Sketch to Image Translation with Generative Adversarial Networks

[1]Sandeep S. Kumar, [2]Ashutosh B. Jena, [3]Robin B. Patoliya, [4]Dr. Jitendra H. Saturwar
[1,2,3]Student, [4]Head Of Department,
Department of Computer Engineering,
Universal College of Engineering, Vasai, India.

Abstract : With the advancements in generation and its technology, humans are finding ways to make their lives faster, efficient and hassle-free. Graphics is a way of portraying text with ease and better understandability. Blender, Adobe Photoshop, Coreldraw, etc. are still used as conventional ways to render graphics. However, the time and effort required to render files in these software is very high. The learning curve of these software is also a limitation for a naive user. These software also need systems with technical specification that are graphic focused and uncompromisable. This professional software also come with a high price tag that is not affordable by most of the users. Besides, these software are not explicitly crafted to convert sketches into equivalent images. Hence they are not a logical solution for converting sketches into images. The proposed system makes the use of Generative Adversarial Networks to overcome the limitations of the existing systems. A generative adversarial network is a class of machine learning algorithms. Two or more neural networks rencounter with one another in a zero sum game framework. At the beginning, the system uses a feature extraction process to recognize the object. Once the object is recognized, the algorithms fill the outlines of the sketch to generate a number of results. These results are then evaluated against the ground truth to produce an output that is as closely accurate to the training data that was fed onto the system during the training process. The system tends to achieve a high level of accuracy in recognizing and filling user fed sketches into realistic viable images. This would eliminate the need of using intensive aforementioned software, thereby improving the user's experience and saving their time and efforts.

Keywords - Sketch to image, Image processing, Generative Adversarial Networks, Graphics rendering, Object recognition, Face Synthesis.

I. INTRODUCTION

With the increasing advancements in technology in this dynamic world, there is a dire need for automation. With the increasing workload on humans everyday, there is a need to inculcate machines to work alongside man. This would result in increased efficiency, brisk work and amplified capabilities. Sketch-to-image translation is an application of image processing that can be utilized as some assistance in different fields. One of which being the use of Generative Adversarial Networks to map edges to photos, with the aid of image generators and discriminators, that work hand-in-hand to produce realistic-looking images. This model can be modified and implemented as software in multiple programs related to image processing.

The system articulates a method for sketch-to-image translation with the use of Generative Adversarial Networks. The project is divided into several modules, in which a user uploads the sketch which then gets transformed into a naturalistic image. The system is trained by uploading the ground truth. In this training process, the system makes use of GANs. A large number of possible outcomes are generated by the system. Subsequently these GANs dispute among different potential outcomes to generate the most conceivable and feasible output nearest to the ground truth. The ground truth is a real world image that the system developer provides, and after training, the machine is expected to anticipate. These modules can be integrated in various aspects of society which include investigation and analysis purposes.

II. LITERATURE REVIEW

The following research articles are chosen for review, keeping in mind the conservative strategy of sketch to image translation.

Chao Feng Chen, suggested an innovative face sketch synthesis technique that takes inspiration from the process of drawing sketches used by artists. The content network precisely portrayed the layout of the face and the sketches were referred for giving a style to the final sketch generated by the framework. It was observed that the running time was high and there was an absence of real time face sketch synthesis. [1]

Hadi Kazemi, articulated a system for unsupervised sketches to face synthesis using geometry of the face. Distinct features of the face were taken into account while computing the geometrical distance between various facial parts. The concluding image was produced by these computed distances. Textures of faces can be used while generating the final image for a more precise prediction. [2]

Phillip Isola, suggested a framework for image to image translation with the assistance of Generative Adversarial Networks to predict outcomes that were exceptionally close to the ground truth. The system had remarkably high complexity due to a generalised approach towards image translation. Inaccuracy of object acknowledgement resulted in false outputs since texture of the surface was not taken into account. [3]

Christian Galea, introduced an automated matching of software generated sketches to realistic pictures. Sketches were software generated which made it simpler for the system to match them to the real face photos since most of the features remained untouched. However, having only one sketch per photo for matching resulted in unsatisfactory results. [4]

Kokila R, articulated a study on matching sketches to mugshot photos for the purpose of investigation. The complexity of the system was significantly low since it was very application centric. A number of sketches were compared to mugshot photos taken from different angles of the face resulting in very accurate outputs. The system was thoroughly dependent on the quality of the sketches that were used, which resulted in mediocre results for inferior quality sketches. [5]

