IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Virtual Try-on Clothing using deep learning

[1]Daud Ibrahim Dewan, [2]Bikal Chapain, [3]Dr.Sandeep Kumar [4]Sneha Misra[5]Samsad Ansari

Software Engineer, Software Engineer, Astt.Professor , Software Engineer, Software Engineer
Computer Science,
Tula's Institute, Dehradhun, India

Abstract : In this era of technology and virtuality, everyone is dependent and inclined towards the virtual means. We even do not want to go outside to purchase a simple good. But if we talk about the fashion sector most of the people are buying it online but it is becoming hard to establish the trust to the seller. Because when we go to the physical market or shop, then we try the clothes before shopping but in online mode in most of the sector there is not provided with the option of the virtual trial. Most of the research on virtual trial based on the 3D pictures taken from the Depth camera. The availability of the depth camera is not economically feasible for every customer. So, our purposed model will take the 2D image of the user and image of the targeted cloth as input and provide the final output of person image with targeted cloth**.**

Keywords **- Virtual Try-on clothing, Human parsing, Unet, Feature mapping, STN, CNN**

INTRODUCTION

As we all know shopping is the basic need of every person. If we look the present scenario of online shopping, then we can see the people's craze towards online shopping. Looking the craze of the people towards it we can easily predict that the future of online shopping is bright, in fact most of the person are now totally becoming dependent towards the online shopping. There is not any greater issue to the goods like Books, Utensil, Electronics, but there is some problem if we talk on the clothing sector. If we go to the physical store than we check the desired cloth, whether it fits properly and if it is looking good on me or not. But if we buy the cloth online then we will not be able to determine whether the targeted cloth suits on us or not. That is because most of the return case on online shopping comes under clothing sector. To resolve this problem, we should provide the user/customer with the experience of the virtual trial. As I already mentioned that most of the research on virtual trial are based on the 3D image of the user. If we talk about the availability and economic feasibility of the Depth camera that is used to take the 3D image than the probability of using the application by the user is minimum. So, we have to find out the way that solves the both problem. For the solution of this issue we purposed a model that accept the image of the user and image of the targeted cloth as the input and provide the output image of the person with targeted cloth image.

RELATED WORK

There have been made lots of efforts and lots of research on implementing the virtual trial on clothing. The aim of transferring the target cloth item onto the reference image that is the image of the user has gain a lot of progress and achievement in these recent years. Although the remarkable progress have been achieved on this field but there remains some challenging task to build up the image-realistic virtual try-on system for real world [23,3,25,2] the main problem is that it remains difficult to partially ascribing the semantic and geometric differences between the aimed clothes and given user reference image as well as the interaction occlusions between the limbs and torso. [9,18,12,13] these research has greatly facilitated the advancement on the sector of image synthesis. Here the generator model learns how to generate the real images to deceive the discriminator, and the discriminator learn to distinguish between synthesized images from the real. Most of existing works focus on clothing compatibility and matching learning [15,10,22], clothing edge/landmark detection [17,24,4,14] and the image analysis of the fashion image [8,6,16]. Where virtual trial is the most challenging in the fashion analysis. In virtual try-on most of the research are based on 3D image. Existing deep learning based approaches [21,1,5,19,20] are the example of 3D image approach for virtual try-on. Some of the 2D image approach are [7,23,3,11]. Most of the virtual trial research focus on the task of keeping the posture and identity. Methods such as VTNFP [25] adopt semantic segmentation [26] for the virtual trial.

Methodology:

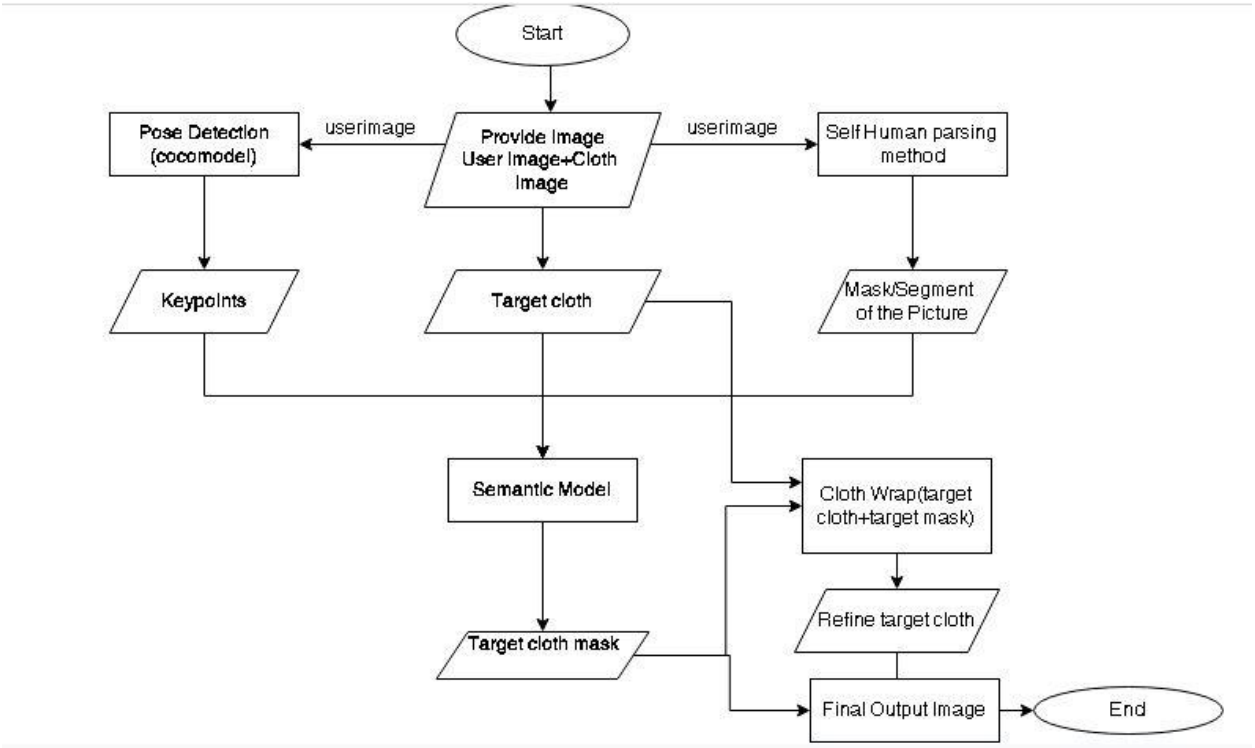

Fig1:Flow chart

The aim of this research is to get the reference image (I) i.e image provided by the user wearing t-shirt and a target clothing item (C) such that the final output of the project is to subject the target cloth image in the body of the reference image. For the functionality of this project we are also going to implement this research and try to give a real virtual try experience by integrating into web-application where an user can try any cloth by simple uploading their image. The whole process are categories into 3 section.

1.Semantic Segmentation module:

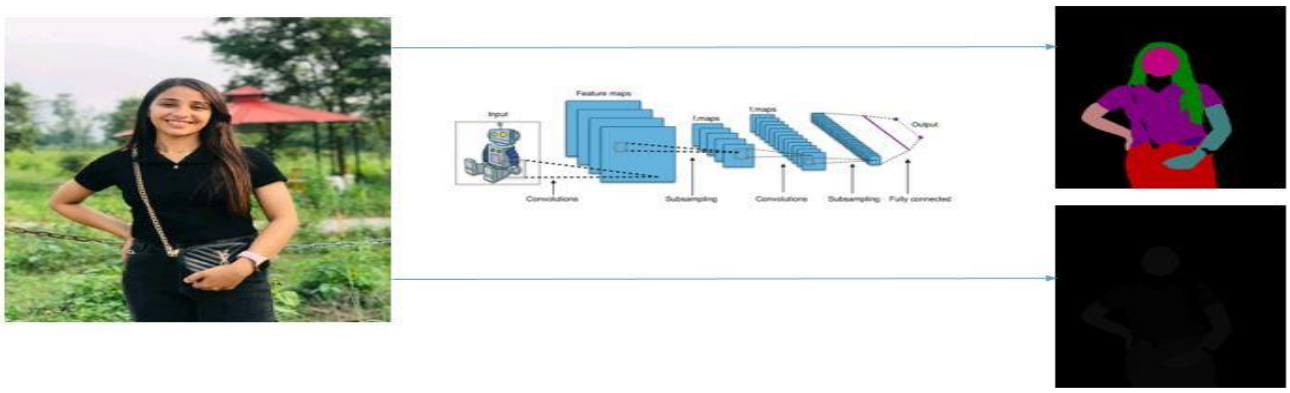

fig2: original Image

fig3: Output given by semanticsegmentation module

The semantic segmentation module (SSM) is to separate the target cloth area by preserving the other bodies parts like hair ,hands and their body shape as well as body pose. Most of the work and research are mainly focused on target cloth only not in other reserved body parts that is hand hair and due to this loss the efficiency of virtual experience was not so good. So to overcome this issue we are going to focus on both target cloth as well as other body parts to give a realistic view. There are lot of mechanism to achieve this but in this paper we are going to use human parsing method. The main aim of human parsing method is for masking the target cloth in person image. In case of human parsing method, there are mostly 3 dataset used for masking purpose. These 3 dataset have different label system and accuracy.

i.LIP(look like person): Lip is mostly used for single person parsing having 59.36% accuracy and this dataset contain 50000+ images having different 20 labels such as Background,hat,hair,glove,sunglasses,uppercloth,Dress,Coat,socks,pants,jumpsuits,scarf,skirt,face,left-arm,right-arm,left-shoe,right-shoe.

ii.ATR: This dataset contains 17000+ image having 82.29% accuracy with 18 labels such as Background, Hat, hair, sunglasses, upper-cloth, dress, pants, scarf, skirt, face, left-arm, right-arm, left-leg, right-leg, leg-shoe, right-shoe, belt, Bag.

iii.Pascal-person-part: This dataset contains 3000+ images having 71.46% accuracy with 7 labels Background, head, Torso, upper-arms, lower arms,upper-legs , lower-legs.

These datasets are mostly used in human parsing and masking purpose but in this paper we have used LIP dataset because it gives more labels and had trained on more image so we decided to use this dataset in this paper. Secondly we have used open pose model to extract the pose (keypoints) of person so that the shape and pose of human image in sample are preserved. There are also different pre-trained model for pose estimation but in this paper we had designed our own posed estimation model to provide a realistic experience. After the results that we obtained from above model, Now we have 2 inputs that is, segmented image of person and his pose keypoints . now we have used heatmap to give human shape using keypoints . After applying heatmap model, we feed this two input to our semantic segment model. The main goal of this model is to map the targeted cloth with person's shape and pose. As from figure3 the output of this modul are i) segemented image ii)label image

