Generation of Realistic Images from Human Drawn Sketches Using Deep Learning

by

Mohammed Julfikar Ali Mahbub
21341031
S. Afsan Rahmatullah
21241075
G M Sohanur Rahman
21341049
Abu Hasnayen Zillanee
18301159
Aknur Kamal Akib
18301209

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Zulfikar

Mohammed Julfikar Ali Mahbub
21341031

Rahmatullah

S. Afsan Rahmatullah
21241075

G M Sohanur Rahman
21341049

Abu Hasnayen Zillanee

Abu Hasnayen Zillanee
18301159

Aknur kamal

Aknur Kamal Akib
18301209

Approval

The thesis/project titled "Generation of Realistic Images from Human Drawing Sketches Using Deep Learning" submitted by

1. Mohammed Julfikar Ali Mahbub (21341031)
2. S. Afsan Rahmatullah (21241075)
3. G M Sohanur Rahman (21341049)
4. Abu Hasnayen Zillanee (18301159)
5. Aknur Kamal Akib (18301209)

Of Fall, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 16, 2022.

Examining Committee:

Supervisor:
(Member)

Md. Ashraful Alam, PhD
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Processing sketches to produce realistic images is an intriguing idea in the world of emerging Artificial Intelligence. We present a Generative Adversarial Network (GAN) based methodology that creates satisfactory images for the most prevalent categories in our approach. The proposed approach is applicable not just to people, but also to animals, objects and foods. The system takes a sketch and analyzes it using a powerful neural engine to produce new photographs that resemble realistic images. We also used a data augmentation method to dramatically increase the variety of data available for training models. The proposed model has achieved approximately 96.36% accuracy over generating sketch to realistic images of people and 40.63% accuracy for objects and animals. Moreover, about 76.63% accuracy on generating sketches from strokes on an average from people class.

Keywords: GAN; Sketch to Realistic Image; Data Augmentation; Diversity of Domain

Dedication

This study is dedicated towards all the researchers who have worked so hard and given their precious time to improve the society and make things easier for general people. Furthermore, we would like to dedicate it to our loved ones for all their inspiration and support.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our supervisor Dr. Md. Ashraful Alam sir for his kind support and advice in our work. He helped us whenever we needed help. It is his continuous support and inspiration that made our work successful.

And finally to our parents. Without their throughout support it may not be possible to reach here. With their kind support and prayer, we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$CGAN$ Generative Adversarial Network

CNN Convolutional Neural Network

$CUHK$ Chinese University of Hong Kong

GAN Generative Adversarial Network

GPU Graphics Processing Unit

PNG Portable Network Graphics

RAM Random Access Memory

$UNet$ A Convolutional Neural Network Architecture

VAE Variational Autoencoders

Chapter 1

Introduction

Traces of drawings may be found dating back to the upper Paleolithic period (Stone Age) [18]. It is one of the various strategies that people employ to convey themselves creatively and effectively. Sketches are utilized in a variety of purposes, varying from children's drawings to detecting criminals. Sketches are easy to make, take little time, and are affordable. However, various individuals will have different interpretations of drawings depending on their background. Except for professionals, it is possible that individuals will be unable to discern the real context of a drawing. In light of the increasing rise of Artificial Intelligence, it is vital that we confront the difficulties by using technology that are accessible to everyone, rather than technologies that were previously only viable for professional human artists.
Despite the significance of drawings, it is difficult to produce a realistic picture through sketches due to the differing viewpoints of artists, the imperfection of sketches via non-artists, and the absence of datasets and trained neural networks for a wide range of categories and subjects.
The use of image retrieval techniques, such as Photosketcher [14] and Sketch2photo [7], to produce realistic photographs from drawings was an earlier method. This is necessary due to the fact that drawings are more sparse than genuine photographs, and hence the characteristics must be planned carefully. A further drawback of such systems is that they need considerable pre-processing in order to eliminate the items from photos that would subsequently match the users' drawings in order to synthesis realistic images from them.
Deep learning [24], [35] has recently been used to synthesize realistic pictures from sketches, which has resulted in some exciting new breakthroughs. It is also apparent that the Generative Adversarial Network (GAN) outperforms all other deep learning algorithms in terms of performance [43], [46], [48], [50]–[52]. While training the model, a GAN-based technique consists of two parts: a generator and a discriminator, which are both used in the same way. As an analogy, one may consider GAN's training process to be a two-person minimax game in which one player attempts to outsmart the other. Instead, the generator's objective is to generate a counterfeit (picture) in order to deceive the discriminator into believing that the created image is legitimate (image). A successful GAN model should train the discriminator to distinguish between genuine and false pictures since the discriminator attempts to distinguish among real and fake images. In order to reduce the time complexity of training the model, it is essential to formulate the generator in such a way that this can recognize responses from the discriminator indicating which segments of

the result appear to be fabricated and operate on all of those particular sections while discarding the portions of the outcome that the discriminator deems to be fraudulent. This being said, it is critical to avoid having one model (generator or discriminator) perform much better than another throughout the training process. Both models should be at the same level of skill and should be able to understand from one another over time. Upon completion of training, the GAN-based model does away with the discriminator function and instead makes use of the generator to generate actual pictures from the user's input, mixing this with the neural network that's been constructed during the training phase of the model.

As a result of the shortage of human-drawn sketches and their corresponding actual pictures, one of the most difficult difficulties in recent research has been the generation of datasets for training the GAN-based model. Cheng and Hays [48] offered a data augmentation technique in order to explore the challenge more effectively. It has shown significant promise in terms of tackling the problem of missing datasets. In this work, we offer a GAN-based model that can synthesis realistic pictures from drawings of more accurate and considerably diversified categories, based on our observations of the difficulties of previous techniques and the fast growth of GAN. Furthermore, we will apply an amazing approach by Ghosh et al. [50] to autocomplete a drawing from a partial or sparse set of strokes from user input. It might be beneficial to those who have minimal sketching abilities or even to those who want to save time when sketching [5]. In addition, Li et al. [52] developed an extension of the SketchyGAN concept that outperformed the original SketchyGAN [48] concept. As a result, we choose to develop and merge the concepts of Ghosh et al. [50] and Li et al. [52] in more detail. The following are the specifics of our contribution to this paper:

1. Using GAN-based synthesis, we offer a methodology for creating realistic pictures from hand-drawn sketches. The limitations of each strategy were identified after a comprehensive examination of previous comparable GAN-based models. The best techniques [48], [50], [52] out of each approach were picked in order to achieve a more accurate model.

2. The range of categories available for sketch to image synthesis has been increased. It is possible to choose from a maximum of 50 categories [48].

3. We merged the two-staged model [52] that uses Masked Residual Unit (MRU) [48] with the concept of partial stroke to realistic images [50].

