Single Image Super Resolution Using Multiple Deep Convolution Neural Network

MSc Research Project

MSc Data Analytics

Khushbu Rajpara

Student ID: x20175671

School of Computing

National College of Ireland

Supervisor: Prof. Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Khushbu Mukeshbhai Rajpara		
Student ID:	X20175671		
Programme:	Msc. In Data Analytics	**Year:**	2021
Module:	Research Project		
Supervisor:	Prof. Vladimir Milosavljevic		
Submission Due Date:	16th December 2021		

Project Title: Single Image Super resolution using Multiple Deep Convolution Neural Network

Word Count: **7832 Page Count:22**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Khushbu M. Rajpara

Date: 16th December 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Single Image Super Resolution Using Multiple Deep Convolution Neural Network

Khushbu Rajpara
X20175671

Abstract

Image super-resolution has long been the subject of controversy in the areas of computer vision and image processing. Deep learning has contributed significant advances in the past several decades in achieving high-resolution images with additional information utilizing deep neural networks. Modern super-resolution (SR) techniques often use convolutional neural networks to read complex non-redirected maps between paired LR and HR images. This study will be focusing on deep neural network usage methods for image enhancing activity. High-resolution images are utilized as the target while degraded images are used as the network's input to get desired outcomes. Convolution neural network (CNN), Auto Encoder model, Multiscale learning model, and enhanced deep residual network (EDSR) are the four deep neural network models employed in this study. To assess the evaluation criteria and gain a comparative analysis among the models. The models are experimenting on public datasets and attempting to use a GPU infrastructure with minimal training times. The efficiency of the model is demonstrated by a comparative analysis of the models which is based on a test image dataset using RMSE, the accuracy of the model, and PSNR metrics. The accuracy of the Auto Encoder model is the highest among the four networks, according to the findings of testing. With 30 epochs, the Autoencoder has an accuracy of 86.37 % and an MSE of 0.0177 with the highest PSNR value which is 34.56 dB.

1 Introduction

1.1 Background

Image-based visual effects models exhibit pixel density independence when zooming images away from the picture sample resolution without losing image quality. A term "single image super-resolution" refers to the process of estimating a high quality of image from a damaged image. The phrase "super resolution" was initially used to explain how the resolution of an optical system could be increased (Gerchberg et al., 1974). The approach of achieving higher resolution images from a corresponding degraded image has been defined as super-resolution for the past two decades. This approach was originally used to mark “spatial resolution enhancement” (Tsai et al, 1984). Although image resolution has increased significantly over time leading to the advancements of high-performance lenses, improved image compression, and display panels, the desire for produce high-resolution photographs from low-resolution photographs is also rising for presentation on high-resolution screens. Multiple-image techniques and Single image super resolution are the two types of image super-resolution technologies currently available. Furthermore, the term "video super-resolution" refers to a form of multiple picture enhancement in which the pictures are portion of a scene with multiple frames. This project solves the common problem of obtaining a high-resolution image from a low-resolution replica.

1.2 Research rationale

Research have been conducted in order to gather knowledge of the different types of combination that has been developed by using a low-resolution sequence of images of any scene which is further used in order to generate a higher resolution image. In the advancement of in-depth learning strategies in the modern time, super-resolution-based learning models have been actively studied and often achieve advanced performance on various benchmark dataset. Image enhancement tasks have been addressed using several in-depth learning methods, from the very first method based on convolutional neural networks to the most effective SR methods based on generative advertising networks. It is a major type of image processing techniques for computer vision and image processing, with real-world applications including medical imaging, satellite images, text and face recognition, surveillance and security, Ultrasound imaging, star photography, fingerprint image enhancement. and many others, this demonstrates the relevance of the topic in recent years (Bashir et al.,2021). (Qin et al.,2020) in the article recognizing face mask-wearing circumstances with image super-resolution as well as a classification network to combat the Covid-19 epidemic. It has been identified that even though the correct face mask is critical to protect oneself from infectious diseases its effectiveness in facial recognition have deteriorated. The advantages include obtaining a better resolution image from that never before or was lost, which might be useful in a variety of situations or perhaps even lifesaving in clinical application. This study provides a comprehensive analysis of four different models that have never been done before in any research work. Previous studies of this model were conducted on a different dataset. To get better outcomes, high-resolution photos are used as the target while degraded images are used to feed the network.

1.3 Research Questions

The following research questions helps us to study the project in detail and provides us with the solution for following descripted research questions.

1) How end to end mapping between high- and low-resolution images can be achieved in image super resolution using deep convolutional neural network?

The design and Implementation chapters describes the answer for the above research question. Deep learning models along with the transfer learning and model weights are trained for the mapping of low-resolution images to high resolution images.

1.4 Research Objectives

Digital cameras are used to collect images in a range of industries, including remote sensing and any type of identification system, such as face and text recognition. The images produced by these devices occasionally have a lack of observation, with components such as noise corruption and illumination, as well as lighting, impacting the low-resolution processes.
The following are the objectives of the study

1. A critical assessment of the associated studies with Single image super resolution
2. Data selection, preparation, and visualization of the data.
3. Implementing the models of deep neural network that can improve the quality of damaged photos.

4. Evaluation, and the outcome of four models (Convolution neural network, Auto Encoder, Multiscale learning, and enhanced deep residual network) based on MSE, accuracy and PSNR.
5. A comparison of the model that has been implemented.

