Image-to-Image Attire Transfer for Virtual Trial Room

Ahsanullah University of Science and Technology

Presented By :

Sifat-Ul-Alam	16-01-04-041
Syed Sanzam	16-01-04-042
Sourav Govinda Das	16-01-04-043
Md. Faisal Ahmed	16-01-04-046

Supervised By :

Mr. Mohammad Imrul Jubair
Assistant Professor,
Department of Computer Science and Engineering,
Ahsanullah University of Science and Technology.

Why Fashion?

Artificial Intelligence in Fashion

- Fashion Embeddings
- Retail Insights
- Fashion Recommendations
- Fashion Object Detection
- Style Transfer

Inspiration

- CycleGAN

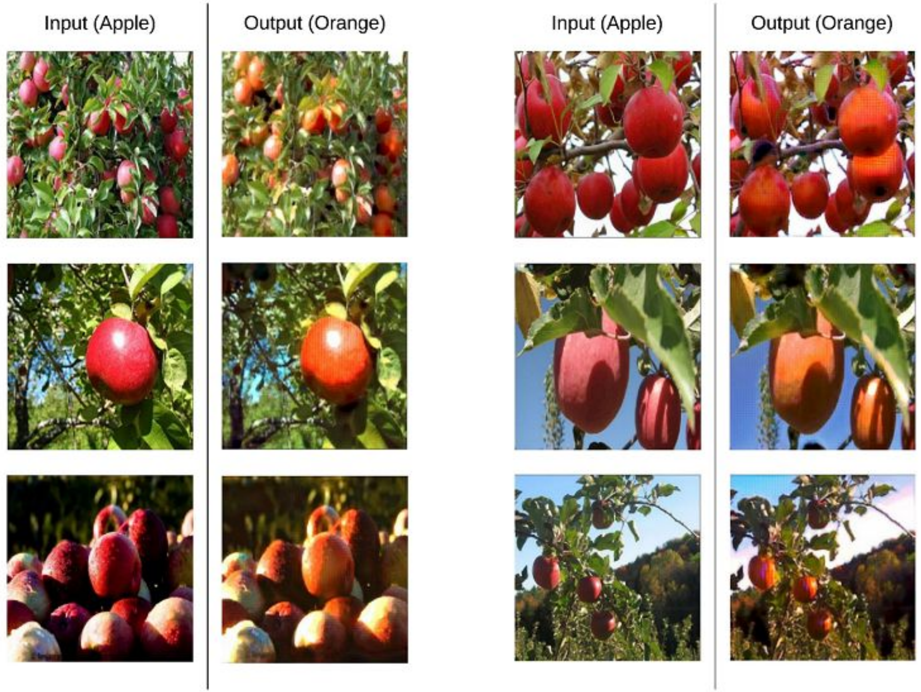
Inspiration

CycleGAN [1]