1.1 Research Problem

Sketches have been used to represent human thoughts and ideas since the beginning of time. Research have revealed that it has existed since the Paleolithic era (around 30 thousands years ago) [15], [18]. Furthermore, a crime scene sketch is a critical component of any criminal inquiry [1]. Because of the enormous use of drawings, we set out to make sketches even more useful by generating a realistic image from the sketches we already had. Before, training an image retrieval system was a time-consuming and expensive procedure that was difficult to put into practice, making it a poor choice for many situations. As a result of the tremendous progress that has

been made in deep learning in recent years, as well as techniques that use GAN-based models, the complexity level has been greatly decreased.
Many studies on the generation of images from drawings using a GAN-based technique have been conducted in recent years. Researchers are working on the advancement of this image creation process on a constant basis. Because the issue is new to computer vision and needs to deal with two entirely distinct domains, such as drawings and photos. Deep learning algorithms are data hungry, and as a result, it is extremely difficult to come up with a viable resolution in the early stages of GAN-based techniques. The hand-drawn drawing is among the most important aspects of this challenge. A hand-drawn sketch can depict anything, including but not limited to human features, animals, plants, and landscapes. Taking into account the vast number of possibilities available, compiling a dataset will be a very difficult task. A multiclass dataset strategy [50] is proposed as a solution to this problem. In this technique, a dataset has various classes of data based on their categories and distinct criteria.
Drawings by hand are classified according to their level of abstraction, as previously stated. For example, a common object may have a variety of distinct looks, making it challenging to develop a true natural image which is not close to the user's expectation, as seen in Figure 1.1. It is sometimes difficult to determine exactly what the artist is attempting to communicate via their sketches.
It is very important to respect this level of ambiguity of a sketch and provide users maximum flexibility to choose what type of image they are trying to sketch. On the contrary, this extra level of input of category/ class can be frustrating for quick sketches. Considering all these points together, we have analysed how a human artist would approach such situations.
From our study, an amateur artist would try to guess by the class of sketch at his/ her first glance of the sketch and the best matching realistic counterpart. For this reason, we have included a mandatory label of class while training our generative model. However, a user is not necessarily required to provide a label for his sketch but a sketch with a high level of abstraction we will provide suggestions of possible classes along with manual entry choice.

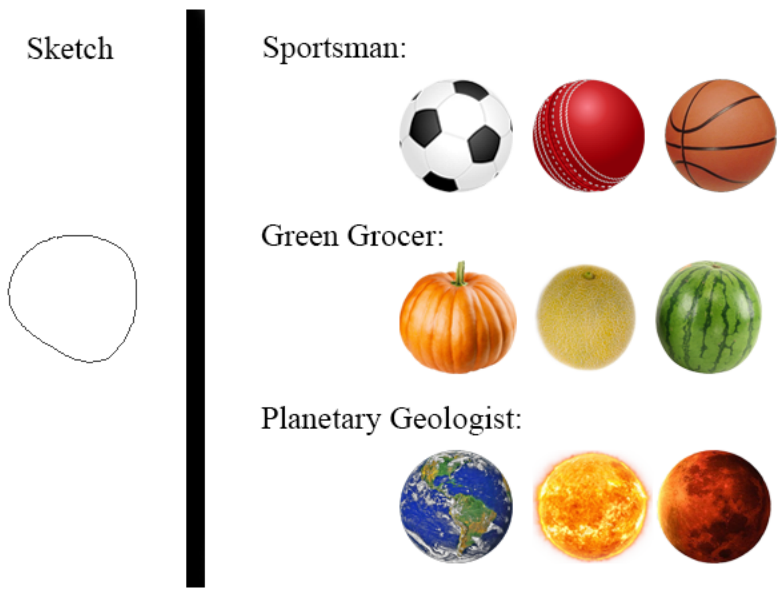

Figure 1.1: Different perspectives on the same sketch

As we know, GAN has accomplished unprecedented success in image generation from sketches with considerable accuracy in recent years [15], [50], [52]. So, in this paper we will try to shed light on how to make realistic pictures from hand drawn sketches of novice artists using GAN-based models. Moreover, generative models were made to produce plausible images from given colors and edge constraints [47]. Hence, our goal is to make image generation work faster with higher accuracy and improve the quality of generated images so that it looks realistic. Additionally, our dataset includes more diverse categories of objects compared to any existing research

1.2 Research Objectives

The significance of sketching cannot be overstated. We have seen how important it is in the manifestation of human expressions. Furthermore, creating a sketch needs little work and does not have to be accurate to the real world. We came up with this concept after realizing the incomparable value and simplicity of a drawing and realizing that it may be made much more beneficial. Following a comprehensive analysis, we discovered that several attempts with great methods had previously been attempted, with some performing really well and others failing miserably. We have identified the shortcomings of previous techniques and have strictly monitored to determine how innovative methods based on deep learning might surpass existing approaches, such as image retrieval, in terms of performance. Furthermore, we intend to challenge the traditional perception of sketches by presenting a realistic image created from drawings produced by rookie artists. As a result, our major goal is to generate realistic pictures from a design that includes borders, lines, and drawing scripts for edge items. The Generative Adversarial Network (GAN), which is a deep learning method, will be used to extract the common elements from drawings and images in order to cross the sketch-photo domain in order to defend our notion and express our desire, as described above. The following are the primary aims of our investigation:

1. In order to generate pictures from hand-drawn drawings, it is necessary to understand how GAN (specifically CGAN) works.

2. Use a partial stroke image generation technique [50] in conjunction with the deployment of a two-staged learning framework [52].

3. The competency of the dataset should be estimated based on past research.

4. Prepare a model that is effective.

5. Compare the proposed model to other models that have already been developed.

6. Extend the concept of Data Augmentation[48] to include increasing the amount of data available for training the model.

1.3 Thesis Orientation

The consequent chapters of the paper have been covered in the following order. Chapter 2 discusses similar work in the field and existing methodologies and gives an intensive analysis of the topics preceding data and information associated with our work. Chapter 3 presents the proposed model in detail. Chapter 4 give the test results and the interrelated discussions. At last, Chapter 5 closes and outlines the the thesis along with future plans.

Chapter 2

Literature Review

Despite the fact that many scholars have previously worked on comparable research as ours, we have observed huge advances in recent years, notably with the release of the GAN system. We needed to look at the previous strategies and approaches in order to comprehend the essence of the research. We have seen the limitations of the Image Retrieval technique, such as the expense of model training.
Furthermore, generic computing does not facilitate the creation of realistic images from sketches. It necessitates a robust application of Artificial Intelligence. And, because deep learning is a data-hungry strategy, it is only lately that these approaches have been able to perform effectively due to the huge data collecting system readily available with the support of the internet and improved computing. For all of these reasons, the most fruitful research has occurred in recent years. As a result, we had to explore for and review previous studies on producing realistic images from sketches incorporating GAN-based techniques. Surprisingly, we acquired a massive quantity of research work on this matter in such a short period of time. However, one of the most significant shortfalls we discovered in the research was a lack of variation in categories.
We observed that majority of the researchers worked in specific categories. With the exception of a few, such as SketchyGAN [48], who claimed to work on 50 categories. Section 2.1 Related Works has a full discussion of our findings from the literature review.

