https://doi.org/159: 10.51153/kjcis.v6i1.159

Evaluation of Image Super Resolution Deep Learning Technique based on PSNR Value

Muhammad Azam[1], Muhammad Nouman [2]

Abstract

Machine learning is a subset of deep learning. Deep learning techniques have become increasingly popular in recent years for a variety of applications. Many techniques are used in image processing for image enhancement, object detection, and a variety of other purposes. Many deep learning techniques are important in single-image super-resolution (SISR). Many models are used for SISR, and they produce good perceptual and adversarial results. The procedure involves creating a high-resolution (HR) image from a low-resolution (LR) image, including some popular SISR methods: "Super-Resolution with Convolutional Neural Network" (SRCNN), "Super-Resolution Residual Network" (ResNet), "Super-Resolution Generative Adversarial Network" (SRGAN), and "Enhanced Super-Resolution Generative Adversarial Network" (ESRGAN). These deep learning models are layered, with each layer performing a different task for an LR image. These methods convert a single LR image into an HR image. A single LR image yields many high-resolution images. This paper creates a dataset of DSLR images for use with the SISR method. Deep learning technology is used to convert an LR image to an HR image. The HR images are then compared to the DSLR images. The "peak signal-to-noise ratio" (PSNR) and the "Structural Similarity Index" (SSI) are used to compare them (SSIM). When compared to the old method, the proposed method produces identical results. SSIM and PSNR are shown for both old and proposed approaches. Both approaches possess almost comparable results. The higher the PSNR value, the better the image quality. The same is the case with the SSIM value

Keywords: PNSR Value, SISR, SRCNN, ResNet, SRGAN, ESRGAN, SSI

1. Introduction

Image super-resolution refers to the sensor's ability to determine the nominal object based on pixel size. High-resolution digital images are always appealing in many applications as two-dimensional signal records. Many imaging techniques have advanced rapidly in recent decades, reaching new resolution levels. The first work on this topic was published in 1984 (Tsai, 1984), and the term "super-resolution" appeared almost immediately after[1] The reality is that, while high-definition displays have advanced in recent years, the need for resolution enhancement in many applications cannot be overlooked[2] . Digital observation products, for example, tend to provide resolution to ensure a suitable frame rate for dynamic scenes. An analogous situation

[1]*Department of Electrical Engineering and Computer Science University of Missouri, Columbia 1 writetoazamkhalid@gmail.com*
[2]*Department of Computer Science University of Agriculture Faisalabad, Pakistan*

exists in the field of remote sensing. There is a trade-off between temporal, spectral, and super resolutions. In medical imaging, 3D models of human anatomy are lowering the level of radiation with high resolution, which is still a challenge[3] ,[4]

1.1 Resolution Enhancement

Based on these facts, demands of high resolutions cannot be satisfied. Enhancement of resolution is still essential, especially in fields like medical diagnosis or remote sensing applications, etc. Enhancement through "hardware" requires a prohibitive cost and has some limitations. Signal processing methods (super-resolution (SR)) become a probable way to get HR images[5]. SR is a process in which a combination of multiple LR images is used to get an HR one. SR is separated into two types:

- Single frame
- Multi-frame

HR images are obtained in a multi-frame by combining multiple LR images. HR images are obtained in a single frame by using a single LR image, which is referred to as a single image super-resolution SISR [6]. Many techniques and algorithms are used for this purpose, but there is still a gap in the texture details in the recreated SR image. The main optimization goal of supervised SR algorithms is typically to minimize the MSE between the ground truth and the improved HR image. In addition to maximizing the MSE, PSNR (a common metric used to compare SR algorithms) also maximizes [6]. However, because PNSR and MSE are based on pixel-by-pixel image differences, their ability to capture perceptually relevant differences is limited [2]

1.2 Basic Structure

A GAN is made up of two networks: a generator and a discriminator. A discriminative algorithm is used to classify input data because, given the characteristics of the data, a discriminator is used for labeling or categorizing the data. It represents the label's features on a map[7]. Generative algorithms, on the other hand, do the opposite. Rather than labeling the prediction with specific features, it attempts to predict the features given a specific label in figure 1. As a result, it can be classified as follows:

- **Discriminative models** learn the boundary between classes.
- **Generative models** are used to simulate the distribution of individual classes.

The neural network generator generates new data instances while the discriminator checks them for authenticity [8]

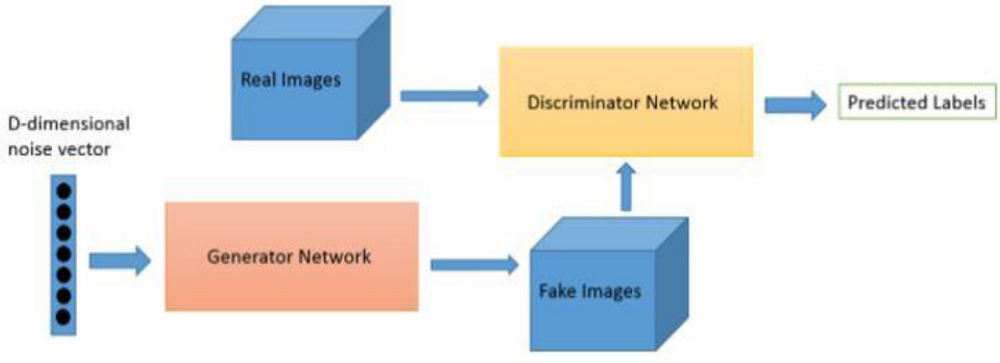

Figure 1: Generative Adversarial Network (GAN)

Both networks optimize a different but opposing goal of loss or objective function. The discriminator and generator's behavior change depending on the running condition. Their loss functions are bumping up against each other [9].

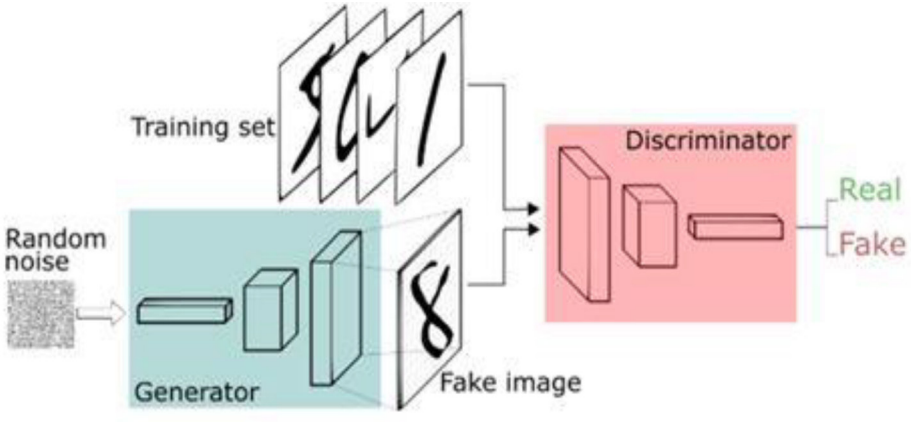

Figure 2: GAN Optimization

1.3 SISR and GAN

SR can obtain a higher-resolution image from a lower-resolution image. The output result of the CNN processing architecture instruction that stepped forward the excellent resolution feats used with ground truth value is over-smoothed without high-frequency information[10]. The Super-resolution Generative Adversarial Network (SRGAN), which uses the previously mentioned perceptual loss technique, improves the visible exceptionality of superb decision results [11].

CNN can be advanced by the introduction of RDDB, which has greater potential and is easier to implement. A discriminator is the second relativistic GAN, and a generator is the third relativistic figure 3 GAN that determines whether a picture is real or fake. GAN's image super-resolution is amazingly simple to configure. A proposed version of the super-resolution convolutional neural network with several layers is compared to the authentic version in terms of accuracy. The peak ratio is calculated by using actual photos [12]. Perceptual loss functions can quantify the first class of visible functions, texture distinction, and image cost with the unique one. In the current research paper, the easy photo outstanding decision has an overly exciting study subject matter. GANs are designed to regenerate a network that generates multiple samples that are distributed and transformed into the vector's noise as well as discriminate against the network[4]. GAN observes conditional generators that give a residual illustration and perceptual discriminators that give a useful residual illustration within the small item detection. A single image with excellent resolution to recover a high-resolution image from a single low-resolution (LR) [13].

1.4 SISR and CNN

Convolutional neural networks (CNN) resemble standard brain networks in that they are made from neurons with learnable loads and predispositions. Every neuron's feedback is recognized, and afterward, a speck item is performed, trailed by non-linearity [14]. With the whole organization, a solitary differentiable score capability is expressed, with some crude picture pixels toward one side and class scores on the other. The loss function is still present on the final layer Convolutional neural networks have increased the number of observers in computer vision tasks by more than a year. There are numerous ones, such as image classification, item detection, semantic segmentation, and visual question answering, among others[15]. The distorted images are also referred to as the adversarial network, which exhibits one-of-a-kind architecture. The first class is the model-specific mechanism, which can assign a specific version of parameters to various edification or test parameters for smoothing functions[16]. Such strategies frequently produce distinct differences that are associated with computationally troubling issues. The second class can define defenses that are version independent. They show the effect of adversarial network dating in the input picture realm collectively by applying multiple adjustments[17]. Super-resolution, in which the technique upscales snapshots while avoiding artifacts or blurring, produces SR images from one or more low-resolution images[18]. Numerous remarkable decision-making strategies can be divided into two lessons and may be based on multiple supply images at sub-images combined with a single upscaled photo. With a single input image, SISR algorithms that produce an excessive decision of the output photograph are provided[19]. Images are taken in the form of input for CNN and some architecture is constrained in a better manner. Neurons are arranged in the layers of CNN in the 3D form: depth, height, and width. Neurons in the layer relate to the smaller region of layers instead of all layer connections. So, the output image changes its dimensions because the full or original image changes its dimensions into some single vector along its depth dimension[20].

