Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 12, December-2019

Survey Paper on Person Attribute Changes using GANS

Priya Charles
HOD, Department of Electronics & Telecommunication
Dr. D. Y. Patil Institute of Engineering Management & Research, Akurdi Pune, India

Mamta Wardhani
Department of Electronics & Telecommunication
Dr. D. Y. Patil Institute of Engineering Management & Research, Akurdi Savitribai Phule Pune University
Pune, India

Onkar Awhad
Department of Electronics & Telecommunication
Dr. D. Y. Patil Institute of Engi-neering Management & Research, Akurdi Savitribai Phule Pune University
Pune, India

G K S Ankit
Department of Electronics & Telecommunication
Dr. D. Y. Patil Institute of Engineering Management & Research, Akurdi Savitribai Phule Pune University
Pune, India

Shubham Gugale
AI & ML Expert
Automaton AI Infosystems Pvt Ltd
Pune, India

Aatish Langhee
AI & ML Expert
Automaton AI Infosystems Pvt Ltd
Pune, India

***Abstract* — Attributes are features that describe a person. Attribute changing basically aims at producing new attributes for the input image. In the field of computers, generating and modifying images with the required attributes along with preserving the realism of the image is a challenging task. It is a difficult task to add facial features like beard, moustache, gender change, etc. in image editing due to the complexity involved. Modern Deep Learning Algorithms like GANS helps to solve the problem without going for manual editing. This paper aims at a problem called attribute manipulation by modifying a facial image in line with a reference facial attribute. If we give a source input image and reference images with a target attribute, our aim is to generate a new image (i.e., target image) that not only possesses the new attribute but also retains the original content of the source image. In order to generate new facial attributes, we train a deep neural network with a combination of losses, which ensure the global consistency of the visual content while implementing the desired attributes often impacting on local pixels. This model automatically adjusts the visual attributes on facial appearances and keeps the edited images as realistic as possible. The evaluation shows that this model can provide a unified solution to both local and global facial attribute manipulation.**

Keywords—Generative Adversarial Networks, facial attribute changing

I. INTRODUCTION

The action of modifying facial attributes (e.g., hair color, eyeglasses, smile) using computer vision techniques attracts numerous research interest due to its potential real-life applications in criminal records identification, entertainment, etc. These facial attributes play an important role in describing a person's look. A person looks different by changing their facial attributes to have different hair color, with beard or without beard, eyebrow shape or color. A person's facial image taken in childhood can be totally different from that taken in adulthood because "Age" is one of the most important facial attributes. Thus, this manipulation of facial attributes is an important task for various applications. For example, if we consider online shopping websites, some of them have the option to select a dress or a frame and to see how it looks on their face or does it suit them. In addition, it can also help facial image search and face recognition by providing precise facial images of different ages.

The existing facial manipulation techniques however are challenging to use because of major two reasons, one of them is the 'realsim' of the image is not maintained during the manipulation process and the other one is learning these tools which work as image editing tools needs a lot of time investment. To overcome these challenges, a generative model with a feed-forward neural network for the purpose of learning the latent representation of an image; a local discriminator which discriminates between the original and generated image; and global discriminator for providing high image realism with respect to the input image. Specifically, in order to modify the attribute, a pairwise attribute loss in the local discriminator is minimized to manipulate the original and generated attribute region.

II. LITERATURE SURVEY

The paper [1], gives an insight of how single or multiple facial attributes can be manipulated by using the Encoder-Decoder architecture i.e., to generate a new face image with desired attributes while preserving other details. They have implemented generative adversarial net (GAN) which consists of encoder-decoder architecture to handle this task with promising results. Facial attribute editing is achieved by decoding the latent representation of given face based on the encoder-decoder architecture. In this paper they applied an attribute classification constraint to the generated image to guarantee the correct change of desired attributes.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 12, December-2019

Meanwhile, the reconstruction learning is introduced to preserve only the attributes excluding details. AttGAN consists of three components (i.e. Generator, Discriminator & Classifier) which cooperate with each other to give high quality facial attribute editing.

