Reviving Black and White Images: Enhancing Colorization with Generative Adversarial Networks (GANs)

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
DATA SCIENCE

Submitted by:

DEVANSHU WALECHA

2K21/DSC/05

Under the supervision of

Mr. Rahul

Assistant Professor

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

MAY, 2023

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

CANDIDATE'S DECLARATION

I, Devanshu Walecha, Roll No. 2K21/DSC/05, student of Master of Technology (Data Science), hereby declare that the Major Project-II Dissertation titled **"Reviving Black and White Images: Enhancing Colorization with Generative Adversarial Networks (GANs)"** which is submitted by me to the Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of requirement for the award of degree of Master Of Technology (Software Engineering) is original and not copied from any source without proper citation. This work has not been previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Dev Walecha

Devanshu Walecha
(2K21/DSC/05)

Place: Delhi
Date: 31st May 2023

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **"Reviving Black and White Images: Enhancing Colorization with Generative Adversarial Networks (GANs)"** which is submitted by Devanshu Walecha, (2K21/DSC/05) to the Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of requirement for the award of the degree of Master of Technology, is a record of project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

31/5/23

Place: Delhi
Date: 31st May 2023

Mr. Rahul
SUPERVISOR
Assistant Professor
Department of Software Engineering
Delhi Technological University

TABLE OF CONTENTS

PAGE NO.

ABSTRACT

One of the more intriguing deep learning applications is colourizing black and white photographs. The technology of automatic image colorization has sparked a lot of attention in the recent decade for a range of applications, including the restoration of aged or degraded photos. This task used to need a lot of human input and hardcoding, but now, thanks to AI and deep learning, the entire process can be automated from start to finish. Using GANs to restore and recolorize historical photos is one answer to this problem. The purpose of this project is to demonstrate GAN functionality and superiority by using a Generative Adversarial Network (GANs) that accepts fixed size black and white images as input and produces corresponding coloured images of the same size as output. Concerning the dataset, I have approximately 3000 rgb photographs from diverse domains such as mountains, forests, and cities, which we will convert to grayscale and use as labels for our model. I used binary cross entropy as the discriminator's loss function and mean squared error as the generator's loss function, & then I used Adam to optimise the generator and discriminator. Moving on to the results, the colourized output from the generator was significantly closer to the original rgb image. The project's future work will include colourizing photographs of our grandparents in order to make the image more remembered, and I also intend to continue this work by colourizing videos of historical heroes such as Charlie Chaplin. Finally, I want this model to be implemented on a web platform so that users all over the world can convert their old black and white images to colourized versions.

ACKNOWLEDGEMENTS

The achievement of Major Project II necessitates the assistance and support of a large number of individuals and an organisation. This project's report writing opportunity allows me to thank everyone who contributed to the project's successful completion.

I would want to express my sincere gratitude to my supervisor, Mr. Rahul, for allowing me to work on my project under his supervision. Thank you very much for your support, encouragement and suggestions; without them, our work would not have been successful. His unwavering support and inspiration helped me to see that the process of learning is more important than the end result.

I want to express my sincere thanks to the faculty and personnel at the institution for providing us with a infrastructure, laboratories, library, suitable educational resources, testing facilities, and a working atmosphere that didn't interfere with our ability to complete our work.

I would also like to thank all of my friends and classmates for their unwavering support. They have assisted me in every way, providing me with fresh ideas, the knowledge I needed, and the will to finish the assignment. I want to express my gratitude to my parents for always supporting me after finishing my task.

Devanshu Walecha
(2K21/DSC/05)
M. Tech (Data Science)
Delhi Technological University

LIST OF FIGURES

LIST OF ABBREVIATIONS AND NOMENCLATURE

GANs:	Generative Adversarial Networks
ANN:	Artificial Neural Network
CNN:	Computation Neural Network
RGB:	Red Green Blue
AI:	Artificial Intelligence
DL:	Deep Learning
RMSE:	Root means squared error
MSE:	mean squared error
MAE:	mean absolute error
ML:	Machine Learning
DCGANs:	Deep Convolutional Generative Adversarial Networks
PSNR:	Peak Signal to Noise Ratio
SR:	Super Resolution
RELU:	Rectified Linear Unit

CHAPTER - 1

INTRODUCTION

1.1 Overview

Images are the primary source for understanding the world and obtaining information, while grayscale images cannot provide sufficient information. Compared with grayscale images, colorful images contain richer information and convey more intuitive feelings. There is a definite tendency towards integrating computer technology with traditional sectors as artificial intelligence advances. One of the more intriguing deep learning applications is colourizing black and white photographs. With the advancement of digital media technology in recent years, numerous techniques for colouring grayscale photographs have been suggested. The technology of automatic image colorization has sparked a lot of attention in the recent decade for a range of applications, including the restoration of aged or degraded photos. This task used to need a lot of human input and hardcoding, but now, thanks to AI and deep learning, the entire process can be automated from start to finish. Using GANs to restore and recolorize historical photos is one answer to this problem. If these grayscale images can have rich and reasonable color information, it can help humans better identify target objects and understand the scene. Potential use cases for this research include restoration of old blackand-white photos [1], automatic colorization of anime sketches [2], grayscale images in medical fields [3], etc. Colorization needs to estimate the missing color channels from a grayscale value, and reasonable colorization solutions are not unique. Colorization is still a challenging task due to its uncertainty and diversity [4].

Concerning the dataset, I have approximately 3000 rgb photographs from diverse domains such as mountains, forests, and cities, which we will convert to grayscale and use as labels for our model.

1.2 Motivation

The main motivation for implementing this project was that we wanted to make recolorization of old black and white images an easy task because it is not easy for everyone to use complex editing softwares to colorize their images. So, if I talk about the overall design of this project, there were two major components one was generator and the other was discriminator so for generator I am going to input original grayscale image and it is going to output generated image which would be rgb image and I am going to pass on the generated rgb image along with the original rgb image to the discriminator and the discriminator would going to figure out whether it was a fake image or a real image.

1.3 Objective

The goal of this project is colorization with deep learning methods. This has been divided into the following sub-goals:

1. First learn the research background in the field of deep learning and image colorization, as well as the application of deep learning methods in digital image processing.
2. Focus on the research progress of generative adversarial networks and corresponding optimization techniques. Starting from its characteristics, the goal is to explore what kind of network is suitable for application in the field of image colorization.
3. In terms of models, the goal is to propose an improved generative adversarial network image colorization algorithm. It can improve the global feature extraction ability on high-resolution images.

The purpose of this project is to demonstrate GAN functionality and superiority by using a Generative Adversarial Network (GANs) that accepts fixed size black and white images as input and produces corresponding-coloured images of the same size as output.

1.4 Organization of Dissertation

Chapter 1 briefly takes you through the introduction of the topic thereafter section 1.2 shows the motivation behind the project, section 1.3 gives the objectives of the project. In Chapter 2, background if the work has been through explained. In sections 2.2 & 2.3, you will find the complete description of the Machine Learning and ANN, thereafter in section 2.4 & 2.5, the Impact of Convolutional Neural Networks on Image Processing and CNN and its layers has been explained thoroughly. In section 2.6, the complete information for understanding GANs has been given. In its subsections you will find the complete details related to Generator & Discriminator and their working principle. A brief understanding of UNET architecture has also been given. Chapter 3 takes you through the summary of previously related studies. Chapter 4 gives the complete information related to implementation & methodology, it also explains the Generator model and discriminator model structure. Chapter 5 shows the results that have been gathered and evaluation of those results. Chapter 6 gives the conclusion on the complete idea of the topics and results; it also discusses about the future scopes related to the idea. In last, some references that have been taken are mentioned.

CHAPTER - 2

BACKGROUND

2.1 Overview

Information exchange is essential in the modern day, and photographs have emerged as a common tool for doing so. In many facets of human existence, including medicine, agriculture, business, and social media, images have proven to be useful. An image's basic building block is a pixel, and the resolution of a picture is determined by the quantity of pixels in a specific area. Resolution, to put it simply, governs the clarity and level of detail that an image may include. A higher resolution makes it possible to distinguish between closely spaced lines and to see more complex details. However, the resultant image's resolution usually falls short of the required criteria for a variety of reasons, such as the use of inexpensive image-capturing equipment, the inability to capture all image elements, or technical difficulties with the capturing procedure. As a result, crucial features that pixels convey are lost. The image processing method known as image super-resolution, which raises the resolution and clarity of the image by adding additional pixels to the same unit area of the image, solves this problem. Deep learning, a well-liked machine learning method that is a subset of artificial neural networks, has made significant progress in the field of image super-resolution.

2.2 Machine learning

Machine learning has become a crucial and important component of this era as a result of our increasing reliance on machines. Depending on the dataset at hand, machine learning entails teaching a computer to carry out particular tasks using either supervised or unsupervised techniques. Machine learning is important in the field of image processing. It is now possible to improve the resolution of new photos by teaching a machine to recognise relationships between low-resolution and high-resolution images. However, the precise picture super-resolution technology used will determine the level of clarity and quality of the final high-resolution image. Artificial Neural Networks

(ANNs), one of the many machine learning techniques, have become extremely popular and widely used. ANNs consistently outperform other machine learning models and give excellent outcomes across various problem domains.

2.3 ANN (Artificial Neural Network)

ANNs, or artificial neural networks, are created to mimic the neuronal architecture of the human brain. They try to mimic the biological interactions and cognitive development seen in our brains. The complex network of billions of neurons in a biological brain is made up of the following components:

a) Axon: Neurons are interconnected through axons.
b) Dendrites: They receive input from the previous layer of neurons.
c) Synapses: They transmit output to the next layer of neurons.

Nodes in ANNs represent neurons, and connections between nodes represent dendrites and synapses. Similar to how connections in the biological brain are adaptive, these connections have movable weights. A node applies an activation function to an input when it is received. The node activates and sends a signal to the associated nodes in the following layer if the calculated value exceeds a predetermined threshold. If the threshold is not met, on the other hand, the node remains dormant and nothing happens. Following the learning process, if a node continuously remains inactive, meaning its activation function output never exceeds the threshold, the node or its associated link is regarded as "dead." This idea is based on the finding that the human brain forms certain connections as it learns a particular task, like telling dogs from cats. Only these specific neurons are active during the discriminating process; other neurons are not. The ability to repeatedly carry out a learnt activity is made possible by these recently formed connections or patterns of connections between the particular neurons.

2.3.1 Advantages of ANN

- ANNs are capable of learning and understanding different relationships between input and output on their own, including complicated nonlinear correlations that are frequently seen in real-world circumstances. This capacity to learn complex mappings between inputs and outputs is facilitated by the network's depth, layers, and structure.

- An ANN can learn and form connections by being trained on large datasets, which enables it to behave intelligently like humans. This makes it possible for the network to operate efficiently when faced with brand-new, untested data during actual usage.

- ANNs are not restricted to a particular set of input data types or formats. In situations like this, they are known to produce better results than other models since they can manage diverse data.

2.3.2 Disadvantages of ANN

- The successful training of ANNs requires a sufficient amount of data. In order to properly saturate the weights and biases based on the input, a larger dataset is required as the model's complexity rises with a greater number of weighted connections. The model's accuracy is improved by the availability of data.

- Due to machine learning's intrinsic complexity, training ANNs takes more time. The complexity grows tremendously when ANNs are used in an effort to mimic the human brain. The method involves a lot of processing power, and the model needs to stabilise after several iterations. Compared to central processing units (CPUs), graphics processing units (GPUs) can aid in reducing training time.

- It is critical to fine-tune the architecture parameters of an ANN. The network includes a wide range of parameters, including weights, biases, learning rate, batch size, and the number of iterations or epochs. To create a reliable and effective network that can deliver precise results, these factors must be adjusted carefully.

