International Journal of Scientific & Engineering Research Volume 10, Issue 12, December-2019
ISSN 2229-5518

Face Formation using Deep Learning Approaches

Deepanshu Mehta

Abstract— In the last few years, a type of generative model known as Generative Adversarial Networks (GANs), has achieved tremendous success mainly in the field of computer vision, image classification, speech and language processing, etc. GANs are the models that are used to produce new samples that have similar data distribution as of the training dataset. In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention because of two primary disadvantages viz. 1) it requires a lot of data, processing time, hyperparameter tuning, and model validation before the start of building the real model, and 2) encoding layers in Autoencoder (Unsupervised learning with CNNs) try to preserve the quantity of information rather than the quality of information. In this work, an endeavor has been made to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called GANs, that have certain architectural assets, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. GAN has been applied to various applications such as computer vision and natural language processing and achieves impressive performance. There are many applications of GANs including Image Synthesis, Image modification, Super Resolution, Assisting Artists by Image Translation, Photo-Realistic Images, Speech Generation, Face Aging, etc. In this work, we investigate the application of GAN to the novel application of face generation for humanoids.

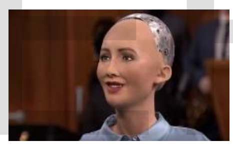
Index Terms— CNN (Convolutional Neural Network), DeCNN (Deconvolutional Neural Network), Discriminator, GAN (Generative Adversarial Network), Generator, Humanoid Robots, Image Synthesis.

IJSER

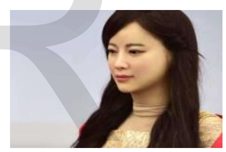
1 Introduction

Humanoid robot is a robot with its body shape built to resemble the human body. The design may be for functional purposes, such as interacting with human tools and environments, or for experimental purposes. Humanoid robots can perform tasks which us humans cannot perform without getting into any risk of losing life in it [1]. Also, researchers like Martin Ford, Carol Reiley, and Boris Sofman predict that they will be able to build humanoid robots who will be as smart as the most intelligent living person on earth by 2030. As we know that the coming era will be of humanoid robots like Sophia of Saudi Arabia, Jia Jia of China, Geminoid DK of Japan, etc. and the production of these kinds of robots will be really high, so this paper proposes a novel idea to generate completely new faces for these future humanoid robots using GANs in order to avoid confusion with the faces of real-world existing people. There are three main techniques for creating images: Autoregressive models, Variational Autoencoders and GANs. Due to the advantages of GANs over the other two methods, namely, quick processing time and use of less hyperparameters, GANs are chosen as the most desirable option for creating entirely new faces that do not exist in the real world for humanoids [2][3][4].

- *Deepanshu Mehta is currently pursuing bachelors of engineering in Information Technology degree program in Panjab University (UIET), India, PH-8283806655. E-mail: mehtadeepanshu2@gmail.com*



Sophia



Jia Jia

Geminoid DK

The basic idea of GANs is that there are 2 models as shown in Figure. 1.

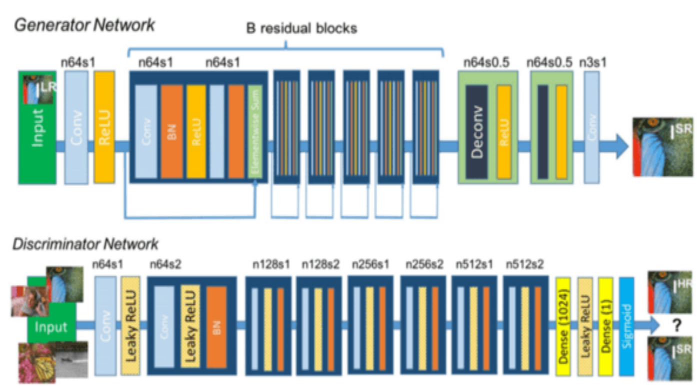


Figure 1: A Generative Network Model and a Discriminative Network Model.

.

IJSER © 2019
http://www.ijser.org

2 Proposed Method Based on GANs

The discriminative model has the task of determining whether a given image looks natural (an image from the dataset) or looks like it has been artificially created. The task of the generator is to create natural-looking images that are similar to the original data distribution. This can be thought of as a zero-sum or minimax two-player game. The analogy used in the paper is that the generative model is like “a team of counterfeiters, trying to produce and use fake currency” while the discriminative model is like “the police, trying to detect the counterfeit currency” [5]. The generator is trying to fool the discriminator while the discriminator is trying to not get fooled by the generator. As the models train through alternating optimization, both methods are improved until a point where the “counterfeits are indistinguishable from the genuine articles”. Image creation by GANs can also be thought of as analogous to human’s imagination [3]. A simplified diagram of GAN is shown in Figure. 2.