2.1 Related Works

We've classified related works into subcategories that describe relevant research arranged by strategy, as well as some of the datasets we discovered when searching for related works.

2.1.1 Sketch-Based Image Retrieval and Synthesis:

Since the introduction of GAN, various articles have focused on sketch-based image retrieval [8]–[13], [16], [19], [21], [22], [26], [27], [32]. The vast majority of strategies that involve word representations and edge detection, on the other hand, provide properties that are consistent across domains. Two significant challenges are the limitations in getting fine-grained data and moving between inaccurately drawn sketch and image edges. By linking drawings and images with the help of deep convolutional network Yu et al. [37] and Sangkloy et al. [34] have tackled these problems. Also, they have used sketch-based image retrieval as a query inside the learnt component embedding space. Furthermore, they have shown that CNNs can do complex and instance retrieval and enhance efficiency dramatically. Aside from recovery, Sketch2Photo [7] and PhotoSketcher [15] combine objects and backgrounds from a specified sketch to generate realistic photographs. PoseShop [17] composites individual photos by allowing users to include a supplementary 2D structure in the query for more reliable results.

2.1.2 Sketch-Based Datasets:

Only a few datasets of hand-drawn drawings exist, and the majority of them are rather modest in relation to the time and work necessary to generate them. Nonetheless, one of the most often used sketch datasets is the TU-Berlin dataset, which contains 20,000 human drawings in 250 distinct genres. Furthermore, despite the availability of only two components – shoes and chairs - Yu et al [37]. presented a new collection of data with connected drawings and images. The CUHK Face Drawings collection [6] also includes 606 modern artists' facial drawings. Finally, the QuickDraw dataset has 50 million amazing sketches [42].
Even then, due to the extreme time limitation of 10 seconds, the sketches are quite rudimentary. The illustrations are sparingly drawn and frequently reflect well-known or iconic topics. In comparison, the Sketchy database [34] appears to have more capacity over a wider variety of positions. It has 75,471 sketches of 12,500 various items arranged into 125 categories. It is, after all, the only large collection of matching images and photos of various kinds, which is why they chose to use it.

2.1.3 Image-to-Image Translation with GANs:

GANs (Generative Adversarial Networks) have demonstrated a huge amount of potential in terms of producing natural, photorealistic visualizations [39], [41], [45]. Rather than directly fixing per-pixel training error, which typically results in fuzzy and traditional output, GANs implement a discriminator to distinguish between false and genuine images, encouraging the generator to produce higher resolution images. In their "pix2pix" study, for example, Isola et al. [43] demonstrate a straightforward way for translating one image to another using conditional GANs, which we have adapted for implementation in our methodology. Other image translation tasks, such as style transformation [36], sketch coloring [46], adjust conditioned and domain adaptation [40] situations in about the same approach. Liu et al. [44] created a non-supervised CoupledGAN [33] image translation system with pairings of variational autonomous encoders [20] instead of conditional GANs and matched datasets. CycleGAN [47] has recently demonstrated outstanding results for unmonitored image translation over cycle consistency losses.

2.1.4 Generative Modeling:

The modeling of metric image propagation is difficult. Autoencoders [4], [20] and Boltzmann [2] machines, on the other hand, are conventional approaches. More modern approaches include Autoregressive models [3], [9], variational autoencoders

(VAEs) [27], and generative adversarial networks (GANs). GANs and VAEs both learn the mappings of a tiny "Latent" code written by a system's feedforward looping to a high-dimensional picture. GANs have lately achieved success [23], [25], [38], and hybrid models contain both image mapping learning and opponent training [28]–[31].

2.2 Datasets

As we couldn't find any such dataset that is suitable for our model, thus we created our own dataset by taking data from various sources. The datasets we have collected far are listed below with proper references.

2.2.1 Male Dataset

We found a sketch to realistic images of more than 100 men from CUHK Face Drawings collection [6]. We further pre-processed them for our model and trained. With about 50 epochs, we had achieved a quite satisfactory result.

2.2.2 Female Dataset

More than 3K images of female faces paired with their sketches were found on the Pretty Face dataset on Kaggle [49]. And it attracted our interest towards training our model with this dataset. The outcome of the train result proofs that it was completely worth it, which we will be showing later in this paper.

2.2.3 Animals and Objects

We used a data augmentation dataset found from Kaggle, Sketch to Image [53] that has at least 6 variations of sketches for each real image of animals and objects. It is the key resource for us to include more and more objects for sketch to realistic image generation.

Chapter 3

Proposed Methodology

The intended sketch to realistic image conversion method is designed to generate scenes from sketches to benefit police, to assist amateur artists in making realistic images, and to contribute to the field of artificial intelligence. To accomplish this, we will apply the GAN-based strategy, which entails creating a process that inputs a human-drawn sketch and a label (object name, to respect the abstraction level in a sketch) as input.

After training the model, the system will generate a realistic image in two steps using the provided input. First, it develops a well-defined sketch from a provided sketch that matches the label. We provide a well-defined meaningful sketch of an item in this process to assist the sketch discriminator in distinguishing between real and false sketches with the sketch class.

In the process of creating this system, we are going to use 2 GAN models. Each model will have 2 neural networks, one is the generator and another one is the discriminator. The sketch generator model will create a complete sketch from a sparse sketch provided by the user and it will give a label to the sketch it has created. And the photo generator model will take that created sketch as input and create a completely colourful figure depending on the training criteria. For both models, the output will be determined by the discriminator and the generator will accept all inputs. Here, the generator will always try to fool the discriminator.

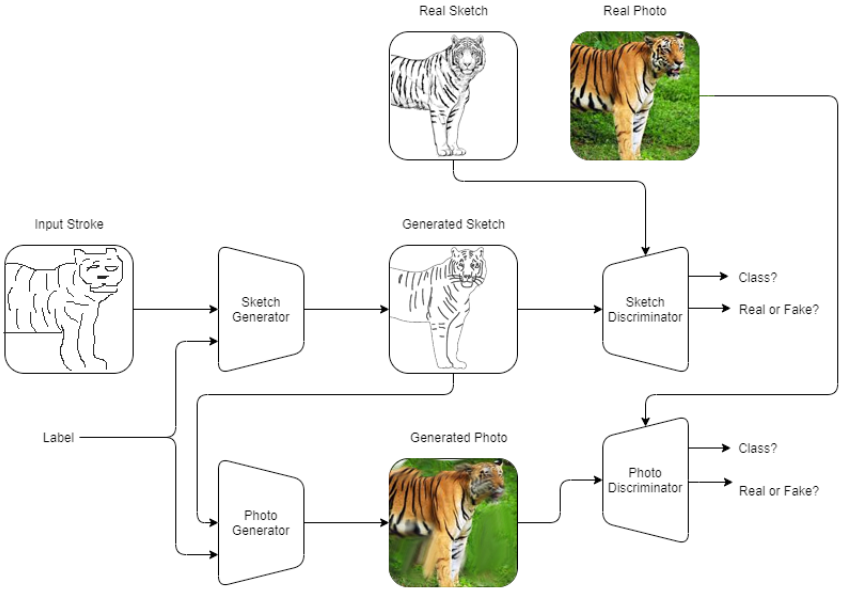

Figure 3.1: Work Plan: A complete overview of our two steps image generation process.