Xing Di, introduced a deep generative framework for the recreation of facial images using the visual attributes. Their technique generated photorealistic images using an intermediate representation. The framework incorporated GANs and VAEs. However, the system produced imprecise results. [6]

III. PROPOSED SYSTEM

Since the conventional methods for converting sketches to images are not efficient, GANs are used to convert sketches to faces by replacing the existing techniques like Photoshop, Illustrator, etc. Different ways to convert a sketch to respective images were explored and implemented. The proposed framework outlives the conventional sketch-to-image transformation capabilities in terms of better time-complexity, automation, lessening human-dependencies and much more. Hence, it combines various components to execute the operations and meet the above benefits. During the training phase, the framework is provided with data to perform the learning process. The sketch is fed to the machine, which with the guidance of the discriminator and generator, learns and finds the image which is the best match to a feasible object. The system then produces a legal object which is compared with the optimal outcome to argue the accuracy from which the system adapts. The result is stored on the system as a part of training and testing. At the time of testing, the user uploads a sketch with the help of a user interface, which causes the GANs to utilise the learned knowledge to first figure out the outlines of the object and then imparts texture on the object. The sketches are translated using Generative Adversarial Networks (GANs).

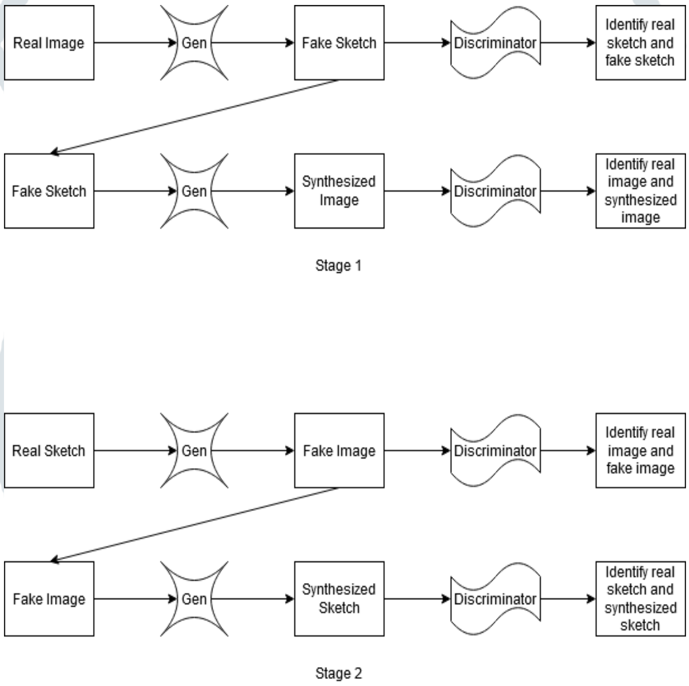

Fig. 1. System Architecture

Let us consider a dataset U, which comprises photos and their corresponding sketches. Let us represent these photos and sketches by P_i and S_i respectively where i is the index number of the image being referred to ranging from 1 to N in which N denotes the total number of photos or sketches contained in the dataset U. The goal of the system is to learn in two stages:
(i) Photo to Sketch synthesis using $S = f_{PS}(P)$
(ii) Sketch to Photo synthesis using $P = f_{SP}(S)$

As seen here, both forward i.e. photo to sketch operation and backward i.e. sketch to photo operation retain equal significance. Therefore, CycleGAN can be used to resolve this problem.
This system consists of two generator networks, $Gen_P$and Gen_S, that aid in converting image to sketch and sketch to image respectively.

Stage (i) : Photo to Sketch synthesis
Considering Gen_P, this generator takes the real photo, i.e. the ground truth, $Real_P$, and converts it into a fake sketch $Fake_S$. Similarly, Gen_Stakes this fake sketch $Fake_S$which was the output previously, and tries to reconstruct an image from this sketch denoted by Syn_P.
This can be expressed as:

$$Fake_S = Gen_P(Real_P)$$

$$Syn_P = Gen_S(Fake_S)$$

Stage (ii) : Sketch to Photo synthesis

Now Gen_S, takes the real sketch of the ground truth, $Real_S$ and converts it into a fake image $Fake_P$. Similarly, the second generator, Gen_P takes this fake image $Fake_P$ which was the output previously, and tries to reconstruct a sketch from this image denoted by Syn_S.