2. Cloth transformation Module(CTM):

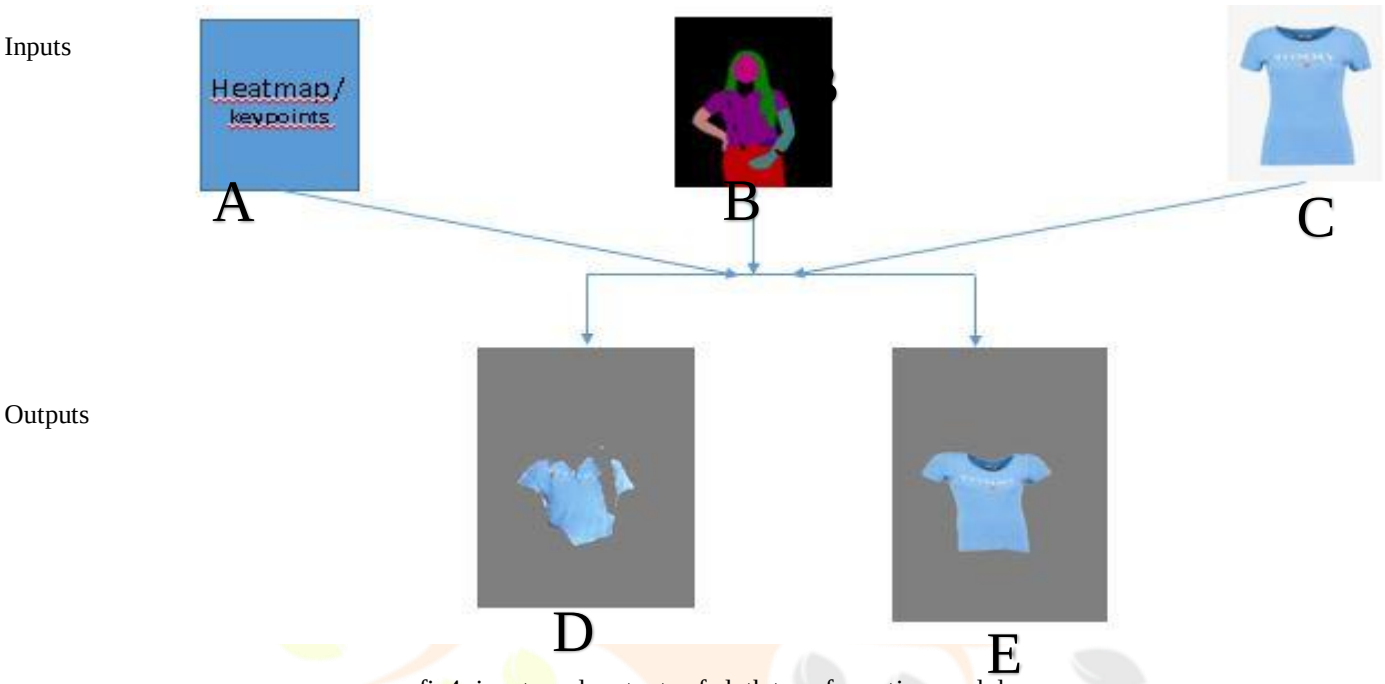

fig4: inputs and outputs of cloth transformation module

The main goal of this CTM is to best fit clothes into the shape of the targeted clothing area with visually natural deformation to human pose as well as to retain the characterists of the cloths. This module transform the cloth into the same size and shape as the target person .Spatial Transformation Network is mostly used .but in real life case simple STN and thin-plate-spline does't seems to work good, so to address this problem , we have used second-order difference on cloth wrapping constrains .This module take two inputs that is cloths image and masked image of person which was obtained from our first module i.e semantic segmentation module. Now this module is trained to execute the output which will be the good fit of cloths having same shape and pose .

Inputs : i) A-Heatmap/keypoints ii) B-Segmented Image iii) C-Targeted cloth image

Outputs: i) E-Wrapped cloth ii) D-refined image

3.Content Fusion module:

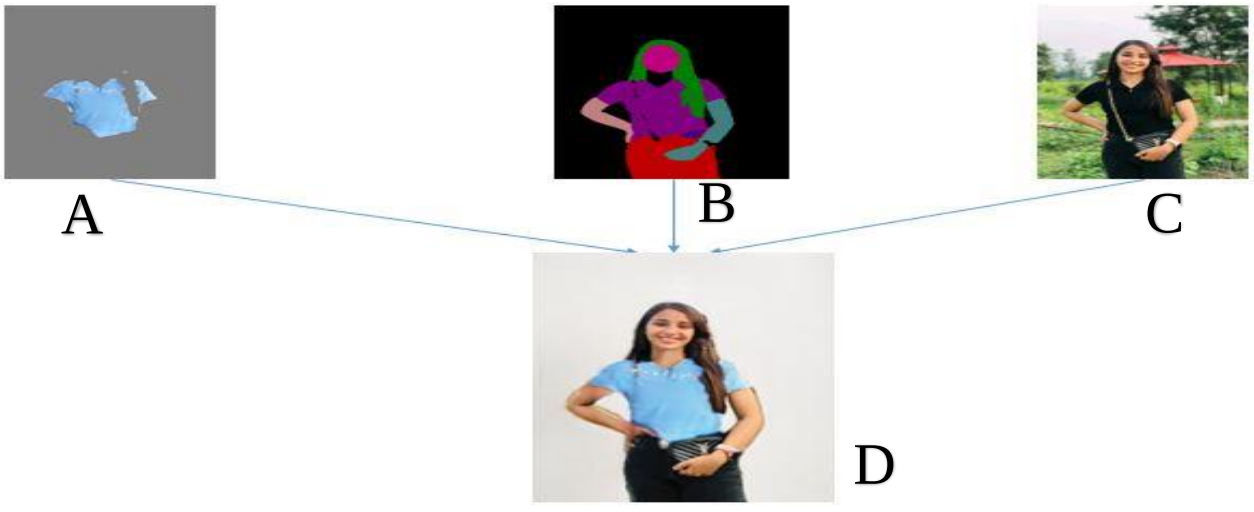

fig5:inputs and output of content fusion module

Now we have obtained the refine cloth having same spatial as the person image and second one is segmented image of person so now the final task is to fuse these two input to produce a realistic output where the refine cloth will be on person body in same texture and shape. there are many way to achieve this goal . like the simple way to replace or overlap the cloth in the same region of the person image having same type of cloth. It has a advantage that the characteristics of the person will be pre-served

and there will be low chance of losing person details like hair, face and arms etc. but this is simple way so it has some disadvantage that is it leads to unnatural appearance at the boundary regions of the image and undesired body parts like hair , arms which are not in proper way to give a natural view. that's why we had choose another way to give natural appearance by using some encoder-decoder network. where encoder reduce the spatial dimensions in every layer and increase the channels where decoder increase the spatial dims while reducing the channels.

As we obtained the wrapped cloth from the previous module, we will integrate human body feature to generate more realistic cloth by the content fusion module. Basically it operates in two step.

Step1:

Constructing a rough try-on result of the previous module and synthesis the mask image of the person obtained from the first module that is semantic segmentation module and wrapped cloth to predict the segmented area where the wrapped cloth should be place.

Step2:

Now Generate the final try-on by fusing these two inputs to give realistic feel.

Inputs : i)refined image ii)segementation image iii) original image

Outputs: i) Final targeted image

Experiments and analysis:

Dataset: we had done all experiments and testing on VITON dataset which contains 19000 images of front view of women and top clothing image pairs. After analysis the images we found that there are invalid images present in this dataset that does not required in our research. so we remove all the invalid images from this dataset and we got 16253 images then we split this dataset into training and testing dataset. we kept 14000 images in training dataset and rest images 2252 images in testing dataset This dataset contains 256*192 resolution of images.

Training setup: In order to design predictive model we have to make sure that we have to setup the network architecture for the training purpose and there are some other parameter that we have to define similarly we had also define some such parameter that make our model robust and effective. In order to train all the model ,we had set the 8 Batch size and ADAM optimizer with B1=0.5 and B2=0.999 and set the maximum number of iteration is 2*10^5 epochs with learning rate 0.0001.In order to achieve the high accuracy you have to take minimum number of batch size and increase the number of iteration epoch with minimum learning rate.

Implementation:

Now we had trained our model and validate with validation dataset .now its time to implement this model to predict on real image and give some realistic experience in fashion world. So to implement this we need two things

1) Server where we can host our model which need GPU configuration Because these all model are trained on gpu system so we have to choose the best server where we can host but for now I am going to use google colab for hosting my all models and decencies because it provide a great experience in case of this kind of project and most important it is free of cost so it is good choose. so in google colab we make a simple rest Api using Flask and create an end-point so any user can make a post request which will accept two image i.e first image of person and the second one is the target cloth image and it will return the one single image having the output by wrapping the target cloth to the person.

2) Now after setting up all the Api, it's time to create some user interface from where user can try various clothes and for this purpose we had design a simple E-commerce website where we posted multiple t-shirt image so that user can try these clothes. So when ever user go to the product details page there is a button called 'try-on' which allow user to upload their image and at the backend part we are using Django framework to handle all these things so whenever user upload their image, a post request is sent to Django framework and then Django convert this image into bytes and prepare 2 input image and call the rest-Api by making a simple post request by passing this 2 image in the form of bytes. and after getting response from the rest-Api , it will again convert that byte into ND Array form and return this image as template so user can preview their virtual-try experience..

