International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-9 Issue-1, November 2019

From 2D Sketches to Photo-Realistic Images using Generative Adversarial Networks

Ekta M. Upadhyay

Abstract*: With increasing technological advancements, there is a need for automation in this ever-evolving world. This may result in improved efficiency, faster work and enhanced capabilities. Sketch-to-image translation is an image processing application that can be used as a helping hand in a variety of fields. One of these is the utilization of Generative Adversarial Networks to guide edges to photographs, with the assistance of image generators and discriminators who work connected to produce realistic images. We have also incorporated Histogram of Oriented Gradients (HOG) as a feature/image descriptor. The HOG technique counts the gradient orientations to differentiate the target and the background. Support Vector Machine (SVM) is the classifier used for classification. This HOG and SVM model can be improved, altered and executed as multi-program software.

***Keywords*: sketch, photo realistic image, Generative Adversarial Network, Histogram of Oriented Gradients**

I. INTRODUCTION

Advancement is an absolute necessity for the cutting edge time that we are living in. Machines that can computerize human assignments are profoundly favored in the present day and age. Image to image translation has a tremendous degree for the ages to come. Image processing assignments that used to expend immense measure of time should be possible in the blink of an eye with the headways in image synthesis. Real images can be rendered effortlessly without utilizing a lot of human help. The utilization of picture manipulation and editors can be made simple with the aid of automation. Gaming industry flourishes to accomplish reasonable designs. Advances in image processing tasks makes generation of real images bother free without the need of high level hardware and software. Image translation can be used for the purpose of inspection in many cases. Sketches can be converted into photo-realistic images for the purpose of fact finding.

The framework proposes a system for sketch-to-image mapping by the utilization of Generative Adversarial Networks. One neural system, called the generator, creates new data cases, while the other, the discriminator, assesses them for authenticity; for example the discriminator chooses whether each example of information that it surveys has a place with the real dataset or not [1]. The generator is making new, images that it goes to the discriminator. It does as such in the expectations that they, as well, will be regarded true, despite the fact that they are fake images.

Revised Manuscript Received on November 05, 2019.

***** Correspondence Author

Dr. Ekta M. Upadhyay*, School of Computer Science, University of Petroleum & Energy Studies, Dehradun, India. Email: eupadhyay82@gmail.com

The objective of the discriminator is to recognize images originating from the generator as fake. In GAN, the generator takes in random numbers and returns an image. This produced image is input into the discriminator along with a series of images taken from the real, ground-truth dataset. The discriminator takes in both genuine and phony images and returns probabilities between 0 and 1, with 1 to a forecast of credibility and 0 to counterfeit.

The proposed work is divided into various modules, to which a user uploads the sketch to be converted into a photo-realistic image. The ground truth is uploaded for training the system. GANs are used for training the system. The system generates a large number of possible outcomes. Then these GANs dispute among various possible outcomes to generate the most possible and feasible output closest to the ground truth. The ground truth is the reality that the system developer establishes, and after training, the machine is expected to predict. These modules can be incorporated in various aspects of society such as investigation and analysis purposes.

II. RELATED WORK

The conventional systems used today to generate images are usually time and space consuming. To overcome this wastage, modules are been implemented for automation. Existing systems also come with a learning curve, which make it difficult and become a hurdle for a new user to work with the software. The hardware accompanying these kind of software is usually high and demanding. The technical requirements of the systems that operate these rendering software are graphic focused and uncompromisable. These software often come with a price tag which make it not-so-affordable for every user of the platform. Additionally, these software aren't specifically designed to convert sketches to corresponding images. Hence they do not stand logical for translating sketches into images.

