Vol-07 Issue 01, January-2023 ISSN: 2456-9348
Impact Factor: 5.004

iJETRM
International Journal of Engineering Technology Research & Management

POSE ESTIMATION FOR ENHANCING USER SELFIE EXPERIENCE

[1]ASHISH CHOPRA, [2]NEERAJ NAGAR, [3]SHABI UL HUSSAN, [4]ANSHUL SUBRAMANIAN, [5]PRASTIK GYAWALI [6]RAJIV MURALI

Samsung R&D Institute, Noida, India

ABSTRACT
Enhancing user selfies in digital photography has always been very imperative. Selfies are images of a person shot typically with a smartphone or digital camera. These are routinely posted on numerous social media platforms and shared with family and friends. While there are many tools for improving selfies, the practice of modifying the selfie pose is however very sporadic. For this reason, we are turning the user's selfie into the best possible "posture" by retrieving and compiling information about similarly looking celebrities from across different social media handles. These extracted images go on to train our Artificial Neural Network (ANN) which generates an intermediate image. The newly generated intermediate image maps with the user image to produce the best-transformed image with the help of a Generative Adversarial Network (GAN). This paper presents a novel way of collecting and extracting closely related celebrity images w.r.t the user image in real time and subsequently produces the best-transformed image.
Keywords:
General Adversarial Network (GAN), Pose Estimation, Selfies, Image Transformation

I. INTRODUCTION

A selfie is one of the most commonly shared types of photographs. It happens to be a self-portrait and mainly captures the face of the subject. As time has progressed, people have discovered a variety of poses, lighting angles, expressions, filters, etc. to enhance the selfie taken. However, the task of taking a picturesque selfie goes from being comically trivial to dauntingly difficult. A majority of users struggle with capturing a good selfie for various reasons related to physical limitations such as the length of the hand, the ability to "click" a photo while the hand is extended, or other factors such as bad lighting or not knowing what poses to strike. We aim to reduce the variability of these factors while taking a selfie by estimating lighting, poses, and environmental factors. Thus, we propose a solution that considers these factors and recommends the most ideal pose along with other aesthetic poses for selfies. The proposed algorithm uses a novel data collection technique that finds the most liked / popular single-person images of celebrities as a database for recommending the best pose.

Fig. 1. Aim of algorithm

The rest of the paper is structured as follows. Section II defines the proposed algorithm. Section III focuses on data collection and pre-processing. Section IV discusses feature learning and modification. Whereas section V explains the ANN model training step while section VI explains the pre-output phase, de-normalization, and transformation phase

iJETRM
International Journal of Engineering Technology Research & Management

done via the GAN. Finally, section VII discusses the final regeneration phase of the algorithm. Further, in section VIII implementation details are mentioned and in section IX all Results are mentioned.

II. PROPOSED ALGORITHM

Our proposed algorithm is as follows:

- **Data Collection:** The User clicks an image. Simultaneously a time-refreshed celebrity image dataset is collected. This assures that the "best pose" recommendations are synchronous with the latest trends across various social media.
- **Data Preprocessing:** Feature extraction and feature vector creation for both the user and celebrity are done. Unique feature vectors for both are prepared which describe their exclusive appearances. The celebrity's feature vector however excludes certain details so that the user's feature vector can be mapped onto it.
- **Data Normalization:** Both feature vectors are normalized to ease the calculations.
- **Feature Learning and Modification:** This step involves generating the various feature vectors from both the user and celebrity image. This is followed by the normalization of these generated images.
- **ANN Training Phase:** An artificial neural network (ANN) is responsible for learning multiple features from the feature vectors of the celebrity's best-posed image and transforming the user feature vector.
- **De-Normalization:** The transformed feature vector is re-scaled to the original value.
- **Intermediate (Pre) Output Phase:** The transformed feature vector is used to generate a partial frame of the new pose.
- **Regeneration & Presentation:** A transformed image is generated using a Generative Adversarial Network.