Along with the label, this resulting sketch is then sent to the photo generator specified at the beginning. Later, the original photo along with the created photo are sent through the photo discriminator, which determines the difference between actual and generated photos. A high-level representation of our suggested system is depicted on figure 3.1.
We will use conditional GAN (CGAN) for training both of our models. Which can be expressed as,

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) = &\mathbb{E}_{x,y}\left[logD(x,y)\right] + \\ &\mathbb{E}_{x,z}\left[log(1-D(x,G(x,z))\right]\end{aligned} \tag{3.1}$$

Where G, generator, tries to minimize the loss function against competitor (adversarial) D, discriminator, which tries to maximize the loss objective.

3.1 Input Data

As our proposed model is a supervised learning model, we will need to provide a label and sketch as input. For training our proposed model, we chose a two stage input system.
The first method takes a rough stroke for a subject with a label. And, this is the method which will be used by the user after finishing the training. Hence, it is the common method for both training and testing. The second method is only used for training the model. It is more like a hidden input of our model. It takes a generated sketch along with appropriate labels to generate the real photo.

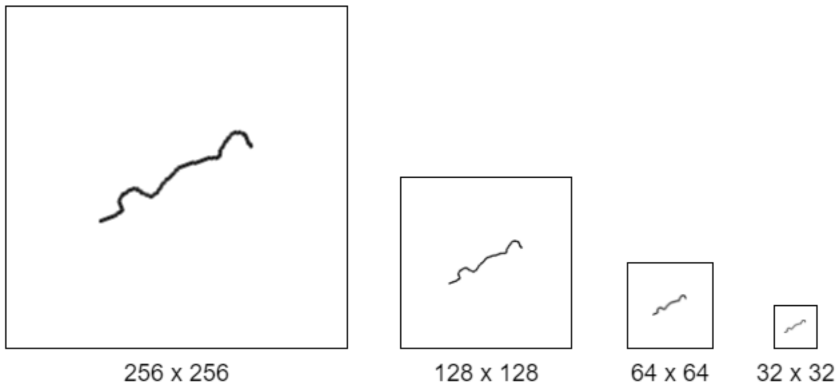

Figure 3.2: Resizing strokes

3.2 Stroke Prediction

Our stroke prediction model, Sketch Generator, is designed to predict a probable sketch from the user's incomplete strokes. Moreover, it will try to generate a sketch after each stroke to match the sketch as much as possible. As described by Ghosh et al. [50], the given stroke is placed on different part of the image grid of 256 $\times$ 256pixels. To achieve this, we will resize the input stroke from a 256 $\times$ 256 image to 128 $\times$ 128, 64 $\times$ 64 and 32 $\times$ 32 pixels images.
In addition to that, to increase the accuracy of prediction for our model, we performed 4 random rotations from 15° to 350° for all the resized images along with the original stroke image. Next, each of the 20 images, resized and resized with random rotation of images, will be projected on the image grid of 256 $\times$ 256 at random positions.
After the projection of the stroke in the images it goes through the Sketch Generator to produce a Generated Sketch. Our proposed model extended the technique proposed by Ghosh et al. [19] with an additional rotation feature to generate sketches from most novice sketches. Multiple possible sketches are handled by the Class Verifier, which takes the class of generated sketch and the label. Class Verifier returns true if the sketch class and label matches, otherwise returns false. Furthermore, once the photo is generated from the sketch, we revert the rotation required to generate the sketch from the given input. As a result, the generated photo of our model keeps the integrity of the user's angle of rotation in their sketch. Hence, more accurate and precise sketches can be generated from novice sketches or strokes. And with more accurate sketches, the realistic image synthesis model, Photo Generator, can produce more accurate results than existing state of the art models.

3.3 Realistic Image Synthesis

To synthesize realistic images from multiple classes, an optimal system should be able to process all the classes using only one model. This has a lot of advantages in terms of efficiency and effectiveness. Apart from these advantages, we believe, a single model system has a very important role in advanced realistic image synthesis. For example, generating scenes of multiple classes where there are common features among different classes. However, that is our future goal of this model that we

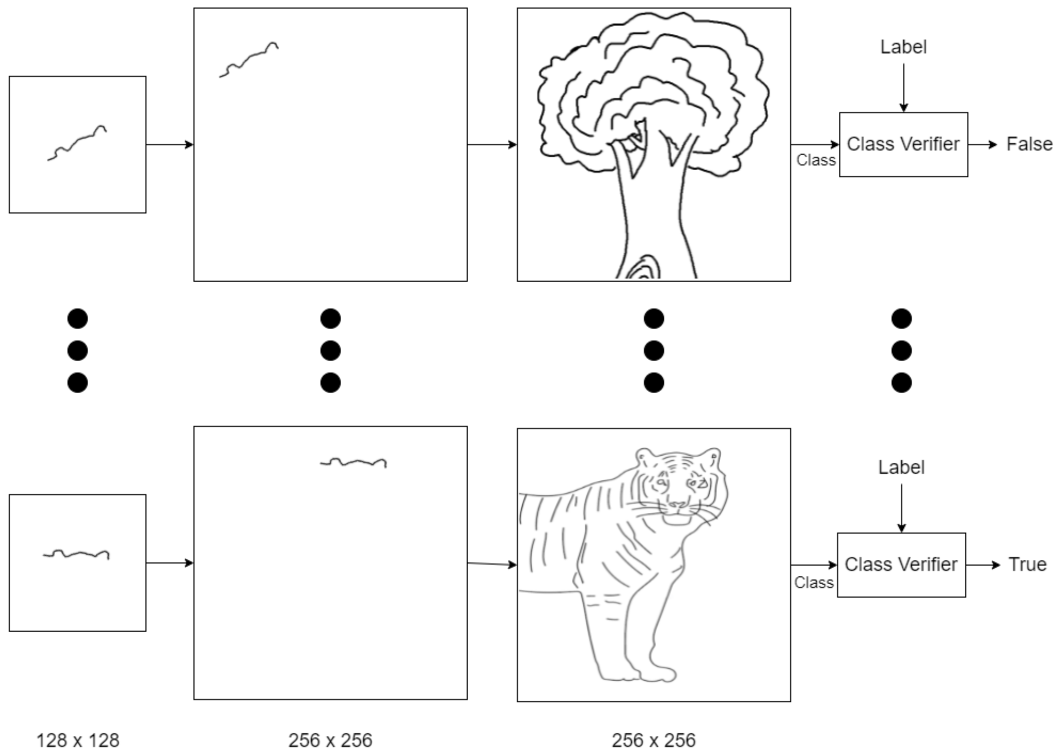

Figure 3.3: Selection of random rotation and projection of strokes of resized stroke images

will discuss later on in this paper. Additionally, common features among different classes can also help in training the model when there are limited resources for training some classes.
We used the UNet architectural model for synthesizing realistic images. However, this architecture requires matching the shape of the input and output images. For that, we had to pre-process the data carefully to obtain uniform image sizes before passing it through the model. To reduce the computational costs and sampling rate, we decreased the size of the image by applying a downsampling stack of layers which has Convolutional layers. Next, we used Transpose layers for upsampling stack of layers to resize the image back to its original 256×256 size. As a result, the resolution of the image has decreased for faster computations to speed up the training process.

