GAN: Era of Image to Image translation

Devidatt Joshi
Computer Science Engineering
JNV University
Jodhpur, Rajasthan
Montujoshi0@gmail.com

Dr. Anil gupta
Computer Science Engineering
JNV University
Jodhpur, Rajasthan
anilgupta@jnvu.edu.in

***Abstract*—Image to image translation is a new sensation in the field of machine learning, artificial intelligence, and vision communities. The principle of an image-to-image translation is to convert an image of the source domain to a target domain image. Recent innovations and techniques have shown remarkable success in the field of an image to image translation for two domains. But the existing methods have limited functionality and methods when we perform translation between more than two domain images. Here, GAN (Generative Adversarial Network) plays a vital role, GAN provides us functional approach that can perform image-to-image translation between multiple domains using a single conversation model. This paper empirically discusses and analyzes image-to-image translation using GAN.**

***Index Terms*—Image-to-Image translation, GAN, Generative models, Image datasets.**

I. INTRODUCTION

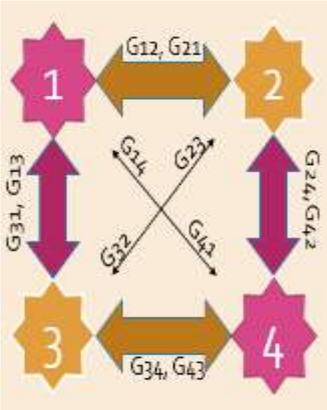
With the rapid growth in the field of computer vision and graphics community, the task and approaches of image reproduction or image translation gain the attention of developers because of the availability of large datasets & extraordinary algorithms. GAN is a powerful technique of deep learning models that use extracted features of images Across the several computer application like medical imaginary[1,2], image classification[3], face detection[4] & object detection[5].The process of image-to-image translation is converted or translate complete aspect of an image with reference to another image. GAN is a powerful approach that can easily convert or change attributes of an image like hair

Figure:1 An example of an image to image translation, Here input image are converted into multiple different domains like a cat, tiger, lion & gorilla. But output images are not sharps due to the limited functionality & scalability of existing methods.

color[6], facial expression, and change scenery of images[7]. In this proposed method we train our model with two different kinds of domain images. Then using the GAN approach we render images from domain to domain. Here we frequently use the following terms:

1. Attributes: feature of an image is known as attributes like an image of a man contains attributes:- hair color, gender, age.
2. Attributes value: Each and every attribute of an image contains some meaningful value that value is known as an attribute value like attribute gender contains male & female as an attribute value.
3. Image domain: Group of images that lies on the same set of attributes are denoted as a domain. For example, images of men & women represent two different domains.

Existing methods are inexpert and untrained in multiple image translation. When we translate an image to multiple domains we required generators. The role of generators in image translation is to map images in multiple different domains. Existing algorithms required n(n-1) generators. Here n is the number of domains.



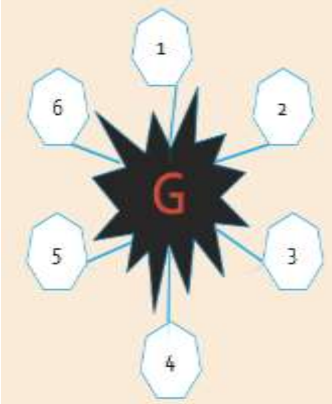


Figure:2 shows the comparison between existing generators and GAN generator. To map images between 4 domains existing method requires 12 generators(Left image), on other hand, GAN uses one generator to map between 6 different domains images(Right image).

In this paper, we present a comparative result of GAN with other existing methods. GAN shows the best result over current methods.

II. RELATED WORK

The remarkable contribution is noted by GAN(Generative adversarial network) in the number of filed computer science like face image rendring[8,9], Image translation [10]. Recent work in **Image-to-Image** translation [8,9] contributes notable results. Conditional GAN (C-GAN) also trends in the research community, It uses discriminator and generator to generate images based on certain class condtions[11,12,13]. The idea is also implemented in photo editing [14]. CycleGAN & DiscoGAN [9,12] translate images by using cycle consistency loss.