Working of Website:

StepI: When the user get logged into the browser the home page will be as:

fig: home page of website

StepII: When the user click on the product the detail of the product along with the virtual trail option is shown as:

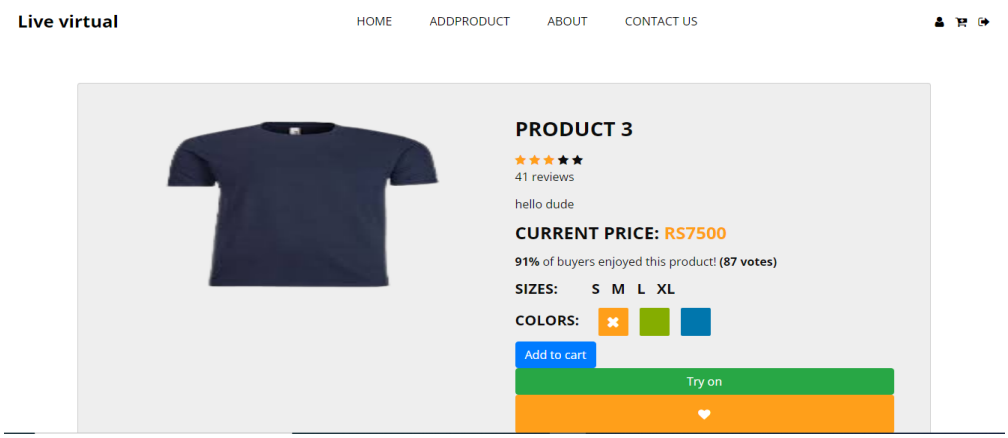

fig: product detail page

StepIII: When the user click on the virtual trial button then the user should provide his/her image.

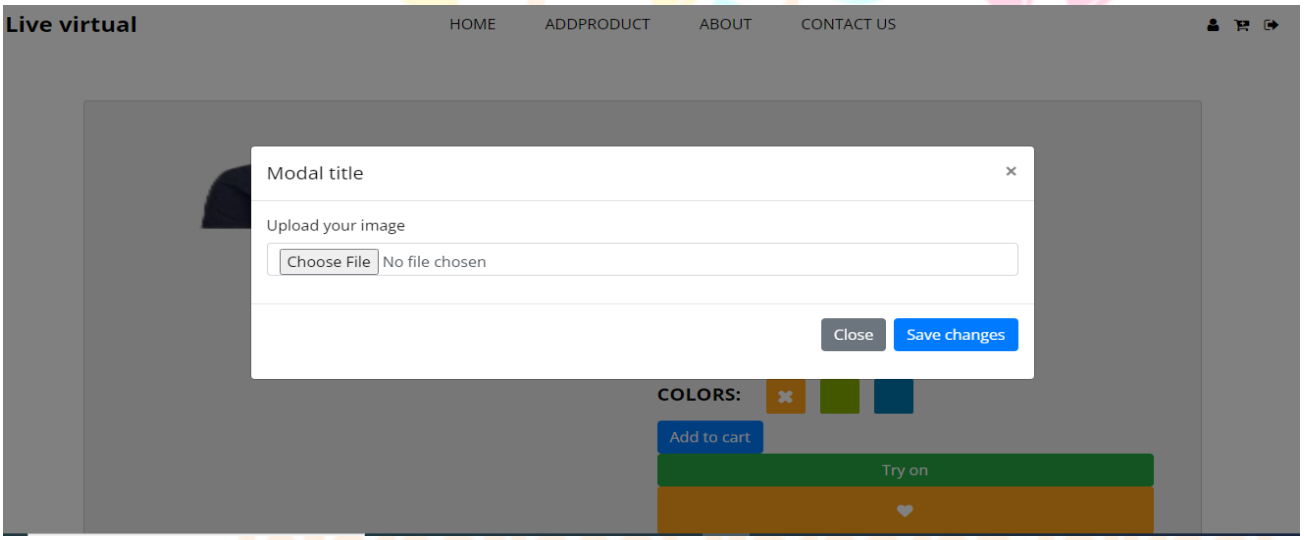

fig16: asking reference image

StepIV:When the user upload his/her image then a final output image is shown.

