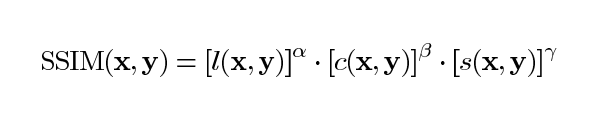
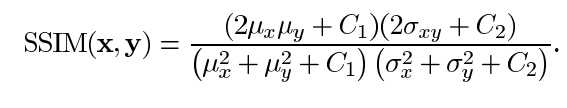
**Structural similarity index measure (SSIM)**

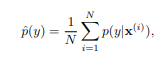
The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a [full reference metric](https://en.wikipedia.org/wiki/Video_quality#Classification_of_objective_video_quality_metrics); in other words, the measurement or prediction of [image quality](https://en.wikipedia.org/wiki/Image_quality) is based on an initial uncompressed or distortion-free image as reference. SSIM is a perception-based model that considers image degradation as *perceived change in structural information*, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms.

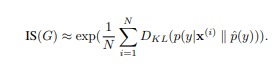




**Inception Score**

The Inception Score is a metric for automatically evaluating the quality of image generative models . This metric was shown to correlate well with human scoring of the realism of generated images from the CIFAR-10 dataset. The IS uses an Inception v3 Network pre-trained on ImageNet and calculates a statistic of the network’s outputs when applied to generated images.

****



**Gaps Identified**

* **Blurriness**

In the previous model the final result images were blur as compared to the original images, because when layers of original image are formed they have to be merged back properly but in previous vton’s there were issues in merging which caused the images to be blurry.

* **Long sleeve dress issues**

In previous vton’s there were issues with long sleeve dresses because of alignment with the users hands.

* **Character Preserving**

In previous vton’s there were issues with alignment of person and clothes, for example if a person is leaning and the cloth is straight then previous vton’s would superimpose the cloth on the person thus distorting the posture hence there were issues with character preservation.

**Dataset**

The dataset used is VITON. 19,000 front view females and top outfits images are there and removed distorted images resulting in 16,253 pairings. Training dataset compromises 14221 pairings and testing dataset compromise 2032 pairing. 18 Key points are being used. Each key point is then converted to a heat map, with an 11x 11 neighbourhood.

**Methodology**

* **Foreground Extraction**

The foremost step in the process of virtual fitting is to extract foreground from background. The algorithm works well on varying light intensities provided that the background has got rid of presence of highly projecting bodies. For reliable extraction, before edge detection, colour clustering by k means is performed. K-means clustering basically converts the image into clusters by quantifying the visual difference according to luminosity and chromaticity as shown in the Fig. 1(a). By doing so, the background is perfectly grouped to a separate cluster. Pre-processing of image is the delicate process which decides the reliability of further output. Hence, after these, pre-processing the image such as to remove small objects and morphological dilation is done.



Fig. 1(a)

* **Extraction Of Human Silhouette**

Extraction of human silhouette from the image in a reliable and efficient way by riding off the disturbances by background is the crucial step for virtual fitting. Background separation process is done by creating a morphological structuring element for applying laplacian filter. Edge detection Algorithm is employed to detect the edges of human. Then to obtain the perfect single line silhouette of human contour, traversal along two axes towards the centre has been carried out. After all, post-processing involves dilation, closing the contour are been performed as in Fig. 1(b).

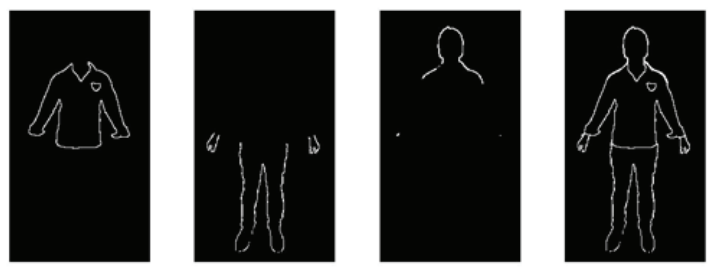


Fig 1(b)

* **Feature Points Extraction**

Feature points extraction is based on the much known morphological facts of human structure. The shoulder and hip points have to be found perfectly and precisely which plays a vital role in accurate fitting of attire. This involves the fact that at the shoulder of human body has the maximum slope considering below the face. So, before doing so, face is detected by Viola-Jones Algorithm. Another fact, being the hip of the human has the maximum width is employed here. Hip points are obtained by finding the maximum of the single line silhouette array as shown in Fig. 1(c). This is performed on both sides from the center giving the reference points for warping.



Fig 1(c)

* **Warping Of Shirt**

The sample shirt selected for fitting must be resized according to the size of the person. Thus the sample shirt is warped to the silhouette by taking feature points as reference. The Fig. 1(d) clearly shows the warping of shirt. This part of work requires tedious algorithm for fine results. Warping is the technique by which image is made to resize and reshaped such that one images superimposes over another image of similar contour. The shirt is divided to regions of distinguished variations and each being resized and concatenated over which interpolation technique is performed.



Fig 1(d)

* **Virtual Fitting**

The next process is to superimpose the warped shirt over the person and to adjust the errors which may occur during the imposing process due to mismatch in hand size, overlay of old shirt, hip correction and kind of. Firstly the shirt is made to impose over the person after warping by the process of image fusion by blending as shown in Fig. 1(e). The error adjustments are done by applying the skin matrix over the difference in sample shirt overlaid and the original images. Finally, the virtually dressed person is fused to the background again, which can be the old background or a new one for the aesthetic view of virtual dressing.



Fig 1(e)

**Results And Discussions**

* **Best Case Result**

Fig 2(a) Fig 2(b) Fig 2(c)

The Cloth (Fig 2(b)) is warped around the model (Fig 2(a)) and will produce the final result (Fig 2(c)).

For the above result.:-

|  |  |
| --- | --- |
| IS | SSIM |
| 0.9671116 | 0.93443 |

* **Combined Dataset Result**

Training and testing done on 2032 images and the final average scores are –

|  |  |
| --- | --- |
| IS | SSIM |
| 3.2976 ± 0.5691 | 0.859997 |

**Drawbacks**

** **

Fig 3(a)Fig 3(b)

Fig 3(b) is distorted after cloth warping around the right hand.

Fig 4(a) Fig 4(b)

In Fig 4(a) , not able to detect whether the subject’s is viewing towards you or away from you, therefore distortion in Fig 4(b)