An Unsupervised Approach towards Varying Human Skin Tone Using Generative Adversarial Networks

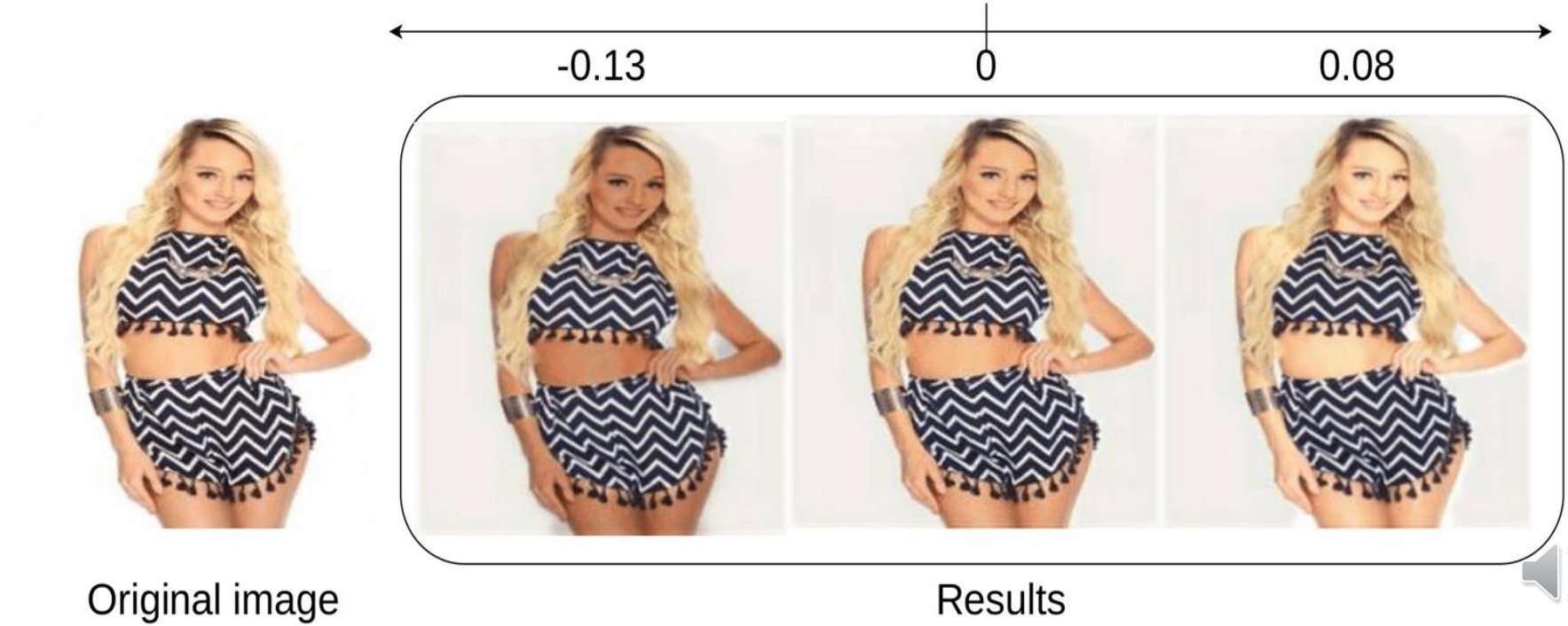
Debapriya Roy, Diganta Mukherjee, Bhabatosh Chanda

Indian Statistical Institute, Kolkata

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Objective

Synthesize images of a person over a varying scale of skin tone, where the tone can vary from dark to fair.



Contributions

- We trained a cGAN which takes the **image as input**, along with the value of a **conditional variable** and synthesizes **a new image** with the **skin tone of the persons in the image** changed in accordance with the **value of the variable**.
- Inspired from the concept of perceptual loss function **we propose a skin color distance based loss function** to train the cGAN.
- Our approach is **unsupervised** and **unconstrained** in terms of **pose**, **illumination**, **number of persons** in the image.