3.4 Train Data Pre-processing

Now, we will impose pre-processing onto our loaded dataset. We have used resize(), random_crop(), normalize() and random_jitter() functions to do these tasks. Let us now explore each function one by one.
resize(): To work with thousands of different images we need to resize them in a uniform image size. It helps to prevent the risk of having different sized images in the dataset. Not only that, it also helps to improve the computational efficiency while training the model. In our model, we resize all the 512×512 pixels images into 256×256 pixels.
random_crop(): It basically crops an image to a specific size at random. This

function provides a cropped image and a true image with the necessary size of 256 $\times$ 256 PX for our project.

normalize(): It helps to normalize all the images to [-1, 1]. In our project, we have normalized both the sketches as well as real images as it makes it more machine readable and reduces residual stresses.

random_jitter(): We have used this function to flip a few images horizontally. After completing all the previous pre-processing steps, we have applied this to correct the orientation of random images.

Chapter 4

Results and Discussions

For experimental results, we start by describing the technique for pre-processing test data to increase the input domain along with the limitations of resources to train the model and achieving satisfactory results from very limited resources. Next, we discuss the implementation of each of our models along with appropriate diagrams. Additionally, appropriate loss function for each model has been explained in detail. And finally, the results of our model are shown and the accuracy of our model was compared with most sophisticated existing models.

4.1 Test Data Pre-processing

Proper testing is one of the key ingredients for machine learning models. It is very important to test a model with variant inputs from the domain range to find the accuracy and performance matrix of a model. It also helps to identify the strength and flaw in a model. Moreover, it should be able to predict the performance of a model after deploying it for public usage. Hence, we used techniques for test data pre-processing in a way that can accept almost any sketches as input to show the performance of the model.

Keeping the goal established to process any sketches as input, our first step in test data pre-processing is to resize the given input as required for our generator model. Additionally, we also needed to ensure the given sketch has the right layers and dimensions which were used while training our model with Tensors. To achieve these goals, we created a new PNG format image of dimension 400×250 that contain both sketch and expected output, using the Image class from PIL library, which will be converted to 256×256 image for both sketch and real image.

It has to be admitted that there are not many well established models to train such sketches to realistic image synthesis on the internet. Most of available models are focused and specific class and that also with a very limited resources. For example, the CUHK Face Drawings collection [6] dataset contains 606 facial sketches of adult human males and females, which are mainly of Asian people. However, we have trained our model with this dataset along with our few custom inputs and obtained quite satisfactory results for the given time of training with limited resources.

Training a GAN model for image processing requires a computer with high specification, such as powerful GPU and high capacity RAM and storage. So, apart from limited training dataset, another challenge we faced to test our model is the lack of high end computers to train our model. Because of the Covid-19 outbreak and

closure of universities, we could not access our Research Lab to train our model with high performance machines. Discussing with our supervisor, we get an option to remotely access the computer from the Research Lab.
However, working from a remote computer is not as of working directly, let alone the slow and unstable internet connection in Bangladesh. Although we got access to research computers but there were a lot of problems working due to dependency issues and we did not had the administrative privileges. Though the lab assistants were helpful in solving the problem but it was quite difficult to manage and for some critical parts, it was not possible to solve. Therefore, we had to have utilize the available resources on the internet to train our model to generate results as accurately as possible.
For training using the online resources. At first we tried to train our model using Google Colaboratory using an organisational account and GPU accelerator. However, the non professional version of Google Colab could not provide the resources we needed. Training GAN model requires a very strong GPU as well as RAM and Storage. For that, next, we trained our model on the Kaggle notebook.
Furthermore, due to the lack of hardware resources, we could not apply the Data Augmentation technique that could increase our training dataset drastically. Despite all these limitations, we have achieved quite reasonable results for our model. Hence, we believe that with proper resources, we can train our model to outperform most of the state-of-art models by predicting/ generating realistic images from most diverse categories with a great fraction.

4.2 Implementation

The major distinction between a regular system and an ideal system is how well each component of the system is implemented. It is very important to implement every part such that the system can generate outstanding results with high accuracy. In our proposed system, we have carefully designed the GAN models for generator and discriminator along with their loss functions to achieve exactly that outstanding performance with maximum efficiency. In the following sections, we will elaborate in detail about each part and model of our system beginning with the generator model.

4.2.1 Building the Generator Model

Both of our two generators accept a black-and-white sketch as input along with a label for the class. From the given sketch/ stroke, the Sketch Generator resembles a most accurate artistic sketch of the input sketch while the Photo Generator resembles the realistic image from the training dataset based on the given class label. As has been described already, we have followed UNet Architecture design to build the Generator which is challenging and complex. Hence, it is required to be designed carefully to match the input and output shapes with the connected layers.

4.2.2 Building the Discriminator Model

Another essential part of GAN is Discriminator. Basically, a discriminator helps to figure out the source of images. We trained our discriminator model using real data instances and fake data instances, where real data instances are real pictures of people and fake data instances are generated image outputs from Generator. While training Discriminator, it uses real data instances as positive examples and fake data instances as negative examples.

4.2.3 Loss function Calculation

As we know, GANs use probability distribution to calculate Loss function of a model. Here, the loss function basically finds out the difference between two distributions. In our paper, we have used two loss functions to calculate the loss of two models independently.

Loss function for Generator: To calculate the loss function of Generator we have used the Sigmoid Cross-entropy Loss of the output generated by Generator and a list of ones. Generator always tries to maximize discriminator output. Here, we are actually trying to trick the discriminator to choose 1, where 1 means discriminator recognizes the generated image as the original image. If discriminator outputs 0, that means the generated image is not good enough to fool the discriminator, therefore the generated image is not structurally identical to the real image.

Loss function for Discriminator: Like the previous one, we have used Sigmoid Cross-entropy loss of the real image and a list of ones in our discriminator. After that, we just simply add it with a cross-entropy loss of output image of the generator model and a list of zeros.

4.3 Results and Accuracy

In this section, at first, we show some exciting results of our model with image figures and then we compare the accuracy and efficiency of our model with existing state-of-the-art models both quantitatively and photographically.

4.3.1 Results of generating sketches of men

Figure 4.1 displays the GAN loss for generators and discriminators for each epoch while training the male dataset. Moreover, figure 4.2 shows a satisfactory prediction of a male image from a given one fourth of the real sketch. Furthermore, figure 4.3 shows the result of synthesizing realistic image of a man from full sketch. We can notice that upon giving full sketch, there are no noticeable losses on the predicted sketch. More interestingly, the predicted image is almost similar to the expected image as well. From the table 4.1, it is important to note that although the accuracy of full sketch, 97.59% is very good but given the portion of partial sketch, 81.97% is just amazing and really satisfactory. It really advances the process of generating realistic images to a next level.