Fig. 2. Overview of the proposed algorithm

iJETRM
International Journal of Engineering Technology Research & Management

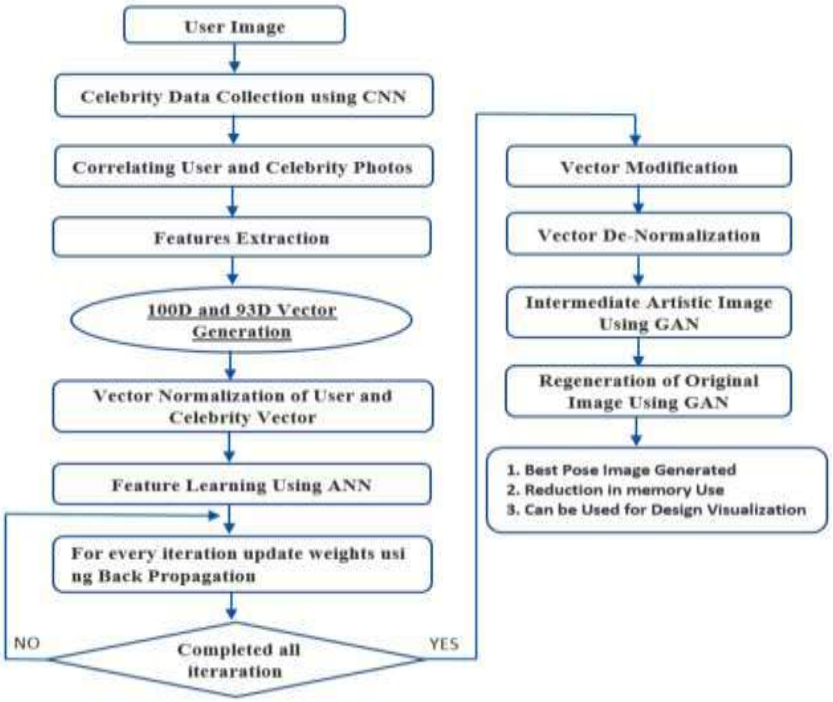

Fig. 3. Flow chart indicating the generation process

III. DATA COLLECTION AND PRE-PROCESSING

Users can either click a selfie or upload an existing selfie. The system saves the user's facial features. To generate the best celebrity pose, we use a novel method that involves the following steps:

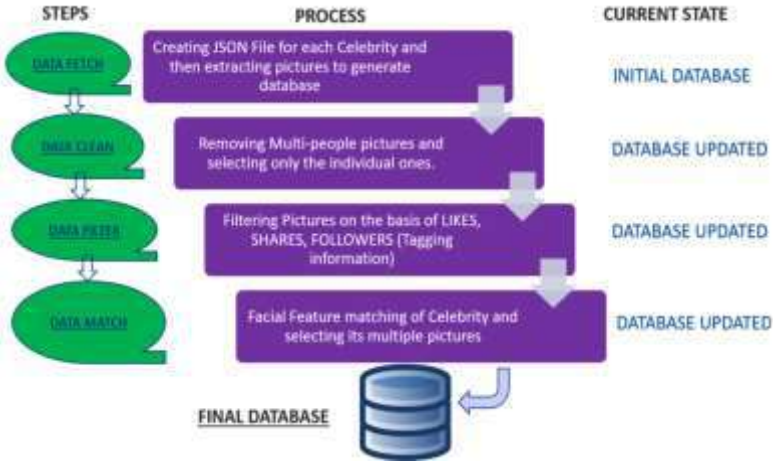

Fig. 4. Process of data pre-processing

iJETRM
International Journal of Engineering Technology Research & Management

1. Data Fetch:
We fetch celebrity's social media handles for the image dataset for our model. We use several websites for tracking celebrity aliases. In our model, we have used the Selenium web scraping framework for gathering data from the various social media handles. It opens each post and retrieves more granular information related to each Image. Ultimately, a final JSON file for each celebrity is populated, which can later be used to match the user image.