Hadi Kazemi, worked on unsupervised sketch to face synthesis using facial geometry. Different features of the face were taken into consideration while calculating the geometrical distance between various parts of the face. The resultant image was produced by these calculated distances. Facial texture can be used while outputting the final image for a more accurate prediction. [2]

Phillip Isola, proposed a system for image to image translation with the help of Generative Adversarial Networks to predict results that were very close to the ground truth. The system had high complexity due to a generalized approach towards image translation. Inaccuracy of object recognition resulted in false outputs since texture of the surface was not taken into consideration. [3]

Christian Galea, worked on automated matching of software generated sketches to face photos. Sketches were software generated which made it easier for the system to match them to the actual face photos since most of the features remained unchanged. However, having only one sketch per photo for matching resulted in unsatisfactory results. [4]

Kokila R, proposed a study on matching sketches to mugshot photos for investigation purposes. The complexity of the system was quite low since it was very application specific. A number of sketches were matched to mugshot photos taken from various angles of the face resulting in very accurate outputs. The system was highly dependent on the quality of the sketches that were used providing unsatisfactory results for lower quality sketches. [5]

III. PROPOSED SYSTEM

The proposed system ranks over the traditional sketch-to-image transformation techniques in terms of better time-complexity, automation, reducing human-dependencies and much more. Hence, it integrates various components to proceed with the operations and meet the above advantages. During the training phase, the machine is fed data to carry out the learning process. The sketch is provided to the machine, which with the help of discriminator and generator learn and finds the image which is best fit to match a legal object. The system then generate a legal object which is compared with the ideal result to argue the accuracy from which the system adapts. The result is stored in the database as a part of training and testing. During the testing phase, the user uploads a sketch through a user interface, which causes the GANs to apply the learned knowledge to first figure out the object boundaries and then impart texture on the object. This system uses Linux OS as the efficiency is better in managing the application.

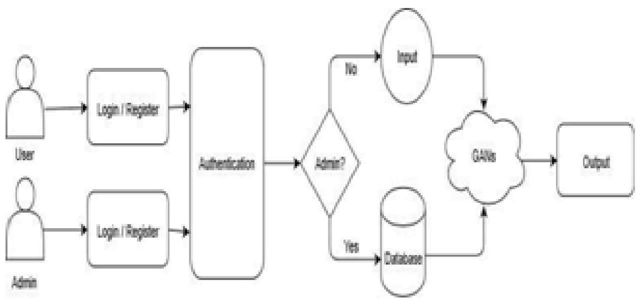

Fig. 1.Proposed Model for sketch to image synthesis

Different modules such as login/registration module, sketch upload module and image output module are integrated to perform sketch to image translation. Generative Adversarial Networks (GANs) are used for translation of sketches.

A. Login/Registration Module

This proposed system as shown in fig. 1 uses login/registration module on both, administrator as well as on the user side. The module is used so that no one can have access to the database that contains the ground truth for mapping purpose. The user can log into the system, upload the sketch and can have a resultant output image for the same. Administrator can log into the system to update ground truth of various images.

B. Sketch Upload/Output Module

Initially, the user uploads a sketch onto the system, which with the help of GANs, maps the sketch onto images and finds 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 job of the discriminator is to identify the real result from the results produced by the generator consisting of both real and fake results. The user then gets the desired output.

IV. ALGORITHM FOR SKETCH TO FACE IMAGE

A. Generative Adversarial Network

Their goal is to synthesize artificial samples, such as images, that are indistinguishable from authentic images [6] as shown in fig. 2. A common example of a GAN application is to generate artificial face images by learning from a dataset of celebrity faces. While GAN images became more realistic over time, one of their main challenges is controlling their output, i.e. changing specific features such pose, face shape and hair style in an image of a face.