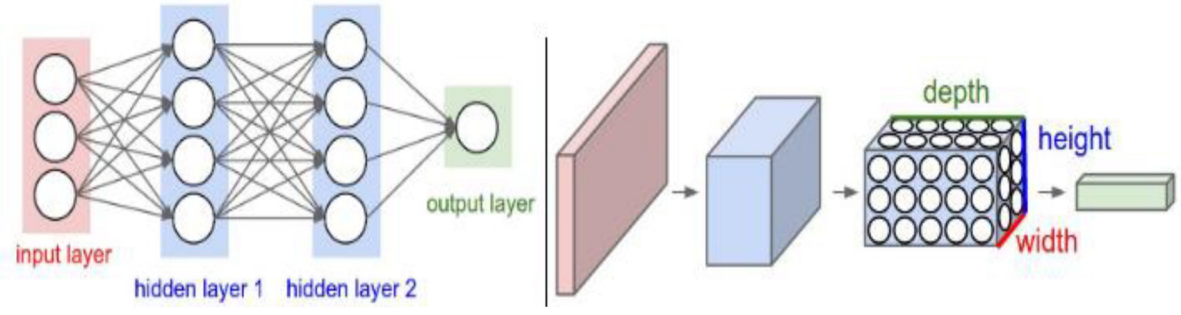

Figure 3: Convolutional Neural Network

1.5 SISR and ResNet

The well-known ResNet issue is the "Vanishing Gradient." This is since when the depth of the community changes, the gradient function used to calculate the loss function also changes to 0 after chain rule implementation[21]. As a result, weight values change, and learning processes slow down. However, with the help of ResNet, backward connections are formed from later layers to the initial layers via gradient skip connections. Allow the ResNet to change size based on the massiveness of the layers and the number of layers. It complies by utilizing the authors' effective way of explaining the structure of these networks[22]. There are few figures and tables available, which makes it difficult to observe. What goes into the element of each step can be easily depicted by figures. In this case, the ResNet consists of a single convolution and pooling step that is tracked by four layers of similar behavior. Each layer follows the same pattern. A 33 convolution is performed with characteristic map sizes F of 64, 128, 256, and 512, and the input is passed after every two convolutions. The width and peak dimension, on the other hand, remain constant throughout the layer [23].

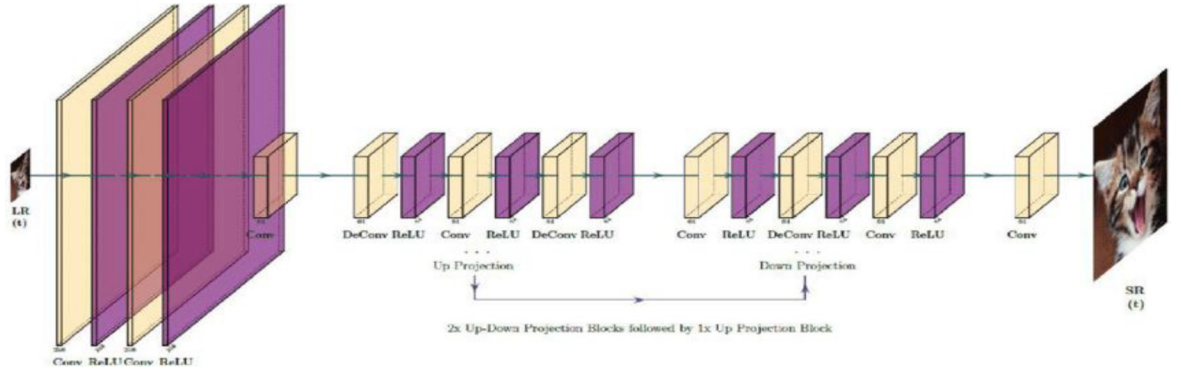

Figure 4: SISR and ResNet Relation

1.6 Deep Learning Techniques

Deep learning strategies rely heavily on super-resolution research. By utilizing several high-goal handling moves toward converting low-goal pictures to high goals, the first striking food pictures are re-established. The results of the test with the difference between a low-resolution image panel and a high-resolution image produced a new high-resolution published image[24]. The binarization procedure is a critical step for OCR algorithms that results in incorrectly finished words and character popularity. The organizers prepared the ground truth photos using a semi-automated technique. The paper that investigates image variability is a fact: through humans, it is mechanically predicted by a binarization set of rules[25]. Customers can extract hidden objects from photographs by removing noisy reflections from their photos and by viewing high-quality photographs. Eliminating reflection from a noisy single image to achieve better visible satisfaction with correct perceptibility More than one image is available to get better strategies for a person's guidance is inconvenient and prone to error connection. Currently, research is being conducted that proposes strategies for removing reflections from a single photograph [26]. The reflection separation, which is a color darkness model that tries to estimate objects in the picture, is the low degree of outstanding single photo decisions. The use of a convolutional neural network with an appreciation for perceptual losses to discover two low- and high-level photograph records is discussed in this paper. However, existing techniques based on maximal and minimal images and image differences can produce accurate overall performance, as evidenced by experimental results [27]. This research can describe a distinct field of resolution enhancement for massive decisions. The conclusion that unique pictures with small portions normally appear repeatedly on an authentic scale that probes for tremendous decisions is known as self-similarity. The best resolution algorithms improve, in most cases, the super-resolution of noisy photos [28]. After the first-rate decision, the alternative side of the denoising post-processing technique appears. The noise for input LR photos is powerful due to the super-resolution method, and its miles are sufficiently robust to deal with this additional noise [29].

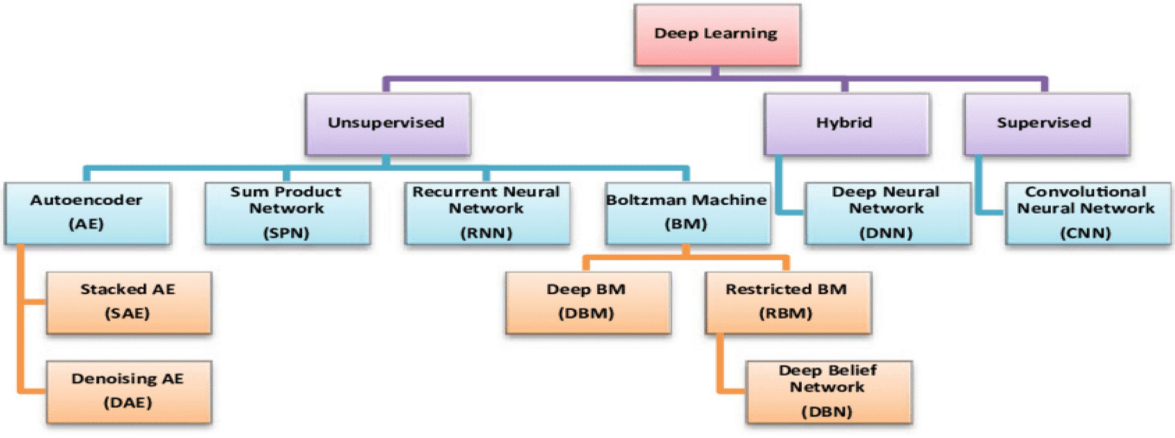

Figure 5: Deep Learning Technique

This work is motivated by the previous work on the ESRGAN model. To produce super-resolved images, Generative Adversarial Network is one of the famous and efficient models. This model is also used for object detection from images, for the generation of an image from many input images, and many more purposes[30]. Here the main goal is to produce a super resolve image from the LR image. Many other techniques are used for super-resolution but in this work, Enhanced GAN is used for super-resolution due to its easy training and low complexity level[31].

1.7 Research Questions

Some research questions are presented here to evaluate the proposed approach and methodology:

- To what extent is the proposed approach superior to the previous approach?
- What is the connection between VGG-19 and ResNet-50?
- How does the number of layers affect the outcome?

To answer the above research questions, ResNet-50 and VGG-19 are used, BN layers are removed, and the network is enhanced. The removal of BN layers has been shown to improve performance and reduce the computational complexity in many PSNR-oriented tasks, including SR and deblurring.

The rest of this paper is as follows: Section 2 stands for related work. The section provides a detailed description of the problem definition and specification, as well as an overview of the proposed approach is explained in section 3. The results of the experiments are discussed in section 4. In the end, section 5 concludes the paper and presents some future directions.

2. Literature Review

Palatino Linotype the performance and minimize the computational complexity for many PNSR-oriented activities with deblurring and SR. Features are normalized with the help of the estimated mean and variance in training. Removing BN-Layers improve the ability of generalization and memory usage. Architecture design is also improved by using novel basic block RRDB. Tan et. al.[32] Proposed innovative approaches for the improvement of GAN for conditional synthesis images. In Art GAN, labels are used so that the discriminator can easily understand and generate a relevant image. It uses the concept of “how humans learn to draw” Art GAN chooses Random label info and noise vector. It gets feedback from the discriminator which discriminates the ring based on the info given in the form of a label for processing.