1. source

Inspiration

CycleGAN



1

1. source

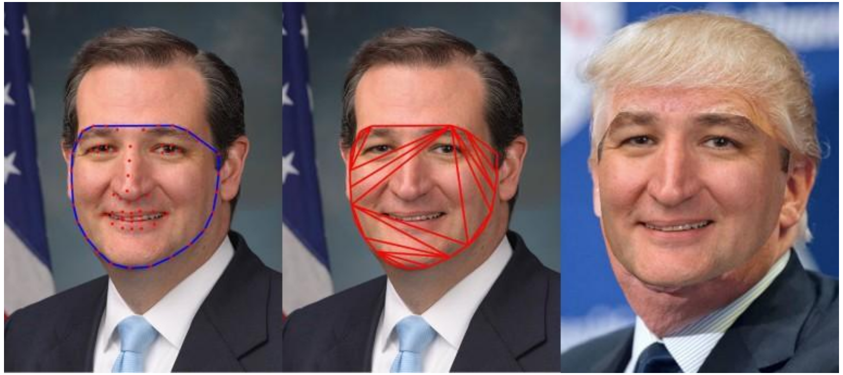
Inspiration

- CycleGAN
- FaceSwap

Inspiration

FaceSwap

1



1. source

Our Objective

Source Image

Person A

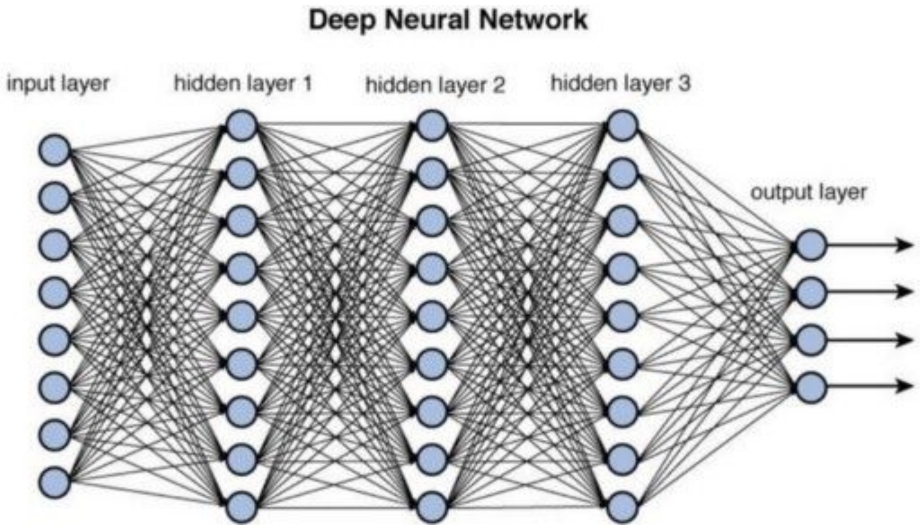
Target image

Person B

1. source

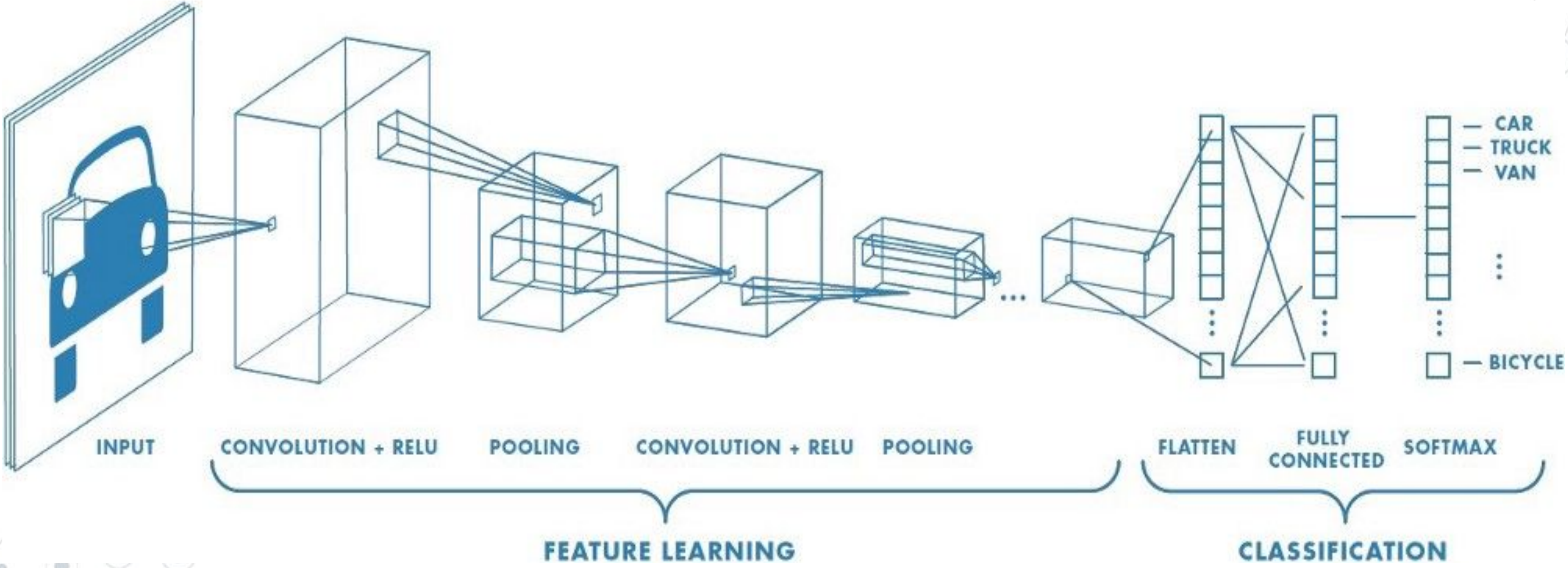
Literature Review - Neural Network

- Neural Network



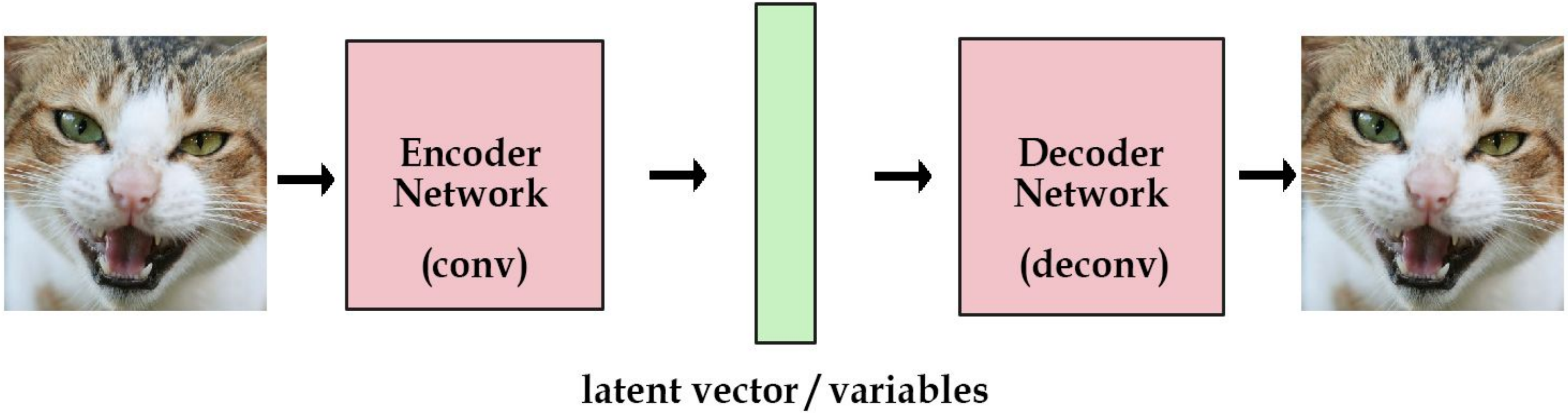
1. source

Literature Review - CNN



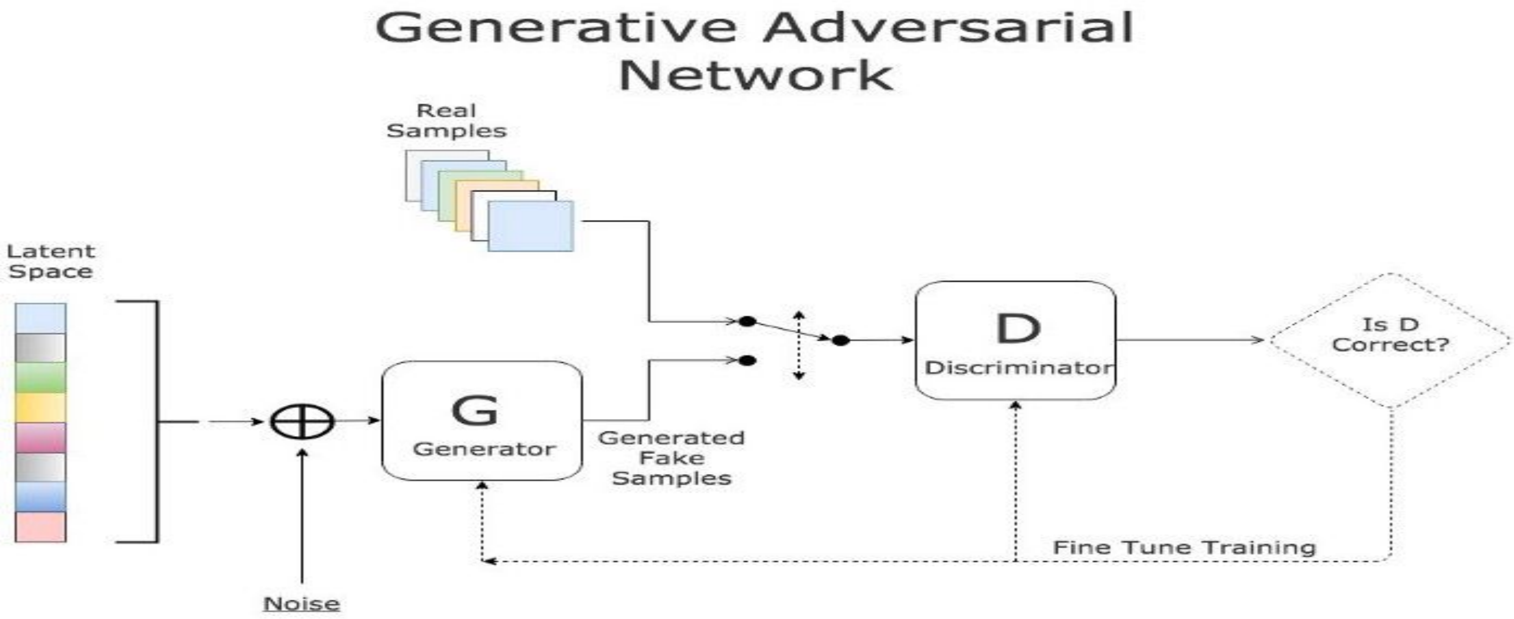