This can be expressed as:

$$Fake_P = Gen_S(Real_S)$$
$$Syn_S = Gen_P(Fake_P)$$

The generators in the GAN network (Gen_P and Gen_S) use the adversarial losses to train themselves that is provided by the discriminator. The loss function in each stage is described below.

Stage 1:

$L_{GAN} = [log\ D_p(S)] + \Sigma_{P\sim photo(P)}[log\ (1 - D_p(Fake_s))]$

Stage 2:

$L_{GAN} = [log\ D_p(P)] + \Sigma_{S\sim sketch(S)}[log\ (1 - D_S(Fake_P))]$

Following is the module of the system:

A. Sketch Upload - Image Output Module

At first, the user provides a sketch onto the system, which with the assistance of GANs, maps the sketch onto images and narrows down the best fitting image as per the ground truth. This has two sides - Generator and Discriminator. Generator produces real and fake results, and then tries to make the fake result as close to real as possible. The function of the discriminator is to identify the real result from the pool of results. This is composed by the generator involving both, real and fake results. The desired output is then obtained by the user.

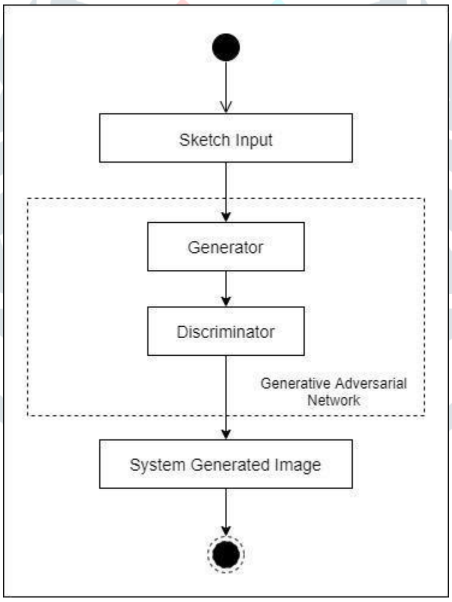

Fig. 2. Sketch Upload - Image Output Module

IV. Results And Discussions

This section illustrates the output produced by the system using Generative Adversarial Networks.

Following are the screenshots in a systematic manner:

Fig. 3. Implementation of Sketch to Image translation

Figure 3 portrays three images, (i) the ground truth, (ii) the user provided sketch and (iii) the system generated output. During the training phase, images (i) and (ii) were provided to the system, which uses GANs to learn the facial features by comparing the two images. Consequently, while testing only the image (ii) i.e. the sketch was provided using which the system generates the image (iii) i.e. the system generated facial image.

V. Analysis

When the existing systems were compared to the proposed system, it was found that the cost of implementation of the existing systems was higher than that of the proposed system. The proposed system was easy to use compared to the existing system which had a steep learning curve. The pace of output generation was more in the proposed system, hence it took less time to generate the output. The quality of output generated by the existing system was highly dependent on human skills. However that is not the case for the proposed system. The proposed system is scarcely dependent on human skills and efforts.

Following are the MSE and SSIM results of a few outputs:

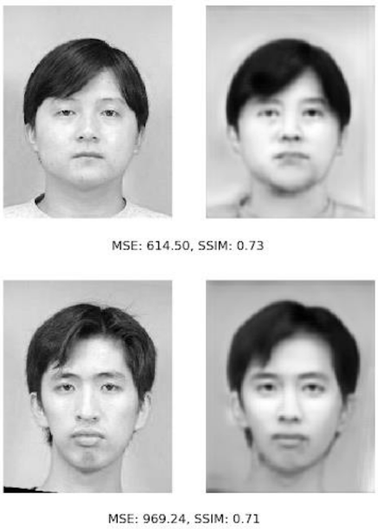

Fig. 4. Efficiency of Generative Adversarial Networks

V. Conclusion

The outcome of this project advocates that using GANs for tasks such as image-to-image translation can be quite efficient. The loss generated by the network during the whole process is used to adapt to the data of the task in order to work in multiple scenarios. Current automated frameworks have realized functionalities that can be used for different purposes.