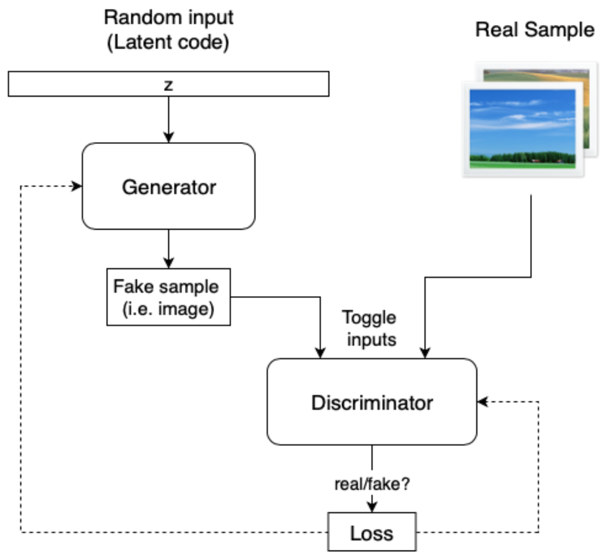

Fig. 2 GAN working model

B. Support Vector Machine (SVM) classifier

Support Vector Machine iterates through the whole image and compares it with face template to classify the region of interest.

C. Histogram of Oriented Gradients

HOG are descriptors used for object detection. In their work, Dalal and Triggs [7] proposed HOG and a stage wise descriptor to identify humans in images. The stages for descriptor are:

1) *Normalizes the image*
2) *Find gradients in both x and y*
3) *Calculate spatial weights*
4) *Contrast the overlapping spatial values*
5) *Form the final vector by collecting all HOG*

To extract HOG descriptors, first count the occurrences of edge orientations in a local neighborhood of an image. This means the image is divided into small connected regions, called cells (e.g., size 8) as shown in fig. 3 and the histogram of edge orientations is computed for each one. Depending on whether the gradient is unsigned or signed, the histogram channels are spread over $0^{\circ} - 180^{\circ}$ or $0^{\circ} - 360^{\circ}$.

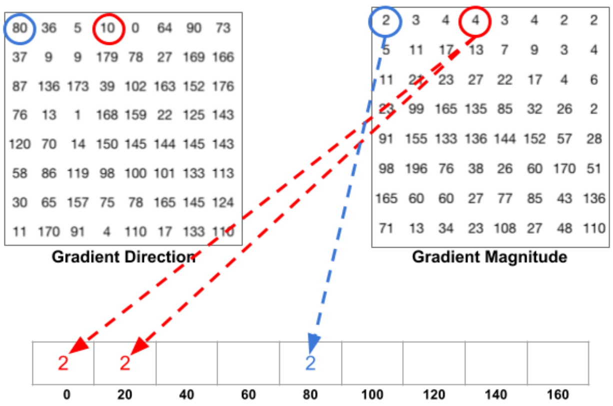

Fig. 3 Histogram of Gradients

D. Face Landmark Estimation

Face land marking is to identify and detect certain key points on the face. Generally, landmarks are corner of eye, nose tip, chin, eyebrow endpoints etc. Landmarks depend on various factors such as face variability, acquisition conditions, and number of landmarks.

We have incorporated transform based land marking method i.e. using HOG features.

E. Phases of project

1) Implementation of sketch to image with the help of GANs.
2) Implementation of sketch to face translation taking textures, highlights and shadows in account with the use of HOG and Face landmark estimation algorithm.
3) Implementation of SVM as classifier.
4) To design a prototype investigation system wherein the user will be able to upload the sketch onto the system to which the system will output possible suspect matches.

V. RESULT AND DISCUSSION

The proposed algorithm was trained and evaluated on face image of size 200*250. The input to the system is the ground truth image and the sketch of the image for training the images. The images were trained using HOG and GAN. We have utilized edge detection algorithm to extract sketches for training as the sketch datasets are very limited.

The training procedure includes, passing an input to the generator to produce a corresponding image output. The generated output and the input are then fed to the discriminator. Then, we adjust the discriminator's weight, to compare the synthesized output and the ground truth image from the sample. So, the generator is trained to fool the discriminator by producing real images. The training phase prepares the sketch-image generator to create faces with the ideal characteristics. At each step, this generator incorporates an image with similar characteristics, as the ground truth image.

The discriminator should recognize the synthesized image as a phony image. At the same time, the sketch-image generator endeavors to trick the discriminators.