Literature Review - VAE

- Variational Autoencoders



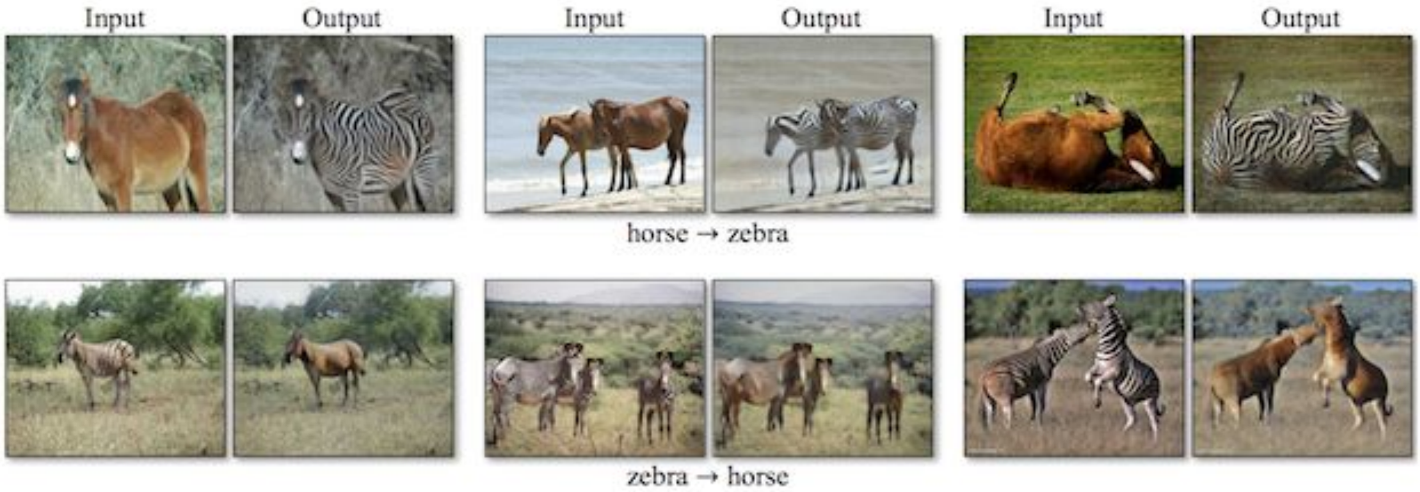
Literature Review - GAN

- Generative Adversarial Networks



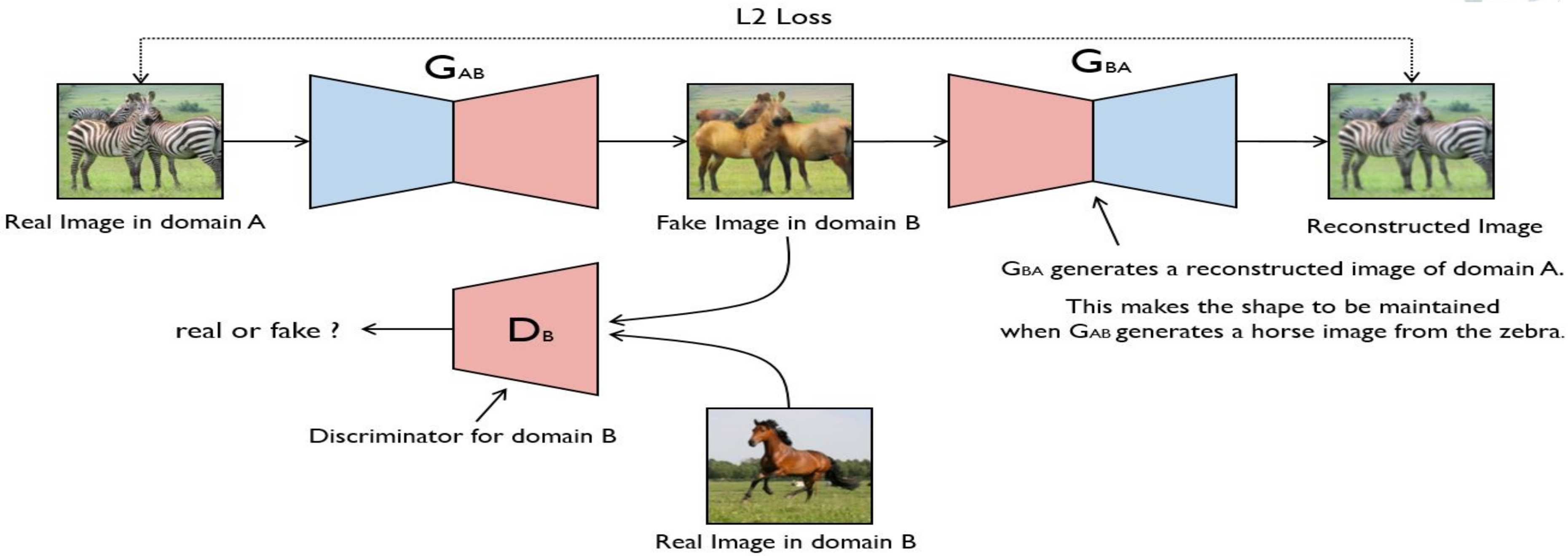
Literature Review - CycleGAN

- CycleGAN



Literature Review - CycleGAN

- CycleGAN (Contd.)



Literature Review - PixelDTGAN

- An image generation model that transfers an input domain to a target domain at a semantic level.
- A domain discriminator is used to achieve realistic images.
- The output is the clothing of a person.
- It could be used to solve the sub-problem of our goal.

Literature Review - PixelDTGAN

Fig: Output generated on the LookBook dataset presented by the author.

Literature Review - U-net with Grabcut

- Grabcut is used to separate objects from the backgrounds in a colored image with certain constraints.
- U-net is used for semantic segmentation.
- Combining both techniques, it’s possible to identify a person’s clothing in an image.