III. GENERATIVE ADVERSARIAL NETWORKS

GAN is a machine learning algorithm-based generative model which is highly compatible with multiple datasets for converting images from one domain to another domain. To implement GAN we use Generator and Discriminator, aim of the generator(G) is to translate an image x to any target domain y. The basic generator function is given by the following function:

$$\boldsymbol{Generator} = \boldsymbol{G(x, l) \rightarrow y} \quad \ldots.(\mathbf{1})$$

$\boldsymbol{l}$ is a label of target domain y in equation 1. The role of the discriminator(D) is to distinguish between real and fake images generated by the generator (G). To differentiate between the real and fake image, a probability distribution is applied on input image x & discriminator classifier(cls) as following:

$$\boldsymbol{D}: \boldsymbol{x} \rightarrow \{\, \boldsymbol{D_{Input}(x)}, \boldsymbol{D_{cls}(x)} \,\} \quad \ldots.(\mathbf{2})$$

Adversarial loss is applied to the discriminator to distinguish between real and fake images.

$$\beth\boldsymbol{adv} = \boldsymbol{\varepsilon_x} [\log \boldsymbol{D_{input}(x)}] + \boldsymbol{\varepsilon_{x,l}} [\log(\mathbf{1} - \boldsymbol{D_{input}(G(x, l))})] \ldots(\mathbf{3})$$

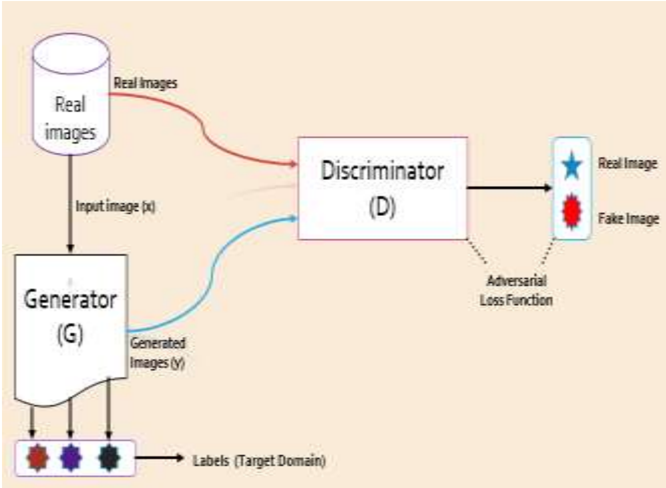


Figure:3 Blueprint of GAN, having two parallel models one is a generator(G) and second is discriminator(D). G takes input image (X) and target domain label l and converts an image into target domain (Y). Here, discriminator (D) distinguishes between real image & generated image.

For a proper translation of image x to target domain y using a label , we have to optimize both generators (G) & discriminator (D) using the following loss function:

$$\beth^{real} = \boldsymbol{\varepsilon_{x,l}} [-\log \boldsymbol{D_{cls}(l|x)}] \ldots(\mathbf{4})$$

$$\beth^{fake} = \boldsymbol{\varepsilon_{x,l}} [-\log \boldsymbol{D_{cls}(l|G(x, l))}] \ldots(\mathbf{5})$$

Real image classification loss(eq. 4) is helping to optimize discriminator D & fake image classification loss is helping to optimize generator G(eq. 5).

VI. DATASETS

(1) CelebFaces Attributes Dataset (CelebA) is a collection of large-scale celebrity faces attributes, around 200 thousand face attributes images of different celebrities are contained inside. CelebA dataset is openly available for research purposes. 202,599 face images of celebrities, each annotated with 40 binary attributes[15]. Here we use images of the CelebA dataset in form of dimension 128 x128 & use reconstruct domain of images based on attributes like age, hair color, and gender.

Figure:4 The CelebA datasets images, 202,599 face images of celebrities. *(Image Source: CelebA)*

(2) The Radboud Faces Database (RaFD) is a collection of 67 models with 8 different kinds of emotional facial expression. RaFD dataset is hosted by the Radboud University Nijmegen(the Netherlands). Here we use images of the RafD dataset in form of dimension 128 x128.