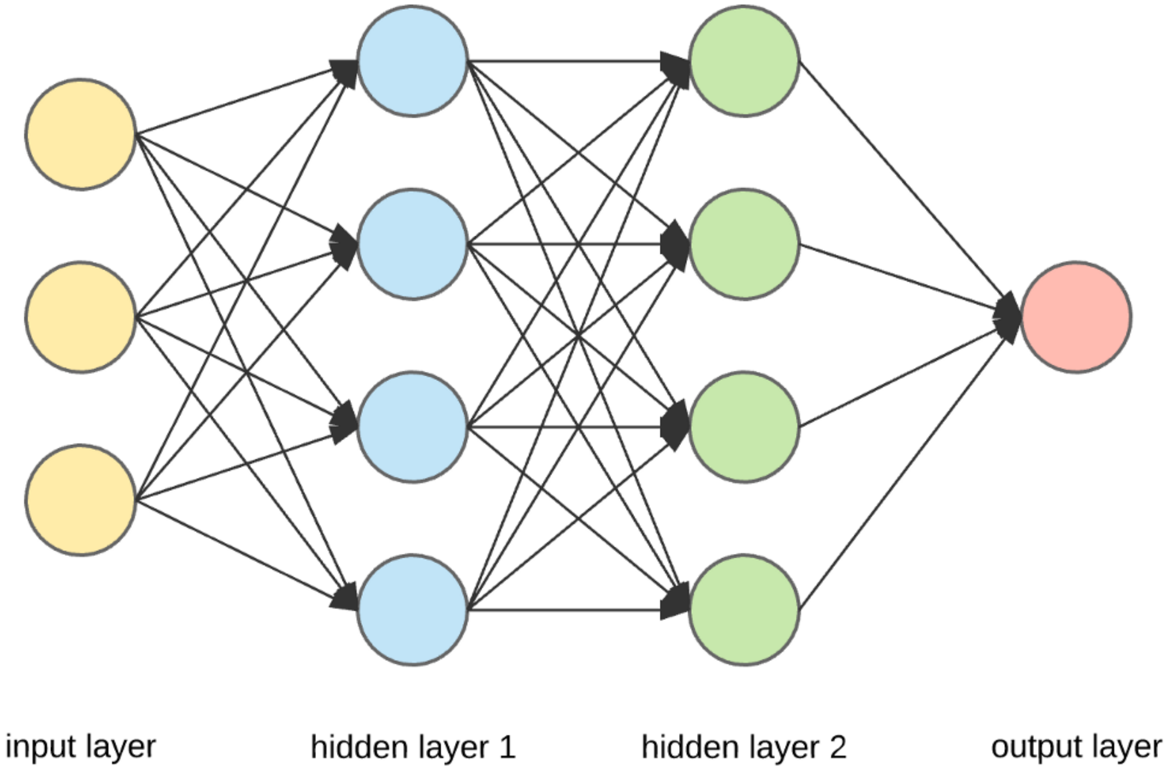


Figure 2. 1: Architecture of ANN

2.3.3 Applications and Importance of Image Super Resolution

Improving image resolution has a wide range of practical applications. A few examples include:

a) Medical Image Processing: For correct diagnosis and prompt treatment, high-resolution and clear images are essential in medical imaging such as MRI or CT scans. While X-ray images may have low contrast, ultrasound images frequently contain noise. Research and treatment in medicine can considerably benefit from increasing the resolution of these images.

b) Satellite Image Processing: Because of the distance and available technology, satellite-derived images frequently exhibit blurriness and lack of clarity. It is therefore essential to increase the pixel count in order to repair and improve these photographs. This procedure enriches the geographic information in the image, minimises distortions, and improves the visual understanding of the image.

c) Multimedia and Video Enhancement: Multimedia apps are important in today's society because they give users engaging experiences. Clearer, higher-definition, and less jarring pictures are provided by high-resolution photographs, which improve visual quality. Old computer games can benefit from super-resolution technology to increase and enhance the visual experience. Since videos are made up of individual images, raising the resolution of each image raises the resolution of the entire video.

d) Microscopic Image Processing: Image super-resolution methods are essential for examining and learning about microscopic objects like cells and their structure. By supplying information at the nanoscale scale, these strategies can enhance aesthetically subpar photos, hence enhancing clarity and resolution. Clearer and images with a greater resolution can be produced by combining many switching mode photos.

e) Processing of Astronomical photographs: Astronomical photographs sometimes lack visual differentiation between closely packed celestial objects and are blurry. These photos can be made more visually clear using super-resolution techniques, which will help astronomical research and study.

f) Additional Uses: Image super-resolution has further uses in the surveillance, military, forensics, automotive, real-time processing, scanning, and object identification industries.

Convolutional Neural Networks (CNNs) are the most utilised and well-liked neural networks. CNNs are crucial for many applications due to the growing emphasis on artificial vision and task automation. They are essential for the development of self-driving automobiles, autonomous robots, and related technologies because they are excellent at processing images, identifying objects, detecting scenes, human faces, and other image features.

2.4 The Impact of Convolutional Neural Networks on Image Processing

Convolutional Neural Networks, or CNNs, are particularly well known as widely applied and very powerful neural networks for image processing tasks. It is excellent at detecting faces, objects, and other image characteristics. CNN is an example of a deep learning model that allows computers to learn from examples in a manner similar to how people learn. With the help of this technology, driverless cars will be able to distinguish between a stop sign and a lamppost and make intelligent decisions under a variety of driving circumstances. The performance of deep learning has been greatly improved by the accessibility of increased computing power, making it possible to achieve remarkable outcomes that were previously unachievable.

In deep learning, the word "deep" denotes the neural network's integration of several hidden layers. Traditional artificial neural networks typically comprise of just two to three hidden layers, however deep learning models can have up to 150 hidden layers. Filters were manually constructed in traditional image processing ways based on the specific problem at hand. This method required a lot of time, effort, knowledge, and energy from the participants. CNN, on the other hand, overcomes this difficulty by automating the feature extraction procedure. It gradually learns which aspects to focus on, if they are present in the image, and how to interpret them. It independently pulls the key features from the photographs. This learning process is made easier by CNN's numerous hidden levels, with each layer building on the abstractions that were extracted by the layers before it.

2.5 CNN (Convolutional Neural Network)

Convolutional neural networks, or CNNs, are frequently used for image processing applications. Height, width, and the number of channels are the three dimensions of the input images supplied into a CNN. The width and height represent the image’s resolution, while the 3rd dimension corresponds to the no. of channels, which can be the pixel values for the RGB - red, green, and blue colors, respectively.

The commonly found layers in a Convolutional NN model include:

1. Convolutional Layer
2. Activation Layer (e.g., ReLU or LeakyReLU)
3. Pooling Layer
4. Batch Normalization Layer
5. Dropout Layer
6. Fully Connected Layer

Each of these layers serves a specific purpose in the overall architecture of the CNN.

2.5.1 Convolution Layer:

The first layer to receive the input image is the convolution layer. To build a feature map, it uses a kernel or filter, which is frequently square in shape. Using fixed gap intervals known as strides, the filter sweeps across the image. To get the desired outcomes, the stride size needs to be properly adjusted. The dot product is computed between the filter's pixel value and the corresponding pixel value from the area of the picture covered by the filter is calculated during the convolution process. The position in the convolved feature map matrix that corresponds to the total of these dot products is then assigned. The result of this procedure is a feature map with fewer dimensions. The filter may extract distinct features from the image by having a variety of shapes and values. One filter might be used to capture curves and lines, while another would be used to recognise particular colour intensities. The intricacy of the filters rises as we move through the network's tiers. Complex filters have the ability to extract features that combine features from earlier filters.

2.5.1.1 Adjustable Parameters for Fine-Tuning CNN Performance:

- Stride: The number of pixels by which the filter is moved as we scan the image is referred to as the stride. It impacts the size of the final feature map and the degree of overlap between adjacent receptive fields. A larger stride may result in a smaller output map, which may result in some information being lost.

- Padding: Padding entails adding zeros to the input image's edges. This is done to guarantee that the convolutional output feature map has the appropriate dimensions. In addition to allowing the convolutional operation to be applied to the whole input, padding aids in maintaining the spatial dimensions of the image.

- Filters: Filters, also referred to as kernels, are of various types and are used to extract particular features from images. Each filter is in charge of identifying a specific pattern or trait. Filters are trainable parameters that are modified to enhance the network's capacity to identify pertinent features in the input data.

2.5.2 Activation Layer

The activation layer is essential in deciding which values should be passed on to the next layer and which ones should be ignored. It often includes an activation function, such as a ‘Rectified Linear Unit (ReLU) or a Leaky ReLU’, that only activates the output connection when a predetermined “threshold” value is surpassed. The network gains non-linearity from the activation function, which enables it to learn complicated correlations and produce more expressive predictions.

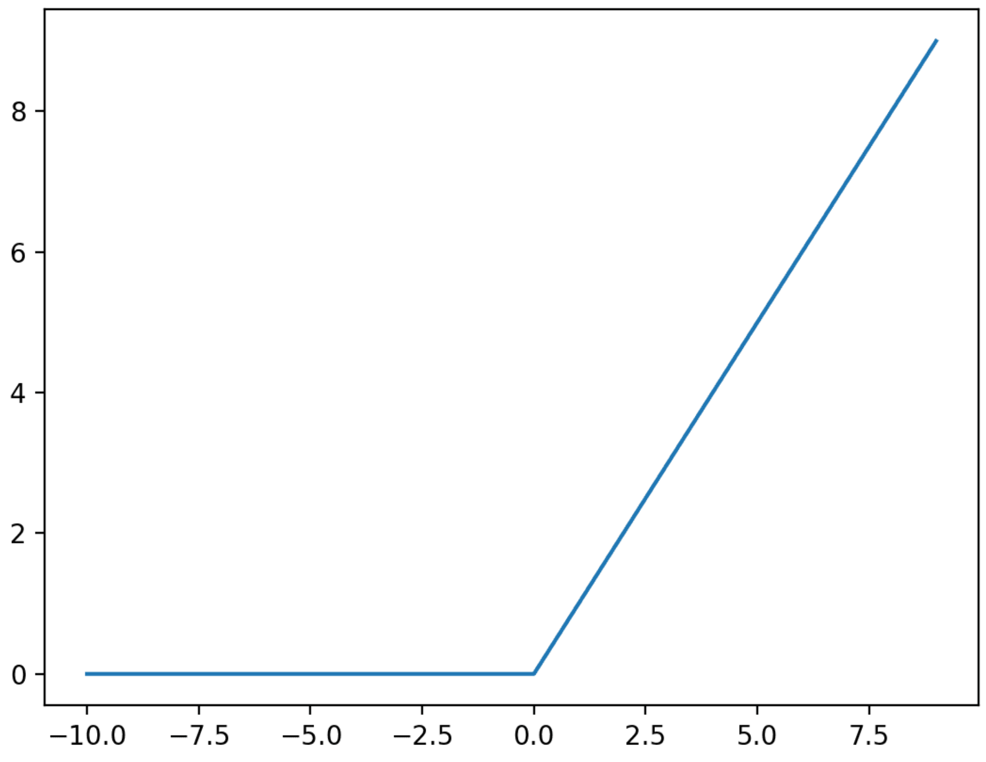


Figure 2. 2: Rectified Linear Unit (ReLU) function

2.5.3 Pooling Layer

The pixel within a set of pixels that contributes the most to the final output is chosen using the pooling layer, with the remainder pixels being ignored. Its primary function is to further reduce the size of the feature-map by eliminating unused sparse cells in the resultant matrix that hold no significant value. Although this matrix size reduction improves the process's speed and efficiency, it also results in information loss. The fundamental idea behind pooling is that the pixel with the most useful data can be used to estimate other nearby or adjacent pixels. Typically, there are three different pooling techniques:

- Max Pooling: This technique chooses the pixel from the collection of pixels with the highest value.
- Average Pooling: This technique determines the average value of each pixel in the group.
- Sum Pooling: This technique computes the sum of all the pixel values in the given group.