Literature Review - U-net with Grabcut

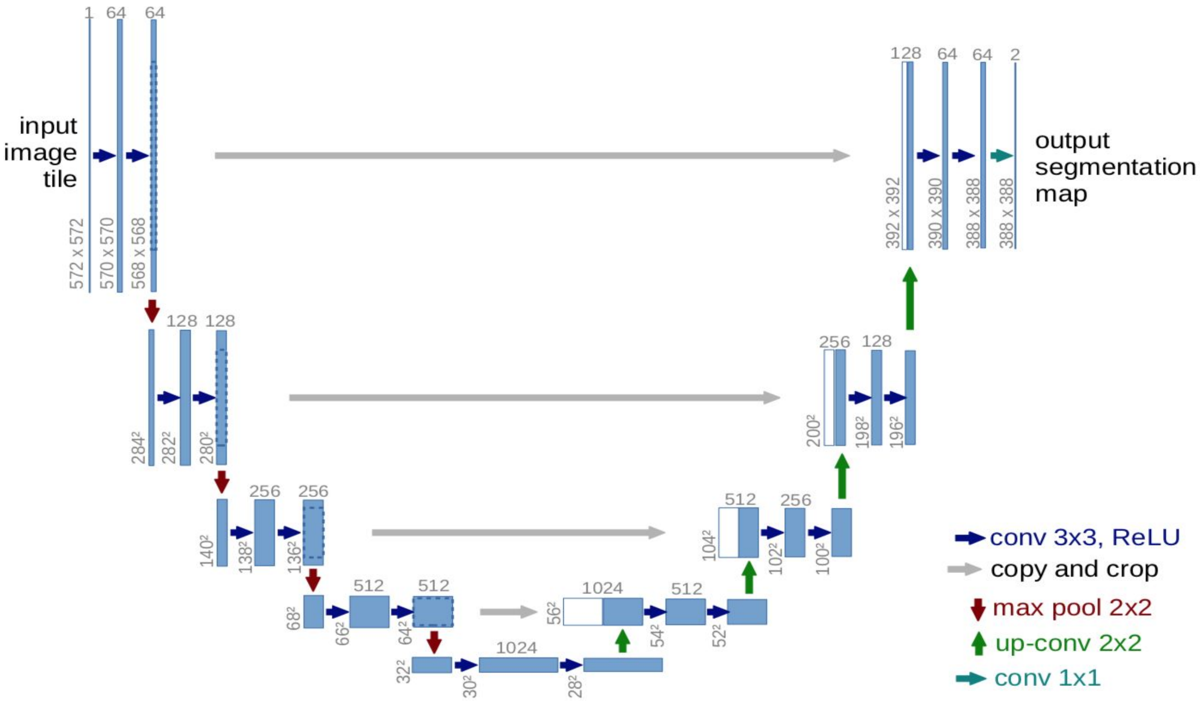


Fig: U-net architecture

Literature Review - U-net with Grabcut

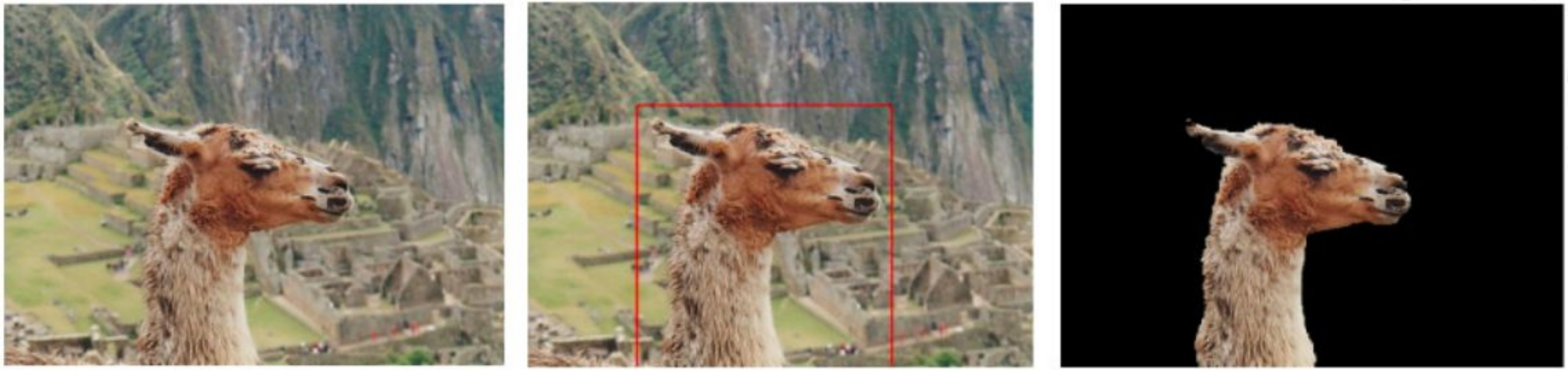


Fig: Foreground and background detection using Grabcut.

Literature Review - Viton and CP-Viton

- Viton is an image-based virtual try-on without using any 3D information.
- Clothing warping network can often generate highly distorted and misaligned clothes.
- CP-Viton improves the clothing warping stage.
- Both of the methods perform on dress to human transfer.

Literature Review - Viton and CP-Viton

Fig: Warping the cloth images to fit on the target image.

Literature Review - Liquid Warping GAN

- A framework that tackles human motion imitation, appearance transfer and novel view synthesis.
- For our study we are interested in ‘Appearance Transfer’.
- Contains Three stages-
  - Body Mesh Recovery
  - Flow Composition Module
  - Liquid Warping Block

Literature Review - Liquid Warping GAN

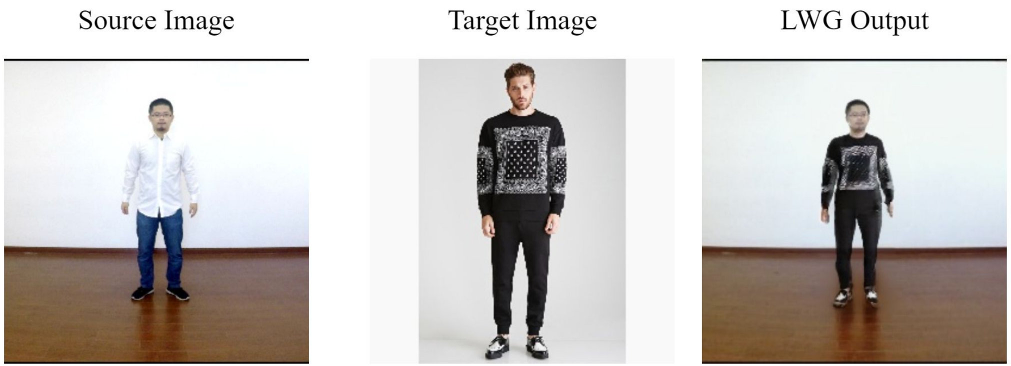


Fig: Output of the ‘Appearance Transfer’ task presented in the original paper and online.

Datasets - LookBook

- The dataset contains 84,748 images.
- 75,016 are of fashion models, rest are of top products.