Figure 4.1: Model loss over training men dataset

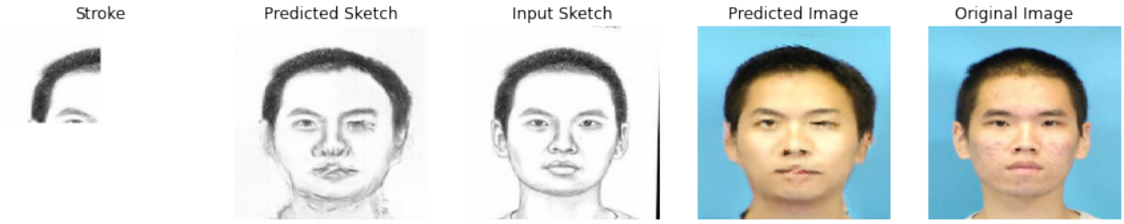

Figure 4.2: Generation of realistic image of a man from a partial sketch

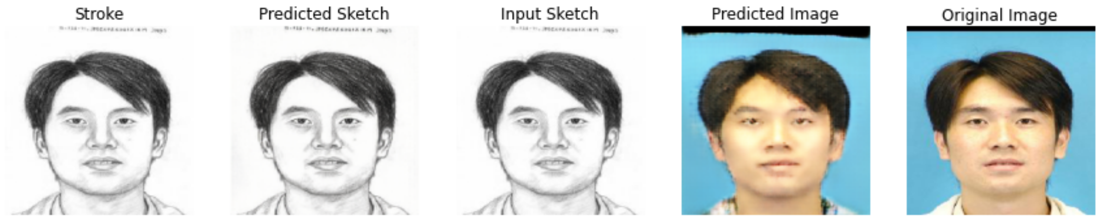

Figure 4.3: Generation of realistic image of a man from a full sketch

Figure 4.4: Generation of realistic image of our supervisor Dr. Md. Ashraful Alam from a full sketch

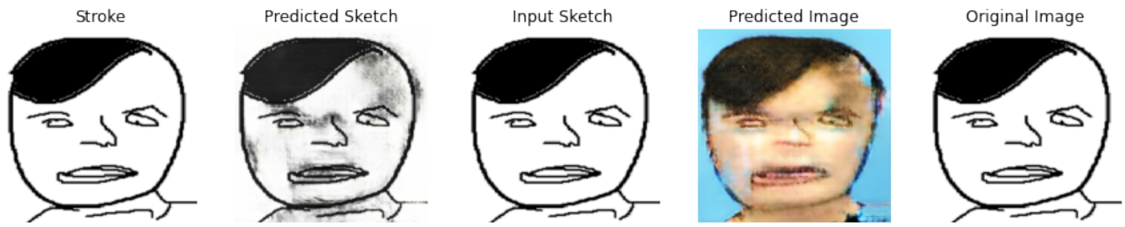

Figure 4.5: Generation of realistic image of a very sparse and unreal sketch

Table 4.1: Accuracy of male

Sketch	Accuracy
Partial	81.97%
Full	97.59%

4.3.2 Results of generating sketches of women

Figure 4.7 shows the synthesis of sketch of a woman from a partial sketch. And figure 4.8 shows the result of synthesizing realistic image of a woman from a full sketch. One interesting thing we can notice here is that compared to sketches of males, accuracy of females is lower. We assume it is because the sketches in the dataset of females are more sparsed compared to the sketches in males dataset. Furthermore, for synthesizing objects and animals we showed how the sparse sketches affects the final output even more.

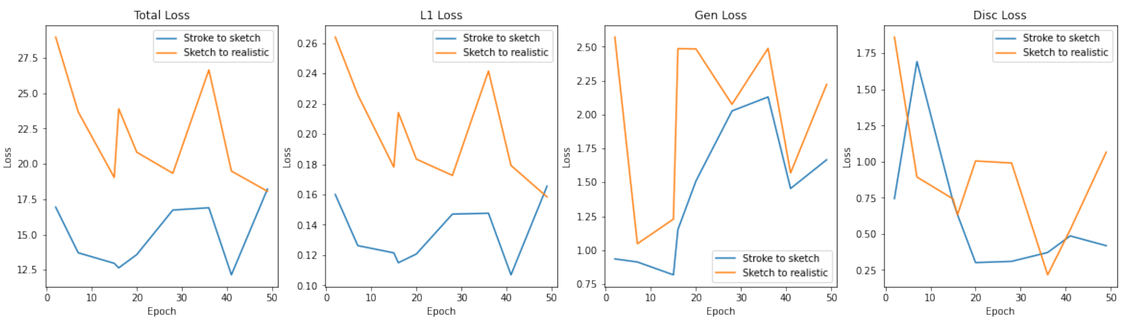

Figure 4.6: Model loss over training women dataset

Figure 4.7: Generation of realistic image of a woman from a partial sketch

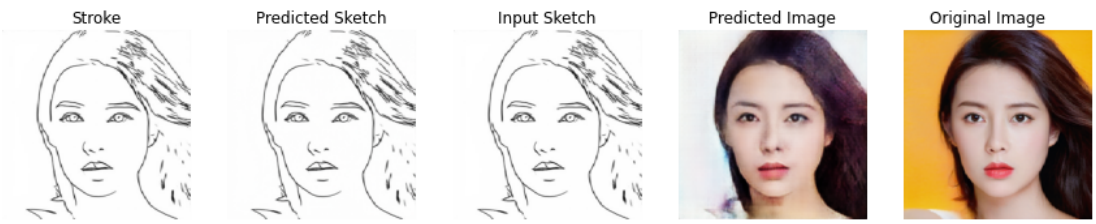

Figure 4.8: Generation of realistic image of a woman from a full sketch

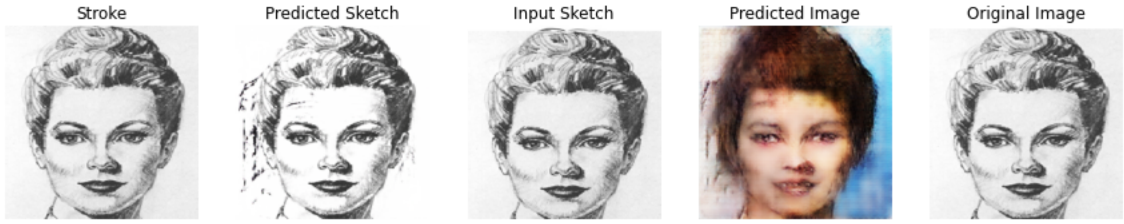

Figure 4.9: Generation of realistic image of a random woman from a full sketch taken from outside the dataset used for training

Table 4.2: Accuracy of female class

Sketch	**Accuracy**
Partial	71.28%
Full	95.13%

Table 4.3: Average accuracy of people class

Mode	**Class**	**Accuracy**
Partial	Male	81.97%
	Female	71.28%
	Average	**76.63%**
Full Sketch	Male	97.59%
	Female	95.13%
	Average	**96.63%**

4.3.3 Results of generating sketches of animals and objects

Figure 4.10 shows the result of synthesizing various objects and animals from a very sparsed sketches. This proves our initial statement about why the earlier approaches are not viable for generating images from sketches. Using a GAN based approach can deal with such problems with much more efficiently by continuously competing two neural networks model with each other. Moreover, we noticed that to include more categories, it will be beneficial if we could componentize elements of each class. We believe it can solve the major problem that we can often see with the current GAN approaches, lots of soft corners and blurry images.