2.5.4 Dropout Layer

An essential element in preventing overfitting in a neural network is the dropout layer. When the model receives similar inputs frequently during the training phase, it may develop a set pattern of learning and inter-node connections. When given many sorts of inputs, this rigid pattern cannot generalise. The Dropout layer chooses nodes at random and sets their connection weights to 0, thus dropping or deactivating those nodes, to solve this problem. This forces the model to develop alternate routes and make use of various nodes and connections in order to create the required mappings. As a result, the network's effectiveness is improved. Dropout makes it possible for the network to actually learn and adapt, allowing it to function successfully even when the input somewhat varies, rather than merely memorizing a direct mapping between specific input-output pairs.

2.5.5 Fully Connected Layer

Typically, an inverted pyramid shape is used to represent the number of parameters in a CNN model, which gradually converge to the intended output classes. This is accomplished by flattening the filter matrix into a single vector and then feeding it into the fully linked layer. This flattened vector is processed by the fully connected layer, which then sends the output through activation functions like softmax or sigmoid. By placing the image into distinct groups, this last step helps to categorise it.

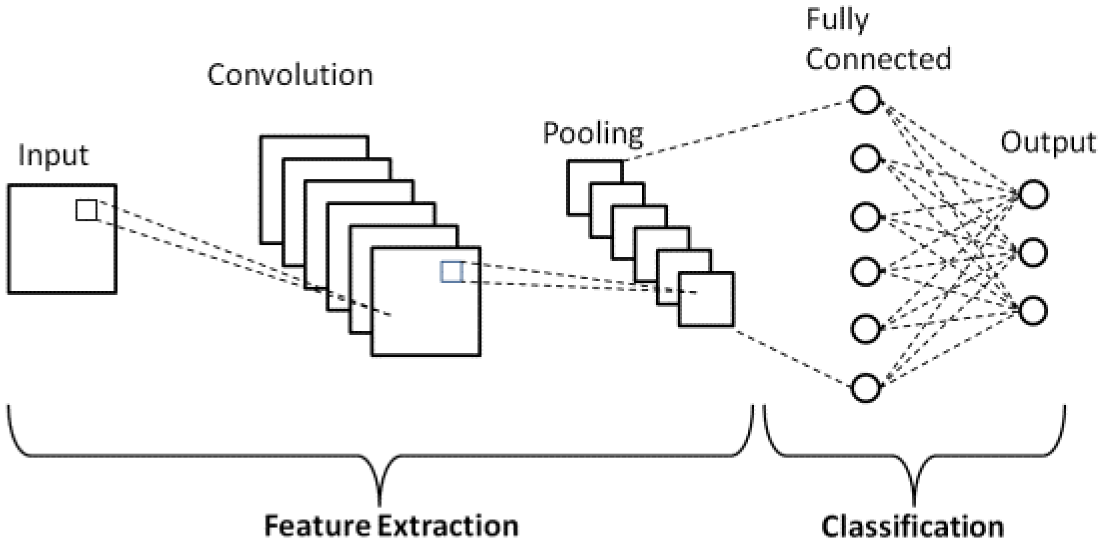


Figure 2. 3: Architecture of CNN

2.6 What are GANs?

GANs (Generative Adversarial Networks) are generative models that are created by competing two neural networks.

The generator and discriminator are the two main components of GANs.

To fool the discriminator, the generator turns noise into a replica of the data.

The discriminator attempts to distinguish genuine data from data generated by the generator.

The generator develops itself to the point where it can mislead the discriminator by decreasing its loss. Fooling the discriminator results in the discriminator producing probability even for generated photographs (closer to 1.0).

In order to maximise the discriminator's loss, created images would need to be accurately classified, and images from the dataset would need to produce good probabilities (closer to 1.0).

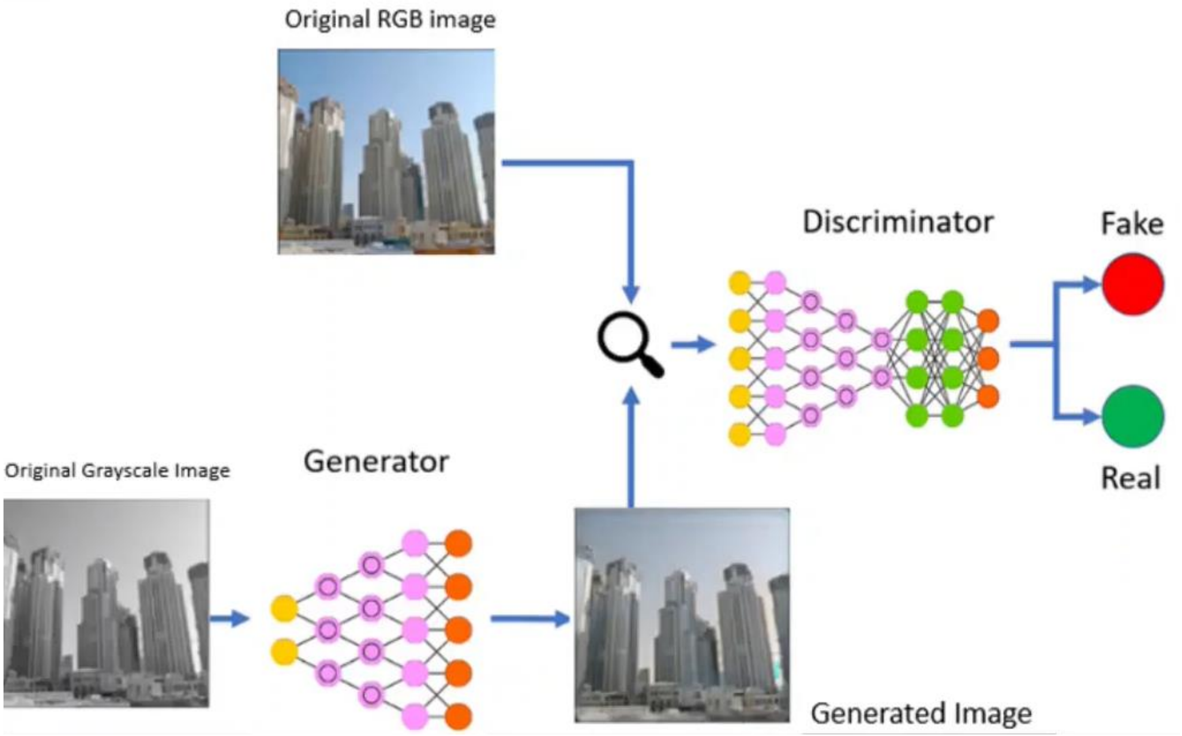


Figure 2. 4: GAN

Courtesy: Coderz Den

The discriminator differentiates between actual and bogus data, and the generator's main purpose is to mislead the discriminator.

2.6.1 Generator

The generator will take a grayscale or black-and-white image and convert it to RGB. The encoder-decoder structure of our generator will be symmetrically arranged layers.
The encoder will create a latent representation (z) of a grayscale image (also called the bottleneck representation).
The decoder's function is to enlarge this hidden representation and produce an RGB image. Most autoencoders and other encoderdecoder architectures employ this strategy.

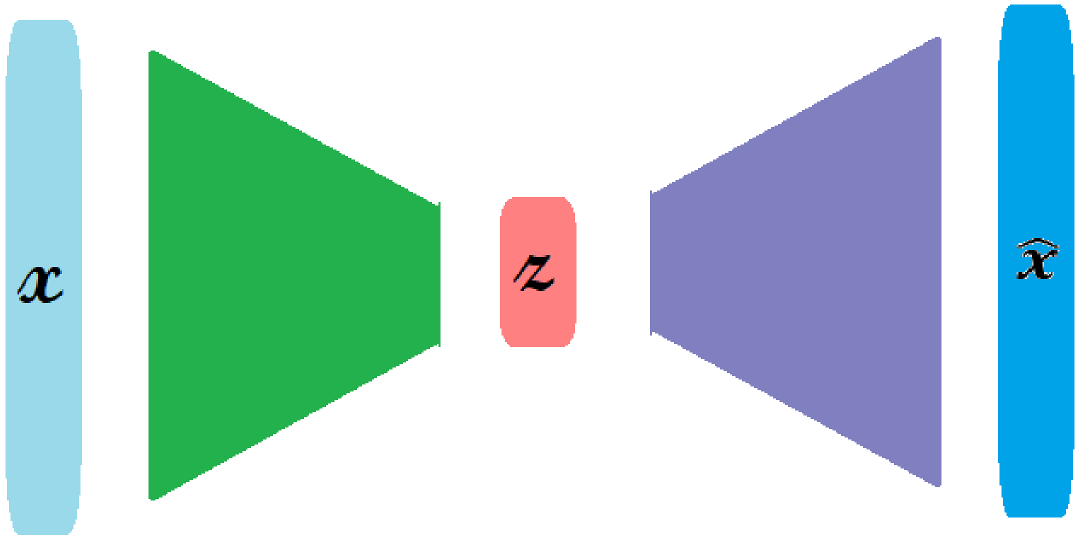


Figure 2. 5: Generator

An important aspect to mention here is that we are utilising a unit architecture to segment the objects from the photos because when we talk about the images being input to the generator, we have an image with different items such as a bird, a flower, or any other vegetation, and these objects are being separated using unit architecture.

2.6.2 UNET Architecture

The architecture consists of two ways.

The first path is the contraction path (also known as the encoder), which is used to record the image's context. A stack of convolutional and maximum pooling layers serves as the encoder.

The second path is the decoder, which uses transposed convolutions to provide precise localization. It is also referred to as the symmetric expanding path.

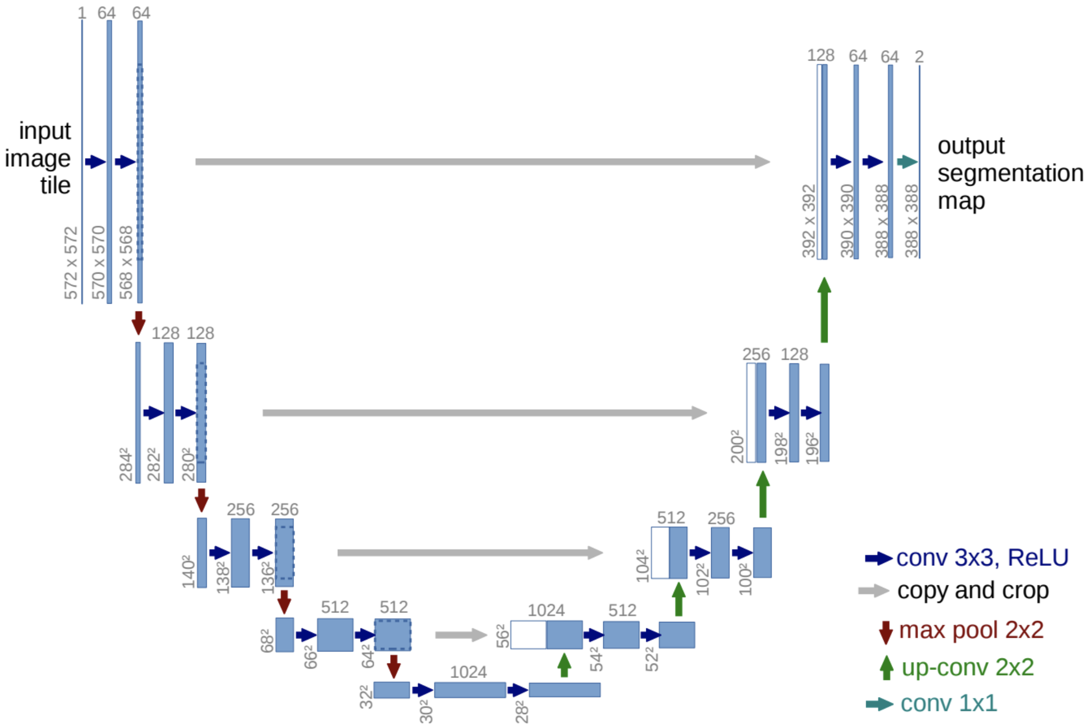


Figure 2. 6: U-net architecture

(Ronneberger, 2015, October)

In the diagram's representation of the U-net architecture (particularly for a resolution of 32x32 pixels), blue boxes stand in for multi-channel feature maps. Each box has a top-mounted channel count indicator and a lower-left edge with an x-y size indication. Feature maps with white boxes indicate copies. Different operations are represented by the arrows in the diagram.