The following is the report's structure. The related study on the image super resolution area is covered in section 2. Section 3 discusses the suggested approach for this study using CRISPDM, followed by modeling, which includes the fundamentals of the model utilized in this study, and finally data evaluation and design specification. Section 5 covers with the research's implementation, which is divided into two subsections: data collecting and preparation. Another data modeling is explained in section 5 in which the implemented model is explained. The evaluation of the implemented model and discussion are included in section 6. The final section of the paper discusses the proposed model's conclusion and future direction.

2 Related Work

Over the recent period, there has been considerable change in image super-resolution techniques, and a range of cutting-edge super resolution methods based on deep learning have been suggested, which have been shown to dramatically increase picture enhancing performance over traditional methods (Nasrollahi et al.,2014). This section summarizes the fundamental aspects of such deep neural networks, such as down sampling methods, up sampling methods, super resolution framework and network designs, to examine those related research accomplished in the image enhancing area.

2.1 Image super resolution

Based on the image basis function, image super-resolution techniques were classified into four categories: edge-based methods, patch-based methods, image statistical approaches, and prediction models. The researcher implemented and evaluated all these methods in their study. The example-based technique performed the best out of all the strategies examined (Yang et al.,2014).

Many research had concentrated on loss functions in order to improve network model training. The highly used loss function for conventional image restoring is mean squared error (MSE), which is a significant measure of the efficiency (PSNR) for such issues. However, (Zhao et al., 2015) the model training by loss function sometimes doesn't achieve better performance in terms of SSIM and PSNR particularly comparing to other loss functions.

The authors of this study examine three super resolution methods: domain specific super resolution, supervised and unsupervised super resolution (Wang et al. 2020). In addition, researchers consider that real-world photographs were prone to deterioration such as additive noise, compression abnormalities, blurring, and illuminating. As a result, models trained on manually generated datasets frequently perform badly in real-world scenarios. They also discovered certain domain-specific super-resolution applications, such as object tracking, scene rendering, and medical imaging.

2.2 Deep learning approach based on traditional method

To achieve super-resolution in the past, fundamental super-resolution approaches such as patch-based methods, sparse representation, prediction-based methods, edge-based methods, and statistical methods were utilized. This was the approach in which geometric characteristics from a high-resolution paradigm were employed as guides to direct the restoration of degraded pictures from the other paradigm utilizing image duplication. As a deep convolution network, they employed sparse-coding-based super-resolution approaches. The important element of their suggested strategy is that it optimizes all layers simultaneously rather than addressing each component separately. Apart from super resolution, another feature of their approach is that it may be utilized for picture deblurring and denoising (Dong et al.,2014).

Sparse coding, a popular learning-based approach, has recently been investigated for single picture SR showing impressive outcomes. (Zeng et al.,2015) developed a data-driven model linked deep autoencoder for single super resolution based on a novel deep architecture with excellent representational capabilities employing sparse coding. Although this hypothesis was excessively restricted or inadequate in generally, this could fail to meet pleasant SR effects. The idea of sparse coding noise is developed (Dong et al.,2012) in the study to improve the efficiency of minimal representation-based picture reconstruction, as well as the focus for image retrieval shifts including how to minimize the SC noise.

2.3 Image super resolution based on neural network

Convolutional neural networks are effective in the area of super-resolution. Traditional CNNs, on the other hand, have certain restrictions in terms of using multi-scale contextual information for image restoration due to the fix kernel in the building modules. The study added a multi-scale convolutional kernel to conventional convolution operation, that also offers multi-range context data for image super-resolution, and even a strategic analysis to CNN, that also dynamically selects the better scale for convolution layers even while significantly lowering the dimensions of intermediate outputs (Du et al.,2018).

Image details are critical for image restoration challenges such as super-resolution and denoising. The input image of network is up sampled using bicubic interpolation before being sent into the deep learning algorithm. Another option is to train the up-sampling module at the end of network state (Szegedy et al., 2015). To address this problem, (Cao et al., 2019) work developed a highly recursive (16 recursive layer) image super resolution approach that uses a much broader context than earlier SR methods that simply used a single recursive layer. The restriction of the model is the fixed number of recursive layers; unless the number of recursive layers is incremented, the performance of the model would be achieved better performance.

Many CNN based models for image enhancement have a feature of shallow channel that frequently loses the image's detailed information. To overcome this problem (Li, et al.,2018) proposed algorithm, that is recognized as s deep and shallow convolutional networks for super-resolution, solves the identified issues. The deep channel, on the other hand, collects precise image features. The suggested technique generates a final mapping among high- and low-resolution images without requiring hand-designed processing. The deconvolution up sampling of the network is included in the two channels, resulting in significantly more effective and efficient training of network, and decreasing the computing cost of the total super-resolution process.

The author concentrated on image super resolution approaches for enhancing automated recognition systems, including facial recognition and optical character identification in real-world situations (Peyrard 2017). The presentation and restoration of spatial higher frequencies is involved in the process of employing neural networks to conclude into the process of image super-resolution, and it will be effective in compensating the artefacts that could be present in the form of blur edges. It helps in ensuring the network is being trained with the presence of the same is addressed simultaneously. It extracts relevant non-linear features, which will be productive in mapping down the low and high-resolution spaces as needed. It could be highlighted that the training has been done in the neural network process held more effective LR and HR images which has been applied to develop blurred kernels. Also, the development of a non-blind set-up and the blind set-up increases the overall observation focus of the model, and it is being reported to the overall SR approaches that are being considered. To overcome the text image and facial feature recognition difficulties, two distinct strategies were presented. OCR accuracy is improved by more than 7.8 points using the offered strategies. Face recognition increased from roughly 6.9 to nearly 8.2 points when compared to basic interpolation.