The proposed framework can be enhanced by refining the framework's ability to recognize the outlines of the object. The absence of texture identification in the proposed system renders the framework in having trouble to recognize the objects correctly. The result is subject to factors such as amount of noise present in the sketch, boundaries of the sketch as well as the precision of the sketch. Due to these factors, the output can be undesirable at times. The observations and findings are at an introductory stage and additional study would disclose the benefits and detriments in an adequate manner.

REFERENCES

[1] Chaofeng Chen, Kwan-Yee K. Wong, Xiao Tan, “Face Sketch Synthesis with Style Transfer using Pyramid Column Feature”, 2018.

[2] Hadi Kazemi, Nasser M. Nasrabadi, Fariborz Taherkhani, “Unsupervised Facial Geometry Learning for Sketch to Photo Synthesis”, 2018, pp.3-9.

[3] Phillip Isola, Jun-Yan Zhu, Alexei A. Efros, Tinghui Zhou, “Image-to-Image Translation with Conditional Adversarial Networks”, 2018.

[4] Christian Galea, Reuben A. Farrugia, “Matching Software-Generated Sketches to Face Photos with a Very Deep CNN, Morphed Faces and Transfer Learning”, 2017, pp. 1-10.

[5] Kokila R, Abhir Bhandary, Sannidhan M S, “A Study and Analysis of Various Techniques to Match Sketches to Mugshot Photos”, 2017, pp. 41-43.

[6] Xing Di, Vishal M. Patel, Vishwanath A. Sindagi, “Gender Preserving GAN for Synthesizing Faces from Landmarks”, 2018.

[7] Yue Zhang, Guoyao Su, Jie Yang, Yonggang Qi, “Unpaired Image-to-Sketch Translation Network for Sketch Synthesis”, 2019.

[8] Xing Di, Vishal M. Patel, “Face Synthesis from Visual Attributes via Sketch using Conditional VAEs and GANs”, 2017.

[9] I. Goodfellow, M. Mirza, J. Pouget-Abadie, B. Xu, S. Ozair, D.Warde-Farley, A. Courville, and Y. Bengio, “Generative adversarial nets”, 2014.

[10] W. Zhang, X. Wang and X. Tang. Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition. “Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)”, 2011.

[11] X. Wang and X. Tang, “Face Photo-Sketch Synthesis and Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31, 2009.

[12] Jitendra Saturwar, Dr. D.N.Chaudhari, “Secure Visual Secret Sharing Scheme for Color Images Using Visual Cryptography and Digital Watermarking”, Published in IEEE digital explore ISBN No. 978-1-5090-3239-6/17 ©2017 IEEE. doi: 10.1109/ICECCT.2017.8117849

[13] J.H.Saturwar, Dr. D.N.Chaudhari, “ Review of models, issues and applications of digital watermarking based on visual cryptography”, Published in IEEE digital explore ISBN No. 978-1-5090-4715-4/17 ©2017 IEEE.

[14] Jitendra Saturwar, Dr. D.N.Chaudhari, “Deciding meaningful number of shares in digital watermarking scheme for secret images”, Published in IEEE digital explore ISBN No. 978-1-5090-4620-1 ©2016 IEEE. doi: 10.1109/SCOPES.2016.7955504

[15] Jitendra Saturwar, Dr. D.N.Chaudhari, “Optimal Number Of Shares For Digital Water-marking Scheme Using Visual Cryptography”, International Journal For Research In Emerging Science And Technology,Volume-2, Issue -1, MARCH-2015, E-ISSN: 2349-7610 pp 407-412

[16] J.H.Saturwar, Dr. D.N.Chaudhari, “Performance Evaluation of watermarking Schemes to Decide Meaningful Number of Shares”, Published in International Journal of Technology & Engineering ISSN 2455-4480 pp. 93-100.2017

[17] Jitendra Saturwar, Dr. D.N.Chaudhari, “ Optimal Number Of Shares For Digital Water-marking Scheme Using Visual Cryptography”, International Journal of Scientific & Engineering Research, Volume 4, Issue 12, ISSN 2229-5518 December-2013 pp 412-418

[18] J.H.Saturwar, Dr. D.N.Chaudhari, “ Analysis of Algorithms to Find Number of Shares For Digital Watermarking Scheme Using Visual Cryptography ”, Published in international Journal of Research In Science & Engineering Volume 2 Issue 1 e-ISSN: 2394-8299 p-ISSN: 2394-8280 ©2015 IJRISE.