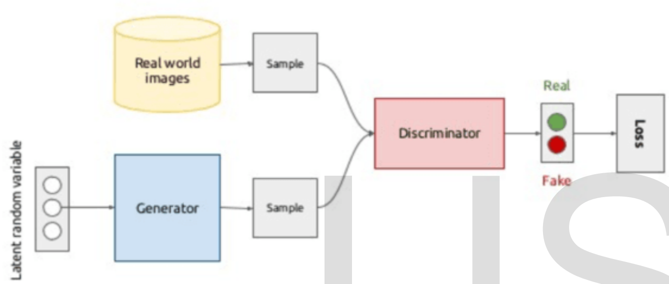


Figure 2: A simplified diagram of GAN

IJSER

The Generator component has a Deconvolutional Neural Network whereas the Discriminator component has a Convolutional Neural Network as shown in Figure. 3. Deconvolutional networks are Convolutional Neural Networks (CNN) that work in a reversed process. Deconvolutional networks strive to find lost features or signals that may have previously not been deemed important to a Convolutional Neural Network’s task [5][6][7].

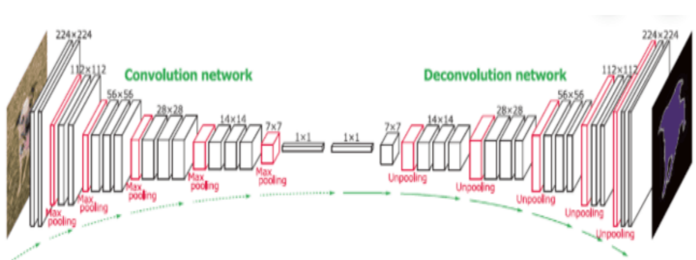


Figure 3: Convolutional and Deconvolutional Network

In Figure. 4, training of discriminator takes place. Random noise is inputted into the Generator and new images are created according to the weights in neural networks. Discriminator does not know anything about the images in the first iteration because of which it is not good at finding the probabilities of whether created images are natural and the probabilities of real images (in this case human images) being real images. Batch of real human photos and Batch of images created by Generator is inputted to the Discriminator. After getting the individual probabilities, the error or loss is calculated by comparing to the true value (i.e. 0 and 1 respectively where 1 means that the image is a real image and 0 means that the image is artificially created and unreal). In the figure, the above group of probabilities should be 0 and below one should be 1. The cumulative loss is backpropagated to the Discriminator and weights inside the Convolutional neural network of the Discriminator gets updated [3]. Hence the Discriminator becomes trained for the next iteration as it learns about the objects like animals, people, cars, etc.

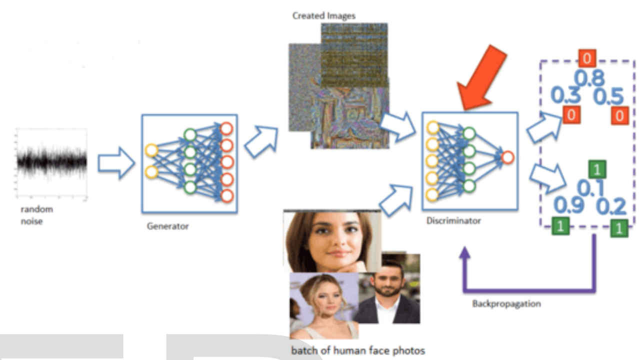


Figure 4: Training of Discriminator

In Figure. 5, training of the Generator takes place. Basically, the Generator learns by trying to trick the Discriminator. Unlike earlier, this time batch of real images or photos is not inputted into the discriminator and images created by Generator are only inputted into the Discriminator. Discriminator now finds the probabilities of whether the created batch of images are natural. This time the probabilities are compared to 1 instead of 0 as in earlier in the training of discriminator. The cumulative loss is calculated and backpropagated to the Generator and weights are updated in it and it gradually gets improved with the number of iterations upon creating better images with clear objects [8][9]. Figure.6 shows what images Generator created after one iteration takes place having Generator and Discriminator both trained just once.