Wang et. al. [33] worked on image super-resolution in an end-to-end manner. The old methods cannot manage the end-to-end learning image from LR to HR in a single model to overcome these issues, a novel idea is generated to ensemble into the SR problem and enhance the performance of basic learning-based SR methods by the suggested general methods. Compared with other

single networks, collections of neural networks are known for the generation of more accurate predictions. This strategy is widely used in image classicization tasks. For the effectiveness of the ensemble strategy, the author forced on the state-of-the-art SR method SCN which helps to propose "Ensemble Based Sparse Coding" (ESCN). Lu et. al (2017) contributed to the field of SISR. The learning methods with sparse coding assume that LR and HR features have the same representation coefficients in their directories. This limits the model's flexibility. To solve this issue, a Sparse Domain Selection (SDS) method is proposed for SISR. In SDS the features of LR and HR have their specific corresponding coefficients with mapping relationships. This gives better reconstruction quality. As the sparse property is an important visual perceptual characteristic. Features of LR-HR are extracted in training pairs.

Dong et. al. [34] proposed a model that deals with start-to-finish planning advancing among LR and HR pictures. This work attempts to show that the meagre coding pipeline is like CNN. This work is the inspiration by scanty coding which straightforwardly gains from start-to-finish planning among LR, and HR sees. This proposed model is of the learning process it does not directly learn from the dictionaries or manifolds for patch space modelling. Melchers et. al [35] proposed an approach that can be applied with or without GAN. It is hard to maintain natural image statistics in image generation and restoration for its natural look. GAN attempts photo realism usually which makes the image natural and put it in the manifolds of natural images. GANs are hard to train and sometimes results are not satisfactory and suffer from artifacts. The author works on this issue and trains a feed-forward CNN which maintains the natural statistics of the image. Feature distribution is focused, and the network is trained that produce images with natural distribution features. Al Shabbily [35] improved the resolution of satellite images with the help of a new algorithm. The goal of improvement is to get an HR image from LR. This algorithm uses interpolation with the filter of two dimensions which is designed specifically for any image with the help of a maximum (PSNR). Laharas, Batavia, and Schindler [36] introduce a method to enhance hyperspectral data resolution. For this purpose, images of multispectral higher resolution are used. Sensors are needed for combining two data sources of information. A model is derived for the estimation of spectral response and relative spatial of two sensors. This formulation, which is proposed, includes the recovery of non-negative remaining registration (shift) errors. Greaves and Winter ([37] described a multi-frame super-resolution technique for videos in which multiple frames are used for improving the quality of the frame.

Chen et al. [38] gave lofty goals of attractive reverberation pictures (MRI) as the clinical application and quantitative picture examination. High-Resolution MRI customarily goes the expense of long output time, littler spatial inclusion, and lower motion-to-commotion proportion (SNR). [39]Later that reviews which single picture super-goals (SISR) is a strategy to recuperate HR subtleties from one single low-goals (LR) input picture are furnished top-notch picture subtleties with the assistance of cutting edges profound learning investigation of convolutional neural systems (CNN). The proposed technique in this paper can talk about the 3D neural system structure of staggered gradually associated super-goals coordinated with the generative ill-disposed system

(GAN) of guided preparing.

3. Materials and Methods

This section provides a detailed description of the problem definition and specification, as well as an overview of the proposed approach, including implementation tools, system requirements, and a framework. The architectures of datasets and networks are also explained. This Section describes in detail the proposed approach's design, scheme, and implementation details. This paper describes in detail all the main steps discussed in section 3.2, namely, pre-processing of the dataset and obtaining the HR image from LR.

3.1 Problem Specification

Many images require enhancement. As a result, super-resolution cannot be ignored. For example, digital observation products typically provide resolution to ensure a suitable frame rate for dynamic scenes. There are some solutions in the field of remote sensing. There is a trade-off between spectral, temporal, and super-resolution. In medical imaging, 3D models of human structures with low radiation levels and high resolution are still a challenge [40].

3.2 System Overview

In this paper, a fully automatic approach to image SR is proposed, based on a Generative Adversarial Network (GAN). The technique is like Enhanced Super Resolution Generative Adversarial Network ESRGAN, but for enhancement purposes, VGG is replaced by the Reset model[32]. The reset is explained further below. This paper is divided into four major phases Figure 3 :

- Pre-processing: Dataset is pre-processed for further processing.
- Generation: Input image is used for the generation of HR image with Enhanced GAN.
- Discrimination: Relativistic Discriminator is used for discrimination of output results.
- Removal of noise: Network Interpolation has used the removal of unpleasant noise Figure 6 shows the overview of the proposed approach:

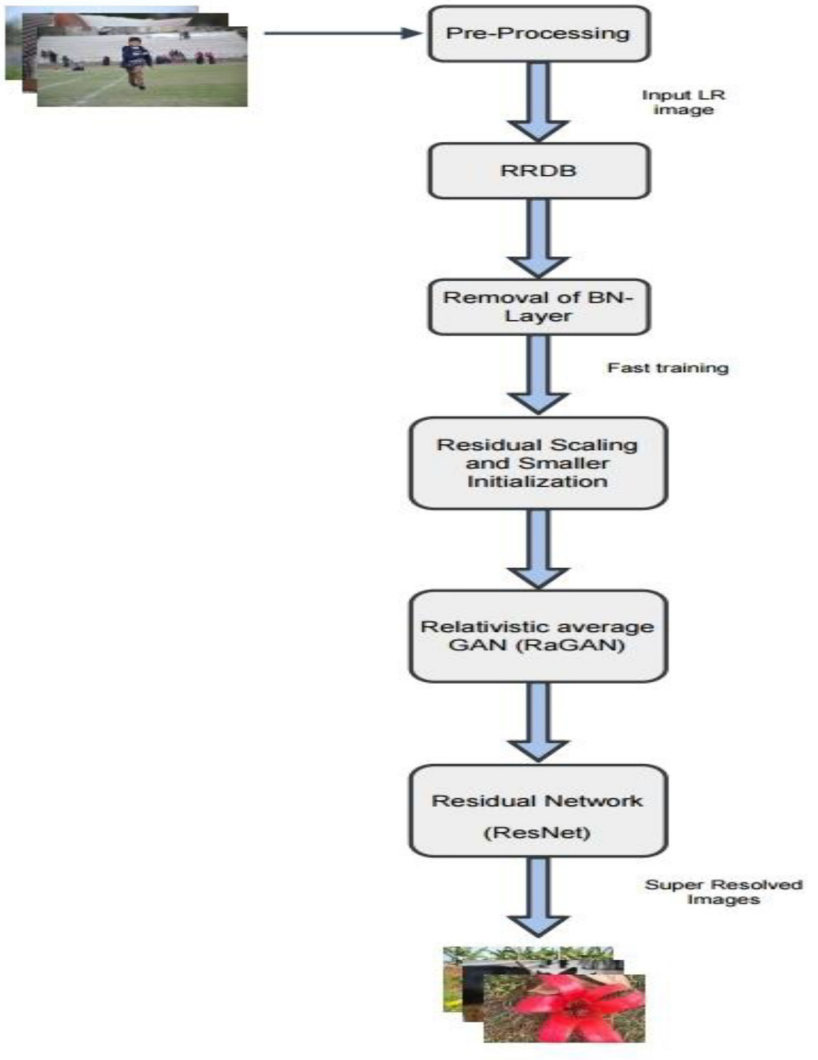

Figure 6: Framework of the Proposed Approach

3.3 Proposed Approach

The proposed method was developed and evaluated on an i7-7500U @ 2.70GHz & 2.90GHz with 8GB of RAM and a 64-bit operating system. The code is written entirely in Python 3.6. Image preprocessing is accomplished by reducing the image size to 230153 and 230306 pixels. Python is used to train and evaluate all the major components, such as generation and discrimination. PYTHON is used for the ResNet-50, which is a skip connection that propagates information across layers.

3.4 Network Architecture:

In deep learning, more complex and deeper networks have been introduced such as ResNet, AlexNet (He et al., 2016), and VGGNet[41]. These networks are based on networks for computer vision applications. VGGNet has different convolutional stages with the pooling layers, which decrease the resolution of feature maps and boost the receptive fields of neurons (Simonian and Zisserman 2014)[42]. Deeper layers learn further intellectual perceptions compared to the local features learned during the primary convolutional stages in the contracting path. In ResNet architecture, there are shortcuts at each layer, which result in a better stream of gradients throughout the network, resolving the classical problems, such as the gradient vanishing problem for DNN[43].

SISR is the fundamental low-level vision problem. SISR aims to recover LR images and convert them to HR. Many network architectures are destroyed and many training strategies are improved for enhancing the SR performance, especially the peak signal-to-noise ratio (PSNR). However, many PSNR-oriented approaches give over smooth results without having sufficient high-frequency details. Many perceptual-driven methods are proposed for the improvement of the visual quality of SR results. For the optimization of the SR model in feature space instead of pixel space, a perceptual loss function is proposed. The GAN model introduced the SR solution that looks more likely to natural images. The semantic images are then used to recover texture details. For this purpose, SR-GAN is proposed to produce usually improved results [44].

The residual blocks in the basic model optimize the image using the perceptual loss function. But there still exists a gap between the result of SRGAN and ground-truth (GT) images. This model is improved in three aspects[45].