The paper [2] aims at modifying the facial image with respect to a reference facial attribute in order to achieve attribute manipulation. The goal of this paper is to generate a new image that not only possesses the new attribute but also preserves the realism with respect to the original image. In order to generate new facial attributes, they have trained a deep neural network with a combination of a perceptual content loss and two adversarial losses, which ensures the global consistency of a visual content while implementing the desired attributes. The evaluation shows that this proposed model can provide a unified solution to both local and global facial attribute manipulation.

The paper [3] shows that facial attributes is an image-to-image translation problem, whose goal is to transfer images from the source domain to the target domain. This paper shows that Facial attributes edit aims only at changing some semantic attributes of a given face image while keeping the contents of unrelated area unchanged. The problem of this method is the lack of paired/labeled data. To train good attributes editing model, we require a great amount of training data which labeled by hand. If this data was reduced, then the editing performance will decrease.

From paper [4], it can be seen that modifying and generating facial images with desired attributes are important and highly related tasks in the field of computer vision. Here they have proposed a model called adversarial regularized U-net (ARU-net)-based generative adversarial networks (ARU-GANs). It is the major part of the ARU-GAN and is inspired by the design principle of U-net. This U-net uses skip connections to pass different-level features from encoder to decoder, which preserves sufficient attribute-independent details for the modification task; it employs an adversarial regularization term which guides the distribution of latent representation to match the prior distribution.

In paper [5], they have shown that realistic image manipulation is challenging because it requires modifying the image appearance in a user-controlled way, while preserving the realism of the result. In this paper, they have proposed a way to learn the natural image manifold directly from data using a generative adversarial neural network. They have defined a class of image editing operations, and constrain their output to lie on that learned manifold at all times. Their model automatically adjusts the output keeping all edits as realistic as possible. All the manipulations are applied in near-real time.

III. EXISTING METHODOLOGY

In this section, we describe the different proposed methods for attribute manipulation.

In [1], AttGAN is introduced for the purpose of face manipulation. The model proposed uses the 'Encoder-Decoder' architecture. This encoder-decoder architecture is the widely used model for the purpose of image manipulation. We can understand the working of this model by dividing it into two parts- the generator and the discriminator. The **generator** is further split into two parts as:

(a) G_{enc}:

The input that is given to the generator is acquired by the G_{enc} in the first stage. This G_{enc} creates a spatial representation of the input image. The input x^a with n binary attribute is provided to the generator. The generated latent representation is given as:

$$z = G_{enc}(x^a)$$

(b) G_{dec}:

The G_{dec} is used to reconstruct the original image from the latent representation that is generated by the G_{enc}. The method of editing the attributes x^a to other attributes b is achieved by decoding z conditioned on b:

$$x^{b`} = G_{dec}(z,b)$$

The second stage of the model is the Discriminator:

Discriminator:

The main aim of the discriminator is to classify the inputs given to it as 'fake' and 'real' images. The Role of generator is to confuse the discriminator. This is an adversarial process because both these entities work against each other. This adversarial process is formulated as a minimax algorithm as:

$$\min_D \max_G Ex\_pdata [\log D(x)] + Ez\_pz [\log(1 - D(G(z)))].$$

In [2], the aim is to generate a photo-realistic version of the input image with the desired attribute. Here, two terms that are considered in this model are content loss and adversarial loss. Content loss is the mean squared error of the feature activations in the given layers in the model, between the input image and the generated image. It is computed at the output of the generator. The content loss in the l-th layer can be given as:

$$Lcontent = \sum_{l=3}^{5} \frac{1}{Cl \times Hl \times Wl} \varphi l(I) - \varphi l(I)^*$$

Adversarial loss is the one that is used in Generative Adversarial Networks. It is used to specify the Eucledian distance loss of two different distributions, i.e., P(G) and P(data). It is the weighted sum of individual loss function.