2.6.3 Discriminator

We will use a conventional CNN as the discriminator to categorise the input. It will analyse an image and determine whether it was created (by the generator) or original.

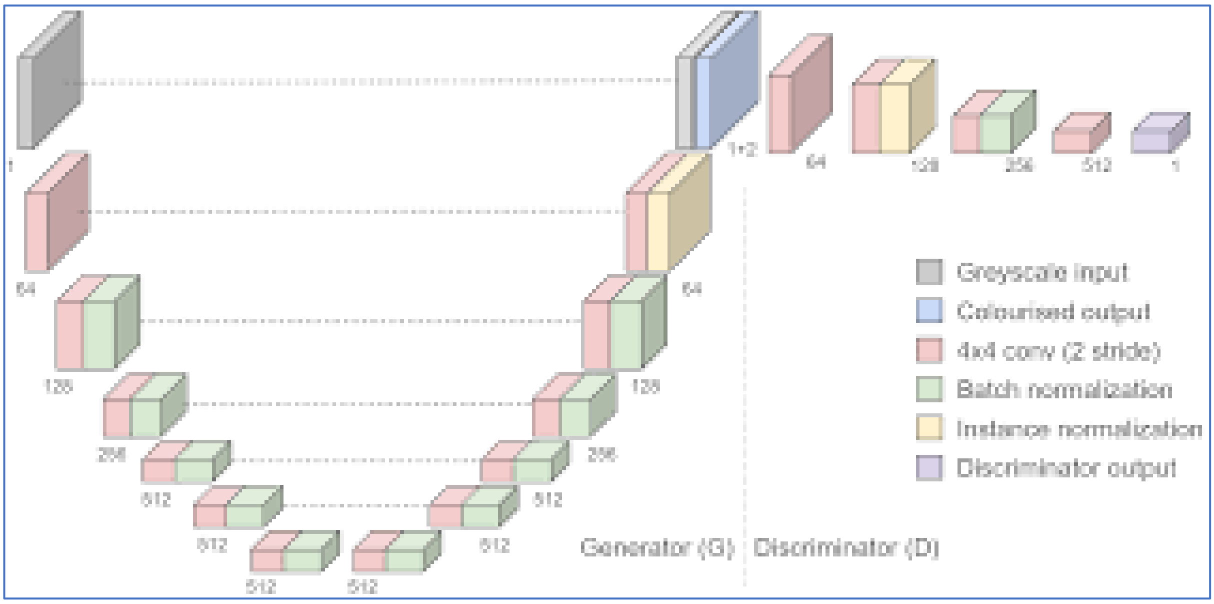


Figure 2. 7: UNET Generator + Discriminator

CHAPTER - 3

LITERATURE SURVEY

Sarmai and Sadhu [6] introduced a novel deep learning approach for image colorization, the networks' goal is to transform black-and-white photos to the appropriate colour format. The proposed method utilizes the RGB color space as the basis for the network implementation. To evaluate the performance of the autoencoder model, metrics such as ‘root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), and colorization accuracy’ were employed. A dataset consisting of 7000 RGB images was utilized to train and implement the networks. The effectiveness of the autoencoder models and generative adversarial network (GAN) was assessed using the mean absolute error metric, revealing that the GAN outperforms the autoencoder models in terms of colorization capability.

In their research, Zhang et al. [7] introduced an innovative approach to image colorization using deep learning. They developed two variations of the CNN network, namely the global hints network and the local hints network. A U-Net architecture is used in the primary branch of their proposed network. The network is made up of ten convolutional blocks. The evaluation of the models was performed using the peak signal-to-noise ratio (PSNR) metric. The local hints network, in addition to colorization, provided predictions for color dispersion. On the other hand, the global hint network did not incorporate spatial information. The global hints network's inputs were processed through four conv-relu layers, each with 512 channels and a kernel size of 11. Based on the authors' analysis of semantic similarities, they concluded that their network effectively predicts user-intended actions.

Madhab Raj Joshi et al.[8] InceptionResnetV2 and deep convolutional neural networks were utilised by the researchers to transform grayscale photos into coloured ones. They use CIE Lab as their colour space. ‘Mean squared error (MSE) and peak signal-to-noise ratio’ are the objective functions they employ to evaluate the colour image's quality (PSNR). 1.2 million old photos were utilised to train their CNN model. The models' MSE, PSNR, and accuracy are 6.08 percent, 34.65 dB, and 75.23 percent, respectively, according to the authors.

Abhishek Pandey et al. [9] conducted a study where they employed a deep CNN autoencoder that incorporated pre-trained layers of InceptionResnetV2. The authors resized the images used in their experiment to dimensions of 256*256 pixels. The evaluation of the model was based on the mean squared error (MSE) metric. Their encoder is split into four sections. The feature extraction generator produces high-level features, whilst the encoder produces mid-level features. These features were subsequently combined in the fusion layer of the autoencoder.

Baldassarre et al. [10] took a similar approach to the previously stated publications in their investigation. They coloured historical photographs as part of their task. They utilized approximately 60,000 photos as validation data, with a 10% data split and a batch size of 100. The writers recolored the grayscale using the CIE Lab colour system images. Several test cases were used to train the model. According to the authors, the model can colourize some photos with an accuracy of up to 80%.

In their research on automatic image colorization, Kotala et al. [11] took a similar technique to Zhang et al. They chose pictures from the ImageNet library, scaled them to 224*224 pixels, and then used those pictures. They used a CNN architecture, similar to Zhang et al., that was modelled after a VGG-style network and consisted of several convolutional blocks. The output colour was chosen by the authors using multinomial log loss as a criterion, and it was then used to colour rebalancing.

Nguyen et al. [12] presented a unique methodology f for colorization of ukiyo-e—a Japanese painting genre—using deep learning techniques, in their study. To distinguish style from content, the authors utilized a pre-trained convolutional neural network that was originally designed for image categorization. They subsequently proposed an approach to colorize grayscale images and merge their content with the separately identified style, resulting in a colorized output.

In their study, Charpiat et al.[13] used a machine learning strategy for automatic image colorization. Support vector machines (SVM) were utilised to do this. The authors employed the CIE Lab colour space to transform grayscale to colour images. Until the prediction step, their method keeps the images' multi-modality. In contrast to previous methods that used manual selection of spatial coherence criteria, this framework provides a logical way of learning local colour predictors along with spatial coherence criteria.

In his study, Stephen Koo[15] attempted to implement automatic image colorization using DCGANs (deep convolutional generative adversarial networks). They used the CIFAR-10 dataset to test their model. The collection contains 60,000 photos, each with a resolution of 32*32 pixels. Their basic method involves learning a mapping from grayscale to colour image space directly. The training process involved using batches of 128 images for up to 125 epochs. The optimizer employed a learning rate decay of 1*10-7. During training, the system achieved a training loss of 0.0071 and a validation loss of 0.0073.

CHAPTER - 4

METHODOLOGY

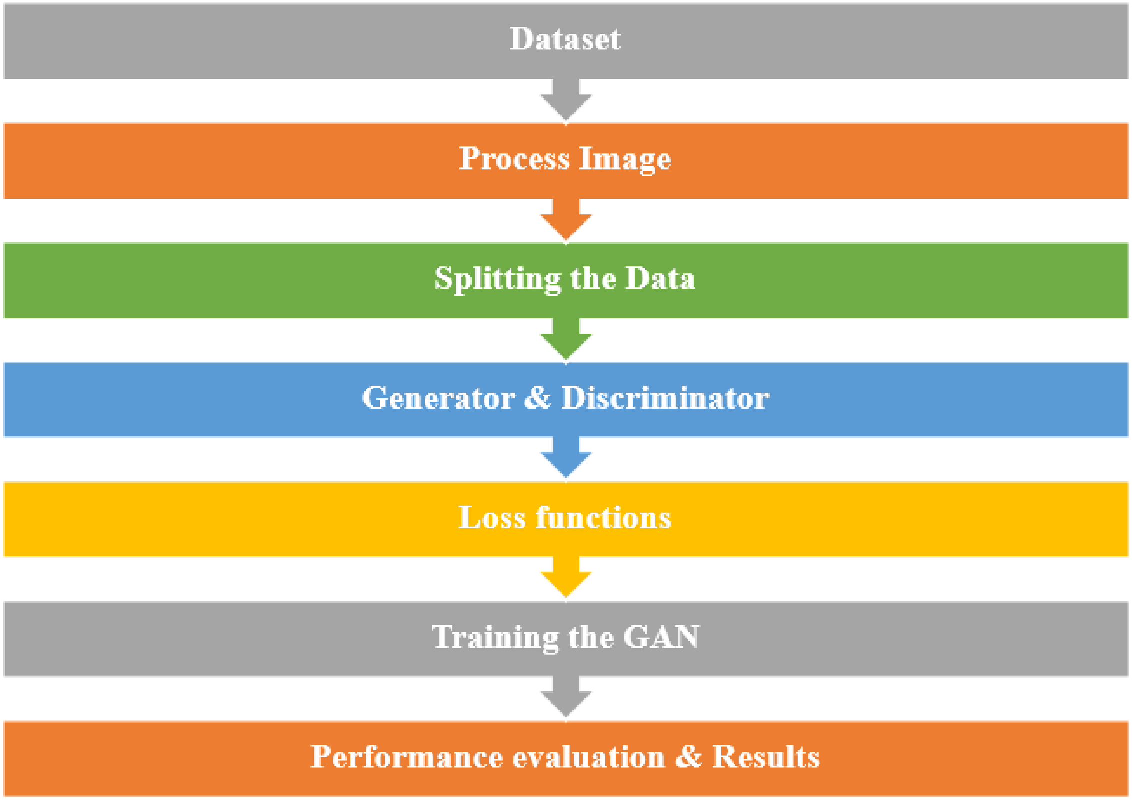


Figure 4. 1: Methodology

Firstly, I downloaded and processed the data so if I talk about the data, the dataset includes 3000 RGB photos from different fields (forests, mountains, cities, etc.,). Thereafter formulated it in a grayscale image and rgb image and kept the batch size of 64, image size of 120 and after that splitted the data to 2500 images because using all 3000 images was in computational overhead so over here what we tried to use using pill library that we convert the rgb images into a grayscale image in order to have both grayscale image and the original rgb image so here we formulated an x and y for our model and then we splitted the data set for training purpose and the testing purpose.

Now let's see how a generator was implemented, it had batch size of 64 and 120 by 120 which was our image size and initially it would be one because the input to the generator is a black and white image and the output is going to have a shape of batch size 64 120 by 120 image size and three because it is going to be an rgb image so we have a three convolutional layers we have convolutional 2d layer on each and the leaky relu activation function is being implemented on each layer subsequently bottleneck the latin representation z that has a stride of 1, activation function tanh and the padding is similar to the convolutional layer 3. After that we are going to concatenate all the layers in order to form our model and here during the concatenation we use the activation function relu moving towards the discriminator structure so if we talk about the discriminator structure for our project we have a standard convolutional neural network which is used for classification because the discriminator is going to find out that whether the data input to the discriminator was real or fake consequently, it will take an image and produce a likelihood indicating whether the image was created by the user or a generator., thus discriminator is a standard convolutional neural network in which we have four convolutional 2d layers the input layer has a size of 32 subsequently followed the hidden layers which have 64 128 and 256 size after that each layer has a max pooling layer after that these all layers are being flattened and throughout these steps value activation function is being used is relu and on the output layer we have sigmoid in order to classify the inputs whether it was a real image or a fake image.