(Albluwi et al., 2018) introduced a novel deep learning strategy which resolves super resolution and deblurring from blurred low-resolution photos at the same time. Their suggested model is based on SRCNN, with the goal of learning an edge mapping that receives the degraded LR picture as input and immediately translates it to the HR image. The model was trained on a standard benchmark dataset known as "set 14" which contained around 21 thousand of images and it was tested using two scenarios: blind and non-blind. The findings demonstrated that the proposed model outperforms the existing model in both scenarios (blind and non-blind), with a PSNR of roughly 32.65 in the non-blind scenario and 31.24 in the blind scenario.

According to (Yang et al. 2021), the layers numbers are improved and increased by means of skipping the connection, which increases the overall nonlinear expression ability of the entire model. Moreover, it is noted that the network used becomes difficult to be trained and converge leading to the need for training smaller models. Multiple attention mechanisms (SR) can be used in the training of smaller models of SR to assist in completing the attention of the channel through the mechanisms and special attention mechanisms included in it. The application of small-scale and improved results SR deep networks is gathered from the reconstruction of the SR model that holds multiple attentional mechanisms, according to the various classifications that have been developed on the application of deep networks in the process of image super-resolution. The suggested method's limitations include parameter optimizations and the amount of network layers, as well as an emphasis on pruning the model to decrease the number of parameters in order to obtain an efficient SR model.

To solve the difficulty of losing information in super resolution method, (Li et al.,2020) presented a model called multiscale residual dense network for super-resolution, in which the network is built on residual dense network. Furthermore, this model incorporates network multiscale information and eliminates information loss at the network's depth level. The model was trained on the DIV2K dataset, and the results show that this approach has the greatest restoration capability, and its training efficiency had not been much lowered in three scale feature learning. It can be identified from the result that the quality of picture restoration deteriorated, and the training of algorithm efficiency drops from roughly 4 to 10 hours.

To increase the flow of information and the east training of the model, multi branch modules such as the residual network and GoogLenet have been developed. The residual learning in super resolution model resulted in a shorter training time and higher accuracy. For that, (Hu et al., 2021) developed a new super resolution model comprised of a collection of cascaded subnetworks, a reconstruction network, and a feature extraction network. cascaded architectures were recommended to ease the difficulties of direct super-resolving the details. A set of cascaded subnets were aimed to rebuild HR parameters in a coarse-to-fine form from retrieved depth features. Comprehensive evaluations were carried out to establish the efficiency and efficacy of the suggested approach. The network parameters were shared throughout the cascaded stages in the proposed MSICF model. The benefit of this approach was that it could be used in recognition fields rather than image improvement.

The models of neural network reconstruction performance are highly susceptible to small structural alterations. In order to train neural networks, properly constructed model architecture and advanced optimization algorithms are required. (Lim et al.,2017) proposed a single scale enhanced deep residual network (EDSR) architecture that accommodates a single super-resolution scale. Furthermore, the theory revealed that increasing the number of network parameters is necessary to improve the model's performance. The suggested model train outperformed the competition on the DIV2k dataset as well as the conventional benchmark datasets in terms of PSNR.

Deep learning based SISR has made considerable progress. Regardless of their high performance, these models are difficult to apply to actual applications due to the large number of factors. To tackle with this issue, (Zhang et al., 2019) presented an improved recursive residual network (ERRN) in their research. To minimize parameters, groups convolution and iterative training are adopted based on residual networks. The performance of ERRN is similar to cutting-edged algorithms with many less parameters, according to the results of assessment on test datasets.

To increase the quality of super-resolution performance, the author suggested "Simplified SR-LNN" and "Intern-layer SRLNN" lightweight neural networks with dense connection network and hybrid residual and trained the model using the DIV2K dataset. In comparison to earlier approaches, the methods were developed to generate equivalent pixel density with lowering the number of network parameters. The data might also be used to examine the possibility of denser and skip connections. When a large number of dense connections between convolution layers is implemented, more network parameters are required in the convolution process. Testing results show that the suggested method minimizes the number of factors by nearly 8.12 % and 4.81 %, correspondingly, with maintaining equal picture quality, when compared to previous approaches (Kim et al., 2021).

2.4 Gap in the review of literature

The study identified a series of literature that help the researcher to devise various mediums that will help improve imaging towards high-resolution. However various drawbacks were found in the literature. Some of the Literature is identified that the device algorithm took a long period of time to provide a high-resolution quality image which was 10 times more than the normal algorithm. Identified efficiency challenges as well coupled mapping help in providing accurate results on constant scenarios however in unconstrained low-resolution scenarios performance accuracy was delayed. Another factor that was identified through the research was the lack in discriminative information as well as pose variations. The domain gap between high resolution and low resolution is huge and does it is critical to debate the problem between heterogeneity and homogeneity image resolutions.

Consequently, based on the whole discussion of the various literature that has been considered, the image super-resolution system has shown to be extremely effective in maintaining a clear administration of the many sorts of activities throughout the entire imaging process.