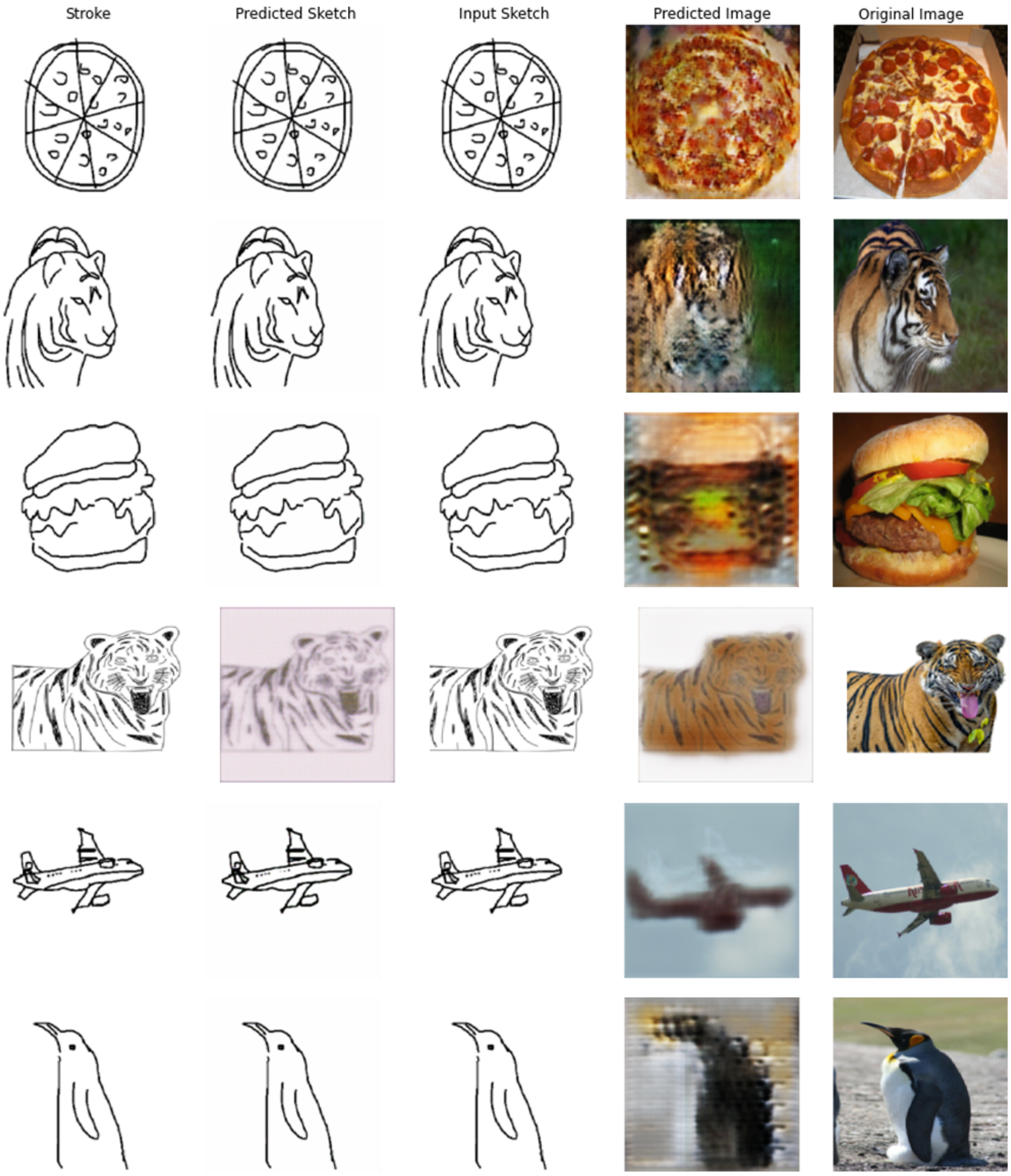

Figure 4.10: Generation of realistic image of animals and object from rough sketches

Table 4.4: Accuracy of synthesizing realistic images for objects and animals

Class	**Accuracy**
Pizza	73.07%
Tiger	30.13%
Burger	53.21%
Royal Bengal Tiger	30.13%
Airplane	40.01%
Penguin	17.23%
Average	**40.63%**

Table 4.5: Average realistic score of the proposed model

Class	Realistic score
Male	97.59%
Female	95.13%
Animal and Objects	40.63%
Average	**77.78%**

4.3.4 Comparisons to state of art models

In our comparisons, we mainly focused on most recent and appreciated GAN models to synthesize realistic images from sketches (Table 4.6). However, we would like to highlight that due to the Covid-19 outbreak we had limitations on proper evaluation that we expected to apply. We could not run all the models due to it's hardware requirements, hence, we compared the results of our work with the accuracy found on their study. But, we tried to keep the testing data as much similar to avoid most of the inconsistencies that can arise in such scenarios.

Table 4.6: Comparisons of realistic scores to state of art models

Model	Realistic score
SketchyGAN	53.70%
Interactive Sketch & Fill	73.12%
Ours	**77.78%**

Although we received a good value for accuracy considering the current advancement, we believe we can significantly improve the accuracy by applying componentization of elements that has not been seen on any related studies. Hence, we aim to continue our study on this topic and implement a component based image synthesis from human drawn sketches.

Chapter 5

Conclusion

Image generation has seen tremendous progress with the introduction of GAN. Not only has the application of GAN been developed in the field of static image generation, but it has also been expanded in a variety of other fields. Notably, GAN integration has been effective in astronomical picture enhancement, fictitious fashion model development, video games, and texture enhancement. Through our work, we hope that we will be able to enhance the sketch to picture generation approach, which will be capable of producing a high-quality and realistic image from a sparse drawing.

Furthermore, from our study, we observed that most GAN based approaches has the blurry and lots of soft corner images are generated because of not componentizing elements of each class. We believe if we can apply this componentizing of elements, we can achieve much accurate and visually satisfied images from sparse sketches.

In addition, this study was to support our main idea of generating realistic images from sketches of a scene rather than just any particular objects. For that, along with elements componentization, we need to identify multiple classes of objects from a sketch.