Figure:5 The RAFd datasets images example, 4,824 face *expressions of 8 models. (Image Source: RaFD)*

V. EXPERIMENT & RESULTS

In this classification experiment, we present empirical outcomes of GAN in the image to image translation from multiple datasets. All models of both datasets are trained on Adam[16]. Here, we take a value of β_1=0.4 & β_2 =0.99. batch size of images for all conversations is 16. The learning rate of the CelebA dataset is 0.0001 for the first 15 epochs & train the RaFD dataset for 100 with a 0.0001 learning rate. We use a mask vector to store the values of attributes of different datasets by which label information of both datasets very well known to each dataset. Mask vector m allows us to focus only on known label values & it ignored unspecified or unknown labels.

$$v = [\boldsymbol{v1}, \boldsymbol{v2} \ldots\ldots, \boldsymbol{vn}, \boldsymbol{m}] \quad \ldots.(\boldsymbol{6})$$

Here vi has represented a vector for i-th dataset. When we trained our GAN model with multiple datasets then domain label v defined in equation 6 is input to the generator. Using this our generator became familiar with known label values and ignore the unspecified labels.

Figure:6 Translation of the input image of domain 1(CelabA) into target domain using attributes label of domain 2(RaFD).

Our proposed model produces a high-quality image - translation as compared to other existing methods. GAN model flexibly translates image according to target domain label.

When we perform translation among attributes of a single domain like hair color, gender & age then the

following results are obtained:

Accuracy of GAN when we perform translation among hair color is **67%,** the accuracy of age translation is **69.02%** & accuracy of gender translation is **38.6%**.

Our model also empirically demonstrates translation between multiple datasets in figure 6. Mask vector also helps the GAN model to utilize attributes of multiple datasets.

VI. CONCLUSION

A flexible and smooth translation of images using GAN are demonstrated. Here, we use only a single Generator and Discriminator to implement the whole system. GAN rendered high-quality images of the source domain into the target domain. Mask vector also enhances the capability of GAN, using mask vector GAN became comapatibile with multiple datasets.

VII. References

[1] Overview of deep learning in medical imaging by- k. Suzuki, 2017, Radiol. Phys. Technol.

[2] Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale by - Lu, J.; Luo, Y.; Zhang, G, Symmetry 2018, VGG 16.

[3] Gradient-based learning applied to document recognition. Proc. By - Y.; Bottou, L.; Bengio, LeCun , 2018, IEEE

[4] You only look once: Unified, real-time object detection by- Divvala, S.; Girshick, R.Redmon, J. Farhadi, A.

[5] Computer Vision and Pattern Recognition, Las Vegas,June 2016, IEEE.

[6] Deep face liveness detection based on nonlinear diffusion using convolution neural network. Signal Image Video Process by- Mahmood, Alotaibi, 2017

[7] Learning to discover cross-domain relations with generative adversarial by- M. Cha, H. Kim, J. K. Lee, T. Kim, , and J. Kim.

[8] Unsupervised VAT(visual attribute transfer) with reconfigurable GAN by M. Cha, J. Kim & T. Kim, B. Kim.

[9] Deep identity-aware transfer of facial attributes by- W. Zuo, and D. Zhang & M. Li,.

[10] Image-to-image translation with conditional adversarial networks by- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, IEEE,CVPR, 2017.

[11] Conditional generative adversarial nets by M. Mirza and S. Osindero,2014.

[12] Semi-supervised learning with generative adversarial networks by A. Odena, 2016.

[13] Conditional image synthesis with auxiliary classifier gans by and J. Shlens , A. Odena & C. Olah,. 2016.

[14] Unsupervised cross-domain image generation by- Y. Taigman, A. Polyak, & L. Wolf., IEEE seminar - CVPR, 2018

[15] Computer Vision and Pattern Recognition-IEEE, 2018 -Paper on- GAN for Multi-Domain datasets Image-to-Image Translation by - Jung-Woo Ha, Yunjey Choi.

[16] A method for SO(stochastic optimization): ADAM approch by D. Kingma and J. Ba 2014

[17] Attention-gan for object transfiguration in wild images by- Yang, X.; Tao, D.Chen, X.; Xu, C.; September 2018,(ECCV) Germany.

[18] domain generative adversarial network for multi-scale face hallucination by- Huang, H.; He, R.; Sun, Z.; Tan, T. Wavelet. Int. J, 2019. Comput. Vis

[19] How good is my GAN? By- Shmelkov, K.; Schmid, C.; Alahari, K. September 2018,(ECCV) Germany.