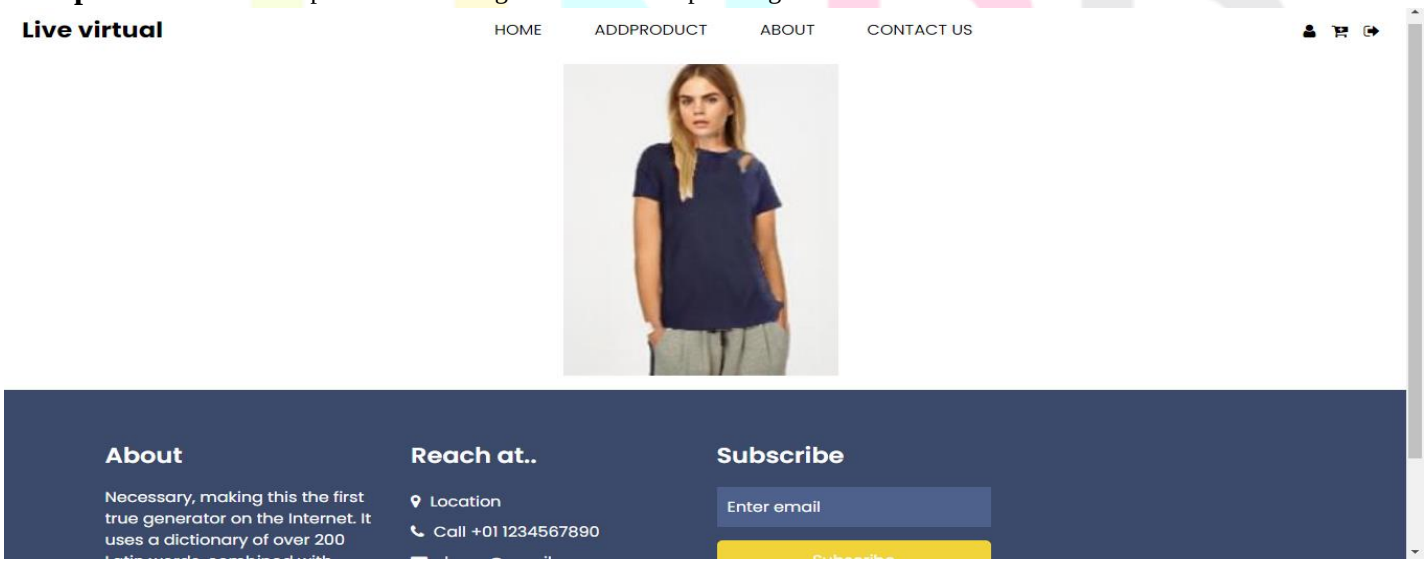

fig: final output

Conclusion:

In this research we have aimed a model that is capable of adaptive content generating and preserving network. The main aims of this research is generating photo-realistic try-on with preserving both the character of the targeted clothes and detain of the human position (posture, body parts, bottom clothes). we have divided this whole module on three sub module that is Mask generation module, cloth wrapping module, and fusion module. The aim of the mask generating module is to extract the mask of the cloth (structure of the cloth) that the user is wearing in the provided image. After finding the mask of the cloth we perform the cloth wrapping module I.e second module. The cloth wrapping module aims to re-structure the targeted cloth image according to the mask obtained by the module (I). finally, we perform the fusion model. The main task of the fusion model is to give the final output i.e image of the user with targeted cloth. The input for the final module will be the output of the wrapping module and the image provided by the user.

.

Refrences:

[1] Remi Brouet, Alla Sheffer, Laurence Boissieux, and Marie- ´ Paule Cani. Design preserving garment transfer. ACM Trans. Graph., 31(4):36:1–36:11, 2012

[2] Szu-Ying Chen, Kin-Wa Tsoi, and Yung-Yu Chuang. Deep virtual try-on with clothes transform. In ICS, volume 1013 of Communications in Computer and Information Science, pages 207–214. Springer, 2018.