3 Research Methodology

The research approach used to address the research question is outlined in this chapter. The Cross-Industry Standard Process for Data Mining (CRISP-DM) technique is employed to conduct this research. This methodology is an effective strategy for this research project since it follows a framework that starts with analyzing the problem and ends with selecting the best model for this research, with a deployment phase that is planned for the future. In addition, a modified CRISP-DM approach will be followed regarding image recognition for resolution which is essence of the flow of a detailed process. This research study follows the use of diagram shown below.

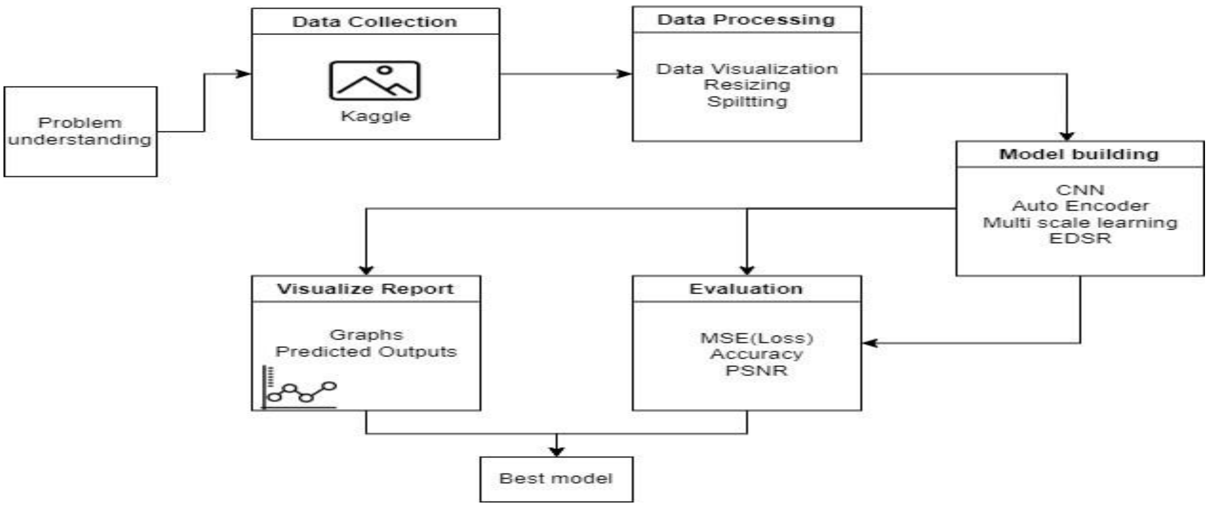

Figure 3-1 research methodology diagram for this study

3.1 Project Understanding

Super-resolution involves up sampling a low-resolution image to generate a high-resolution counterpart and improving the details within an image. “Quality” refers to the way the image presents details contained in pixels, color, shadows, contrast, etc. It is commonly used in the following cases: Satellite and aerial photos, medical imaging, facial recognition devices, sign and number plate reading, and text recognition are all examples of where digital image devices are employed. Because of the capacity of the devices and latent observation in which poor resolution is affected by elements such as illumination, noise corruption, and lighting, the images acquired from these devices contain some deteriorated images. By employing a deep convolution neural network model that can upscale an image while also adding details to increase image quality. In general, increasing the resolution of a low-resolution image would improves image quality. The theory involved, models, various strategies used to achieve the research goal, and relevant dataset will be discussed in this work.

3.2 Data Understanding

The dataset understanding stage from the CRISP-DM model is one of the vital stages in the research study analysis. While all previous research on image super resolution has concentrated on large datasets, this research is based on a relatively small dataset. The dataset is collected from the publicly available source, Kaggle which had one major motive of solving the computer vision issue of dealing with the low-quality image dataset. The dataset can be utilized to improve the resolution of the images.

3.3 Data Preparation

The data preparation stage of the model includes various steps that are necessary to clean and model the data according to the deep learning models requirement based on the convolutional neural network. OpenCV and NumPy has array of libraries that can be used for the data preparation and data analysis to obtain the project research goal.

Accordingly, data pre-processing is performed on the Kaggle dataset and fed to the model, which is explained in the following section.

3.4 Modelling

After separating data into train and test sets and constructing a range of models for the data utilized, the data modelling phase started. The following are four models for picture super resolution that have been implemented for this study. This section provides a summary of the model that was employed, whereas the implemented section of the paper provides detailed information.

3.4.1 Convolutional Neural Network

Deep learning, which is based on a deep convolution neural network, is one of the current solutions for single picture super-resolution. In image processing and classification, CNN has become increasingly used. The CNN model has a variety of advantages, including lowering the number of parameters by using locality qualities and learning hierarchical representations, which allows reducing the error at the network's end. Another benefit of CNN is that it has a fully connected layer that allows the model to take input and output of various sizes, which is then stabilized by network architecture. CNN is used to achieve high level image data interpretation. Learning in neural networks involves weights and biases. CNN's architecture consists of basic building blocks, such as configuration layers / filters, activation functions, pooling layers, and fully connected layers. It is basically produced by stacking a few layers one after the other. In addition, the spatial knowledge of the CNN image is robust enough, with the practical advantage of using fewer parameters, thereby reducing the computational time and data required for model training (Kim P.,2017).