2. Data Clean:
The database is cleaned by removing the images which contain more than 1 person. Only images of individuals are retained.

3. **Data Filter:**
The pictures are then filtered based on likes, shares, and other tagging information.

4. Data Match:
This process compares the facial features of users and celebrities in a database, then filters out the best pictures of celebrities based on the feature similarity scores. The database consists of five photographs of each celebrity, sorted by popularity. To find facial landmarks in an image, we use Open CV and the dlib library. The photographs are filtered based on a similarity score. As shown in the figure below, three of the most relevant celebrity photographs are chosen as the model's input. The mapping of the facial features is done as shown below:

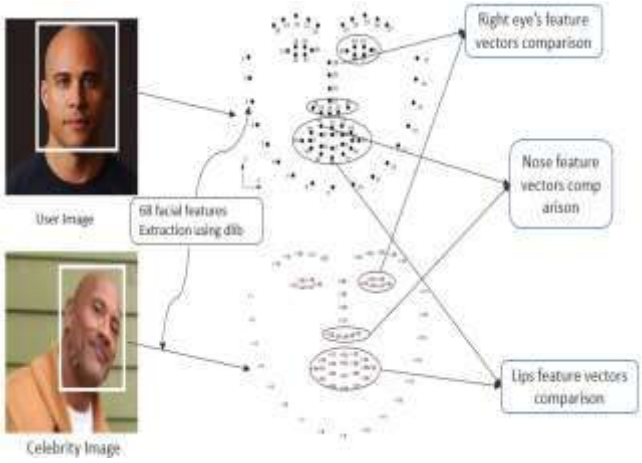

Fig. 5. Feature mapping of user image and celebrity image

IV. FEATURE LEARNING AND MODIFICATION

This step involves generating feature vectors for the user's image and the celebrity's image. It is done as follows:

A. Feature Vector Generation for User's Image

We create a 100-dimension vector matrix that stores information for various feature points that are required for detecting the user's pose. The feature points include:

iJETRM
International Journal of Engineering Technology Research & Management

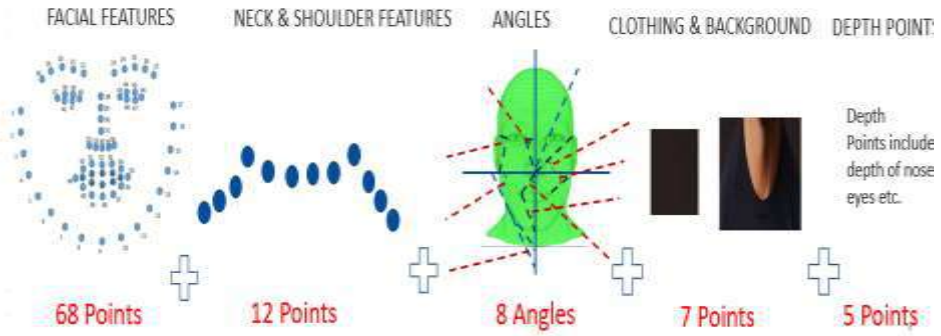

Fig. 6. Feature points of User image.

Face: 68 Points that will be in ordered pairs (x, y) determining the location of that point on that face w.r.t the origin and on the X and Y axis.

Neck & Shoulder: 12 Points that will again be in ordered pairs and show the location of that point on the neck or shoulder**.**

Facial Angles: 8 angles that determine the roll, yaw i.e, the direction of tilt of that part in the face or the neck-shoulder region.

Background and Clothes: 7 Points in the matrix will store the information on the clothing's color, and the texture required at the regeneration time.

Depth Points: 5 points in the matrix will store information about the depth of the nose, eyes, etc.