The objective of the loss function of the GAN is to optimize the generator and the discriminator in a min-max fashion such that Generator tries to fool the Discriminator by generating images that look real while the Discriminator tries to distinguish the fake images from the real ones.

Generator: An input image I along with the feature vector is provided as input to the generator encoder. The feed forward network is trained with the dataset available. The decoder part of the generator then produces the modified image. Thus,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 12, December-2019

this block generates a realistic image with new desired attributes. The objective function of the training encoder and decoder is as follows:

$$\text{Gloss}(\theta P, \theta G) = \min_{\Theta} \sum_{i=1}^{n} \text{Lcontent}(G(P(xi, \theta P), \theta G), xi)$$

Where, xi is the i^{th} image, $\Theta = \{\theta P , \theta G\}$ *represents parameters.*

Discriminator:

Here, the local attribute discriminator is introduced which is trained using the pairwise loss. It is used to generate images with new desired attributes. The proposed model has two parts of discriminator:

(a) Spatial Transform Network (STN): These networks are a generalization of differentiable attention to any spatial transformation. These allow the neural network to learn how to perform spatial transformations on the input image. This results in enhanced version in terms of the geometric invariance of the model.

(b) Pairwise Attribute Loss Network (PALN): It detects the output of the STN as pairwise label of two similar attribute regions. If in case, the output of STN has similar attribute then it is treated as label 0 or else it is treated as label 1.

In [3], the proposed model is made up of three major parts, which are Encoder, Decoder and the Residual Attributes Extractor. These can be denoted as Enc, Dec and ResAttr respectively. The encoder and the decoder form a part of the generator, whose main aim is to generate a 'fake' image. The encoder produces the spatial representation of the input image and the decoder generates/reconstructs the image from this representation. The discriminator is used to distinguish between the data. The main aim of the residual attribute extractor is to make correct attribute predictions on the input face image.

In the proposed model, the encoder-decoder architecture is used for the generation purpose. The generation process can be formulated as:

$$Z^a = Enc(X^a)$$
$$X`^a = Dec(Enc(X^a), Y^a)$$
$$X`^b = Dec(Enc(X^a), Y^b)$$

where $X`a$ and $X`b$ denote the desired editing output and the reconstructed image respectively.

The problem here is that paired data is not available, thus if there are two faces with same identity, learning the residual attributes in such case is difficult. In order to overcome this, the residual attribute extractor is trained using un-paired data. This can be formulated as:

$$S^\alpha, S^\beta, Res\_Attr\_\alpha\beta = D(X^\alpha, X^\beta)$$

Where, X^α *and* X^β represent unpaired images with attributes α and β. S^α and S^β represent the high level features.

The function of the discriminator model used is to distinguish if the images are real or fake.

In [4], a model named U-net (ARU-net) is proposed. ARU-net stands for Adversarially Regularized U-net. The main aim of this proposed model is to modify and generate facial images with the desired attribute. These two tasks are challenging in computer fields, as a result of which the ARU-net is proposed. The skip connection technique is used to pass on the features from encoder to decoder, in order to preserve the attribute independent details. The input taken for this model is celeb faces dataset (CelebA). The ARU-net is incorporated with GAN which results in ARU-GAN to perform the facial attribute manipulation.

The ARU-GAN is made up of four major components namely, a U-net-like generator *G*(*Genc* &*Gdec*), an adversarial network *Dz* on latent space, a discriminator *Dx* on samples and an auxiliary classifier *C*. The loss functions associated in this model are:

(a) Reconstruction Loss: It can be defined as the loss between the images generated by the generator. It can be formulated as:

$$L_{rec} = E_{x^a,a,b}[||x^a - G_{dec}(G_{enc}(x^b), a)||_1],$$

(b) Adversarial Loss on Latent Space: For the generation of new facial images, the main aim is to generate new faces by sampling from prior. The encoder- decoder is trained by using the loss which is defined by:

$LDz = Ez\_p(z)[\log(Dz(z))] + Ez\_p(zjxa)[\log(1 - Dz(z))]$

In [5] a real photo is given as an input then it projects it into their approximation of the image by finding the closet latent feature vector to the original image. It proposes a real – time method for smoothly updating the latent vector so that is generates a desired image that satisfy the user's edits and stays close to the natural image manifold.