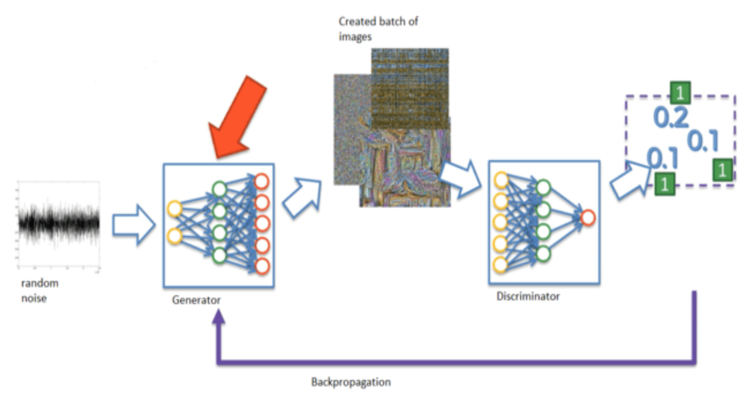


Figure 5: Training of Generator

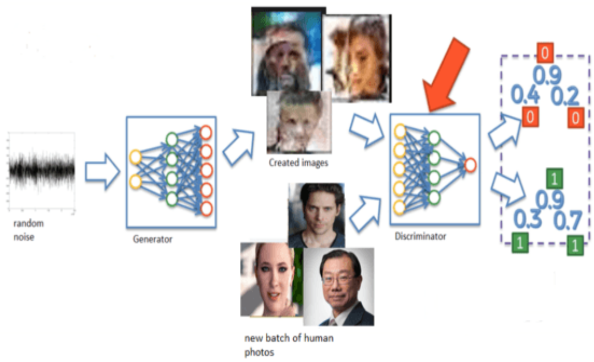


Figure 6: Images created by Generator after 1st iteration of training of Generator.

This process of training of GANs happens repeatedly with each iteration until it starts creating natural-looking human face images. For instance, the images of people in Figure. 7 are all created by training the GANs repeatedly and no such person exists in the real world [10][11].

Figure 7: Facial Images created after repeated training of GAN

3 Applications of GANs for Face Modification

3.1 Arithmetic Operations on Facial Images

Besides generating objects, arithmetic on abstract ideas like removing glasses from a face can also be done, as shown in Figure. 8. This arithmetic can help us in creating furthermore new images. For example, what we did with glasses in Figure. 8 can also be done for lips, eyes, nose, or any other human face feature to get a desired face [12].

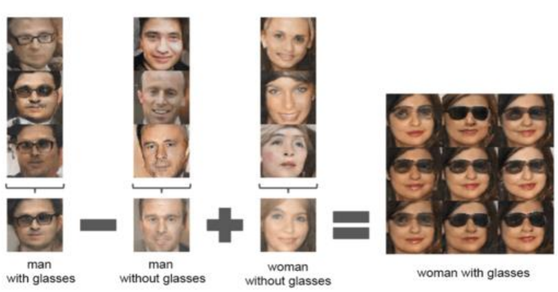


Figure 8: Arithmetic operations performed on Facial Images

3.2 Face Aging

Face Aging software uses GANs and the results are shown in Figure. 9. This is often used for fun purposes by people. But in our case, it can help us in the generation of faces for old human-looking robots old or newborn baby looking robots [13][14].

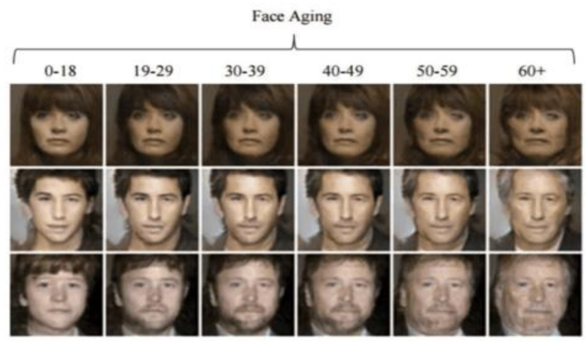


Figure 9: Face Aging performed using GANs

3.3 Super Resolution

Super Resolution is one of the applications of GANs. Researchers recently compared the performance of bicubic, ResNet and GAN on improving the resolution of a faded image [15].

In Figure. 10, the results are shown about which model performed the best. The original image was faded and all three models were applied to improve the faded image's resolution and the one which proved to the best was GANs. This can also be used as a tool in enhancing image quality of faces if the generation of new faces by GAN produces a low-resolution image [15].

Figure 10: GANs performed the best at improving Image Quality

3.4 Raw Portrait to Real Life Image

GANs also can assist artists by filling there drawn image with colors and make them look real and in 3D. This is shown in Figure. 11. This gives artists the power to imagine their real looking face and GANs will help them to create a human image which they can use for the robots thus, giving scientists or researchers the ability to create a robot's face according to their desire with the help of an artist who can make a raw drawing of it [16][17].