The intermediate training images of the proposed work are as shown in fig.4 and fig. 5. The first image is the training image, the second is the HOG output.

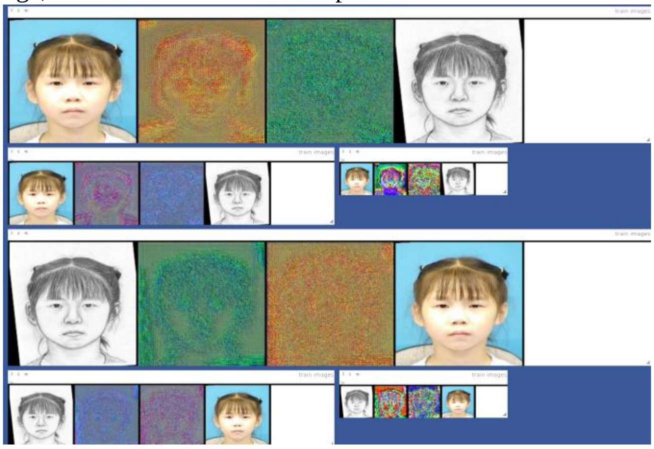

Fig. 4 Intermediate results of training on sample image 1

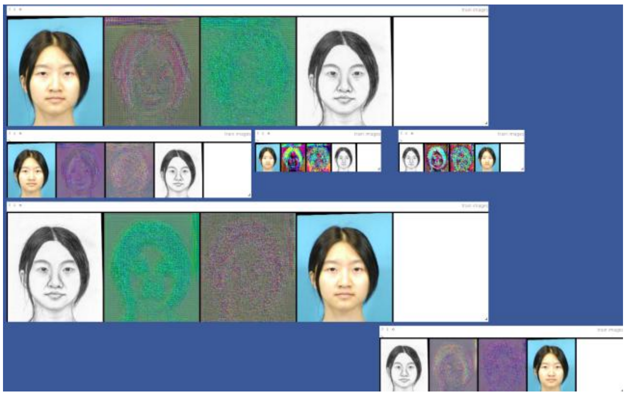

Fig. 5 Intermediate results of training on sample image 2

Fig.6 shows the final result of proposed work on various images after training them using sketch and ground truth images. The first image is the ground truth image, the second image is the image to sketch for training, the third image is the output after training and the fourth image is the sketch of the image.

We can see that, indeed, our selected faces were retrieved correctly. In short, this works by putting images through a generator and training it to give the generated images that will in turn reproduce similar images.

Because the generator ordinarily expects samples from a multivariate normal distribution, we have to train the encoder to output values with a similar distribution if we want realistic outputs. The time consumption for training the GAN depends on the size of the image as the features are extracted from it, but if we reduce the size it has a negative impact on the accuracy of the classifier.

As the sketch datasets are rare, extracted sketches from edge detection algorithms often contain lots of details and their low level insights are altogether different from hand drawings.

Lastly, a major drawback, ground truth data (for training) are the most important but it is very difficult to find it. It is the integral part of GAN based image translation, since it is not easy to identify translated images.

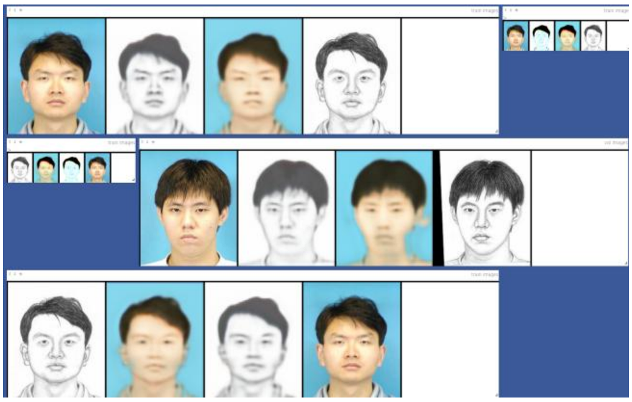

Fig. 6 Synthesized images after training the model

VI. CONCLUSION

The results shown by the project suggest that generative adversarial networks are a promising approach for many image-to-image translation tasks, especially those involving highly structured graphical outputs. These networks learn a loss adapted to the task and data at hand, which makes them applicable in a wide variety of settings. The drawbacks of conventional systems have been reduced to a minimum. Automation on the existing system and technology has been achieved with functions that can be used for various purpose.