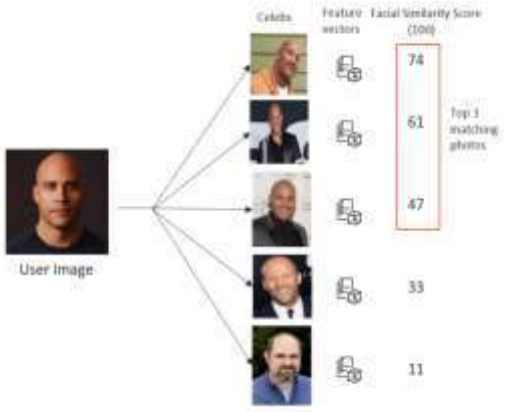

Fig. 7. Top matching celebrity images

iJETRM
International Journal of Engineering Technology Research & Management

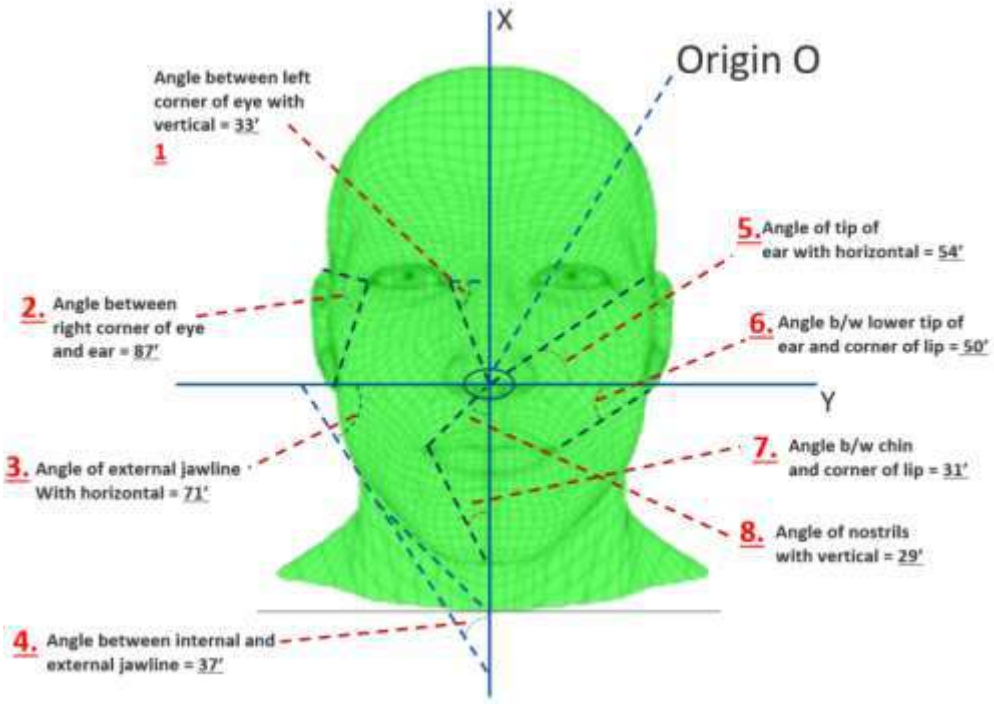

Fig. 8. Angles for the User image

B. Feature Vector Generation for Celebrity's Image

We create a vector matrix that stores information for feature points of the celebrity image. The feature points include the same as the user's but do not include information related to background and clothes - 7 points. As a result, this is a 93-Dimension feature vector. This feature vector is produced for N-different top images of Celebrities. These feature vectors contain the necessary information to map the user's appearance in the best pose.