3.4.2 Auto Encoder Model

The auto encoder model uses a latent vector to turn the degraded picture into an improved image and then reassembles the input data with the best image quality. Originally, the Autoencoder model was employed for image compression since it takes original data as input and give outputs as a vector. The decoder then uses this vector to reassemble the original data. Previous learning-based SR approaches focused solely on the strong mapping function for SR, ignoring the image's compositional originality, which revealed small structures or textures. Moreover, this architecture also faces additional challenges, such as limited memory and power. By considering, making use of previous information of an image to direct the upgrading of the deep neural network's output towards super-resolution (Huang et al., 2018).

3.4.3 Multi Scale learning Model

The multi-layer scaling approach is a part of a convolutional neural network that solves numerous learning tasks at the same time., taking advantage of similarities and differences in tasks. It is suggested a multiscale learning neural network with two different convolution channels. Their findings show that in the initial stages of feature extraction, the two layers are independent of one another. The two layers' features are then merged in the subsequent step. The classification capacity of standard CNNs is increased by increasing the number of input channels and the dimensionality of the convolution kernel level. On the other hand, this increases overfitting and computations. This difficulty was avoided by using depth-wise separable convolution in the network. After that, there's a full-connection layer, a max pooling layer, and a feature map. the final features, which served as the input of classifier. As the last layer of classification, SoftMax regression was used. Advantage of this model is that this model is work better when the associated task is learned equally (Wang et al., 2019). Additional advantage of multiscale approaches is that they integrate deep and shallow networks, resulting in novel training methods which continuously expand the depths of the CNN while reusing starting parameters. This study will apply this model with customized function for improving the quality of image by fine-tuning and changing the hyperparameters of this model.

3.4.4 Enhanced Deep Residual Network

To clarify the network model, EDSR is improved by evaluating and eliminating redundant elements as it is based on SRResNet. This model also called as single scale model. Residual blocks have been employed widely in numerous models over the years. By utilizing skip connections, Residual Learning transports feature maps across different convolution layers. During training the network for residual learning, the issue of degradation concerns and sluggish convergence of very deep neural networks. (Lim et al.,2017) suggested a unique model based on the upgraded SRResnet to increase the efficiency of the existing approach. The suggested model training approach transfers information from a model trained at other sizes to a model developed at this scale. Furthermore, the parameters in their suggested model are shared across scales. The advantage of using EDSR is that it improves performance of the model in terms of memory, cost, and execution speed.

3.5 Data Evaluation

This stage based on the CRISP-DM model is considered just after to the model training and testing. This stage is basically cross check and analyze the model parameters for model training, testing and evaluation. This stage helps in selection of the best suited model for the selected dataset research study.

4 Design Specification

The flow in which this research is done is the focus of the Design Specification. The architectural structure is designed to develop a functional framework that helps to determine the resolution of the image.

1. The dataset is selected and gathered based on the requirements of the project.
2. After that, the data is saved to the drive for further processing.
3. The implementation phase is carried out with the help of Google Colab by using GPU server.
4. This research focuses on convolution neural network-based approaches to achieve the research aim to work on deep neural networks.
5. The next step is to import the necessary libraries and do pre-processing, which includes scaling, data visualization, and data separation for train and test data.
6. The implementation of the model is done utilizing four different CNN-based models in keras.
7. The final stage is to evaluate the four models that have been implemented for their predicted images using loss and accuracy and PSNR.

5 Implementation

5.1 Data Collection

This is the most essential phase, where the data is selected based on the research purpose. For image super resolution, standard benchmark datasets such as "DIV2K," "set4", and "set5" are available. According to the literature review done on this area, these datasets contain hundreds of photos, which may demand the use of standard hardware resources, as well as a longer period to train the model which is perhaps more than 5 to 10 days. Considering all of the factors, the data for this study was gathered from Kaggle, which is publicly available on Google. There are three subfolders in this dataset: Train Data, and Validation Data. High_res and Low_res are two subdirectories in the folder that include high-resolution images in High_res and low-resolution photos in Low_res. The dataset contains 800 photos of various area from across the world, which will be utilized for training, validation, and testing which is in the resolution of 256*256 pixel. The image below represents how a dataset looks in real life.

Figure 5-1 Image of the dataset

5.2 Data Pre-processing

Data pre-processing is required for any machine learning technique. It is the procedure for processing raw data for use in a machine learning model. It entails data cleaning, which includes dealing with corrupt data, missing data, deleting unnecessary data via feature selection, and establishing new parameters out from existing information. This collection includes pictures of cars, nature, roads, animals, and a variety of other things. In order to acquire cleaned data for the model, the following processes are used in this study.

Image resizing: The raw data is gathered from Kaggle, which contains images of various sizes. As previously stated, the photos are available in a variety of sizes. As a result, it is transformed to form of 256*256 pixels.

Image Splitting: Training and validation samples should be segregated from the data. On all datasets, the dataset contains a total of 855 pictures, 700 for training, 130 for validation, and 25 for testing the model.

After pre-processing, it is noticed that none of the photos in the dataset are distorted. Exploratory analysis is also not performed on the datasets because there are no classifications.