Datasets - LookBook

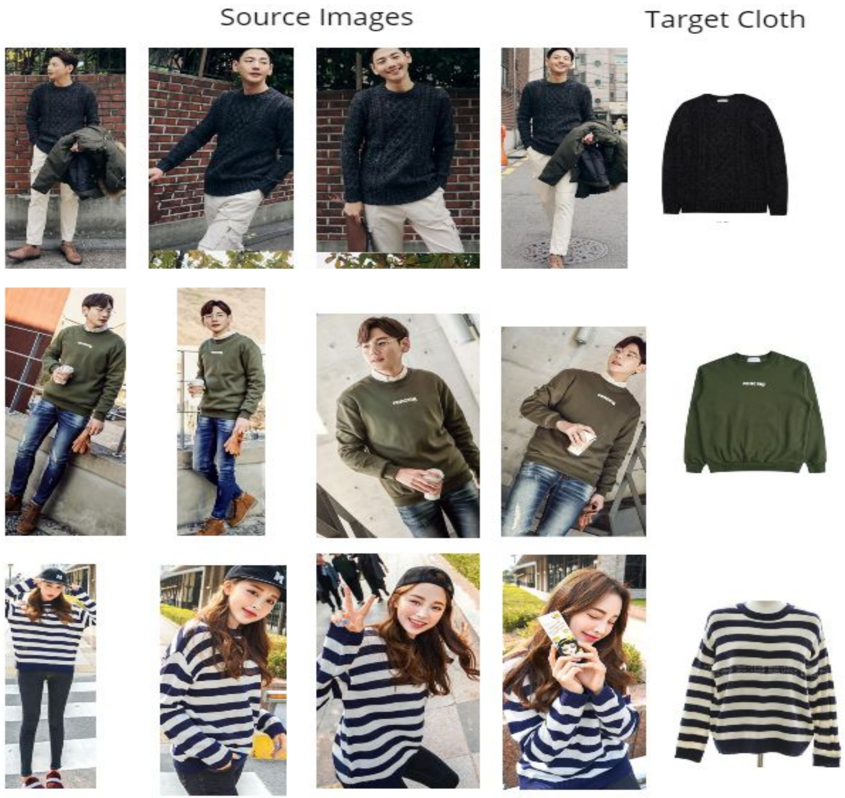


Fig: Images from the LookBook dataset.

Datasets - Impersonator (iper)

- The dataset contains 30 subjects of different conditions of shape, height, gender.
- 103 total clothes and 206 video sequences.

Datasets - Impersonator (iper)

Fig: Images from the Impersonator dataset.

Datasets - D-Cut-Dataset

- Developed by us to train the U-net model.
- Total of 150 images.
- Divided into 3 categories.

Datasets - D-Cut-Dataset

Fig: Images from the D-Cut-Dataset.

Our Journey-CycleGAN

The CycleGAN is a technique that involves the automatic training of image-to-image translation models

- Applied to LookBook Dataset
- Model is trained on 84784 images
- The output image is distorted and not satisfactory

CycleGAN Input/Output

Our Journey-FaceSwap

FaceSwap works on 68 landmark points on human faces and focuses on facial geometry.

- There is no fixed landmark points for human body shape
- The algorithm should not work on variety of clothes

Our Journey-PixelDTGAN

Used to transfer an input domain to a target domain in semantic level and generate the target image in pixel level.

- One Generator and two discriminators
- Works on LookBook dataset
- The generated results are distorted

PixelDTGAN Input/Output

Our Journey-Liquid Warping GAN

A framework focused on human motion imitation, appearance transfer, and novel view synthesis as the principal component

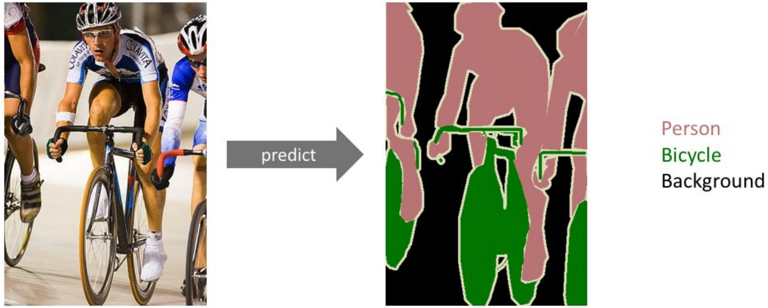
- Works with Impersonator Dataset
- Resulting images are comparatively better than other GAN-based approaches
- In most cases, output is distorted and human shape is not preserved

Liquid Warping GAN : Input/Output

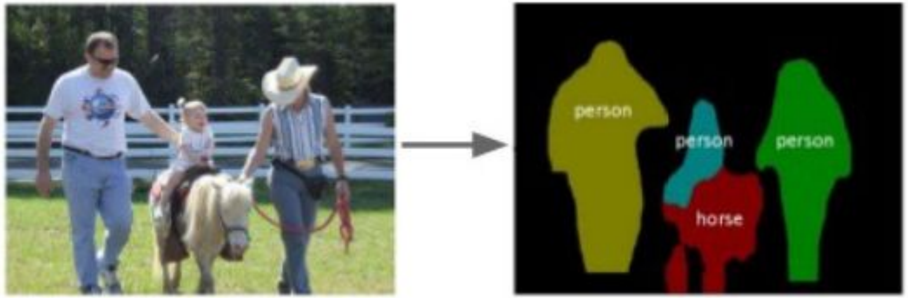
Our Journey-Image Segmentation

- Typically of two types. Semantic Segmentation and Instance Segmentation
- Semantic Segmentation labels each pixel of an image with a corresponding class.
- Instance Segmentation classifies each instance of a class separately.

Image Segmentation: Semantic vs Instance



1



2

Image Segmentation-Grabcut

It is used to separate Foreground and Background in various images.