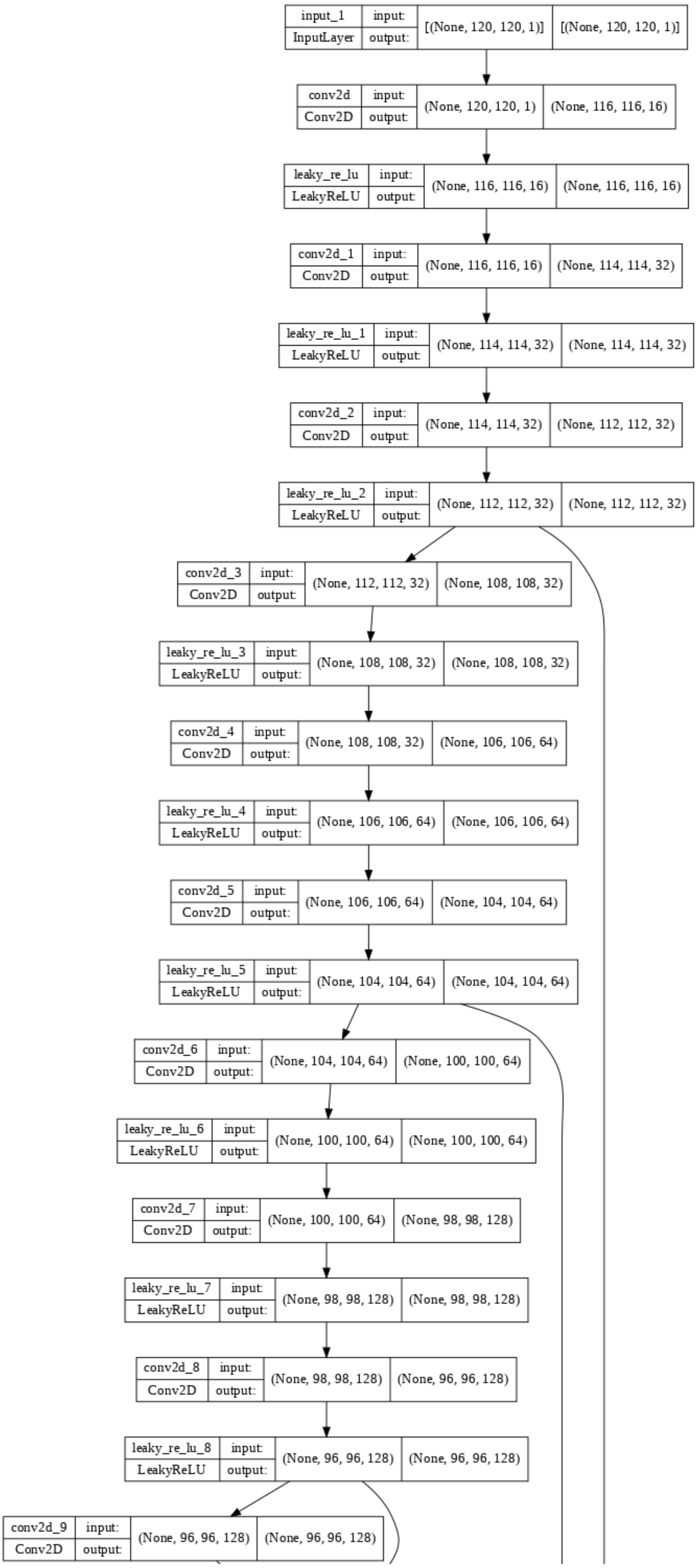
For a single step, we ran the generator once and the discriminator twice because reading different literature on gans we found out state of the art technique that discriminator needs to be run a greater number of times than generator.

The generator develops itself to the point where it can mislead the discriminator by decreasing its loss. Fooling the discriminator results in the discriminator producing probability even for manufactured photographs (closer to 1.0).

In order to maximise the discriminator's loss, created images would need to be accurately classified, and images from the dataset would need to produce good probabilities (closer to 1.0).

The discriminator will be trained to yield probabilities that are closer to 1.0 for real photos (from our dataset) and closer to 0.0 for images created by the generator. The output probability for genuine photos for a discriminator that is "smart" will be closer to 1.0. (coming from our dataset). Therefore, even if the images are fake, we are educating our generator to produce realistic images that will increase the discriminator output probabilities near 1.0. (not from our dataset, but from the generator).

4.1 Generator model



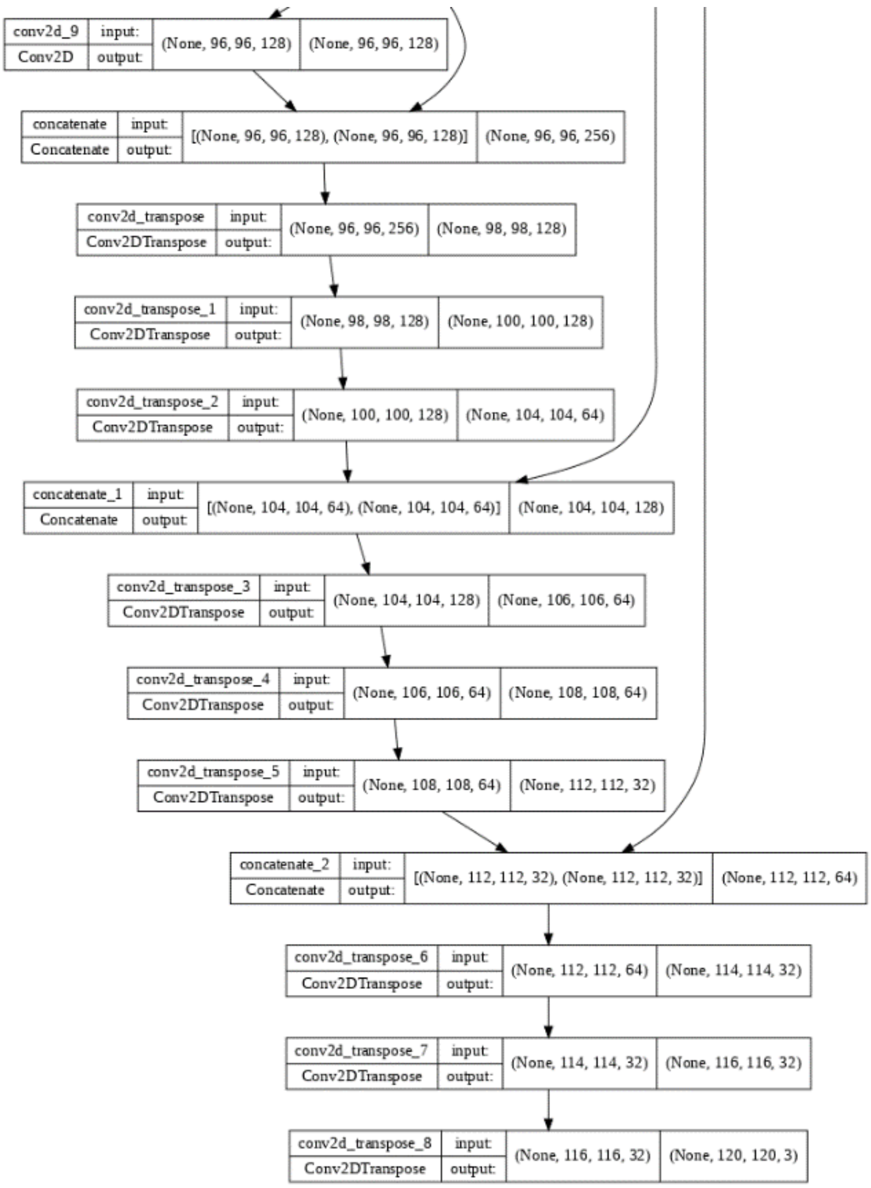


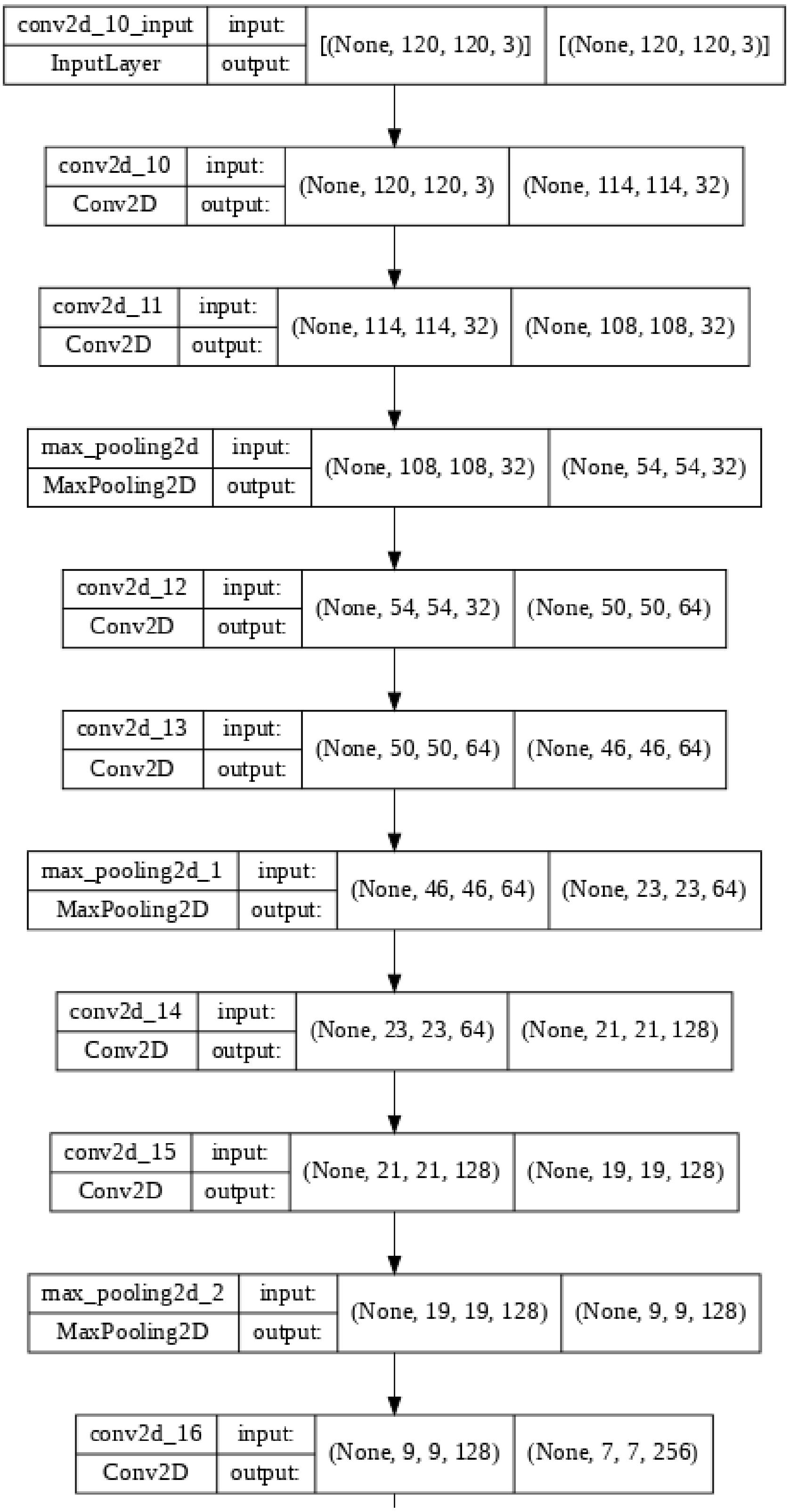
Figure 4. 2: Generator model

The generator (denoted by G) will take a grayscale image x and generate an RGB image G (x). It should be noted that x will be a tensor of shape (batch size, 120, 120, 1), and the output G(x) will have a shape (batch size, 120, 120, 3).