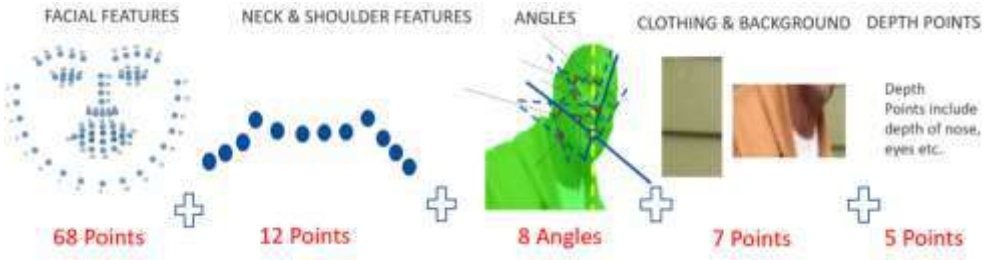

Fig. 9. Feature points of Celebrity image.

iJETRM
International Journal of Engineering Technology Research & Management

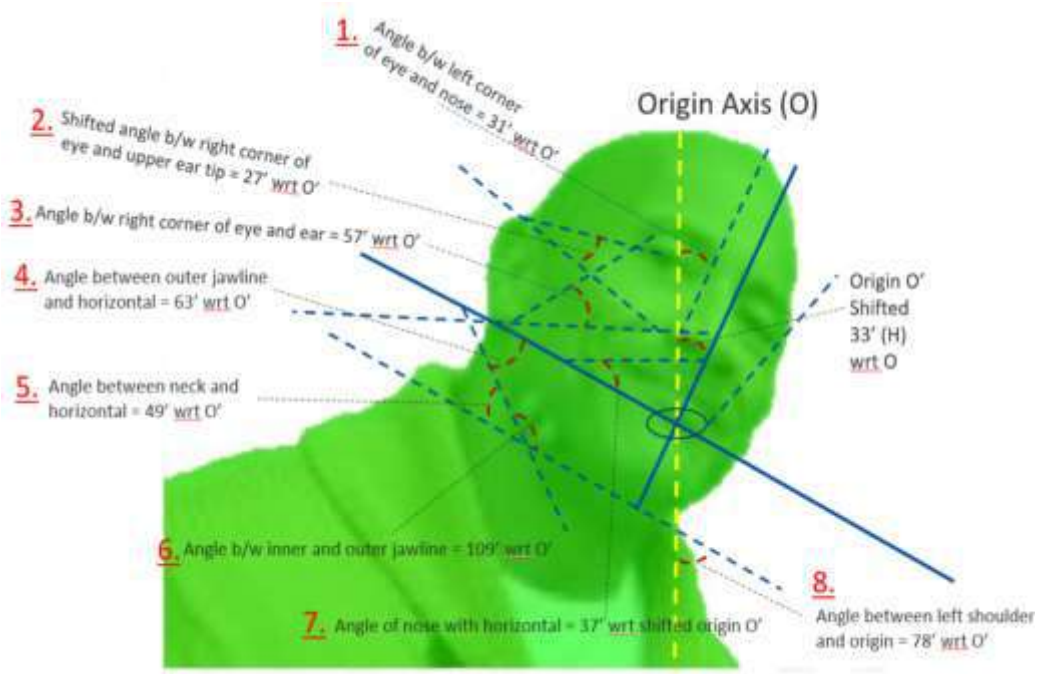

Fig. 10. Angles for the Celebrity image.

C. Vector Normalization Phase

The feature vectors are in the range of (-1000, 1000). This increases computation while calculating Euclidian distances between data points. Therefore, normalization is performed on the feature vectors, which brings the range between (0, 1). We use min-max normalization to normalize the feature vectors.

$$\frac{X - X_{minimum}}{X_{maximum} - X_{minimum}}$$

V. ANN Training Phase

This phase succeeds the Vector normalization phase. The normalized feature vectors of "N" celebrity images are then fed into the ANN model which trains to render a fresh new 93N feature vector matrix by incessantly adjusting the weights and biases with the help of back-propagation. This newly generated 93N matrix on treatment with the 100D User vector matrix finally gives the transformed image of the user by the best celebrity pose image.