In this transformation, the generative model usually loses some of the important low-level details of the input image. Therefore, [5] propose a method that estimates both per-pixel color and shape changes from the edits applied to the generative model. It then transfers these changes to the original photo using an edge-aware interpolation technique and produces the final manipulated result.

Its goal is to reconstruct the original photo using the generative model G by minimizing the Euclidean error.

Projection via optimization: We can directly optimize the model because the feature extractor and the generative models are differentiable. Using this method, we can start from multiple random initializations get the solution with minimal cost. However, the number of random initializations required to obtain a stable reconstruction is prohibitively large, which makes real-time processing impossible.

Projection via a feedforward network: In this we train a feedforward neural network. The training objective for the predictive model can be given as:

$$\theta_P^* = \arg\min_{\theta_P} \sum_n \mathcal{L}(G(P(x_n^R; \theta_P)), x_n^R).$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 12, December-2019

Where x_n^R denotes the n-th image in the dataset. The architecture of the model P is equivalent to the discriminator D of the adversarial networks, and only varies in the final number of network outputs.

A hybrid method: This method combines the advantage of both the approaches mentioned above. Given a real photo x^R, we first predict $P(x^R; \Theta P)$ and then use it as the initialization for the optimization objective. So, the learned predictive model serves as a fast bottom-up initialization method for a non-convex optimization problem.

Manipulating the Latent Vector –
In this paper it updates the initial projection x_0 by simultaneously matching the user intentions while staying on the manifold, close to the original image x_0.

Each editing operation is formulated as a constraint on a local part of the output image. The editing operations include color, shape and warping constraints. Given an initial projection x_0, we find a new image x € M close to x_0 trying to satisfy as many constraints as possible

$$x^* = \arg\min_{x \in \mathbb{M}} \Big\{ \underbrace{\sum_g \|f_g(x) - v_g\|^2}_{\text{data term}} + \underbrace{\lambda_s \cdot S(x, x_0)}_{\substack{\text{manifold} \\ \text{smoothness}}} \Big\}$$

The above equation simplifies to the following on the approximate GAN manifold ~M:

$$z^* = \arg\min_{z \in \mathbb{Z}} \Big\{ \underbrace{\sum_g \|f_g(G(z)) - v_g\|^2}_{\text{data term}} + \underbrace{\lambda_s \cdot \|z - z_0\|^2}_{\substack{\text{manifold} \\ \text{smoothness}}} + E_D \Big\}$$

Edit Transfer – Give the original photo, its projection, and a user modification by its method. The generated image captures the change we want, that is the quality is degraded w.r.t the original image.

Transfer the modifications to the original photo: After estimating the color and shape changes in the generated image sequence, we apply them to the original photo and produce an interesting transition sequence of photo-realistic images. As the resolution of the flow and color fields is limited to the resolution of the generated image, it up samples those edits using a guided image filter.

IV. PROPOSED METHODOLOGY

Now-a-days, the data that is available for the purpose of training a model is very limited. This is because of the privacy issues that are raised with respect to the data. The data required for the training and testing purpose should be in a huge quantity for providing better results. In order to deal with this, we are implementing a model of GAN that will provide facial images with desired attributes at the output side. The attributes that we have considered are eyeglasses, facial expressions, hair color, facial hair, etc. The implemented model of GAN consists of Generator that uses the encoder-decoder architecture, Discriminator and a Classifier.