Improvisation upon the proposed system can be done by improving its object outline recognition features. The system lacks features for texture identification which makes it difficult for the system to identify objects. The output is highly dependent on factors like noise in the sketch, accuracy of the fed sketch and it's well defined boundaries. One major important factor observed after experimenting is the alignment of faces in the images. For precise results the faces should be aligned and the size to be kept uniform. Deviation leads to the many problems. This, sometimes result in an output that is not desirable by the user. The experiment conducted in this proposed work was on a limited dataset and a constrained training time. The findings and observations are at a preliminary stage and further study would reveal the pros and cons in an effective manner.

REFERENCES

1. Creswell, A. et al. “Generative Adversarial Networks: An Overview” IEEE Signal Processing Magazine, 35(1), pp.53—65, 2018.
2. H. Kazemi, F. Taherkhani, N. M. Nasrabadi,“Unsupervised Facial Geometry Learning for Sketch to Photo Synthesis,” in International Conference of the Biometrics Special Interest Group (BIOSIG), 2018, 10.23919/BIOSIG.2018.8552937, IEEE.
3. Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1125-1134. 2017.
4. Galea, Christian, and Reuben A. Farrugia, "Matching software-generated sketches to face photographs with a very deep CNN, morphed faces, and transfer learning." IEEE Transactions on Information Forensics and Security, vol 13, no. 6, pp: 1421-1431, 2017.
5. Kokila, R., M. S. Sannidhan, and Abhir Bhandary, "A study and analysis of various techniques to match sketches to Mugshot photos." In 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 41-44. IEEE, 2017.
6. Wu, Xian, Kun Xu, and Peter Hall, "A survey of image synthesis and editing with generative adversarial networks." Tsinghua Science and Technology, vol 22, no. pp: 660-674, 2017.
7. Dalal, Navneet, Bill Triggs, and Cordelia Schmid, "Human detection using oriented histograms of flow and appearance." In European Conference on Computer Vision, pp. 428-441. Springer, Berlin, Heidelberg, 2006.
8. Xu, Fen, and Ming Gao, "Human detection and tracking based on HOG and particle filter." In 2010 3rd International Congress on Image and Signal Processing, vol. 3, pp. 1503-1507. IEEE, 2010.
9. Li, Yi, Timothy M. Hospedales, Yi-Zhe Song, and Shaogang Gong, "Free-hand sketch recognition by multi-kernel feature learning." Computer Vision and Image Understanding, vol 137, pp: 1-11, 2017.
10. Ian J. Goodfellow , Jean Pouget-Abadie , Mehdi Mirza , Bing Xu , David Warde-Farley , Sherjil Ozair , Aaron Courville , Yoshua Bengio, “Generative adversarial nets” Proceedings of the 27th International Conference on Neural Information Processing Systems, p.2672-2680, December 08-13, 2014, Montreal, Canada.
11. Ronneberger O., Fischer P., Brox T, “ U-Net: Convolutional Networks for Biomedical Image Segmentation” In Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015, Lecture Notes in Computer Science, vol 9351. Springer, Cham.
12. Tao, L. et al., “LLCNN: A convolutional neural network for low-light image enhancement”, 2017 IEEE Visual Communications and Image Processing (VCIP).
13. Kazemi, Hadi, Mehdi Iranmanesh, Ali Dabouei, Sobhan Soleymani, and Nasser M. Nasrabadi. "Facial attributes guided deep sketch-to-photo synthesis." In 2018 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 1-8. IEEE, 2018.