Proposed Method

- Skin segmentation
  - Objective -
    - Segmentation - skin, non-skin classification of given image
    - Skin color estimation - Estimating rgb value indicating skin tone
- Image synthesis
  - Objective -
    - Synthesizes images of varying human skin tone given a source image as input
    - Leverages the skin segmentation result from the previous stage
    - Employs the skin color estimation network as part of the loss function

Skin segmentation

- The skin segmentation network is a Convolutional Neural Network (CNN)
- Contains two subnetworks
- First subnetwork - skin, non-skin classification of image pixels.
- Loss functions - Count loss, VGG Perceptual loss, SSIM loss
  - Count loss -
    - measures the absolute difference between the counts of the skin pixels of the predicted and the ground truth image.
    - Observed to be effective in predicting binarized output

Input image

Segmentation results

Skin segmentation

- Second subnetwork - skin color estimation network
  - Training is unsupervised

Proposed Method

- Skin segmentation
  - Objective -
    - Segmentation - skin, non-skin classification of given image
    - Skin color estimation - Estimating rgb value indicating skin tone
- Image synthesis
  - Objective -
    - Synthesizes images of varying human skin tone given a source image as input
    - Leverages the skin segmentation result from the previous stage
    - Employs the skin color estimation network as part of the loss function

Image Synthesis

- We formulate the problem as a conditional image generation problem.
- We trained a **conditional GAN** (cGAN) with inputs **- source image**, **skin segmentation**, **control variable z**
  - z = 0 -> no change of skin color
  - z < 0 -> change towards darkness
  - z > 0 -> change towards fairness

Loss function for training the cGAN

$$L_{cGAN} = l^1 + l^2 + \lambda(m \times z + l^3 - \epsilon) + L_{ADV}.$$

Where considering $\hat{x}_{z=0} = f_g(x, z = 0, \hat{x}_{seg})$ and $\hat{x}_{z\neq 0} = f_g(x, z \neq 0, \hat{x}_{seg})$, we define,

$$l^1 = L_p(\hat{x}_{z=0}, x)$$

$$l^2 = L_p(\hat{x}_{z=0} \times \hat{x}'_{seg}, x \times \hat{x}'_{seg})$$

$$l^3 = \log(0.5 - L_{color})$$

$$L_{color} = L_p^{color}(\hat{x}_{z\neq 0}, x).$$

Generator network

VGG-perceptual loss

Similar to perceptual loss but the underlying network is the skin color estimation network

Quantitative Analysis

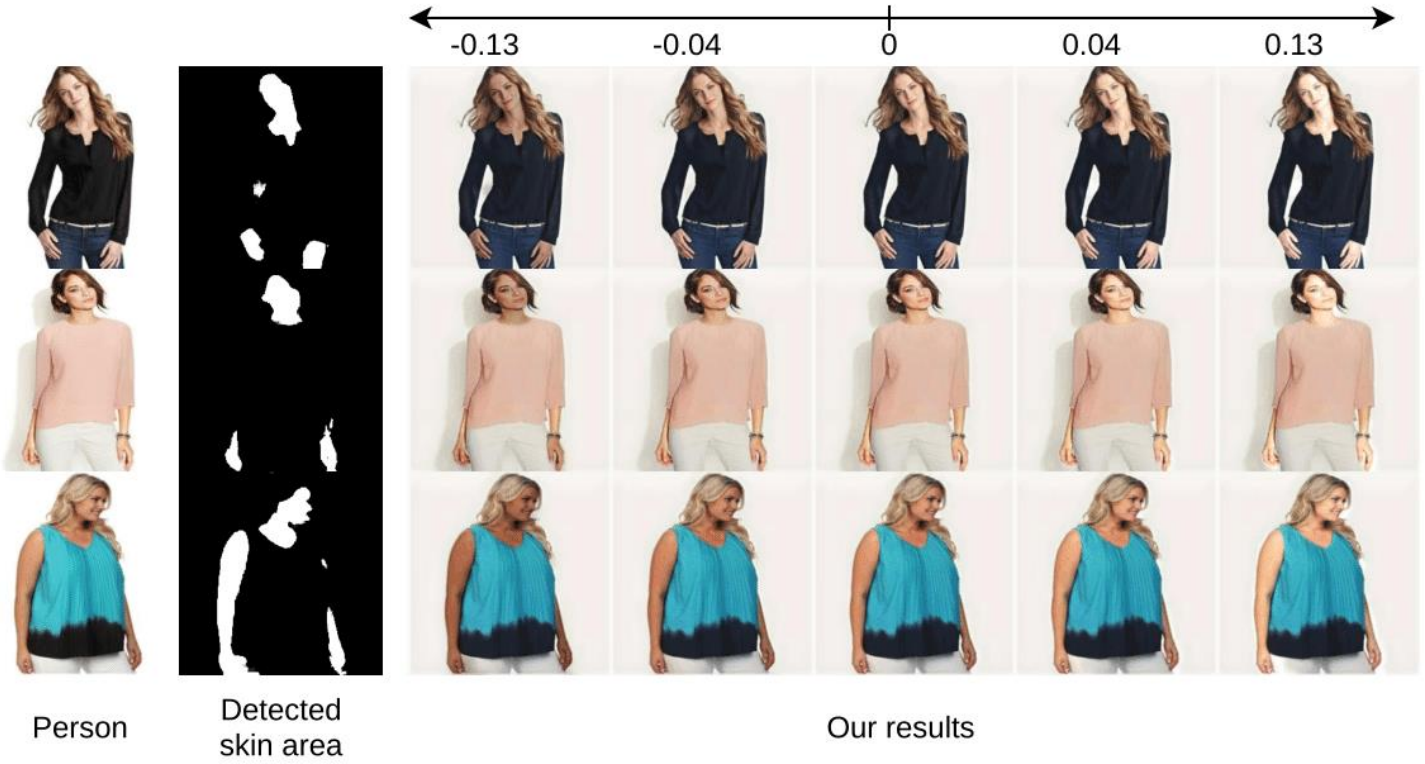
TABLE I: Values of Incepion Score (IS) and Frechet Inception Distance (FID) and SSIM on results of different data sets.