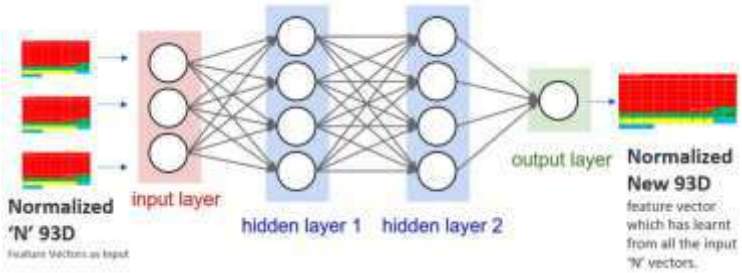

Fig. 11. : Artificial feature learning neural network

iJETRM
International Journal of Engineering Technology Research & Management

VI. DE-NORMALIZATION & INTERMEDIATE PHASE

We get a modified and normalized feature matrix of the user's image with respect to the modified celeb's matrix. We have to de-normalize the matrix now so that we can work on original values during the regeneration of the user's transformed image.

$$X = X_{normalized} \times (X_{maximum} - X_{minimum}) + X_{minimum}$$

After de-normalization, we proceed to generate an intermediate with the help of stored information in the modified feature vector. The modified feature vector contains information about the transformed pose applied to the user's image. We achieve this using our novel Artistic GAN Model, which takes input from random noise and modifies the feature vector of the user's image to generate an artistic image that depicts a modified pose (target pose).

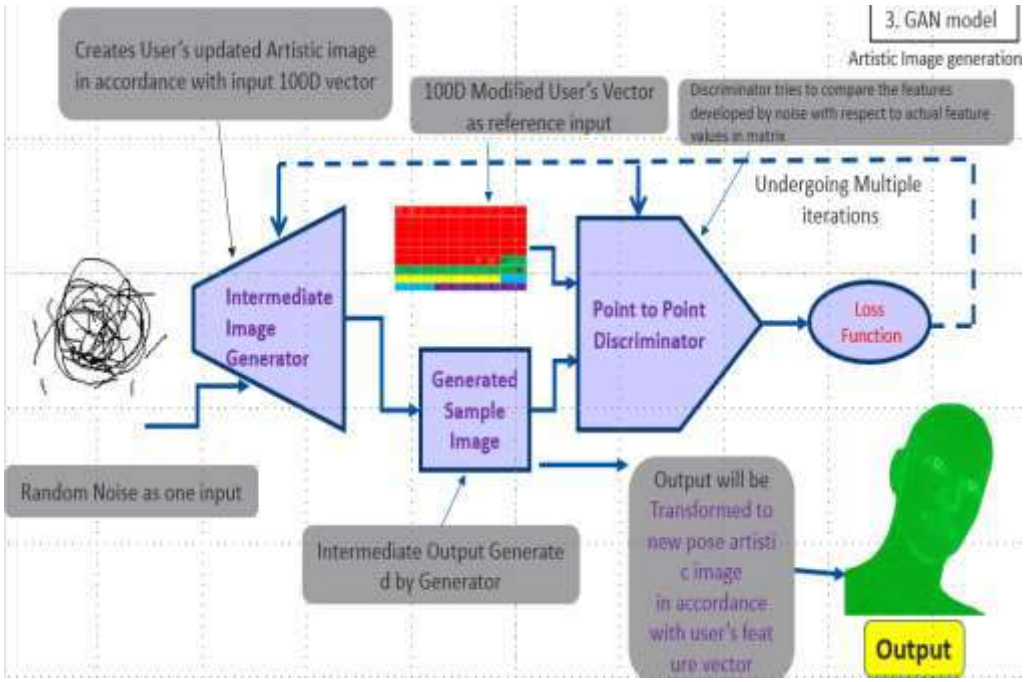

Fig. 12. Intermediate artistic image generation

VII. REGENERATION & PRESENTATION

This is the final phase of the algorithm wherein the actual reconstruction of the best-pose image happens. This GAN focuses on regenerating the original texture and background along with the updated pose of the user.