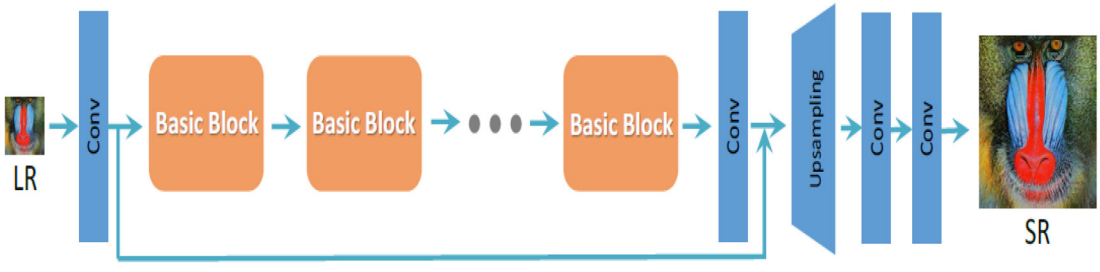

Figure 7: Network Architecture

Firstly, the network structure is improved by introducing Residual-in-Residual Dense Block (RRDB) having high capacity and is easy to train. Batch Normalization (BN) layers are removed, residual scaling is used, performance is increased by improving the BN layers and computational complexity is reduced in many PSNR-oriented tasks including deblurring and SR[46]. The layers of BN utilize mean and fluctuation to standardize the component in clump preparing and the assessed change and mean of that preparing dataset are utilized during testing. At the point when

there exists a contrast between measurements of preparing and testing datasets, horrendous curios and restricted speculation capacity are presented. These antiques are delivered when the organization is prepared further under the GAN structure. So, BN layers are removed from the consistent and stable performance[47].

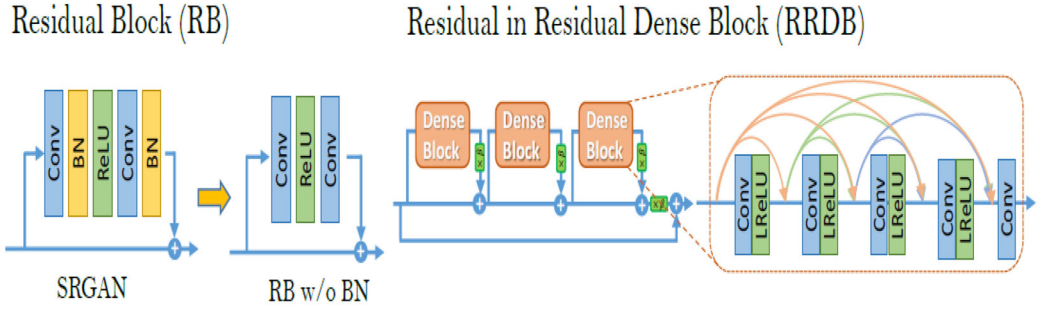

Figure 8: Removal of BN-Layers

The large-stage engineering plan of SEGAN is saved and an original fundamental block of RRDB is applied. A few perceptions are made that the growth of additional layers and institutions can constantly guide execution. The RRDB is proposed with an extra complex and more profound construction. This RRDB has a leftover in-lingering shape wherein final gaining knowledge is applied for one-of-a-kind tiers. A few specific processes are moreover used to work with the guidance interaction like the last scaling which will increase the constant someplace within the variety of 0 and 1 to downsize the lingering before including them to the manner which is the important manner for imperceptibility counteraction. A more modest introduction is additionally used to assist the instruction cycle which unearths the last engineering simpler to put together when the change for the underlying boundary decreases [48].

3.5 Dataset Collection

For this project, the dataset is collected by using a DSLR camera. Images of various locations, persons, and flowers are captured for this purpose. This dataset contains two hundred pictures captured from NIKON D3100, SONY DSC-H9, and SEMC X10i. Images are captured randomly to check the goodness of the project. Pre-processing is applied to capture images to reduce the size of images and to make them LR images for further processing.

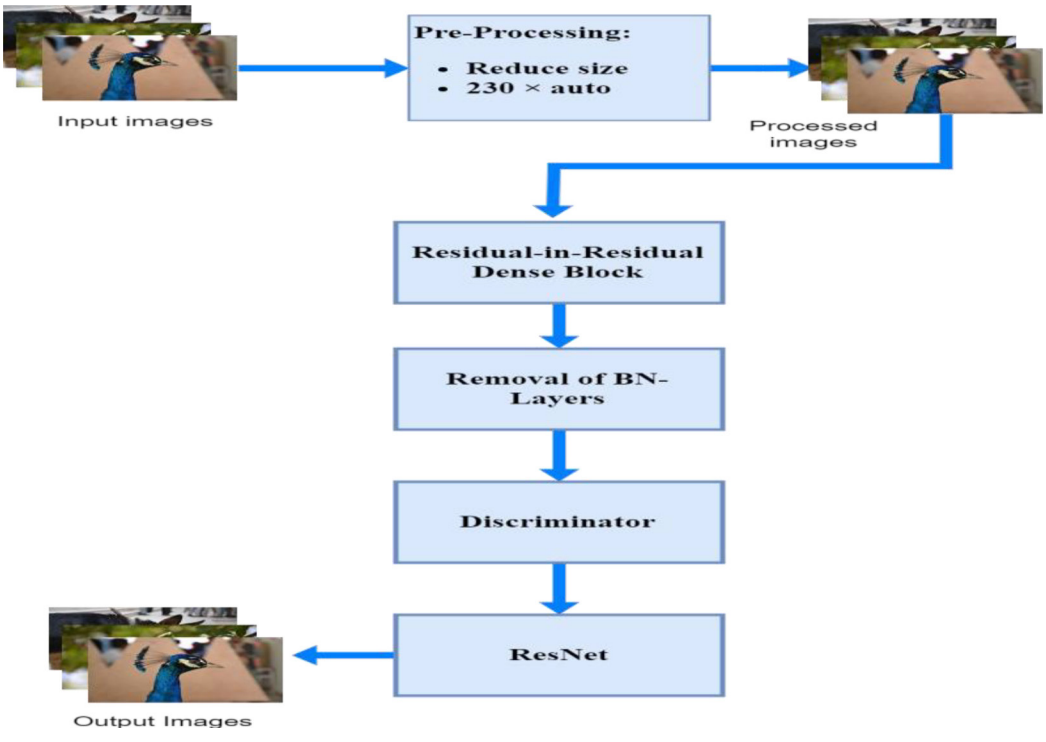

Figure 9: Proposed approach (Flow Diagram of Work)

3.6 Algorithm

The algorithm shows the structure of the proposed approach. In the first step, pre-processing is applied to reduce the size of 230 autos. In the second step, a relativistic discriminator is used to discriminate the results and to make them more enhanced. Both generator and discriminator are used for the generation of the super-resolved image by the input of the LR image and discrimination is then applied to discriminate the results in terms of realness and fakeness. Next, the perceptual loss function is defined. More detail is discussed in the following section.

#Algorithm: Super Image Resolution

#Input: LR-Images

#Output: HR-Images

function pre-processing (*images)*
    for *x* in all images **do**
        set pixel ← x, y
        x ← horizontal 230.
        y ← vertical auto.
    end for
    return G(x_i)
end function

function relativistic discriminator (r_{eal}, f_{ake})
    for L_D^{RaD} in all images **do**

$$L_D^{RaD} \leftarrow -\mathbb{E}x_{real} * \left[\log\left(D_{RaD}\left(x_{real}, x_{fake}\right)\right)\right] - \mathbb{E}x_{fake} * \left[\log\left(1 - D_{RaD}\left(x_{real}, x_{fake}\right)\right)\right]$$

$$L_G^{RaD} \leftarrow -\mathbb{E}x_{real} * \left[\log\left(1 - D_{RaD}\left(x_{real}, x_{fake}\right)\right)\right] - \mathbb{E}x_{fake} * \left[\log\left(D_{RaD}\left(x_{real}, x_{fake}\right)\right)\right]$$

$$D_{RaD}\left(x_{real}, x_{fake}\right) \leftarrow \left(c(c_{real}) - \mathbb{E}x_{fake}\left[c\left(x_{fake}\right)\right]\right)$$

    end for
    return $L_D^{RaD} and\ L_G^{RaD}$
end function

function perceptual_loss ($G(x_i)$, L_{Gen})
    for $G(x_i)$ in all images **do**

$$L_{Gen} \leftarrow L_{per} + \lambda * L_G^{RaD} + \eta * L_1$$

$$L_1 \leftarrow \mathbb{E}x_i ||G(x_i) - y||1$$

    end for
    return $\mathbf{L_{Gen}}$
end function

3.7 Dataset Training Pre-Processing:

As the dataset is captured by DSLR cameras, the images are of high resolution and high quality. So, the first task is to convert these high-quality and high-resolution flow-resolution and quality. So, that it can be used for the project. The need or inputs for the project are low-resolution images with less pixel size so that the model works faster and generates output with less complexity. The trained model is used model is trained on the DIV2K dataset containing 800 high-resolution images. Other datasets are also used like Flickr2K obtained from the Flickr website having 2650

x 2000 high-resolution images. To enrich the training outdoor scene training (OST) data set is used. Using these high-resolution datasets with richer textures helps the generator to produce more natural results [49].