ISSN 2229-5518

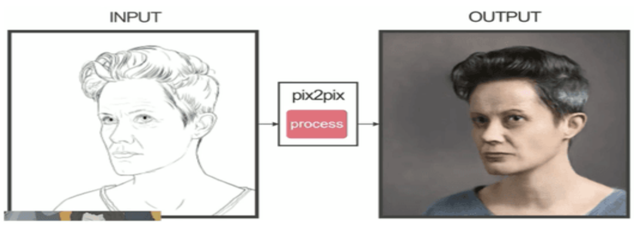


Figure 11: A portrait drawn is changed to Real Life photo using GANs

4 Related Works

This section discusses some variants of GANs which are as follows:

1. DCGAN (Deep Convolutional GAN) [18]
2. CGAN (Conditional GAN) [19]
3. LSGAN (Least Square GAN) [20]
4. ACGAN (Auxiliary Classifier GAN) [21]
5. InfoGAN [22]

4.1 DCGAN

In this method, deep convolutional networks are used at both the discriminator and the generator. Discriminator D is a set of convolution layers with stridden convolutions so it downsamples the input image at every convolutional layer. Generator G is a set of convolution layers with fractional stridden convolutions or Transpose convolutions so it upsamples the input image at every convolutional layer.

At Discriminator D,

$Dloss_{real} = \log(D(x))$ (1)

$Dloss_{fake} = \log(1-D(G(z)))$ (2)

$Dloss = Dloss_{real} + Dloss_{fake}$ (3)

Total Cost $= (1/m) \sum_{i=1}^{m} [\log(D(x^i)) + \log(1-D(G(z^i)))$ (4)

At Generator G,

$Gloss = -\log(D(G(z)))$ (5)

Total Cost $= (1/m) \sum_{i=1}^{m} [-\log(D(G(z^i)))$ [18] (6)

4.2 CGAN

In this method, both the generator and discriminator are conditioned on some extra information 'y' which can be a class label or a tag. The core idea is to train a GAN with a conditioner. We can perform the conditioning by feeding **'y'** into both the discriminator and generator as an additional input layer. We concatenate this vector 'y' with real 'x', which is then fed into the discriminator (D)**.** Also, we concatenate this vector 'y' with noise 'z', then we feed it to the generator (G). Loss functions are as given below [8][19]**,**

$Dloss = E[\log(D(x,y))] + E[\log(1 - D(G(z),y))]$ (7)

$Gloss = E[\log(D(G(z),y))]$ (8)

Total loss $= \min_G \max_D V(D,G) = E[\log(D(x|y))] + E[\log(1 - D(G(z|y)))]$ (9)

4.3 LSGAN

GANs may lead to the vanishing gradients problem which slows the learning process. LSGAN attempts to overcome this problem by adopting the least-squares loss function instead of the sigmoid cross-entropy loss for the discriminator. The objective function of LSGAN yields minimizing the Pearson divergence as given below,

$\min_D V(D) = (1/2)\, E[(D(x) - b)^2] + (1/2)\, E[D(G(z) - a)^2]$ (10)

$\min_G V(G) = (1/2)\, E[D(G(z) - c)^2]$ (11)

here a and b are labels for fake data and real data, and c denotes the value that G wants D to believe for fake data[20].

4.4 ACGAN

In it, every generated sample has a corresponding class label, c in addition to the noise z. G uses both noise z and class c to generate images. The discriminator gives both a probability distribution over sources and a probability distribution over the class labels. The objective function has two parts: the log-likelihood of the correct source, LS, and the log-likelihood of the correct class, LC**.**

$LS = E[\log P(S=real \mid X_{real})] + E[\log P(S=fake \mid X_{fake})]$ (12)

$LC = E[\log P(C=c \mid X_{real})] + E[\log P(C=c \mid X_{fake})]$ (13)

D is trained to maximize LS + LC while G is trained to minimize LC – LS [21].

4.5 InfoGAN

As discussed above, after training in GANs, some noise z is added and the trained generator G is used to produce the fake images. This learning is called entangled representation. InfoGANs are able to learn disentangled representation in an unsupervised way. In conditional GANs (CGAN), we give the condition c (label y) manually above but InfoGANs try to make it learned automatically. Since c is unknown, it is solved by maximizing the Mutual Information between the latent variable c and generator's output and use that information to train InfoGAN.

$\min_G \max_D V(D,G) = V(D,G) + \lambda I(c; G(z,c))$ (14)

where I is the mutual information.

$I(c; G(z,c)) = H(c) - H(c|G(z,c))$ (15)

where H is the entropy [22].