5.3 Data Modelling

The models for image super resolution that have been implemented and compared in this paper. These implemented models are fine-tuned by altering the number of parameters, as well as adding or removing other layers (up sampling, down sampling, convolution layer, residual block, max pooling). It is not just a case of altering the batch size and kernel. The deep learning models such as EDSR (Enhanced deep super resolution network) model, CNN (Convolutional neural network) model, Auto-encoder model and multi-scale learning model is used for enhancing the quality of the degraded images. The image sized are resized to maintain the uniformity in the data folder as per the requirement. For each model the training data is in the shape of (700, 256, 256, 3), testing data is in the shape of (25, 256, 256, 3) and the validation data is in the shape of (130, 256, 256, 3). The batch size of all the models is considered to be 1 along with the dropout function value of 0.4 and 64 filter layers. The filter layers and dropout layers should be necessary to change due to model overfitting with the usage of training dataset. The below table shows the parameters used for the model building.

Table 1 Model Configuration Table

Model Name	Optimizer	Epochs	Input shape	Batch Size	Trainable parameters
CNN	Adam	30	(256,256,3)	1	1073923
AutoEncoder	Adam	30	(256,256,3)	1	9600206
Multiscale learning	Adam	30	2x (256,256,3)	16	300483
EDSR	Adam	30	(256,256,3)	1	2947331

5.3.1 Convolution Neural Network

It is one of the models which is used as an image super resolution. The basic work done by this model is convolution. It can also be described as Transpose convolution model (deconvolution model). The customized two function is used to enhance the quality of images which is work as a up sampling and down sampling in the model. This model Given some input feature map, attempts to predict the input image which was responsible for creating it. The CNN model contains a down sampling block and an up-sampling block. It conducts convolution and Batch normalization if needed for the down sampling block. Basically, this model performs Up and Down - sampling and offers appropriate precision to the required outputs. The model at the end concatenated all layer to make a 2D convolutional layer. The model use Adam as an optimizer, and the loss is calculated using mean square error. The model has a batch size of one and includes 30 epochs. The model's total trainable parameter is 9,600,206. The result of this model is explained in evaluation section.

5.3.2 Multi-scale Learning

It is basically used for multi-scale resolution of the layer This model can be utilized to dynamically scale a layer to a different possible resolution rather than a fixed one. The Multiscale model can have several branches that, when combined, give us a result for the final layer. Each branch has three layers, each of which is made up of a series of layers, such as a 2D convolutional layer coupled to a dropout layer (This helps in the prevention of overfitting) and then down sampled using a 2D Maxpooling layer. The dropout has set to 0.4. In this model, there are two inputs, each of which performs convolution followed by a dropout before being pooled together. Conv2D-Dropout-MaxPooling2D has three levels, with the number of filters doubling after each layer (32-64-128). The model accepts two images as inputs (left and right) and output as the second argument. The pixel of the image size for both input layers is 256*256. Additionally, this model includes 30 epochs and a batch size of 16. This model's total trainable parameter is 300,483.

5.3.3 Auto Encoder

This Auto-encoder model mainly utilizes residual learning. The residual functional block contains of a convolution layer followed by a batch Normalization layer. The input image is moved through a series of convolution layers before being pooled. The auto - encoder residual is comprised of two inputs to the model: the input image and the decoder value. The input of the decoder is output of the encoder layer. The network is released of the work of learning the redundant information already present in the degraded picture by using residual learning. Residual learning tries to discover only the high-frequency image features and afterwards merge them with the low-resolution image's pre-existing low-frequency image data. The batch size is 1 and the model has a total of 30 epochs.

5.3.4 EDSR

Enhance deep Super-resolution network is based in SRResNet architecture. This model utilizes residual learning in association with a convolution layer. The input layer has a defined size, while the subsequent layers are a fixed set of layers in a series. The model includes one customized function that increases the dimensionality of the dataset. The input image is processed by the convolution layer, which is followed by the Leaky Rectified Linear Unit. This is repeated several times, and other blocks like as pooling and batch Normalization are performed when needed. The output of this layer is down sampled using the Maxpooling layer, and then a 2D convolutional layer with 1D stride is added. Batch normalization is used to normalize the inputs, and the output of this layer is then up sampled and sent to Maxpooling, which sends the output to the 2D convolutional layer, which predicts the image. The input size of this model is also same as three model with same epochs. This model's total trainable parameter is 2,947,331 and its non-trainable parameter is 1408.
The summary of the implemented models is attached in the configuration manual because of the large size of the images.

6 Evaluation

The results of the models implemented on a dataset are presented in this stage. The loss function, accuracy on training and validation data are used to evaluate the results after training the data with the above models. The PSNR is used to evaluate the testing data. PSNR stands for Peak Signal-to-Noise Ratio. A high PSNR score suggests that the image is of good quality, whereas a low value suggests poor image quality. The mathematical function mean square error is used to compute the loss. The loss with the lowest value indicates good model performance, while the loss with the highest value indicates inadequate the performance of the model. The model’s performance is also analyzed using a graph that depicts training and validation accuracy per epoch. According to a survey of literature on the subject, it can be concluded that deep learning-based algorithms benefit from large data as training. In comparison, this study used a small amount of data, consisting of 800 images in total. The result of the models after training data are explained in the next section.

6.1 Convolution Neural Network

Input size of this model is (256x256) and its batch size is 1. After 30 epochs, the CNN model achieves an accuracy of 87.26 % on the training dataset and 84.56 % on the validation dataset and the PSNR value is 29.70 dB as shown in the below table.