The preparation interaction is partitioned into two phases. The PSNR-situated model is prepared at a pace of 2 x 104 and rots by an element of two after each 2 x 105 smaller than expected group refreshes. The prepared PSNR-situated model is then used to instate the generator. The misfortune capability is utilized to prepare the generator: The pace of learning is set to 1 x affirmative and split at 50K, 100K, 200K, and 300K emphasis with = 5 x 10-3 and = 1 x 10-2. This pre-preparing with pixel-wise rate misfortune helps the GAN-based technique in creating more outwardly engaging outcomes by staying away from bothersome nearby optima for the generator. After preparing, the discriminator creates great, super-settled pictures as opposed to outrageous, counterfeit ones. Focusing more on surface discrimination is advantageous. Adam (39) is utilized in advancement with/31 = 0.9 and/32 = 0.999. The generator and discriminator are on the other hand refreshed until the model merges. For the generator, there are two principal settings. One of them has 16 leftover blocks and a limit like SRGAN, while the other is executed utilizing the Poarch structure.

3.8 Perceptual Loss

The loss function is also enhanced or improved by constraining the features before activation rather than after, as in GAN. The idea of perceptual misfortune resembles the idea of perceptual similitude in past work; the perceptual misfortune is characterized by the enactment layers of a more profound organization that has been pre-prepared in which the distance between two actuated highlights is limited. In this model, highlights are utilized before the enactment layers to address the deficiencies of the past plan. In the first place, the enactment highlights are extremely scanty, particularly after an exceptionally profound organization. Second, when contrasted with the ground-truth picture, includes the utilization of the after-actuation capability brings about capricious remade brilliance[50]. As a result, the total loss for the generator is

$$L_{Gen} = L_{per} + \lambda L_G^{RaD} + \eta L_1 \qquad \text{Were,} \qquad L_1 = \mathbb{E}x_i||G(x_i) - y||1 \qquad (1)$$

Equation depicts the content loss used to calculate the 1-norm distance between ground truthy and recovered image. A and t denote the coefficients used to balance the various loss terms, A fine-learned VGG network is used to investigate the variant loss.

3.9 Relativistic Discriminator

Aside from the enhancement of the generator structure discriminator, the relativistic GAN is also used. The probability is calculated based on how much more realistic the real image is than the fake one. The discriminator in the older SRGAN only checks that the input image is real and natural[51].

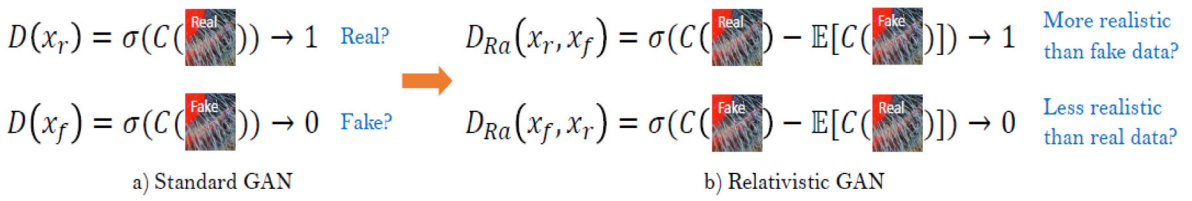

Figure 10: Discriminator

The standard discriminator is supplanted by Relativistic normal Discriminator Rd. This standard discriminator D is unique to SRGAN's discriminator, which is utilized to assess the likelihood of information picture x concerning realness and effortlessness. This relativistic discriminator D predicts the probability that the real image xreal¬ is more realistic than the fake image x-fake, as shown in Figure 10.

Relativistic discriminator D is replaced by relativistic average discriminator RaD as denoted by D_{RaD}. The variable

$$D(x) = \sigma(C(x)) \qquad (2)$$

shows the standard discriminator. As x is the input image. Where are the sigmoid function and C(x) is the output of the non-transformed discriminator? Hence

$$D_{RaD}(xreal, xfake) = \sigma(C(x_{real}) - Ex_{fake}[C(x_{fake})]) \qquad (3)$$

is used for RaD formulation. Where Exfake [.] shows the operations for taking the average for all the fake data in the min-batch. The data loss is defined as:

$$L_D^{RaD} = -Ex_{real}[\log(D_{RaD}(x_{real}, x_{fake}))] - Ex_{fake}[\log(1 - D_{RaD}(x_{fake}, x_{real}))] \qquad (4)$$

The symmetrical form of adversarial loss for the generator is:

$$L_G^{RaD} = -Ex_{real}[\log(1 - D_{RaD}(x_{real}, x_{fake}))] - Ex_{fake}[\log(D_{RaD}(x_{fake}, x_{real}))] \qquad (5)$$

Were $x_{fake} = G(x_i)$

as xi shows the input LR image. As observed that adversarial loss of the generator contains both xreal and xfake. So, the generator takes the benefit from both values, But in SRGAN only generated part takes place.

3.10 Network Interpolation

A flexible and effective strategy called network interpolation is proposed for removing unpleasant noise in GAN-based methods while maintaining good perceptual quality. First, a PSNR-oriented network, GPSNR, is trained, and then a GAN-based network, GGAN, is obtained by fine-tuning the interpolation of GPSNR and GAN parameters [52]. Network interpolation has two advantages. For starters, this interpolated model produces feasible results without introducing artifacts. Second, perceptual quality and fidelity are criticized[53].

3.11 Residual Network (Reset)

ResNet is a CNN architecture that can aid in the computation of loads or additional convolutional layers. ResNet can add many layers with robust performance, whereas previous architectures saw a drop in effectiveness with each additional layer. Deep convolutional and neural networks have resulted in a slew of image classification breakthroughs. It serves as the foundation for many computer vision tasks. It enables the successful training of extremely deep neural networks with 150+ layers. The reset's main strength is its skip connections. The skipped connections are depicted in the diagram below in figure 11 [54].

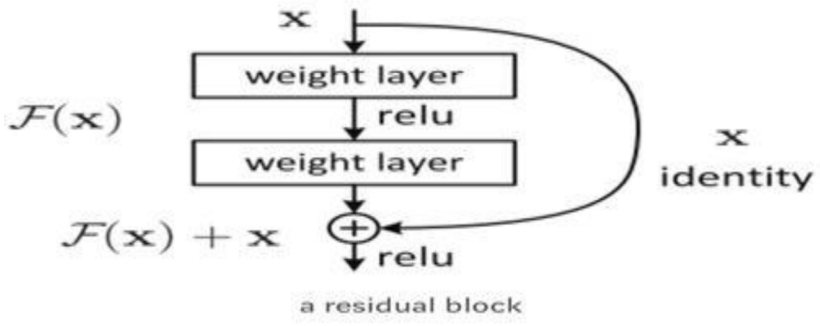

Figure 11: ResNet

The label "Identity" allows the network to bypass the other weighted layers, bypassing the input through the block [55].

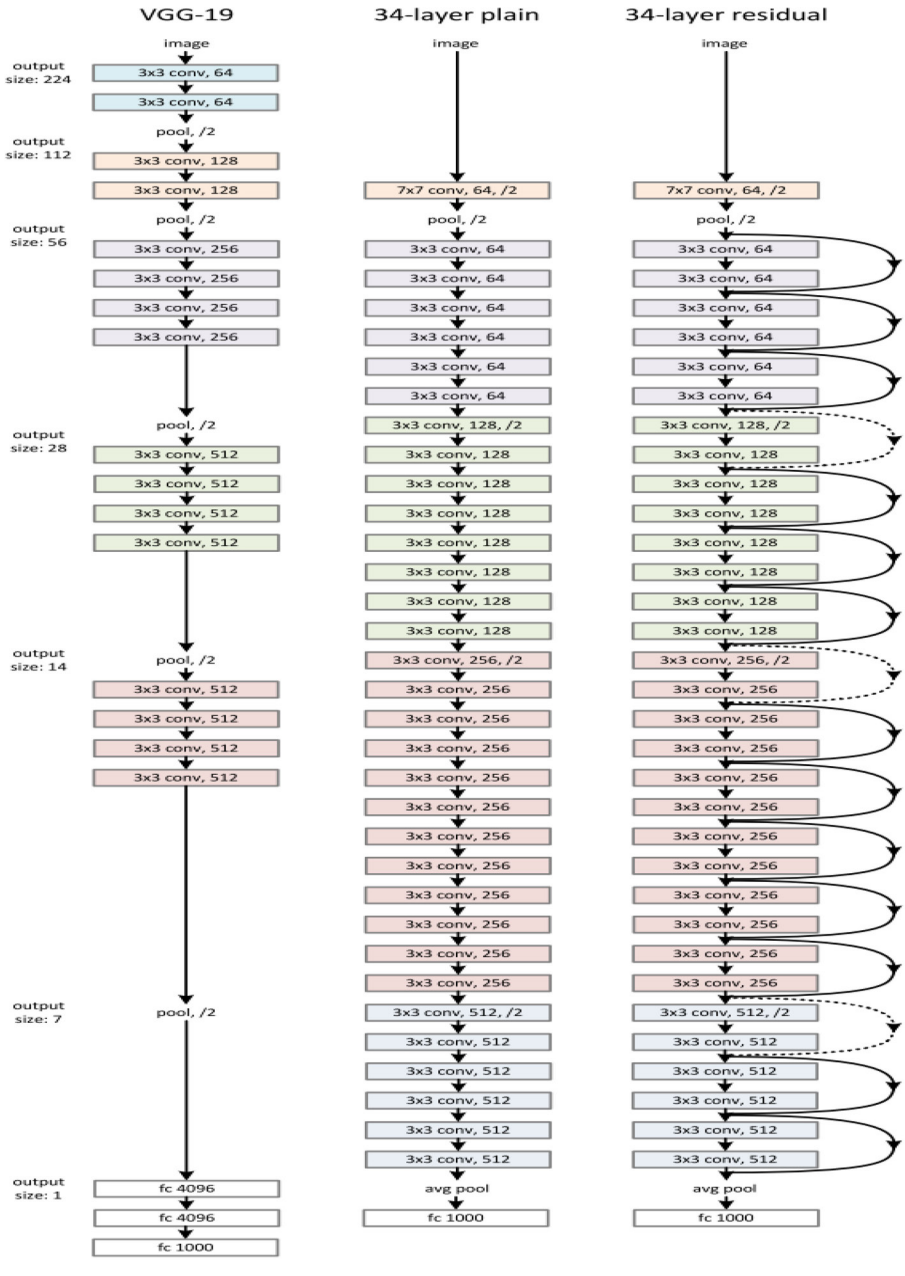

Figure 12: ResNet Skip Connections

This permits extra layers to be stacked and a more profound organization to be constructed, balancing the evaporating slope by permitting the organization to skip layers [56].

