© February 2025 | IJIRT | Volume 11 Issue 9 | ISSN: 2349-6002

Comprehensive Survey On Forensic Sketch Transformation To Realistic Image

Onkar Nigdikar[1], Manisha Darak[2], Harsh Patil[3], Aakash Patil[4], Aniket Margale[5]
[1,2,3,4,5]*Department of Computer Science, RMD Sinhgad College of Engineering Pune, India*

***Abstract*—In this survey paper, we overview the recent advances of the sketch-to-photo transformation especially using deep learning techniques with basic concepts from computer vision and generative modeling. The paper includes details regarding how sketch to photo tasks is accomplished using various techniques like image translation as well as use of Generative Adversarial Networks (GANs) in this context. Transforming sketch drawing into clear and photo-realistic images to assist with crime solving has become important issue when it comes to identify criminals based on limited visual data. We will examine the main techniques used and also discuss their advantages, limitations and suitability for real world application. The purpose of this paper is to combine much of available experience and knowledgebase.**

***Index Terms*—Cycle Generative Adversarial Network (CycleGAN), Deep Convolutional Generative Adversarial Network (DCGAN), Generative Adversarial Network (GAN), Pix2Pix.**

I. INTRODUCTION

Forensic Sketch-to-image generation can be a important tool for identifying suspects. Traditionally, a witness describes someone to forensic artist who sketches based on description of suspect. But this process heavily depends upon skills of artist and how well the witness remembers the face of suspect. Minor details can be remembered incorrectly which makes the sketches less accurate.

Generative Adversarial Networks(GANs) can be used to transform Sketches in to realistic images. GANs work by having two neural networks which compete with each other. GANs consist of Generator and Discriminator[9], generator generates image and discriminator tries to differentiate between real and generated image. A specific GAN model called pix2pix is useful for converting sketches in to images. pix2pix uses supervised dataset for its training[1]. Another model like CycleGAN which can be trained on unsupervised dataset. which works with unpaired data but may struggle to keep details consistent, DCGAN is another type of GAN which can be used for this sketch-to-image translation purpose but can suffer from problem of overfitting.

Therefore, advancements in this field can help forensic department to turn through rough sketches in to realistic images, speeding up investigations and improving accuracy in identifying individuals.

II. METHODOLOGIES

A. Generative Adversarial Network(GAN)

GAN was invented by Ian Good fellow [9] which comprises of 2 nodes as shown in fig.1 The node D is Discriminator which is simply a classifier which tries to differentiate between actual image and factual image. It acquires input from two sources one from sample of factual images and another generated images from node G. While Node G generate images which will be given to node D as input. It is two player competition. The discriminator acts as cop and generator acts as forger which generates mock photo from forensic sketch and second net discriminator (Police) tries to differentiate between real and copy image generated by generator. In the game both the networks generator and discriminator try to improve their methods until generated images becomes indistinguishable from real image until game comes to an end and at last generator fool's discriminator by exactly creating similar image as real image.

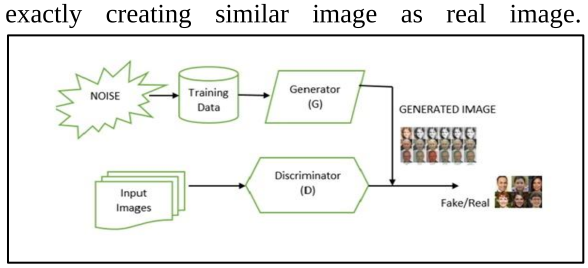

Fig. 1. GAN Architecture [3]

B. Deep Convolutional Adversarial Network (DCGAN)

The Concept of DCGAN was first suggested by Radford et al. in 2015 There are two nodes in this DCGAN architecture both the nodes G and as well as D uses a deep convolutional network architecture, the generator (G) which uses transpose convolutions instead of forward convolutions. Techniques like Batch Normalization and Leaky Rectified Linear unit are used to enhance the performance of DCGAN. The basic concept of transforming sketch to photo is done as follow. The Generator takes sketch as input and generates realistic fake photographs as an output and gives it to discriminator.

A Discriminator is used to train the generator to increase the accuracy of realistic fake photographs which are being generated by generator. Both the networks are trained at same time. At some point of time both the networks become accurate enough in achieving their goal. The discriminator becomes smarter in differentiating between real and fake image and generator becomes smarter in generating more realistic photos that discriminator is unable to find difference between real and fake image.