The encoder-decoder structure of the generator will be comparable to the UNet architecture. In addition, to have a large receptive field, we use Dilated convolutions.

In order to improve the flow of information from the encoder to the decoder, we implement skip connections into our model.

4.2 Discriminator model



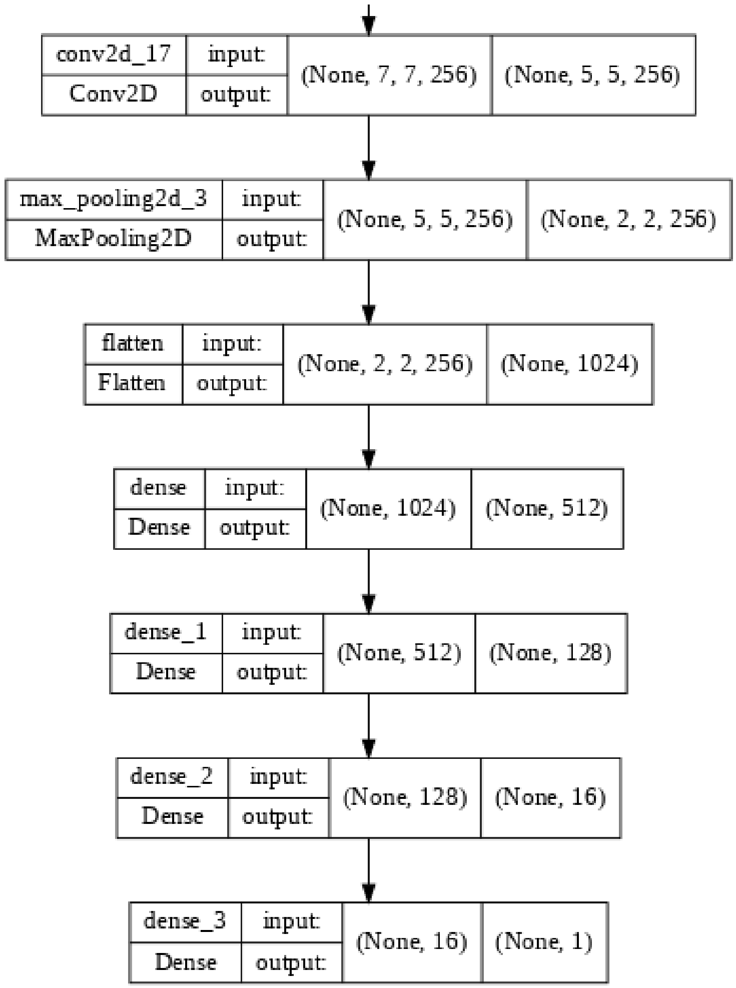


Figure 4. 3: Discriminator model

The discriminator model, denoted by D, will output two probabilities from the real picture y (from the training data) and the generated image G(x) (from the generator).

We train the discriminator in such a way that it can distinguish between actual and produced photos. As a result, we train the model to produce a 1.0 output for y and a 0.0 output for G(x).

We utilise soft labels that are close to 1 and 0 instead of hard labels such as 1.0 and 0.0. So, given a hard label of 1.0, the soft label will be $(1-\epsilon)$, with ϵ being randomly chosen from (0,0.1].

CHAPTER - 5

EVALUATION AND RESULTS

For the evaluation of the model, I have implemented the loss functions for our GAN model. As you might know that we have two loss functions, one for the generator and another for the discriminator.

For discriminator, we'll use the binary cross entropy as a loss function and
For generator, we'll use the L2/MSE loss function.

For optimization, The Adam optimizer is utilized, with a learning rate of 0.0005.

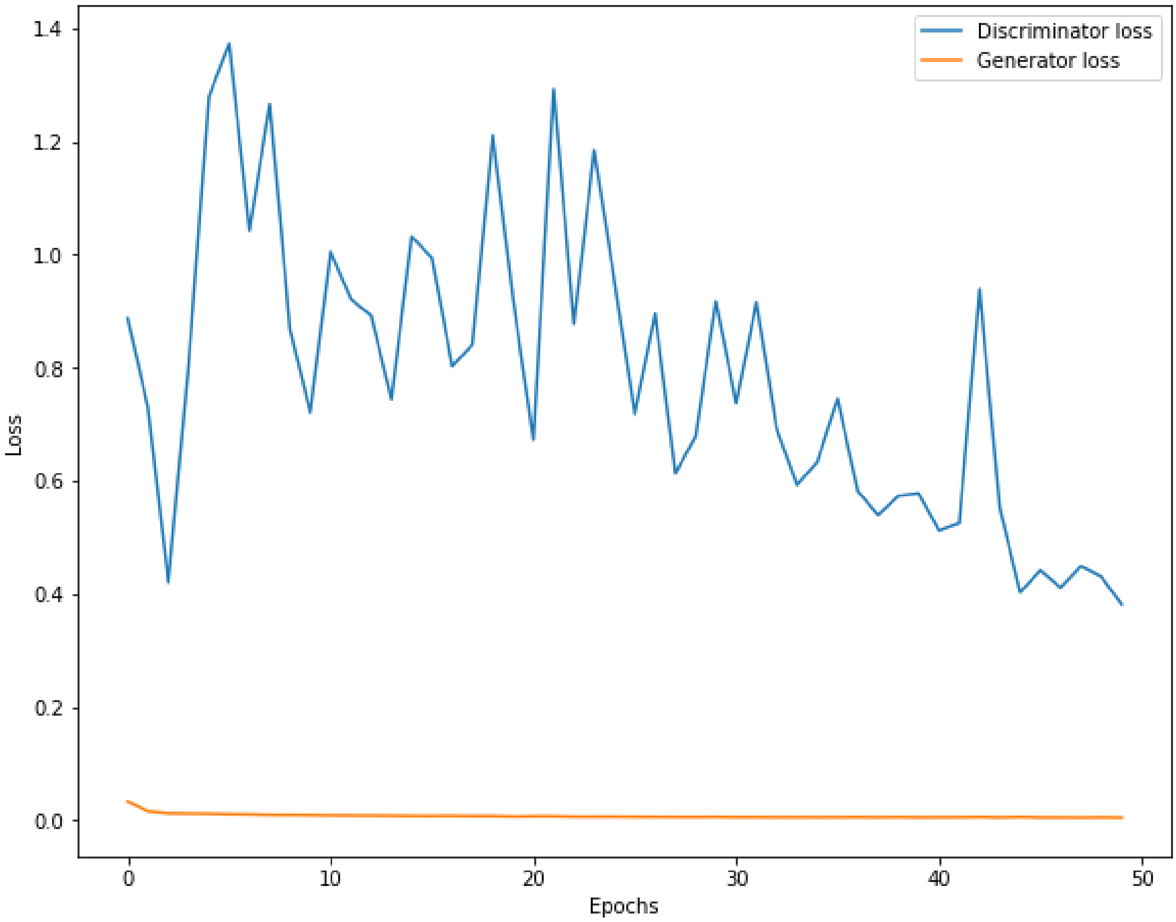


Figure 5. 1: Loss vs Epochs

Moving to the results, the colorized output from the generator was really closer to the ground truth of the rgb original image.