3.12 Variations in ResNet:

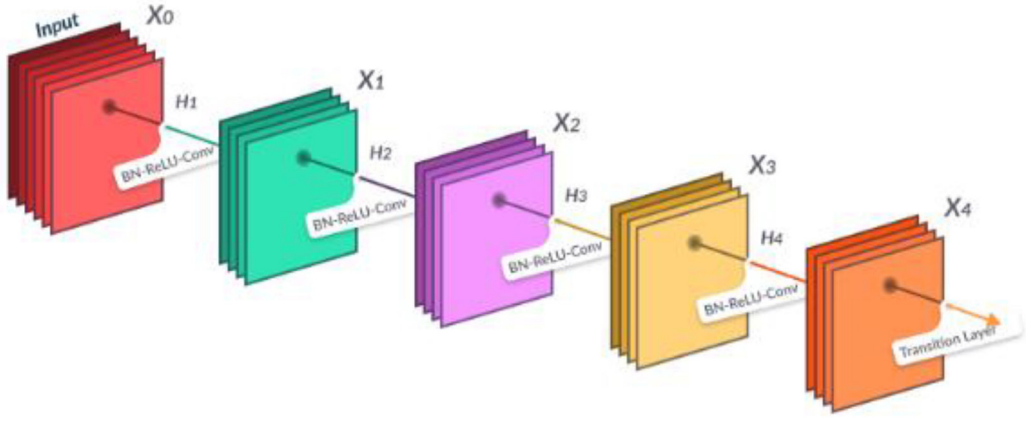

Figure 13: Layers working

3.13 ResNet Porch Implementation

Reset became the name of the country's artwork in computer vision in 2015 and is still hugely famous. It can teach masses or thousands of layers without a "vanishing gradient “displayed in Figure 14 Porch makes it simple to create ResNet fashions; it includes several pre-trained ResNet architectures and allows you to create your own [57].

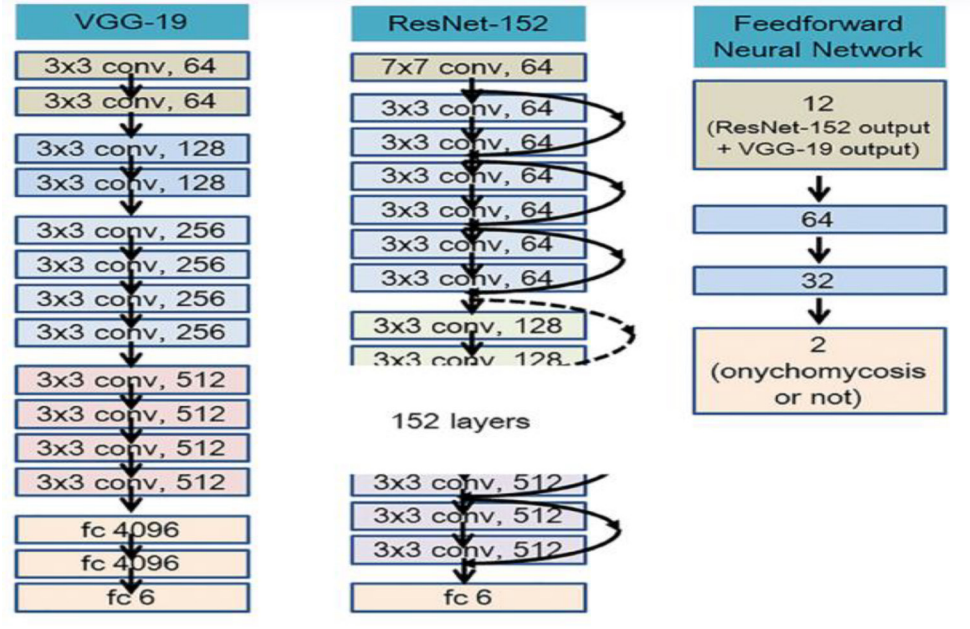

Figure 14: VGG-19, ResNet, and Feedforward Neural Network

4. Results and Discussion

4.1 Process:

The evaluation of the approach is executed as follows. From the dataset of 500 DSLR images, 200 images are selected randomly, and the model is run on these images after resizing. After getting the result, the input set is again resized according to the output images. This will clear the visual difference in output concerning input images.

4.2 Evaluation Metrics:

To access the performance of the proposed approach, PSNR and SSIM are measured. These metrics are described here:

- PSNR: Peak signal-to-noise ratio, often abbreviated PSNR, is a designing term for the proportion between the most extreme conceivable force of a sign and the force of tainting commotion that influences the constancy of its portrayal. Since many signs have an extremely wide powerful reach, PSNR is generally communicated as far as the logarithmic decibel scale.
- SSIM: The structural similarity (SSIM) file is a strategy for anticipating the apparent nature of computerized TV and true-to-life pictures, as well as different sorts of computerized pictures and recordings.

4.3 Peak Signal-to-Noise Ratio (PSNR):

The term top sign-to-commotion proportion (PSNR) is an articulation for the proportion between the greatest conceivable worth (force) of a sign and the force of contorting clamor that influences the nature of its portrayal.

4.4. Mean Square Error:

For PSNR, Mean Square Error (MSE) is measured. This is measured by subtracting the test signal from the reference. The average energy of the error signal is computed. MSE is the simplest and most widely used quality measurement for images. This metric is widely used in signal processing and is defined as:-

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x(i,j) - y(i,j)\right)^2 \qquad (6)$$

where $x(i,j)$ represents the original image and $y(i,j)$ represents the modified image and i and j

are the pixel positions of the M×N image. The value of MSE is zero when $x(i,j) = y(i,j)$.

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{\sqrt{MSE}} \quad (7)$$

The mean squared Error (MSE) permits us to analyze the "valid" pixel upsides of the first picture to the corrupted picture. The MSE addresses the normal of the squares of the "mistakes" between the genuine picture and the boisterous picture. The blunder is the sum by which the upsides of the first picture contrast from the debased picture.

The suggestion is that the higher the PSNR, the better the corrupted picture has been changed to match the primary picture, and the better the reconstructive computation. This would occur considering the way that restricting the MSE between pictures concerning the most outrageous sign worth of the image.

4.5 Structural Similarity Index (SSIM):

SSIM is a method to predict the perceived quality of digital and cinematic pictures, as well as for other kinds of images and videos. The SSIM is measured with the following equation: -

$$SSIM = \frac{(2 \times \bar{x} \times \bar{y} + C1)(2 \times \sigma_{xy} + C2)}{(\sigma_x^2 + \sigma_y^2 + C2) \times ((\bar{x})^2 + (\bar{y})^2 + C2)} \quad (8)$$

Where C1 and C2 are constants. $\bar{x}, \bar{y}, \sigma_x^2, \sigma_y^2$ and σ_{xy} are given below:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{1=1}^{N} (x_i - \bar{x})^2$$

$$\sigma_y^2 = \frac{1}{N-1} \sum_{1=1}^{N} (y_i - \bar{y})^2$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

4.6 Comparative Analysis of Old and Proposed Approach:

Results are evaluated for both old and proposed approaches. PSNR and SSIM are calculated via PYTHON. Results are shown in the following table.