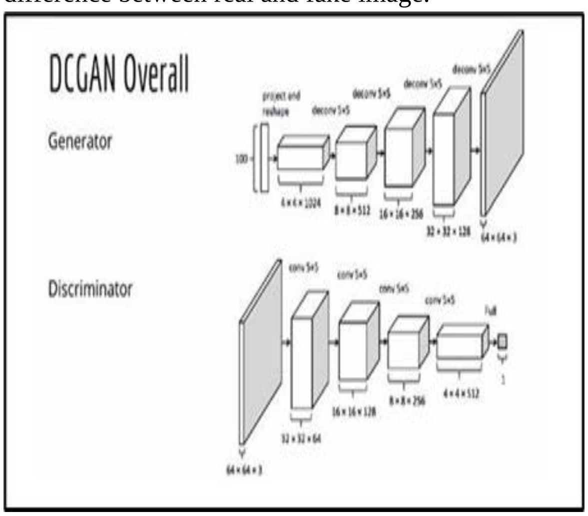

Fig. 2. DCGAN Architecture [3]

C. Cycle Generative Adversaeial Network (CycleGAN)

CycleGAN is one type of deep learning architecture designed to unpaired image-to-image translation. Basically, it means that it can change images in one domain into another through an unsupervised learning method, assuming the pairs do not come together readily in real world data; rather, we find either side. Here, there is the approach through two generator models: the process includes adversarial training and generates a transform between a given source image in the one domain to its related version in another.

The two major networks present in CycleGAN are generator and discriminator. The former transforms images in the source domain into images in the target domain, while the latter evaluates the generated images as real or fake. The model consists of two cycles: one to convert images from domain A into domain B and another for converting them back from domain B into domain A [2]. The cycle consistency in this cycle ensures that if an image is translated to the other domain and then back again, it should look similar to the original image [2].

At the training, CycleGAN uses two types of loss: adversarial loss that urges the generator to generate pictures that have a probability greater than random noise in discriminating with the discriminator; cycle consistency loss, forcing the restored images to correspond to their original inputs, if going through a round-trip over both generators. The latter allows learning mapping between the domains efficiently and takes account of the underlining feature and character of both to handle the issue of the unpaired data.

D. Pix2Pix

Pix2pix is a type of conditional GAN (CGAN). Generation of image is done by Generator and production of target image is dependent on source image. The Generator and Discriminator comprise the network. The Generator generates the picture from the input image. The discriminator compares the generated image with real image and tries to differentiate between them. In Pix2pix U-net architecture is used as Generator and PatchGAN is used as discriminator. Discriminator is simply a classifier which is trained before training of generator. U-net architecture comprises of encoder and decoder, encoder does down sampling in which it extracts important features. Decoder does up sampling which can be done by using trans convolutions. Skip connections are connected between encoder and decoder. The Discriminator in Pix2pix which is PatchGAN evaluates image patches instead of entire image. Pix2pix combines adversarial loss (from the GAN) with L1 loss between generated and real image [1].

III. LITERATURE REVIEW

Methodology	Description	Strength	Limitation	Performance Metrics
Pix2pix [1]	A conditional GAN model that learns a mapping from paired input to output images through supervised training. It uses U-Net as the generator and PatchGAN as the discriminator.	Produces realistic image details and fine textures; effective with paired training data	Require paired data, which can be limited in real world forensic cases; may produce blurred edges without additional refinements	SSIM:78.60%
DCGAN [3]	A GAN model utilizing convolutional layers, known for generating high-quality images with a generator and discriminator network. Enhanced with dense layers for realistic face generation from sketches.	Effective in generating high-resolution images; robust for sketch-to-photo synthesis tasks in constrained environments	Limited by its reliance on structured data; may struggle with capturing subtle details without additional training enhancements	SSIM: 0.587
CycleGAN [2]	An unpaired image-to-image translation model based on GANs, enabling transformations without paired data using cycle consistency loss.	Ideal for unpaired datasets; learns transformations even without direct image mappings	Can produce artifacts and inaccurate results in detailed or complex forensic sketches	SSIM: 0.345

Raghavendra Shetty Mandara Kirimanjeshwara, Sarappadi Narasimha Prasad suggested a method to convert Sketches in to photo realistic images using Pix2pix architecture. In that research they have used 3 different datasets which are The Chinese University of Hong Kong (CUHK) and Indraprastha Institute of Information Technology Delhi's (IIIT-D's) students face sketch. They used 88 subjects from the CUHK student dataset as a training set and the remaining 518 subjects as a testing set, which consists of 123 images from the augmented reality (AR) dataset, 295 images from the XM2VTS dataset, and the remaining 100 images from the CUHK student dataset and corresponding digital images. Extensive paired data experiments prove that their strategy outperforms the alternatives they investigated. Their suggested method has been shown to produce a pixel accuracy of 82.7% in experimental settings.