Image Segmentation-Grabcut Input/Output

Input Output

Input Output

Image Segmentation-Grabcut + U-Net

U-Net is used for Semantic Image segmentation.

Image Segmentation-Grabcut + U-Net

U-Net is used for Instance Image segmentation.

- We created our own dataset namely D-Cut-Dataset
- This contains images of three category i.e ”Original”, ”Body” and ”Dress”
- The output is satisfactory

Image Segmentation: Grabcut + U-Net Input/Output

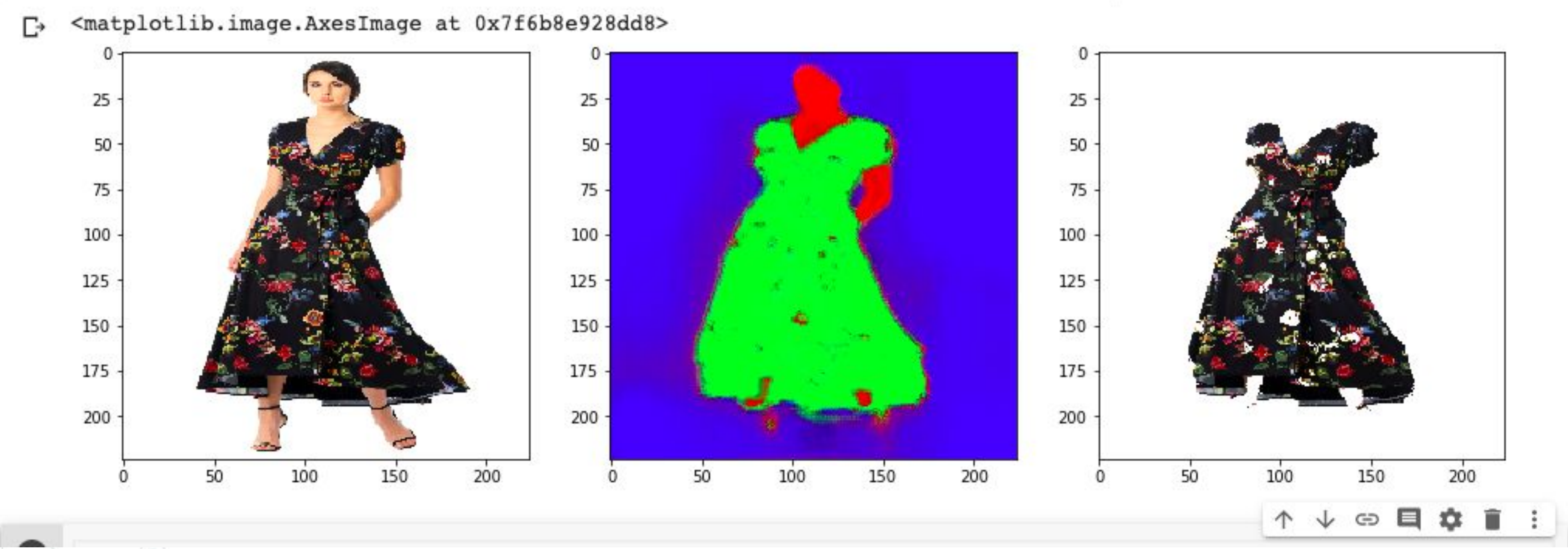
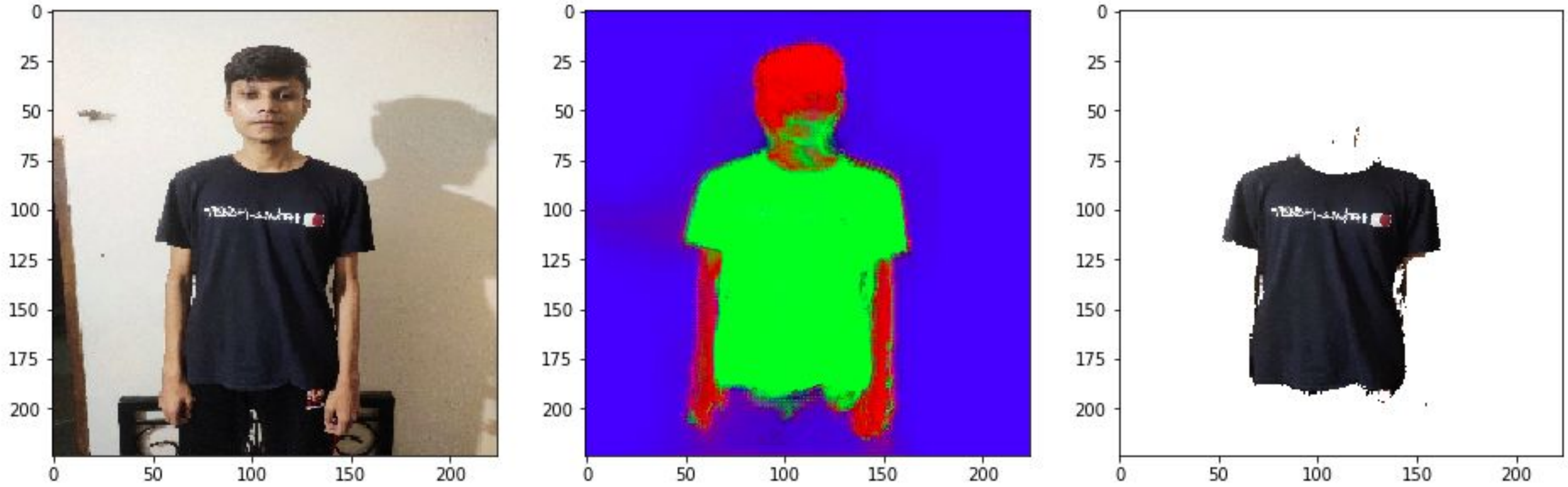