iJETRM
International Journal of Engineering Technology Research & Management

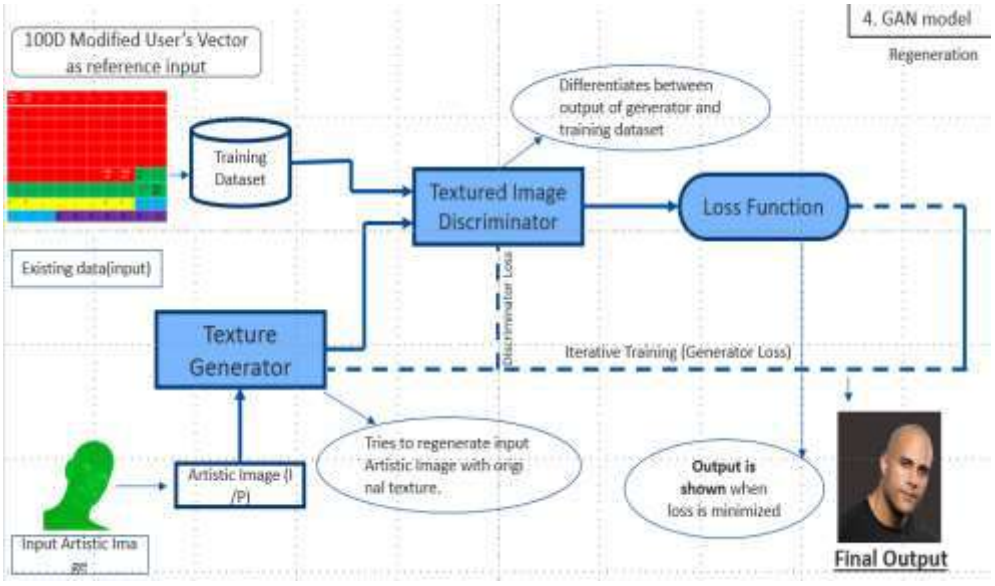

Fig. 13. GAN treatment

VIII. IMPLEMENTATION DETAILS

The full loss function is described below:

$$\alpha_{full} = \arg\min_G \max_D \alpha_{LGAN} + \alpha_{cL1} \quad (7)$$

Where α_{LGAN} stands for adversarial loss and α_{cL1} stands for combined L1 loss. The total adversarial loss is calculated using D_A and D_S:

$$\alpha_{GAN} = \beta\{\log[D_A(P_c,P_t).Ds(S_t,P_t)]\} + \mu\{\log[(1\text{-}D_A(P_cP_g)).(1\text{-}Ds(S_t,P_g))]\} \quad (8)$$

Take note that, β, μ distribution is described below:

$$\beta = \mathbb{E}_{st} \in p_{s,(Pc,Pt)} \in p$$
$$\mu = \mathbb{E}_{st} \in p_{s,Pc} \in p_{,Pg} \in \hat{p}$$

Where, p, $\hat{p}$, p_s represents the distribution of the user's image, celebrity image, and person poses, in that order. The combined L_1 loss can be expressed further as:

$$\alpha_{cbL1} = \lambda_1\alpha_{L1} + \lambda_2\alpha_{pL1} \quad (9)$$

Here, α_{L1} is the pixel-wise α_{pL1} loss calculated between the generated and user's image and $\alpha_{L1} = \| Pg - Pt \|_1$.

iJETRM
International Journal of Engineering Technology Research & Management

IX. RESULTS AND DISCUSSIONS

Fig. 14. Final conversion of User image

Table 1: Market-1501 and DeepFashion state-of-the-art comparison. Here *represents the outcomes of our test set

Model	Market-1501			DeepFashion		
	SSIM	IS	DS	SSIM	IS	DS
Ma et al. [11]	0.099	3.483		0.614	3.228	
Siarohin et al. [20]	0.29	3.185	0.72	0.756	**3.439**	0.96
Ma et al. * [10]	0.261	**3.495**	0.39	0.773	3.163	0.951
Siarohin et al. * [20]	0.291	3.23	0.72	0.76	3.362	0.967
Ours	**0.297**	3.128	**0.73**	**0.788**	3.115	**0.969**
Real Data	1	3.89	0.74	1	4.053	0.968