Nischal Tonthanahal, Sourab B R, Dr Sharon Christa proposed a method to transform sketches to realistic images using cycleGAN. They used cycleGAN as their base model with changed hyperparameters. They used Pytorch to train the model on CHUK student dataset which has 88 images and IIIT-D student and staff dataset which has 72 images where input image is of 256 X 256 pixel size. The CycleGAN model is trained for 294 epochs. In the Sketch-to-Face transformation, the obtained face photos map the sketches accurately. It is seen that the facial tone is sensitive to the shades provided in sketches.

Sreedev Devakumar proposed a method which transforms forensic sketch in to realistic image using DCGAN. The dataset used in the proposed method is CHUK face sketch FERET Database(CUFSF). The Structural Similarity Index (SSIM) of their proposed DCGAN is found to be 0.587 in their research. Their proposed method has been compared with the other methods in terms of SSIM and identification rate. After their addition of dense layer, the SSIM and identification rate of the proposed method is higher in comparison with the previous techniques.

IV. CONCLUSION

Forensic Sketch-to-image translation remains a challenging yet promising area in field of computer vision. Our survey explored several key methodologies like Pix2pix, DCGAN and CycleGAN. Pix2pix, with its supervised learning approach , excels in controlled datasets but may struggle with the diverse range of real-world forensic sketches. DCGAN can generate high quality images ,shows promise but requires further adaptation for sketch based applications. CycleGAN, with its unsupervised learning paradigm, provides greater flexibility in unpaired data translation. Overall, future work could focus on enhancing image quality and robustness and it could improve image realism to further bridge the gap between forensic sketches and photographic images.

REFERENCES

[1] Raghavendra Shetty Mandara Kirimanjeshwara, Sarappadi Narasimha Prasad, "Photo-realistic photo synthesis using improved conditional generative adversarial networks", Apr. 2023, DOI: 10.11591/ijai.v13.i1.pp516-523

[2] Nischal Tonthanahal, Sourab B R1, Dr Sharon Christa, " Forensic Sketch-to-Face Image Transformation Using CycleGAN ",International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 07 ,July 2021.

[3] Sreedev Devakumar, Greeshma Sarath, "Forensic Sketch to Real Image Using DCGAN", 2023, Available online at www.sciencedirect.com

[4] Zhenglin Li,Bo Guan,Yiming Zhou,Yuanzhou Wei,Jingyu Zhang and Jinxin Xu, "Mapping New Realities: Ground Truth Image Creation with Pix2Pix Image-to-Image Translation", May2024, arXiv:2404.19265v2

[5] M.Shetty K. Raghavendra, Prasad Narasimha Sarappadi, "TRANSFER LEARNING WITH PIX2PIX GAN FOR GENERATING REALISTIC PHOTOGRAPHS FROM VIEWED SKETCH ARTS", Aug.2022,DOI : 10.35741/issn.0258-2724.57.4.17

[6] JIAN ZHAO , XIE XIE, LIN WANG , MENG CAO,MIAO ZHANG, "Generating Photographic Faces From theSketch Guided by Attribute Using GAN", March 5, 2019, DOI : 10.1109/ACCESS.2019.2899466

[7] Nishiket Waghmode, Pravin Bansode, Digambar Chalkapure and Ms. Uttara Varade, "Solve the Mystery: DCGAN-Based Sketch to Real Face Conversion", July 2024, https://doi.org/10.47392/IRJAEM.2024.0328

[8] Xingfang Yuan,"Translation from sketch to realistic photo based on CycleGAN"2023,DOI: 10.54254/2755-2721/50/20241168 Phillip Isola, Jun-Yan Zhu ,Tinghui Zhou Alexei,A. Efros,"Image-to-Image Translation with Conditional Adversarial Networks",26 Nov 2023 https://www.tensorflow.org/

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), 2672–2680