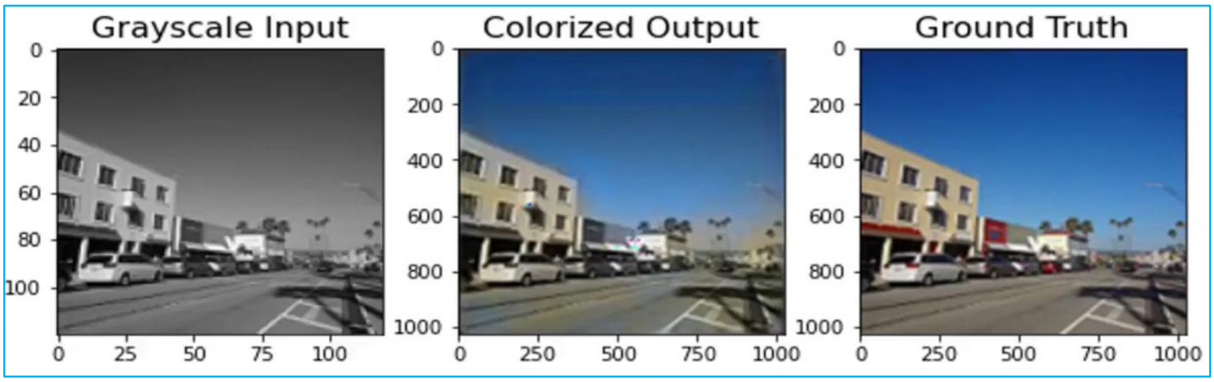


Figure 5. 2: Output 1

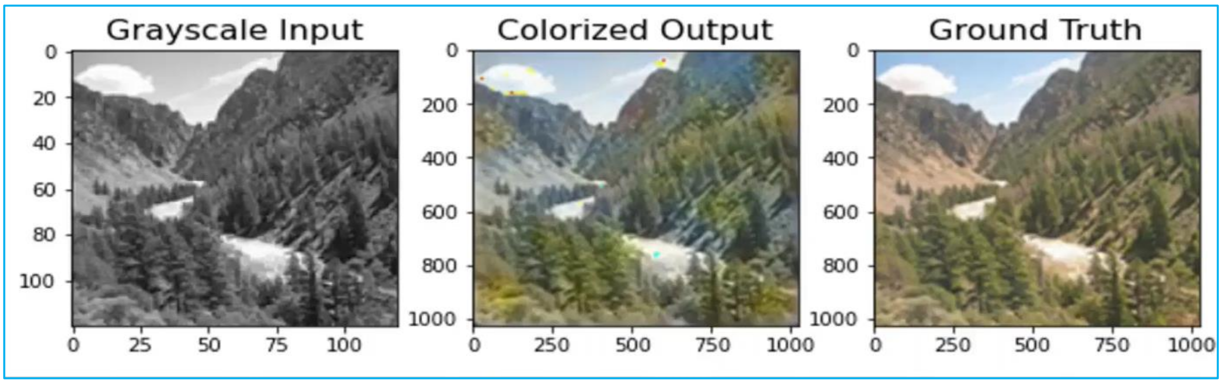


Figure 5. 3: Output 2

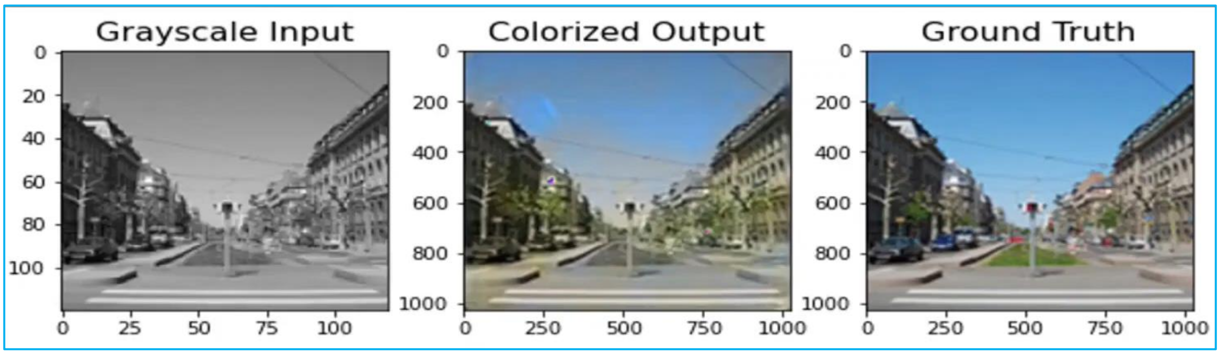


Figure 5. 4: Output 3

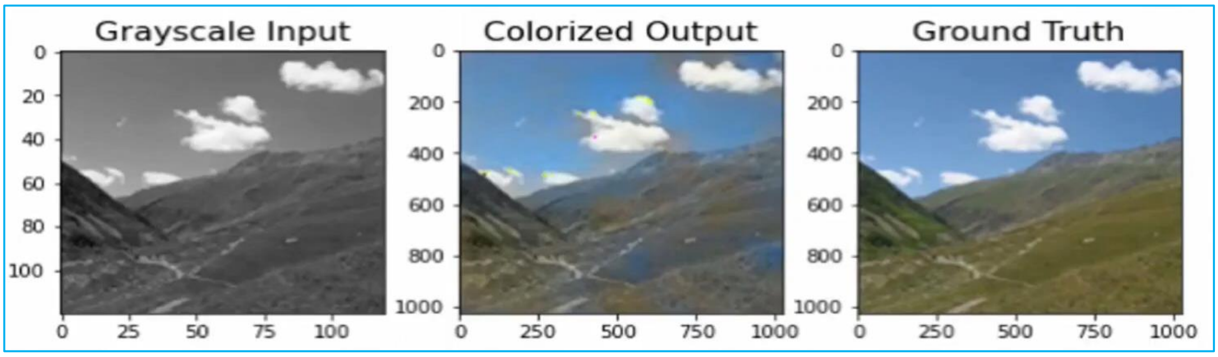


Figure 5. 5: Output 4

The four state-of-the-art strategies that we used in this gan model are also advised by researchers for implementing the deep learning model. First, we used soft and noisy labels. Second, Adam was employed as an optimizer. Third, we trained the discriminator twice as much as the generator, and fourth, we chose relu as an activation function, which would greatly aid us in avoiding sparse gradients due to its lu decomposition.

CHAPTER - 6

CONCLUSION AND FUTURE WORK

The overwhelming results depicts the power of GANs and the disruption which can be brought through them.

The colorized output produced by the generator was really closer to the ground truth of the rgb original image.

By using this project, we can convert old b/w photo of our grandparents into colorized photo making that picture more memorable.

We also aim to carry our work further by colorizing the videos too.

Lastly, we want our model to be deployed on web platform, so that it is accessible for users worldwide.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[2] Blanch, M. G., Mrak, M., Smeaton, A. F., & O'Connor, N. E. (2019, September). End-to-end conditional gan-based architectures for image colourisation. In 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR. 2016. arXiv preprint arXiv:1512.03385.

[4] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[5] Ronneberger, O. F. (2015, October). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention* (pp. pp. 234-241). Springer, Cham.

[6] Sarmai, A., & Sadhu, A. Image Colorization using Convolutional Autoencoders and Generative Adversarial Networks

[7] Zhang, R., Zhu, J. Y., Isola, P., Geng, X., Lin, A. S., Yu, T., & Efros, A. A. (2017). Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999*.

[8] Joshi, M. R., Nkenyereye, L., Joshi, G. P., Islam, S. M., Abdullah-Al-Wadud, M., & Shrestha, S. (2020). Auto-colorization of historical images using deep convolutional neural networks. *Mathematics*, *8*(12), 2258.

[9] Pandey A., Sahay R., Jayavarthini C. (2020). Automatic Image Colorization using Deep Learning. International Journal of Recent Technology and Engineering, Volume-8 Issue-6, 1592-1595.

[10] Baldassarre, F., Morín, D. G., & Rodés-Guirao, L. (2017). Deep koalarization: Image colorization using cnns and inception-resnet-v2. *arXiv preprint arXiv:1712.03400*.

[11] Kotala S., Tirumalasetti S., Nemitha V., Munigala S. (2019). Automatic Colorization of Black and White Images using Deep Learning. IJCSN Journal Volume 8, Issue 2, 125-131.

[12] Nguyen, T., Mori, K., & Thawonmas, R. (2016). Image colorization using a deep convolutional neural network. *arXiv preprint arXiv:1604.07904*.

[13] Charpiat, G., Bezrukov, I., Altun, Y., Hofmann, M., & Schölkopf, B. (2010). Machine learning methods for automatic image colorization. *Computational photography: methods and applications. CRC Press, Boca Raton*, 395-418.

[14] Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F. S., & Muzaffar, A. W. (2020). Image colorization: A survey and dataset. *arXiv preprint arXiv:2008.10774*.

[15] Koo, S. (2016). Automatic colorization with deep convolutional generative adversarial networks. *CS231n*.

[16] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

[17] Liang, Y., Lee, D., Li, Y., & Shin, B. S. (2022). Unpaired medical image colorization using generative adversarial network. Multimedia Tools and Applications, 81(19), 26669-26683.

[18] Wang, W., Chen, Z., & Yuan, X. (2022). Simple low-light image enhancement based on Weber–Fechner law in logarithmic space. Signal Processing: Image Communication, 106, 116742.

[19] Madhusudana, P. C., Birkbeck, N., Wang, Y., Adsumilli, B., & Bovik, A. C. (2022). Image quality assessment using contrastive learning. IEEE Transactions on Image Processing, 31, 4149-4161.

[20] Fernandez-Maloigne, C., Robert-Inacio, F., & Macaire, L. M. (2023). Digital color imaging. Wiley Online Library.

[21] Huang, S., Jin, X., Jiang, Q., & Liu, L. (2022). Deep learning for image colorization: Current and future prospects. Engineering Applications of Artificial Intelligence, 114, 105006.

[22] Chen, S. Y., Zhang, J. Q., Zhao, Y. Y., Rosin, P. L., Lai, Y. K., & Gao, L. (2022). A review of image and video colorization: From analogies to deep learning. Visual Informatics.

[23] Wang, N., Chen, G. D., & Tian, Y. (2022). Image Colorization Algorithm Based on Deep Learning. Symmetry, 14(11), 2295.

PUBLICATIONS

[1] Rahul & Devanshu (2023). Reviving Black and White Images: Enhancing Colorization with Generative Adversarial Networks (GANs). In 14th International Conference on Computing Communication and Networking Technologies – 2023 (ICCCNT). IEEE. [Scopus Indexed] [Accepted]

[2] Rahul & Devanshu (2023). A Study of Advancements in Deep Learning-Based Image Colorization Techniques. In 5th IEEE International Conference on Advances in Computing, Communication Control and Networking (ICAC3N–23). IEEE. [Scopus Indexed] [Accepted]

turnitin

Similarity Report ID: oid:27535:36340399

PAPER NAME

[THESIS] Reviving Black and White Images Enhancing Colorization with Generative Adversarial Networks

WORD COUNT

7445 Words

CHARACTER COUNT

41230 Characters

PAGE COUNT

41 Pages

FILE SIZE

1.1MB

SUBMISSION DATE

May 28, 2023 2:10 PM GMT+5:30

REPORT DATE

May 28, 2023 2:11 PM GMT+5:30

31/5/23

● 18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 11% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 14% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Small Matches (Less then 8 words)

Gmail

Devanshu Walecha <devanshu99walecha@gmail.com>

14th ICCCNT 2023 submission 2128

14th ICCCNT 2023 <14thicccnt2023@easychair.org> Thu, Jun 8, 2023 at 12:34 PM
To: Devanshu Walecha <devanshu99walecha@gmail.com>

Dear authors,

we like to inform you that your submission
to 14th ICCCNT 2023 entitled

Reviving Black and White Images: Enhancing Colorization with Generative Adversarial Networks (GANs) got accepted

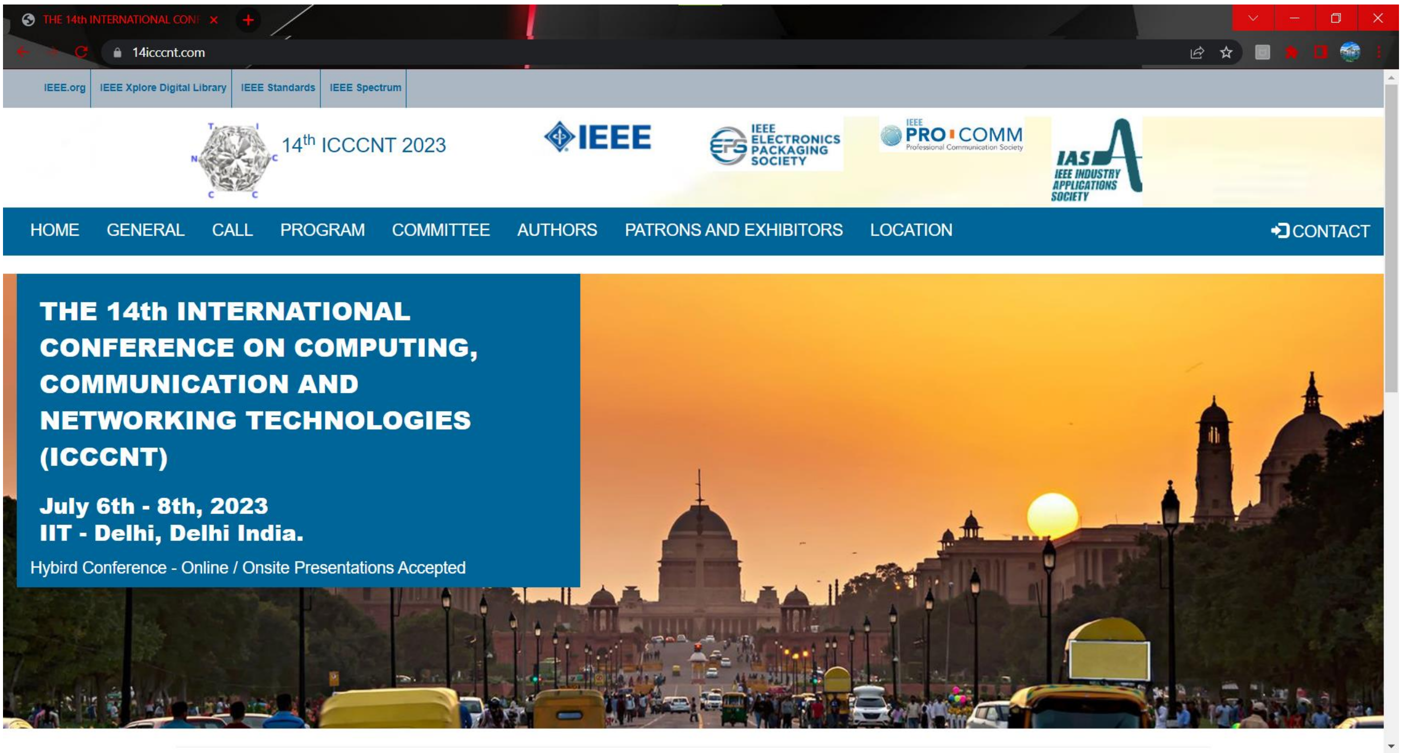
paper must be strictly in ieee conference template

paper looks poor in format

https://www.ieee.org/conferences/publishing/templates.html

at the end of section II, talk about what is missing in the past and what is new in this paper