5 Conclusion and Future Works

In this paper, an extension of the generative adversarial net framework is introduced for generating faces for humanoid robots. The ability to condition on arbitrary external information can be added to both the generator and discriminator components.

The faces generated by the proposed method will be unique and have extremely low possibility to match with any face of a living human being in this world. The reason behind this is that several facial photos have been used to train proposed GAN models and the facial images which are created by the generator will have a number of different facial features coming from many different faces. So, it is impossible to find such a face in the real world, as it will require the genes of several humans to artificially go through gene transmutation, which is currently banned in all the countries. Creating many unique faces can become a very tedious task and it requires a lot of innovation. Therefore, we contend and demonstrate how the

IJSER

proposed method can prove to be effective for face generation for humanoid robots as well as for some other purposes like creating unique characters for animated movies, etc.

As future directions for research, it is noteworthy to point that GAN and its variants have improved face generation with time, but still, there are some faces which are distorted and not created up to mark. Some sort of elimination process before the training process which removes defected faces and weird posed faces can increase the productivity of quality faces which are difficult to be classified as unreal. Adam optimization algorithm can be used instead of gradient descent as it can reduce the training time period of GANs by combining the best properties of AdaGrad and RMSProp. Also, further reduction of hyperparameters can increase the efficiency and reduce the training time of GANs. The type of GAN having quickly computed loss function with the least number of hyperparameters required for its working is the best-suited alternative.

REFERENCES

[1] Akhtaruzzaman, M. & Shafie, A. A. Evolution of humanoid robot and contribution of various countries in advancing the research and development of the platform. ICCAS 2010 - Int. Conf. Control. Autom. Syst. 1021–1028 (2010).

[2] Hitawala, S. Comparative Study on Generative Adversarial Networks. (2018).

[3] Goodfellow, I. J. et al. Generative Adversarial Networks. 1–9 (2014).

[4] Wang, Z., She, Q. & Ward, T. E. Generative Adversarial Networks: A Survey and Taxonomy. 1–16 (2019).

[5] Schmidhuber, J. Deep Learning in NN: An Overview. Expert Syst. Appl., vol.72, 221–230 (2017).

[6] Ma, F., Ayaz, U. & Karaman, S. Invertibility of convolutional generative networks from partial measurements. Adv. Neural Inf. Process. Syst. 2018-Decem, 9628–9637 (2018).

[7] Chaplygin, A. & Chacksfield, J. Stochastic Deconvolutional Neural Network Ensemble Training on Generative Pseudo-Adversarial Networks. (2018).

[8] Gauthier, J. Conditional generative adversarial nets for convolutional face generation. Tech. Rep. (2014).

[9] Kaneko, T. Generative Adversarial Networksの基礎と応用. 日本音響学会誌, vol.74, 208–218 (2018).

[10] Lu, L. & Crunch, L. L. 4 Years of Generative Adversarial Networks (GANs) Chinese New Year 2018 Best Wishes for the Year of the Dog. (2018).

[11] Huang, H., Yu, P. S. & Wang, C. An Introduction to Image Synthesis with Generative Adversarial Nets. 1–17 (2018).

[12] Shen, W. & Liu, R. Learning residual images for face attribute manipulation. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 1225–1233 (2017).

[13] Antipov, G., Baccouche, M. & Dugelay, J. L. Face aging with conditional generative adversarial networks. Proc. - Int. Conf. Image Process. ICIP 2017-Septe, 2089–2093 (2018).

[14] Liu, S. et al. Face aging with contextual generative adversarial nets. MM 2017 - Proc. 2017 ACM Multimed. Conf. 82–90 (2017). doi:10.1145/3123266.3123431

[15] Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 105–114 (2017).

[16] Fu, Q., Hsu, W.-T. & Yang, M.-H. Colorization Using ConvNet and GAN. arXiv (2017).

[17] Isola, P., Zhu, J. Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 5967–5976 (2017).

[18] Chapagain, A. DCGAN--Image Generation. (2019). doi:10.13140/RG.2.2.23087.79523

[19] Osindero, S. Conditional GAN Tutorial (2014).

[20] Mao, X. et al. Least Squares Generative Adversarial Networks. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, 2813–2821 (2017).

[21] Odena, A., Olah, C. & Shlens, J. Conditional image synthesis with auxiliary classifier gans. 34th Int. Conf. Mach. Learn. ICML 2017, vol.6, 4043–4055 (2017).

[22] Chen, X. et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. Adv. Neural Inf. Process. Syst. 2180–2188 (2016).

IJSER