Table 1: PSNR and SSIM of both models

A	B	C	D
Input Image	**Metrics**	**ESRGAN with VGG-19**	**ESRGAN with ResNet**
1	PSNR	13.89	14.01
	SSIM	0.8803	9.8544
2	PSNR	18.59	18.7
	SSIM	0.8923	0.8711
3	PSNR	17.17	17.89
	SSIM	0.7896	0.7453
4	PSNR	∞	∞
	SSIM	1	1
5	PSNR	19.12	19.87
	SSIM	0.9236	0.8675
6	PSNR	16.78	17.03
	SSIM	0.8734	0.8346
7	PSNR	16.31	16.79
	SSIM	0.9543	0.9112
8	PSNR	14.59	15.21
	SSIM	0.7643	0.7345
9	PSNR	16.43	17.11
	SSIM	0.8711	0.8222
10	PSNR	20.04	19.78
	SSIM	0.8122	0.7999
11	PSNR	21.22	20.18
	SSIM	0.9632	0.9322
12	PSNR	20.87	`19.67
	SSIM	0.9111	0.8665
13	PSNR	17.89	17.77
	SSIM	0.7112	0.6998
14	PSNR	18.09	17.99
	SSIM	0.9042	0.8554
15	PSNR	15.9	15.04
	SSIM	0.8923	0.8212
16	PSNR	18.67	18.99
	SSIM	0.7812	0.7236
17	PSNR	19.54	19.03
	SSIM	0.9146	0.8854
18	PSNR	20.34	19.89
	SSIM	0.7922	0.7543
19	PSNR	14.56	14.22
	SSIM	0.8544	0.8.321

The table shows the results of ESRGAN with VGG-19 and ESRGAN with ResNet. The values of PSNR and SSIM are shown here. As in PSNR, MSE is in the denominator that estimates the error. Therefore, if the error is high (i.e., low image quality), then the value of PSNR is lower and if the error is less (i.e., better image quality), then the value of PSNR is high. So, lower PSNR values indicate better image quality. On the other hand, SSIM is used to measure the similarity index. If the value is one, it means that both images have the same quality, but if the values are different then the similarity between the input and the resulting image is low.

Figure 15: Color Space

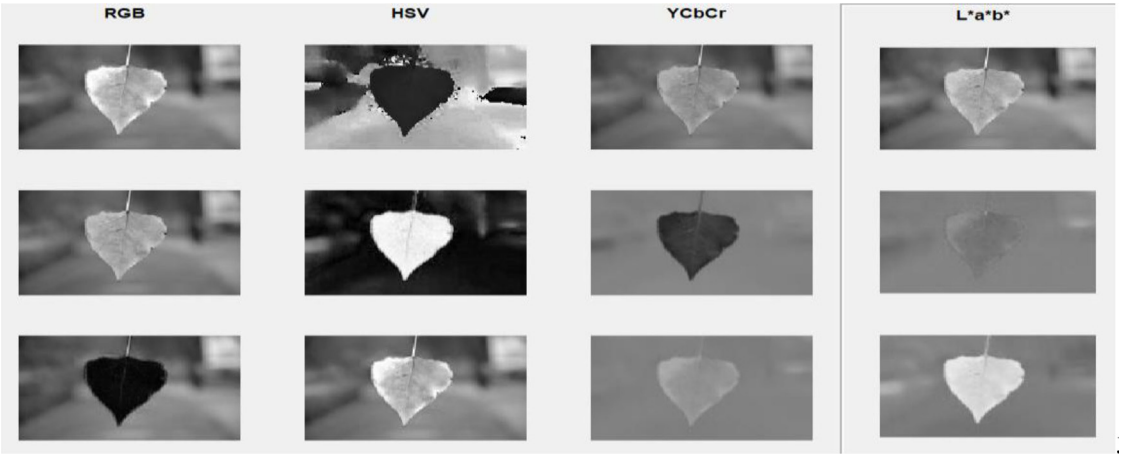

Figure 16: Color Space

The above figures show the color space of a picture. RGB, HVS, YCbCr, and L*a*b color spaces are shown here.

Figure 17: Comparison of reference and blurred image in SSIM

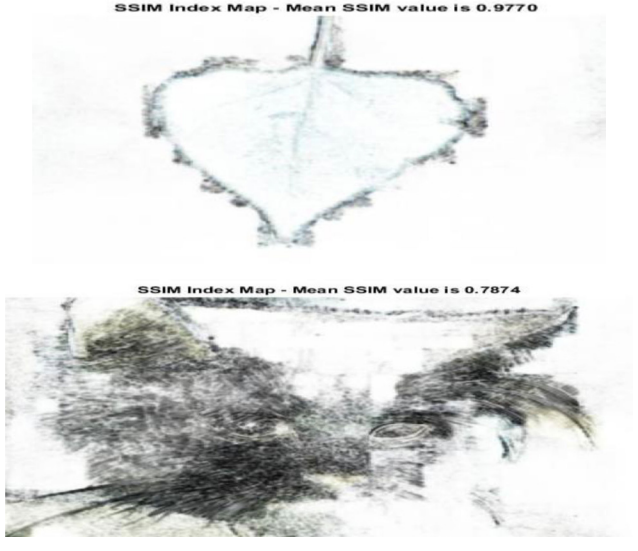

Figure 18: SSIM index map

The above figures show the SSIM index Map of the images. The higher the value is, the better the results are. Where the value is 0.9770, it means it is closer to one and the similarity index is high which shows that the image possesses high quality after processing. Where the value is o.7874, it means that this image possesses excellent quality but has some difference from the original image.

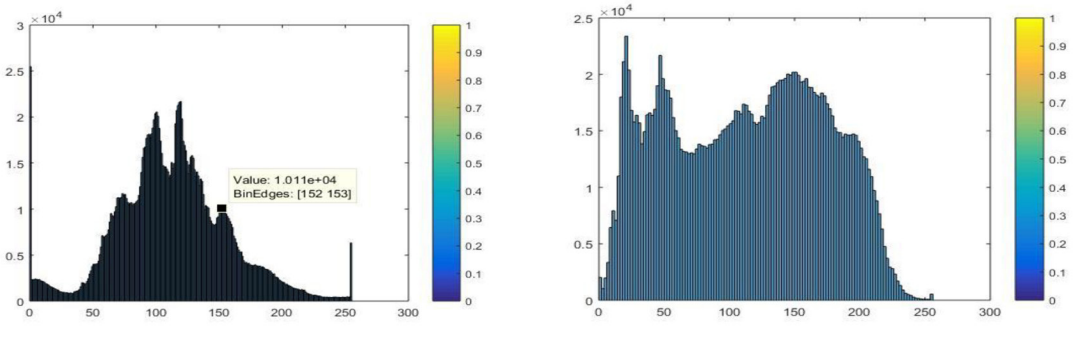

Figure 19: Histograms

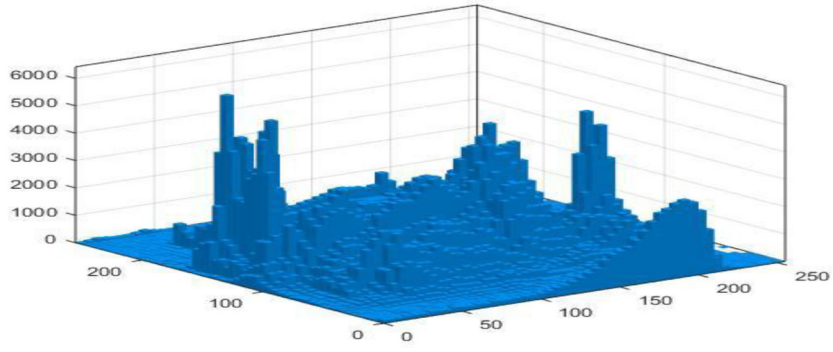

Figure 20: Comparison of two images

The above figures show the histograms of images. It is a graphical representation of the tonal distribution in digital images. The number of pixels is plotted for each tonal value. The x-axis represents the tonal variation while the y-axis represents the number of pixels in that tone.

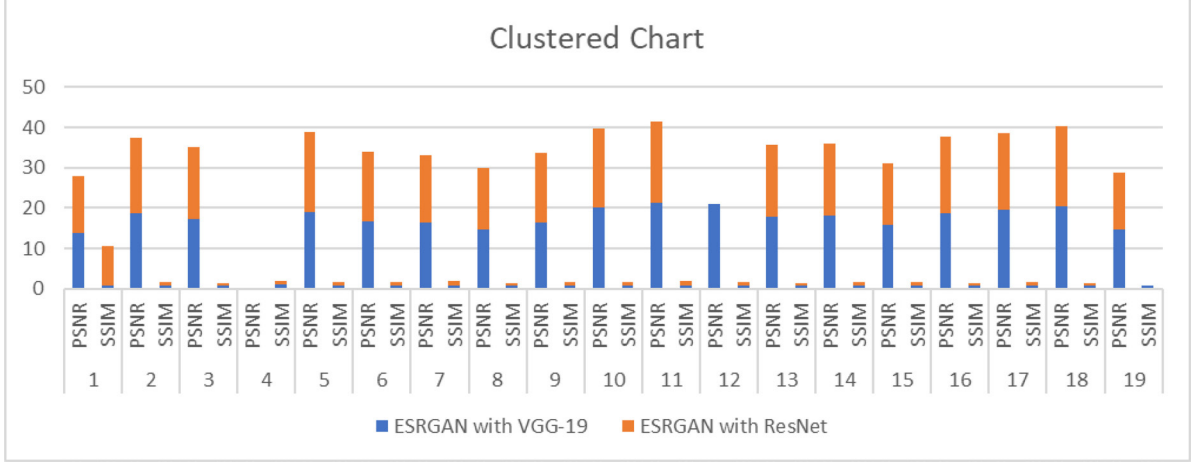

Figure 21: Clustered chart for the comparison of models

The above figure shows the different values of the processed dataset. SSIM and PSNR are shown for both old and proposed approaches. Both approaches possess almost comparable results. The higher the PSNR value, the better the image quality. The same is the case with the SSIM value.