V. ARCHITECTURE

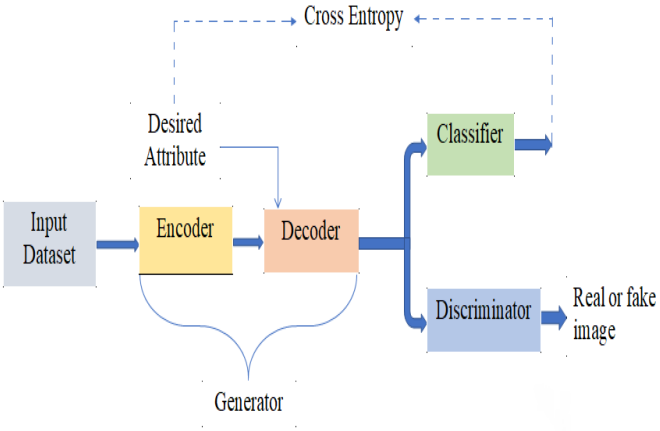

Fig. 1 – Block Diagram of the proposed model

Generator – It learns to create fake data by incorporating feedback from the discriminator. Generator takes random noise as its input. It then tries to transform this noise into a meaningful output. The Generator block consists of two parts Encoder and Decoder. The Encoder is used to encode the input images into latent representation and the Decoder produces an original image from the spatial representation generated by the Encoder.

Classifier – It is used to constrain the generated image to correctly own the desired attributes.

Disriminator – The discriminator tries to distinguish real images from the fake images created by generator.

Loss Functions – GAN uses loss functions to reflect the distance between the distribution of data that is generated by it and distribution of the real data.
Adversarial Loss: It is basically the distance of the two distributions. The generator tries to minimize the following function while the discriminator tries to maximize it, this formula is derived from the cross entropy between the real and generated distributions.

$$E_x\,[log(D(x))] + E_z\,[log(1 - D(G(z)))]$$

Where, D(x) is the discriminator's estimate of the probability that real data instance x is real. E_x is the expected value over all real data instances. G(x) is the generator's output when the noise given is z. D(G(z)) is the discriminator's estimate of the probability that a fake instance is real.

VI. ADVANTAGES

Editing facial attributes aims at modifying single or multiple attributes on a given face images, i.e. creating a new face image with desired attributes while retaining other information.
GANs are easy to implement and to code as per the requirements.
The data generated can be used for training different models.
GANs can take images of any pixel size.
Represent and manipulate high-dimensional probability distributions

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 12, December-2019

Generative models can be used in reinforcement learning in many ways. They can be trained with missing data and can provide predictions on inputs that are missing data.
Adversarial learning is employed for visually realistic editing.

VII. DISADVANTAGES

The major disadvantage of the existing methods is that, the facial images taken as input should be facing straight. If any side face images are taken as inputs, the modification of attributes is not carried out properly. Another disadvantage is that, in the existing facial modification methods, new attribute is imposed on the attribute that is already present in the image. This increases the unnecessary computation time.

VIII. FUTURE SCOPE AND CONCLUSION

In future, the model can be trained in such a way that is can generate a new image out of some particular attributes that are available. Also, the model can be trained to modify the attributes of a facial image which is facing sideways.
This could be useful to people to generate a large number of samples for a dataset. It solves the problem of the limited availability of datasets.

IX. REFERENCES

[1] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan And Xilin Chen, “Attgan: Facial Attribute Editing By Only Changing What You Want ”
[2] Yilin Wang, Suhang Wang, Guojun Qi, Jiliang Tang And Baoxin Li, “Weakly Supervised Facial Attribute Manipulation Via Deep Adversarial Network” In IEEE 2018 Winter Conference On Applications Of Computer Vision.
[3] Rentuo Tao, Ziqiang Li, Renshuai Tao And Bin Li, “Resattr-GAN: Unpaired Deep Residual Attributes Learning For Multi-Domain Face Image Translation” In IEEE Access Digital Object Identifier 10.1109 / ACCESS.201922941272
[4] Jiayuan Zhang, Ao Li, Yu Liu And Minghui Wang, “Adversarially Regularized U-Net-Based Gans For Facial Attribute Modification And Generation” In IEEE Access Digital Object Identifier 10.1109 / ACCESS.2019.2926633
[5] Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman And Alexei A. Efros, “Generative Visual Manipulation On The Natural Image Manifold” On 16 Dec 2018