This paper presents a novel way to choose the best celebrity image corresponding to a user's feature vector. In this implementation, 100 feature vector points from the user image and only 93 points from the celebrity are selected to preserve the user background and then are normalized to ease the entire calculation process. Then the feature vectors of "N" celebrity images are then extrapolated and fed into the Artificial Neural Network (ANN) which in the end renders an intermediate artistic image that is later de-normalized. Subsequently, the de-normalized artistic image is then fed into the GAN to create the best-modified image accompanying both the features of the user and the celebrity. Ultimately the pose of the user was transformed with respect to the celebrity's image. This way the aim of the algorithm was achieved.

REFERENCES

[1] Z. Cao, T. Simon, S. E. Wei and Y. Sheikh, "Real-time multi-person 2d pose estimation using part affinity fields", *arXiv preprint arXiv:1611.08050*, 2016
[2] A. Jain, J. Tompson, M. Andriluka, G. W. Taylor and C. Bregler, "Learning human pose estimation features with convolutional networks", *arXiv preprint arXiv:1312.7302*, 2013.
[3] Z. Cao, T. Simon, S. E. Wei and Y. Sheikh, "Real-time multi-person 2d pose estimation using part affinity fields", *arXiv preprint arXiv:1611.08050*, 2016.

iJETRM
International Journal of Engineering Technology Research & Management

[4] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I. J., Bergeron, A., Bouchard, N., and Bengio, Y. (2012). Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop.
[5] Bengio, Y., Thibodeau-Laufer, E., Alain, G., and Yosinski, J. (2014b). Deep generative stochastic networks trainable by backprop. In Proceedings of the 30th International Conference on Machine Learning (ICML'14).
[6] F. Bach, R. Jenatton, J. Mairal and G. Obozinski, Optimization for Machine Learning, 2011.
[7] Ahmed M. Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In Proc. CVPR, pages 681–688, 2004.
[8] Ahmed M. Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In Proc. CVPR, pages 681–688, 2004.
[9] Ahmed M. Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In Proc. CVPR, pages 681–688, 2004.
[10] an Ma, Xu Jia, Qianru Sun, Bernt Schiele, Tinne Tuytelaars, and Luc Van Gool. Pose guided person image generation. In Proc. NIPS, pages 405–415, 2017.
[11] Liqian Ma, Qianru Sun, Stamatios Georgoulis, Luc Van Gool, Bernt Schiele, and Mario Fritz. Disentangled person image generation. CoRR, abs/1712.02621, 2017.
[12] Ajith Abraham, "Artificial Neural Networks" in Stillwater, OK, USA, 2005.
[13] Carlos Gershenson, "Artificial Neural Networks for Beginners" in , United Kingdom.
[14] K Anil, Jain, Mao Jianchang and K. M Mohiuddin, "Artificial Neural Networks: A Tutorial", *Michigan State University*, 1996.
[15] I. Masi, A. T. Tran, J. T. Leksut, T. Hassner, and G. G. Medioni. Do we really need to collect millions of faces for effective face recognition? In ECCV, 2016.
[16] Miao, S.; Xu, H.; Han, Z.; and Zhu, Y. 2019. Recognizing facial expressions using a shallow convolutional neural network. IEEE Access 7:78000–78011
[17] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4295–4304, 2015.
[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
[19] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition, 2008.
[20] Aliaksandr Siarohin, Enver Sangineto, Stephane Lathuili ´ ere, ` and Nicu Sebe. Deformable gans for pose-based human image generation. CoRR, abs/1801.00055, 2018.
[21] I. Masi, A. T. Tran, J. T. Leksut, T. Hassner, and G. G. Medioni. Do we really need to collect millions of faces for effective face recognition? In ECCV, 2016.
[22] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1891–1898, 2014.