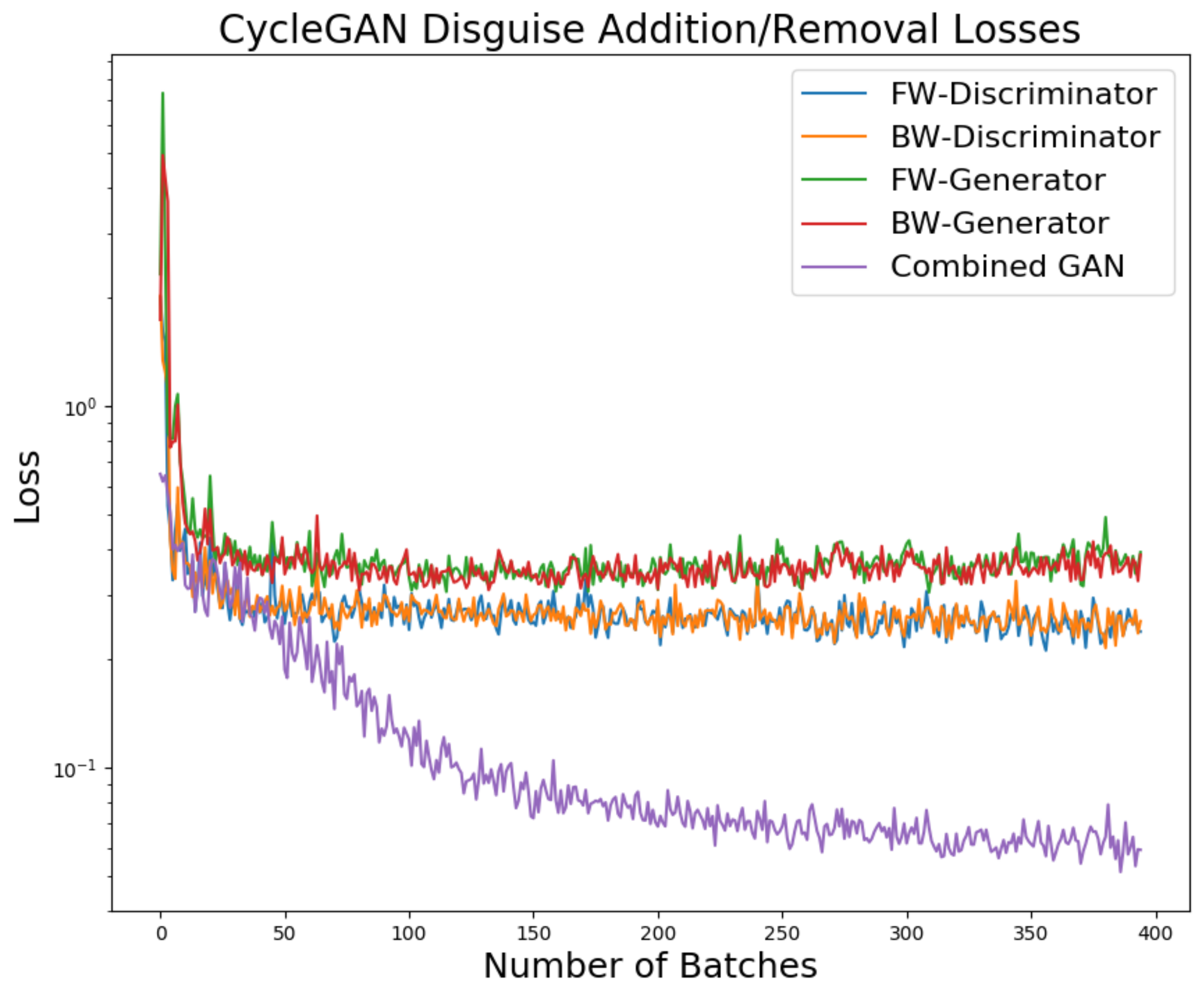


Fig 4. Losses of generator, discriminator and combined GAN for black and white beard removal/addition task.

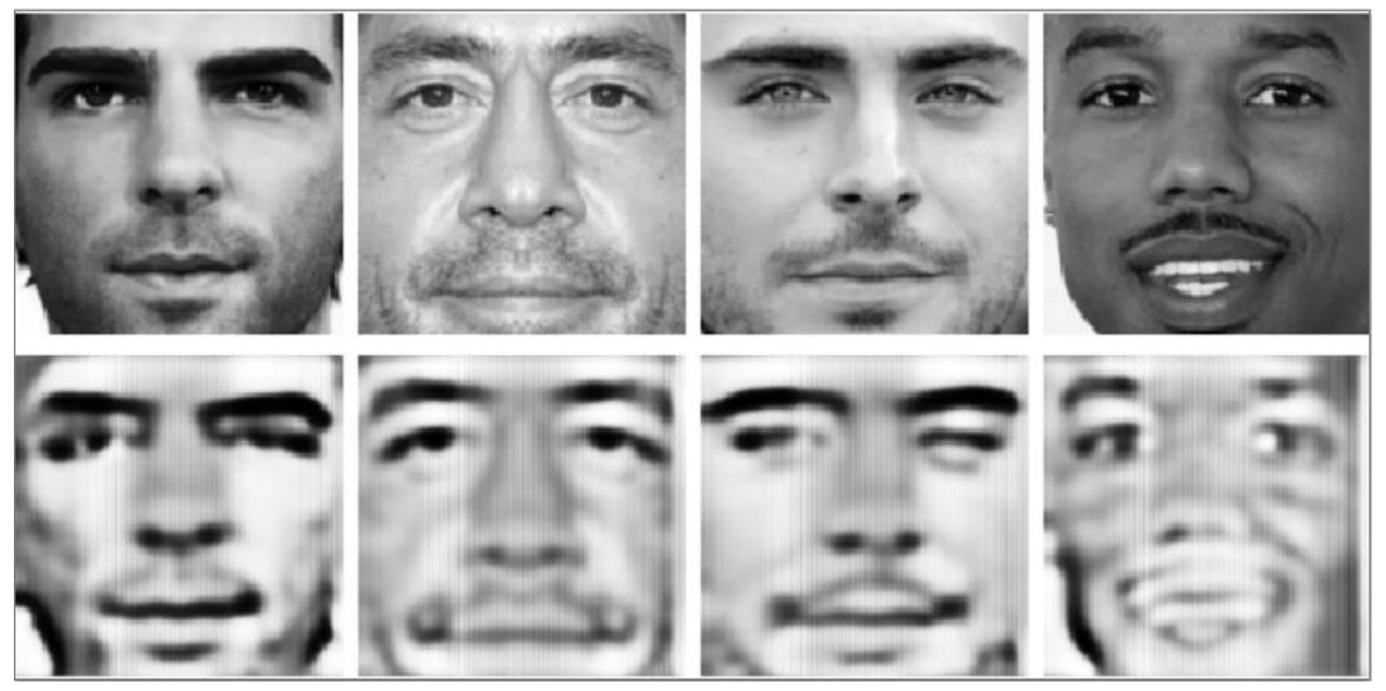


Fig 5. Images generated using a simple ResNet-based GAN.

Future Works

We can obtain higher image quality using more sophisticated network structure. There also exists methods that can numerically evaluate quality of generated images, such as *Inception Score.* It is also desirable *to* support more facial features. Potentially make use of semi-supervised or unsupervised methods to enable training with unlabeled data.

References

[1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *arXiv preprint* (2017).
[2] Hansen, Lars Kai, and Peter Salamon. "Neural network ensembles." *IEEE transactions on pattern analysis and machine intelligence* 12.10 (1990): 993-1001.
[3] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.