Training accuracy	Validation accuracy	PSNR
87.26	84.56	29.70 dB

Table 2 Results of CNN

The graph shows that the loss value of the training dataset decreases as epochs increase, whereas the loss value of the validation dataset begins with low values and fluctuates continuously as epochs increment. As the number of epochs increases, the graph of accuracy on testing data starts with higher accuracy and exhibits some ups and downs. Because of the limitation of testing data, accuracy on testing data is displaying significant fluctuation.

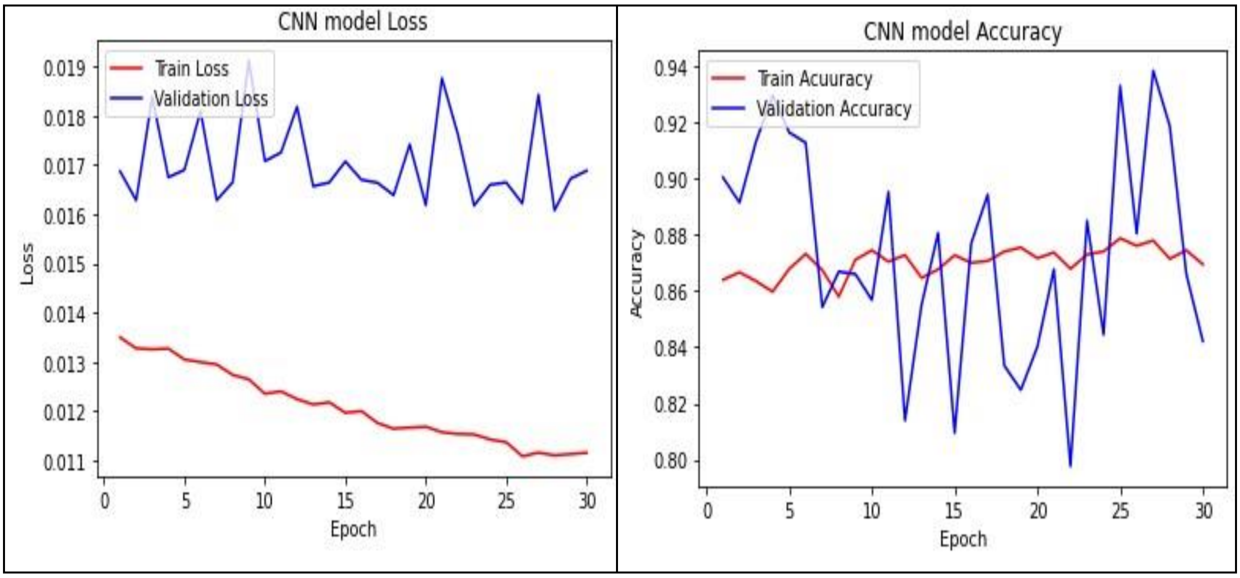

Figure 6-1 graph of accuracy and loss vs. epoch for CNN model

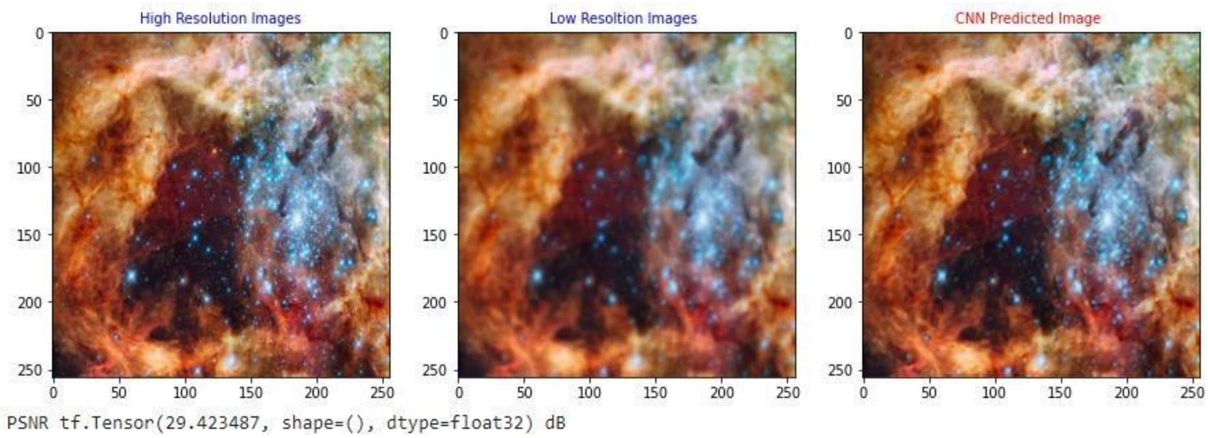

Figure 6-2 predicted results of CNN model

6.2 Auto Encoder

Auto encoder has an accuracy of 85% on training data and 86 % on validation data with the 30-epoch size. The PSNR score of 34.56 dB indicates that the model performed best on image quality.

Table 3 Auto Encoder Results

Training accuracy	Validation accuracy	PSNR
85.17	86.37	34.56 dB

The model loss graph on the training dataset below indicates that it starts at the highest loss value and slightly declines as the epoch increases. Unfortunately, a loss on the testing dataset begins with around 0.02 and rises significantly after 3 epochs, then gradually decreasing as the number of epochs increases. However, as seen in the accuracy graph, accuracy on the training set begins at around 70% and steadily increases until the 5th epoch, after which it progressively rises and ends at around 85%. For the validation set is remains same as CNN model.