Finally, we believe this study will have a powerful impact for next generation image synthesis from strokes or sketches and in a bigger picture, an improved scene analysis from sparse sketches for the rising of Artificial Intelligence.

Bibliography

[1] J. L. Dowling, *Criminal investigation*. Harcourt Brace Jovanovich New York, NY, 1979.

[2] P. Smolensky, “Information processing in dynamical systems: Foundations of harmony theory,” Colorado Univ at Boulder Dept of Computer Science, Tech. Rep., 1986.

[3] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in *Proceedings of the seventh IEEE international conference on computer vision*, IEEE, vol. 2, 1999, pp. 1033–1038.

[4] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[5] F. Cole, A. Golovinskiy, A. Limpaecher, *et al.*, “Where do people draw lines?” In *ACM SIGGRAPH 2008 papers*, 2008, pp. 1–11.

[6] X. Wang and X. Tang, “Face photo-sketch synthesis and recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 11, pp. 1955–1967, 2008.

[7] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, “Sketch2photo: Internet image montage,” *ACM transactions on graphics (TOG)*, vol. 28, no. 5, pp. 1–10, 2009.

[8] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang, “Mindfinder: Interactive sketch-based image search on millions of images,” in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1605–1608.

[9] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, “An evaluation of descriptors for large-scale image retrieval from sketched feature lines,” *Computers & Graphics*, vol. 34, no. 5, pp. 482–498, 2010.

[10] ——, “Sketch-based image retrieval: Benchmark and bag-of-features descriptors,” *IEEE transactions on visualization and computer graphics*, vol. 17, no. 11, pp. 1624–1636, 2010.

[11] R. Hu, M. Barnard, and J. Collomosse, “Gradient field descriptor for sketch based retrieval and localization,” in *2010 IEEE International conference on image processing*, IEEE, 2010, pp. 1025–1028.

[12] C. Wang, Z. Li, and L. Zhang, “Mindfinder: Image search by interactive sketching and tagging,” in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 1309–1312.

[13] Y. Cao, C. Wang, L. Zhang, and L. Zhang, “Edgel index for large-scale sketch-based image search,” in *CVPR 2011*, IEEE, 2011, pp. 761–768.

[14] M. Eitz, R. Richter, K. Hildebrand, T. Boubekeur, and M. Alexa, "Photosketcher: Interactive sketch-based image synthesis," *IEEE Computer Graphics and Applications*, vol. 31, no. 6, pp. 56–66, 2011.

[15] ——, "Photosketcher: Interactive sketch-based image synthesis," *IEEE Computer Graphics and Applications*, vol. 31, no. 6, pp. 56–66, 2011.

[16] R. Hu, T. Wang, and J. Collomosse, "A bag-of-regions approach to sketch-based image retrieval," in *2011 18th IEEE International Conference on Image Processing*, IEEE, 2011, pp. 3661–3664.

[17] T. Chen, P. Tan, L.-Q. Ma, M.-M. Cheng, A. Shamir, and S.-M. Hu, "Poseshop: Human image database construction and personalized content synthesis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 5, pp. 824–837, 2012.

[18] J. Ruskin, *The elements of drawing*. Courier Corporation, 2012.

[19] R. Hu and J. Collomosse, "A performance evaluation of gradient field hog descriptor for sketch based image retrieval," *Computer Vision and Image Understanding*, vol. 117, no. 7, pp. 790–806, 2013.

[20] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[21] Y.-L. Lin, C.-Y. Huang, H.-J. Wang, and W. Hsu, "3d sub-query expansion for improving sketch-based multi-view image retrieval," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 3495–3502.

[22] S. James, M. J. Fonseca, and J. Collomosse, "Reenact: Sketch based choreographic design from archival dance footage," in *Proceedings of international conference on multimedia retrieval*, 2014, pp. 313–320.

[23] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," *arXiv preprint arXiv:1506.05751*, 2015.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[25] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[26] D. Turmukhambetov, N. D. Campbell, D. B. Goldman, and J. Kautz, "Interactive sketch-driven image synthesis," in *Computer Graphics Forum*, Wiley Online Library, vol. 34, 2015, pp. 130–142.

[27] F. Wang, L. Kang, and Y. Li, "Sketch-based 3d shape retrieval using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1875–1883.

[28] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 2180–2188.

[29] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

[30] V. Dumoulin, I. Belghazi, B. Poole, *et al.*, "Adversarially learned inference," *arXiv preprint arXiv:1606.00704*, 2016.

[31] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *International conference on machine learning*, PMLR, 2016, pp. 1558–1566.

[32] K. Li, K. Pang, Y.-Z. Song, T. Hospedales, H. Zhang, and Y. Hu, "Fine-grained sketch-based image retrieval: The role of part-aware attributes," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2016, pp. 1–9.

[33] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," *Advances in neural information processing systems*, vol. 29, pp. 469–477, 2016.

[34] P. Sangkloy, N. Burnell, C. Ham, and J. Hays, "The sketchy database: Learning to retrieve badly drawn bunnies," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–12, 2016.

[35] O. Seddati, S. Dupont, and S. Mahmoudi, "Deepsketch2image: Deep convolutional neural networks for partial sketch recognition and image retrieval," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 739–741.

[36] D. Yoo, N. Kim, S. Park, A. S. Paek, and I. S. Kweon, "Pixel-level domain transfer," in *European conference on computer vision*, Springer, 2016, pp. 517–532.

[37] Q. Yu, F. Liu, Y.-Z. Song, T. Xiang, T. M. Hospedales, and C.-C. Loy, "Sketch me that shoe," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[38] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

[39] D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.

[40] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3722–3731.

[41] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," *arXiv preprint arXiv:1704.00028*, 2017.

[42] D. Ha and D. Eck, "A neural representation of sketch drawings," *arXiv preprint arXiv:1704.03477*, 2017.

[43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[44] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in neural information processing systems*, 2017, pp. 700–708.

[45] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, "Plug & play generative networks: Conditional iterative generation of images in latent space," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4467–4477.

[46] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5400–5409.

[47] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[48] W. Chen and J. Hays, "Sketchygan: Towards diverse and realistic sketch to image synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9416–9425.

[49] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *arXiv preprint arXiv:1812.04948*, 2018.

[50] A. Ghosh, R. Zhang, P. K. Dokania, *et al.*, "Interactive sketch & fill: Multiclass sketch-to-image translation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1171–1180.

[51] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, and K. He, "Fusegan: Learning to fuse multi-focus image via conditional generative adversarial network," *IEEE Transactions on Multimedia*, vol. 21, no. 8, pp. 1982–1996, 2019.

[52] Z. Li, C. Deng, E. Yang, and D. Tao, "Staged sketch-to-image synthesis via semi-supervised generative adversarial networks," *IEEE Transactions on Multimedia*, 2020.

[53] A. Sheoran, *Sketch to image*, version 1, 2020. [Online]. Available: https://www.kaggle.com/ankitsheoran23/sketch-to-image/version/1 (visited on 10/09/2021).