add limitations of the paper in the conclusion section


THE 14th INTERNATIONAL CONF
14icccnt.com

IEEE.org | IEEE Xplore Digital Library | IEEE Standards | IEEE Spectrum

14th ICCCNT 2023

IEEE

IEEE ELECTRONICS PACKAGING SOCIETY

IEEE PRO COMM Professional Communication Society

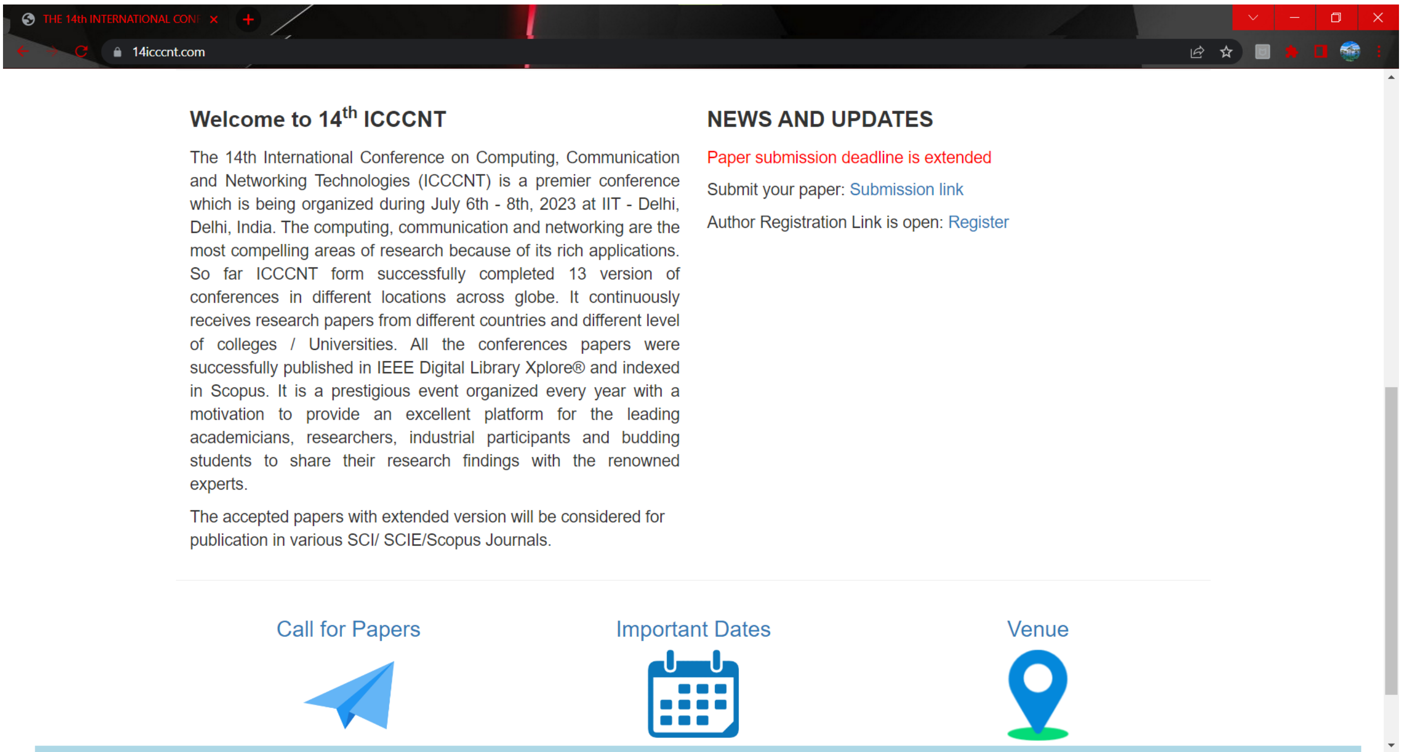
IAS IEEE INDUSTRY APPLICATIONS SOCIETY

HOME GENERAL CALL PROGRAM COMMITTEE AUTHORS PATRONS AND EXHIBITORS LOCATION CONTACT

THE 14th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)

July 6th - 8th, 2023
IIT - Delhi, Delhi India.

Hybird Conference - Online / Onsite Presentations Accepted


THE 14th INTERNATIONAL CONF
14icccnt.com

Welcome to 14th ICCCNT

The 14th International Conference on Computing, Communication and Networking Technologies (ICCCNT) is a premier conference which is being organized during July 6th - 8th, 2023 at IIT - Delhi, Delhi, India. The computing, communication and networking are the most compelling areas of research because of its rich applications. So far ICCCNT form successfully completed 13 version of conferences in different locations across globe. It continuously receives research papers from different countries and different level of colleges / Universities. All the conferences papers were successfully published in IEEE Digital Library Xplore® and indexed in Scopus. It is a prestigious event organized every year with a motivation to provide an excellent platform for the leading academicians, researchers, industrial participants and budding students to share their research findings with the renowned experts.

The accepted papers with extended version will be considered for publication in various SCI/ SCIE/Scopus Journals.

NEWS AND UPDATES

Paper submission deadline is extended

Submit your paper: Submission link

Author Registration Link is open: Register

Call for Papers | Important Dates | Venue

THE 14th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)

July 6th - 8th, 2023

Venue: IIT Delhi, Delhi, India

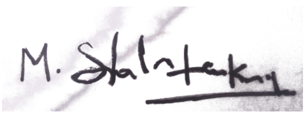
Payment Receipt

Date: 2023-06-13

Receipt No.	333		
Author Name.	Devanshu Walecha	**Author Address.**	C-19/2, Guru Ram Das Nagar, Laxmi Nagar, Delhi - 110092
Paper ID	2128		
Payment Reference ID	316328077014		
Payment Type	upi		
Description	Conference Registration Fee		
		Amount Paid	**Rs. 7900.00**

Thank you for the payment.

Print



(Authorized Signatory)

Gmail **Devanshu Walecha <devanshu99walecha@gmail.com>**

Acceptance Notification 5th IEEE ICAC3N-23 & Registration: Paper ID 1286

Microsoft CMT <email@msr-cmt.org> Wed, Jun 7, 2023 at 12:34 AM
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Devanshu Walecha <devanshu99walecha@gmail.com>

Dear Devanshu Walecha,
Delhi Technological University

Greetings from ICAC3N-23 ...!!!

Congratulations....!!!!!

On behalf of the 5th ICAC3N-23 Program Committee, we are delighted to inform you that the submission of "Paper ID- 1286 " titled " A Study of Advancements in Deep Learning-Based Image Colorization Techniques " has been accepted for presentation and further publication with IEEE at the ICAC3N- 23 subject to incorporate the reviewers and editors comments in your final paper. All accepted papers will be submitted to IEEE for inclusion into conference proceedings to be published on IEEE Xplore Digital Library.

For early registration benefit please complete your registration by clicking on the following Link: https://forms.gle/7VkMHWYFsQ7XREgc6 on or before 13 June 2023.

Registration fee details are available @ https://icac3n.in/register.
https://drive.google.com/file/d/1RGu6i4eGUI5B07zOfgRmDRfyjOhOhiC6/view?usp=share_link

You can also pay the registration fee by the UPI. (UPI id - icac3n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1Ry-sF0apvy_0zUjM8lNW02lZgWAsE0YD/view?usp=sharing

You must incorporate following comments in your final paper submitted at the time of registration for consideration of publication with IEEE:

Reviewers Comments:
The theme and title of the article "A Study of Advancements in Deep Learning-Based Image Colorization Techniques" is suitable and appropriate for publication.
Paper is well written and technically sound.
Overall flow of paper is very good.
Result are well explained
English language is satisfactory.

Editor Note:
1. All figures and equations in the paper must be clear. Equation and tables must be typed and should not be images.
2. Final camera ready copy must be strictly in IEEE format available on conference website www.icac3n.in.
3. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However the author will be given fair and sufficient chance to reduce plagiarism.
5. Change in paper title, name of authors or affiliation of authors will not be allowed after registration of papers.
6. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
7. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment, you can Contact @ 8168268768 (Call) / 9467482983 (Whatsapp/UPI) or write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N - 2023

Gmail

Devanshu Walecha <devanshu99walecha@gmail.com>

Confirmation Request: Scopus Indexing of ICAC3N - 2023

ICAC3N-2023 Dr. Vishnu Sharma <icac3n23@gmail.com> Sat, Jun 10, 2023 at 10:38 AM
To: Devanshu Walecha <devanshu99walecha@gmail.com>

Dear Author,

It is to confirm to you that proceedings of the conference 5th ICAC3N-23 will be published by IEEE and all papers will be available online on IEEE Xplore Digital Library and will be scopus indexed. The IEEE Conference Record Number of Conference is #60023, also available on the website. Further, all IEEE conferences are now scopus indexed.

We are also delighted to inform you that all first four conferences of this series are scopus indexed and also available online on IEEE Xplore Digital Library.

The proceedings of 4th IEEE ICAC3N-22 is available on **IEEE Xplore (Scopus Indexed)**

https://ieeexplore.ieee.org/xpl/conhome/10073967/proceeding

The proceedings of 3rd IEEE ICAC3N-21 is available on **IEEE Xplore (Scopus Indexed)**

https://ieeexplore.ieee.org/xpl/conhome/9725360/proceeding

The proceedings of 2nd IEEE ICAC3N-20 is available on **IEEE Xplore (Scopus Indexed)**

https://ieeexplore.ieee.org/xpl/conhome/9362509/proceeding

The proceeding of 1st IEEE ICAC3N-18 is available on **IEEE Xplore (Scopus Indexed)**

https://ieeexplore.ieee.org/xpl/conhome/8742891/proceeding.

Regards

Team ICAC3N-23

[Quoted text hidden]

ICAC3N-21: 3rd IEEE International Conference on Advances in Computing, Communication Control and Networking

Galgotias College of Engineering & Technology (GCET),Greater Noida U.P. (India) -201306

Greater Noida, India, December 17-18, 2021

Conference website	https://icac3n-21.in/
Submission link	https://easychair.org/conferences/?conf=icac3n21

Topics: evolutionary computing big data machine learning networking

About ICAC3N-21:

3rd IEEE International Conference on Advances in Computing, Communication Control and Networking (ICAC3N-21) will be held during **December 17-18, 2021** in **Galgotias College of Engineering and Technology, Greater Noida, India**. The conference is an international forum which aims to bring together leading academician, researchers and research scholars to exchange and share their experiences and hard-earned technological advancements about all aspects of based on their research related to Computing, Communication Control & Networking. We invite all leading researchers, engineers and scientists in the domain of interest from around the world. We warmly welcome all authors to submit your research papers to ICAC3N-21, and share the valuable experiences with the scientist and scholars around the world.

IEEE Conference Record No. #53548

Publication and Indexing

ICAC3N-21 is indexed in SCOPUS and Google Scholar. All registered & presented papers are available on IEEEXplore DigitalLibrary.

Gmail

Devanshu Walecha <devanshu99walecha@gmail.com>

Registration Confirmation, 5th IEEE ICAC3N-23, Paper ID 1286

Microsoft CMT <email@msr-cmt.org> Mon, Jun 12, 2023 at 10:16 PM
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Devanshu Walecha <devanshu99walecha@gmail.com>

Dear Devanshu Walecha,
Delhi Technological University

Greetings from ICAC3N-23 ...!!! Thanks for Completing your registration...!!

Paper ID- "1286 “
Paper Title- “ A Study of Advancements in Deep Learning-Based Image Colorization Techniques “

This email is to confirm that you have successfully completed your registration for your accepted paper at ICAC3N-2023. We have received your registration and payment details. Further, your submitted documents will be checked minutely and if any action will be required at your end you will be informed separately via email.

For further updated regarding conference please keep visiting conference website www.icac3n.in or for any query please write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N – 2023

Note:
1. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE. (For this you will be informed separately via email well before conference)
2. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However, author will be given sufficient and fair chance to reduce the plagiarism.
3. Change in paper title, name of authors or affiliation of authors is not allowed now.
4. Violation of any of the above point may lead to cancellation of registration.
5. Registration fee once paid is non-refundable.