Dataset	IS↑	FID ↓	SSIM↑
In Shop	3.21 ± 0.17	38.33	0.93
Category-and-Attribute	3.58 ± 0.19	36.19	0.95
MPV	3.03 ± 0.23	42.56	0.92

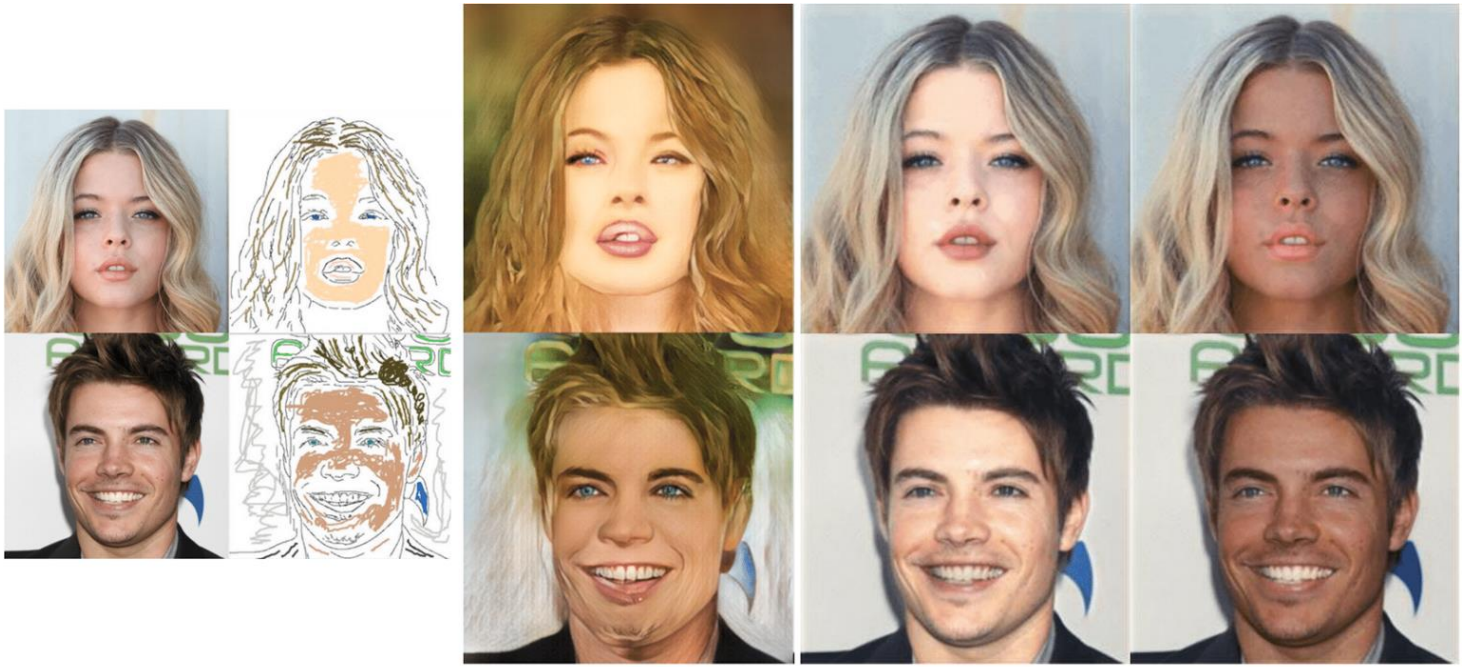
TABLE II: Values of Kolmogov-Smirnov test (KS test) statistic along with the corresponding P-values on results of different data sets.