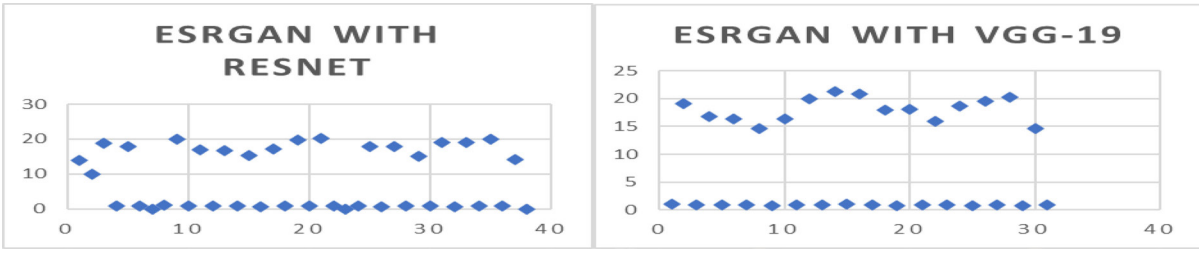

Figure 22: PSNR and SSIM for ResNet and VGG-19

The above scatter chart shows the scattered values of PSNR and SSIM. The higher the values are, the better the results are. The difference in values for both approaches can be seen here easily. To evaluate the proposed approach and methodology, some research questions are demonstrated here:

1) If the proposed approach is better than the older approach, then to what extent?
2) What is the relationship between VGG-19 and ResNet-50?
3) How does the quantity of layers affect the results?

To answer the above research question 1, ResNet-50 and VGG-19 are used, BN layers are removed, and the network is enhanced. Removal of BN layers have proven to increase the performance and reduce the computational complexity in many PSNR-oriented tasks including SR and deblurring. To answer the question 2, models are run, and results are evaluated. VGG-19 consists of 138 million parameters. On the other hand, in Residual Network ResNet, skip connections are used. These skip connections are also known as gated units or gated recurrent units. To answer the question 3, qualitative and quantitative analysis is done. PSNR and SSIM are calculated. Visual results are seen.

5. Conclusion

Image super-resolution is the fundamental task of improving the visual quality of the image. This task is done by taking a single LR image and converting it into an HR one with the help of some suitable models of neural networks. This task of image processing is in the computer vision area. The primary point is to upgrade or work on the perceptual nature of picture SR. SR strategies in which perceptual misfortune capability apply by the profound learning CNN. CNN concentrates on profound learning ways to deal with resolving the nature of single-picture super-goal. Various papers can portray various ways to deal with picture super-goal which recuperate the high goal from the low-goal picture. The primary point is to upgrade or work on the perceptual nature of picture SR. SR strategies in which perceptual misfortune capability apply by the profound learning CNN. CNN concentrates on profound learning ways to deal with resolving the nature of single-picture super-goal. Various papers can portray various ways to deal with picture super-

goal which recuperate the high goal from the low-goal picture/. We can estimate the multiple CNN approaches in which architecture training that improved the SR of the image calculated the ground truth value to show the output result with features information. Enhanced super-resolution is an especially important part of the resolution techniques which describe the perceptual loss function and improved the visual quality of the super-resolution result. We can estimate part one by one first pre-process the input image to convert the residual into the residual block. They remove the BN layer with fast training. Residual scaling feature of the image estimate with convolutional and smaller initialization. SISR is the fundamental low-level vision problem. SISR aims to recover LR images and convert them to HR. Many network architectures are destroyed and many training strategies are improved for enhancing the SR performance, especially the ground truth value. Relativistic average GAN calculates the average value of the ground truth image. Then ground truth image shows the result of the VGG model for the high-resolution image to perform different network architecture. RaGAN that we use to better relativistic discriminator which estimated the features for sharper edges and multiple detailed textures in deep learning. The GAN structure is improved to improve them based on the Relativistic GAN. ESRGAN which compares the multiple models that represent each column configures ratio. The residual in the residual block is the deep network. We can define in this research a deeper learning network with RRDB. We define the graph of the super-resolution image dataset as compared to different values in content loss and perceptual loss results for the GAN-based network techniques. The method of the perceptual driven model enhances super-resolution and super-resolution of the GAN is defined by the multiple image dataset and ground truth values are calculated in the resultant image and converted into multiple graphs.

Reference

[1] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." *European conference on computer vision*. Springer, Cham, 2016.

[2] Zhang, Lvmin, et al. "Two-stage sketch colorization." *ACM Transactions on Graphics (TOG)* 37.6 (2018): 1-14.

[3] Al Azzeh, Jamil, et al. "Creating a color map to be used to convert a gray image to color image." *International Journal of Computer Applications* 153.2 (2016): 31-34.

[4] Kumar, Smriti, and Ayush Swarnkar. "Colorization of gray scale images in lαβ color space using mean and standard deviation." *2012 IEEE Students' Conference on Electrical, Electronics and Computer Science*. IEEE, 2012.

[5] Haldankar, Govind, Atul Tikare, and Jayprabha Patil. "Converting gray scale image to color image." *Proceedings of SPIT-IEEE Colloquium and International Conference, Mumbai, India*. Vol. 1. No. 189. 1989.

[6] Ramesh, Balasubramaniam, et al. "Can distributed software development be agile?." *Communications of the ACM* 49.10 (2006): 41-46.

[7] Kock, N., and M. Mayfield. "PLS-based SEM Algorithms: The Good Neighbor Assumption." *Collinearity, and Nonlinearity* (2015

[8] Karlik, Bekir, and Mustafa Sariöz. "Coloring gray-scale image using artificial neural networks." *2009 2nd International conference on adaptive science & technology (ICAST)*. IEEE, 2009.

[9] Yu, Lu, et al. "Beyond eleven color names for image understanding." *Machine Vision and Applications* 29.2 (2018): 361-373

[10] Al-Azzeh, Jamil, Ziad Alqadi, and Mohammed Abuzalata. "Performance Analysis of Artificial Neural Networks used for Color Image Recognition and Retrieving." *international Journal of Computer Science and Mobile computing* 8.2 (2019): 20-33.

[11] Chary, R., D. Rajya Lakshmi, and K. V. N. Sunitha. "Feature extraction methods for color image similarity." *arXiv preprint arXiv:1204.2336* (2012).

[12] Kuo, Cheng-Hao, Sameh Khamis, and Vinay Shet. "Person re-identification using semantic color names and rankboost." *2013 IEEE workshop on applications of computer vision (WACV)*. IEEE, 2013.

[13] Ansari, Mohd, and Dushyant Kumar Singh. "Significance of Color Spaces and Their Selection for Image Processing: A Survey." *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* 15.7 (2022): 946-956.

[14] Malviya, Ashwini V., and Siddharth A. Ladhake. "Copy move forgery detection using low complexity feature extraction." *2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)*. IEEE, 2015.

[15] Anwar, Saeed, et al. "Image colorization: A survey and dataset." *arXiv preprint arXiv:2008.10774* (2020).

[16] Huang, Shanshan, et al. "Deep learning for image colorization: Current and future prospects." *Engineering Applications of Artificial Intelligence* 114 (2022): 105006.

[17] Dhir, Rashi, Meghna Ashok, and Shilpa Gite. "An overview of advances in image colorization using computer vision and deep learning techniques." *Review of Computer Engineering Research* 7.2 (2020): 86-95.

[18] Royer, Amelie, Alexander Kolesnikov, and Christoph H. Lampert. "Probabilistic image colorization." *arXiv preprint arXiv:1705.04258* (2017).

[19] Varga, Domonkos, and Tamás Szirányi. "Fully automatic image colorization based on Convolutional Neural Network." *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016.

[20] Nguyen, Tung, Kazuki Mori, and Ruck Thawonmas. "Image colorization using a deep convolutional neural network." *arXiv preprint arXiv:1604.07904* (2016).

[21] En.wikipedia.org. (2019). *"Mughal-e-Azam"*. [online] Available at: https://en.wikipedia.org/wiki/Mughal-e-Azam#Colourisation, 2019.

[22] Bao, Bin, and Hongbo Fu. "Scribble-based colorization for creating smooth-shaded vector graphics." *Computers & Graphics* 81 (2019): 73-81..

[23] Reddit.com. (2019). " *Colorization - The colorization of old black & white photos*". [online] Available at: https://www.reddit.com/r/Colorization/, 2019.

[24] He, Mingming, et al. "Deep exemplar-based colorization." *ACM Transactions on Graphics (TOG)* 37.4 (2018): 1-16.

[25] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." *European conference on computer vision*. Springer, Cham, 2016.

[26] "Colorize Black and White Photos", Algorithmia, 2019. [Online]. Available: https://demos.algorithmia.com/colorize-photos, 2019.

[27] Reinhard, Erik, et al. "Color transfer between images." *IEEE Computer graphics and applications* 21.5 (2001): 34-41.

[28] Tai, Yu-Wing, Jiaya Jia, and Chi-Keung Tang. "Local color transfer via probabilistic segmentation by expectation-maximization." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. IEEE, 2005.

[29] Welsh, Tomihisa, Michael Ashikhmin, and Klaus Mueller. "Transferring color to greyscale images." *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 2002.

[30] Alim, Affan, et al. "The most discriminant subbands for face recognition: A novel information-theoretic framework." *International Journal of Wavelets, Multiresolution and Information Processing* 16.05 (2018): 1850040.

[31] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." *European conference on computer vision*. Springer, Cham, 2016.

[32] "CNN Architectures, a Deep-dive", Medium, 2019. [Online]. Available: https://towardsdatascience.com/cnn-architectures-a-deep-dive-a99441d18049, 2019.

[33] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60.6 (2017): 84-90.