Image Segmentation: Grabcut + U-Net Input/Output

<matplotlib.image.AxesImage at 0x7f3f903eadd8>



Our Journey – A summarization

Paper	Dataset	Our Findings
CycleGAN	LookBook	Works mostly for unpaired data, for example, Horse to Zebra or Apple to Orange. The images present in the LookBook contain images of various types. For our purpose the output images do not meet the expectation at all.
PixelDTGAN	LookBook	After finishing the training on the entire dataset, the resulting images look extremely distorted.
Liquid Warping GAN	Impersonator	Output is far better than the rest of the methods. However, the shape of the human body is not preserved.
GrabCut	N/A	Can differentiate the foreground from the background using segmentation techniques. This method alone can not be used for extracting the clothes. However, it can be used as a tool and integrated to a pipeline of techniques to achieve our goal.

Terminology

Target Image : Image of the person whose clothes are being moved

Source Image: Image of the person on whom the clothes will be moved

Proposed Methodology

First we feed both source and target images to the LWG framework and receive an initial output.

Source Image

Target Image

LWG output

Proposed Methodology

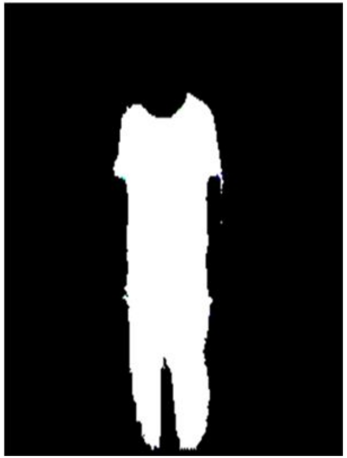
Second, we create masks for both Source image and LWG output using U-Net and Grabcut.

Source Image

Source Image Mask

LWG Image

LWG Image Mask



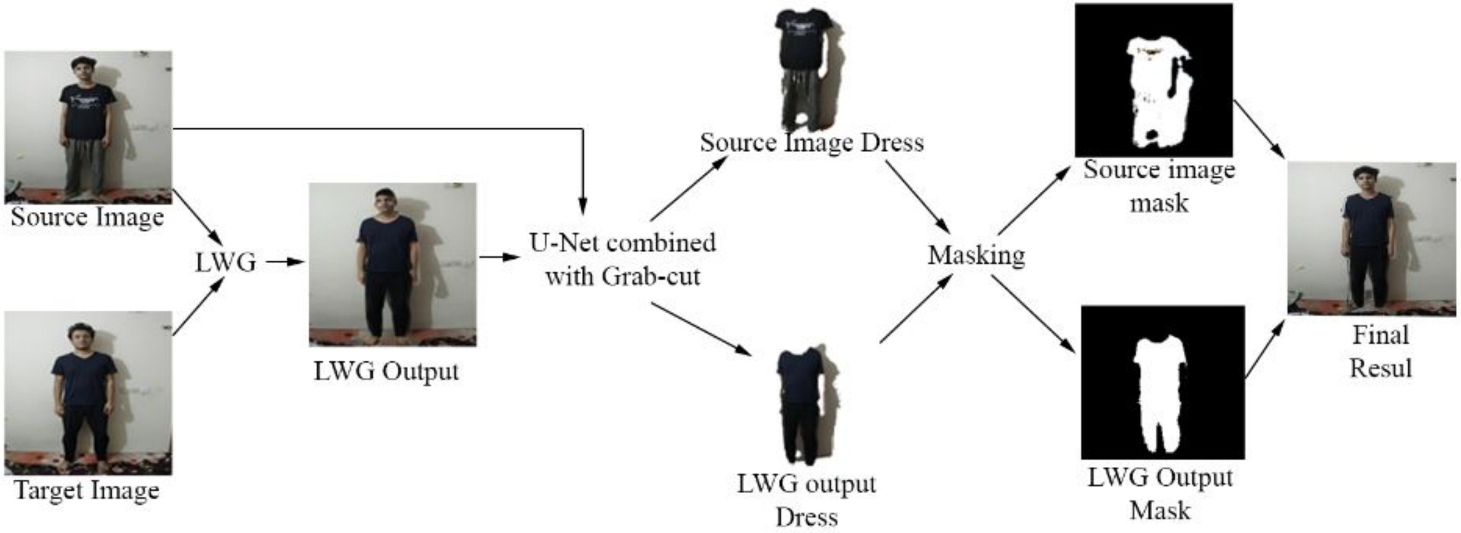
Proposed Methodology

Finally, we are combining the two masks to avoid selecting unnecessary pixels

Source Image | Target Image | LWG output | Our Output

Proposed Methodology

The following is a detailed demonstration of how our method works



Result

Source Image

Target Image

Output

Result

Source Image

Target Image

Output

Comparison with LWG Output

Source Image

Target Image

LWG Output

Our Output

Limitation

Due to the shortage of samples in our dataset, there exists an issue with consistent performance of our U-Net model.

LWG Output

GrabCut + Unet output

Limitation

Source Image	Target Image	Output
		

Limitation

Source Image

Target Image

Output

Limitation

Source Image	Target Image	Output
		

Limitation

Source Image

Target Image

Output

Future Work

Source Image

Target Image

Output

Thank you!