Dataset	KS statistic ↓	P-Value↑
DeepFashion (Category-and-Attribute)	0.0249	0.5545
MPV	0.0450	0.0837

Results

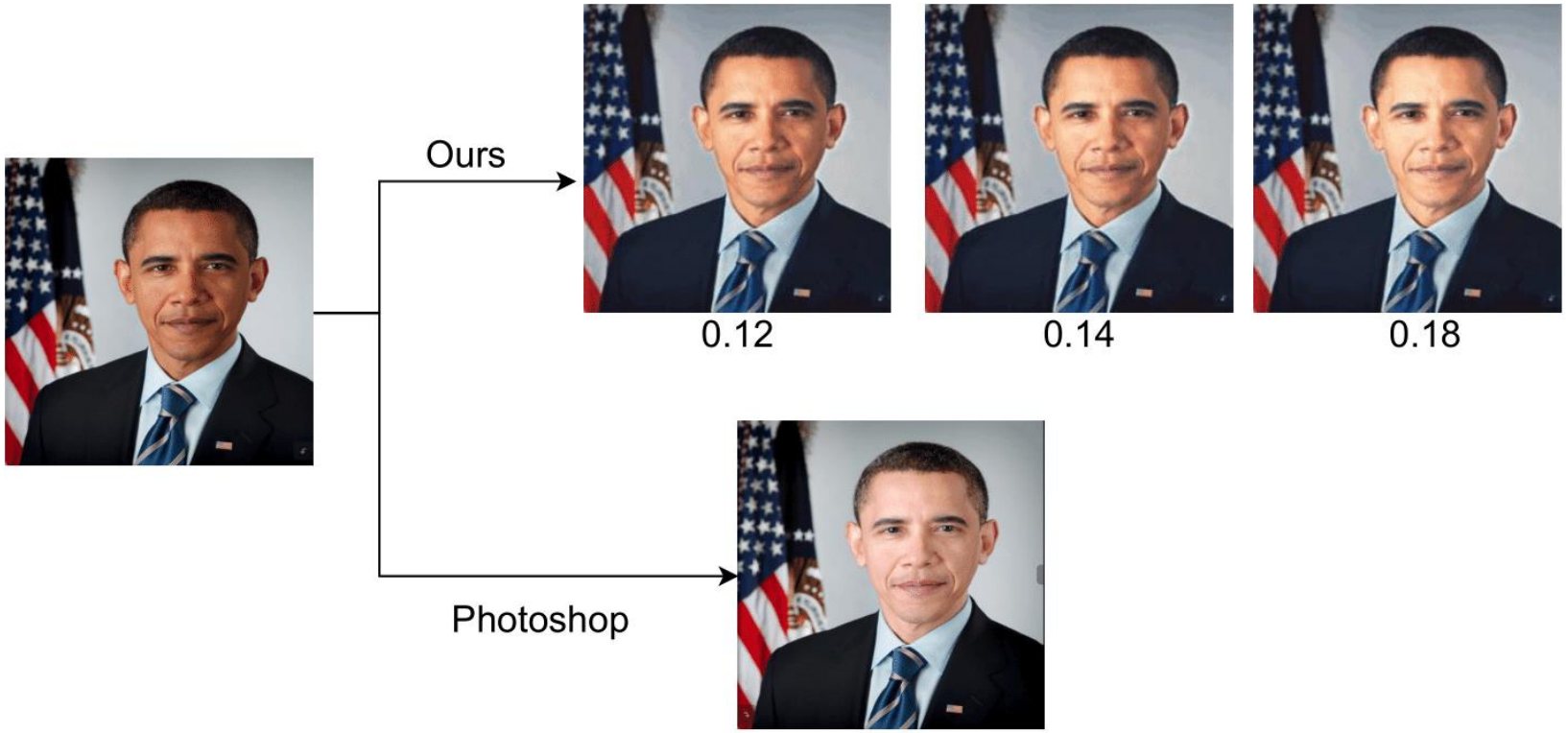


Comparison with SC-FEGAN



Person | Free-form Input | SC-FEGAN | -0.11 | Ours | 0.11

Comparison with the results of photo editor



Result on images with persons of different skin tones

Result on in-the-wild images

Source Image

Segmentation results

Result (-0.11)

Result (0.11)

Failure Case

Source Image

Segmentation result

Result (-0.13)

Thank You