[3] Haoye Dong, Xiaodan Liang, Bochao Wang, Hanjiang Lai, Jia Zhu, and Jian Yin. Towards multi-pose guided virtual try-on network. CoRR, abs/1902.11026, 2019.

[4] Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, and Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5337– 5345, 2019.

[5] Peng Guan, Loretta Reiss, David A. Hirshberg, Alexander Weiss, and Michael J. Black. DRAPE: dressing any person. ACM Trans. Graph., 31(4):35:1–35:10, 2012.

[6] Xintong Han, Zuxuan Wu, Weilin Huang, Matthew R Scott, and Larry S Davis. Finet: Compatible and diverse fashion image inpainting. In Proceedings of the IEEE International Conference on Computer Vision, pages 4481–4491, 2019.

[7] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. VITON: an image-based virtual try-on network. In CVPR, pages 7543–7552. IEEE Computer Society, 2018

[8] Wei-Lin Hsiao, Isay Katsman, Chao-Yuan Wu, Devi Parikh, and Kristen Grauman. Fashion++: Minimal edits for outfit improvement. arXiv preprint arXiv:1904.09261, 2019.

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, pages 5967–5976. IEEE Computer Society, 2017.

[10] Tomoharu Iwata, Shinji Wanatabe, and Hiroshi Sawada. Fashion coordinates recommender system using photographs from fashion magazines. In IJCAI, pages 2262–2267. IJCAI/AAAI, 2011.

[11] Nikolay Jetchev and Urs Bergmann. The conditional analogy GAN: swapping fashion articles on people images. In ICCV Workshops, pages 2287–2292. IEEE Computer Society, 2017

[12] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In ICLR. OpenReview.net, 2018.

[13] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In CVPR, pages 4401–4410. Computer Vision Foundation / IEEE, 2019.

[14] Sumin Lee, Sungchan Oh, Chanho Jung, and Changick Kim. A global-local embedding module for fashion landmark detection. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 0–0, 2019.

[15] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. Mining fashion outfit composition using an end-to-end deep learning approach on set data. IEEE Trans. Multimedia, 19(8):1946–1955, 2017.

[16] Jingyuan Liu and Hong Lu. Deep fashion analysis with feature map upsampling and landmark-driven attention. In Proceedings of the European Conference on Computer Vision (ECCV), pages 0–0, 2018

[17] Ziwei Liu, Sijie Yan, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Fashion landmark detection in the wild. In European Conference on Computer Vision, pages 229–245. Springer, 2016.

[18] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In CVPR, pages 2337–2346. Computer Vision Foundation / IEEE, 2019. [19] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. Clothcap: seamless 4d clothing capture and retargeting. ACM Trans. Graph., 36(4):73:1–73:15, 2017.

[20] Damien Rohmer, Tiberiu Popa, Marie-Paule Cani, Stefanie Hahmann, and Alla Sheffer. Animation wrinkling: augmenting coarse cloth simulations with realistic-looking wrinkles. ACM Trans. Graph., 29(6):157, 2010.

[21] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Learning-based animation of clothing for virtual try-on. Comput. Graph. Forum, 38(2):355–366, 2019

[22] Andreas Veit, Balazs Kovacs, Sean Bell, Julian J. McAuley, Kavita Bala, and Serge J. Belongie. Learning visual clothing style with heterogeneous dyadic co-occurrences. In ICCV, pages 4642–4650. IEEE Computer Society, 2015.

[23] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristicpreserving image-based virtual try-on network. In ECCV (13), volume 11217 of Lecture Notes in Computer Science, pages 607–623. Springer, 2018.

[24] Sijie Yan, Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Unconstrained fashion landmark detection via hierarchical recurrent transformer networks. In Proceedings of the 25th ACM international conference on Multimedia, pages 172–180. ACM, 2017.

[25] Ruiyun Yu, Xiaoqi Wang, and Xiaohui Xie. Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. In The IEEE International Conference on Computer Vision (ICCV), October 2019.

[26] Ruimao Zhang, Wei Yang, Zhanglin Peng, Pengxu Wei, Xiaogang Wang, and Liang Lin. Progressively diffused networks for semantic visual parsing. Pattern Recognit., 90:78– 86, 2019