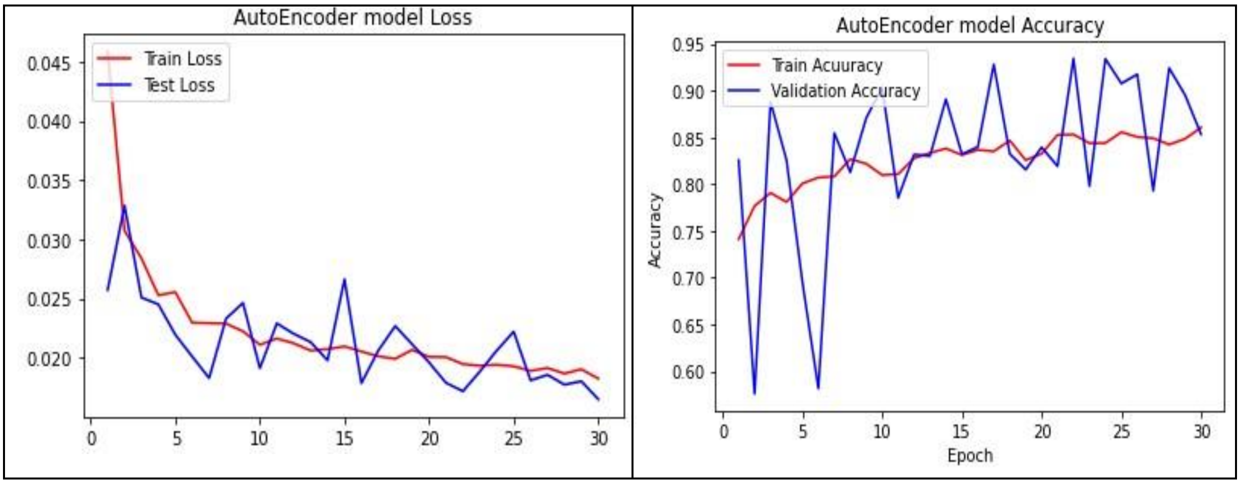

Figure 6-3 graph of accuracy and loss vs. epoch for Auto Encoder model

Figure 6-4Predicted results of Auto Encoder Model

6.3 Multiscale Learning

Multi scale model training using 30 epochs and a batch size of 16 because the loss of this model is quite substantial, therefore increasing the batch size is necessary. This model achieves 84 % accuracy on the training dataset and 86 % accuracy on the validation dataset, with a PSNR value of 22.19 dB for the predicted image.

Table 4 Multiscale results

Training accuracy	Validation accuracy	PSNR
84.59	86.37	22.19 dB

The graph of the multiscale model for loss shows that the loss is quite high for the validation data, that is not an acceptable model result. In terms of accuracy, the model provides estimable performance in both the training and testing datasets. The training accuracy starting low and gradually increases until the final epoch, when this model outperforms the other two models in testing accuracy.

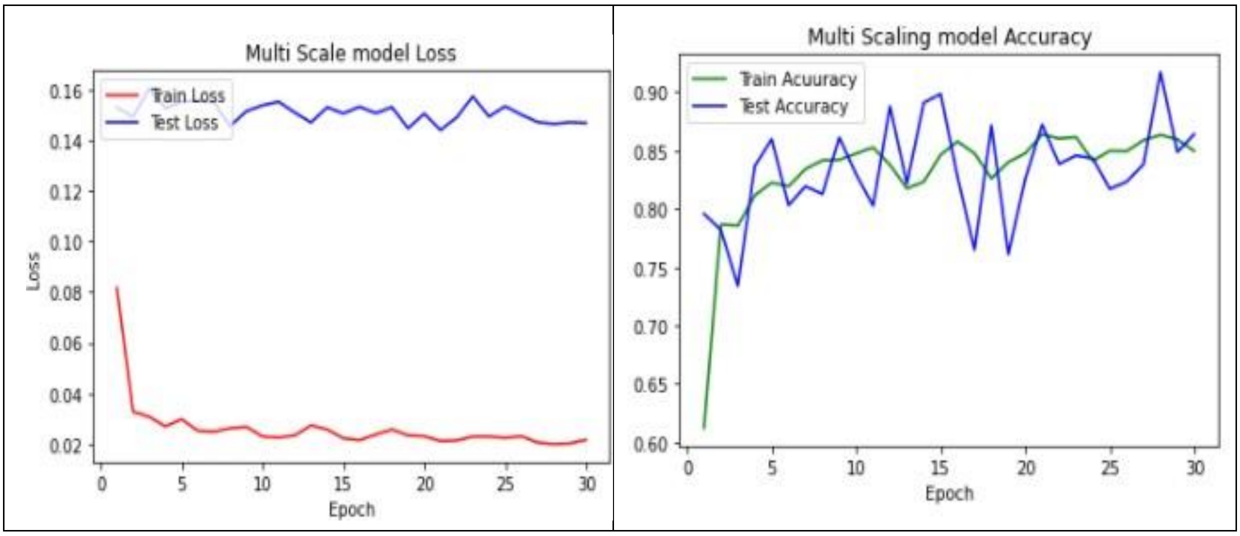

Figure 6-5 graph of accuracy and loss vs. epoch for Multi Scale model

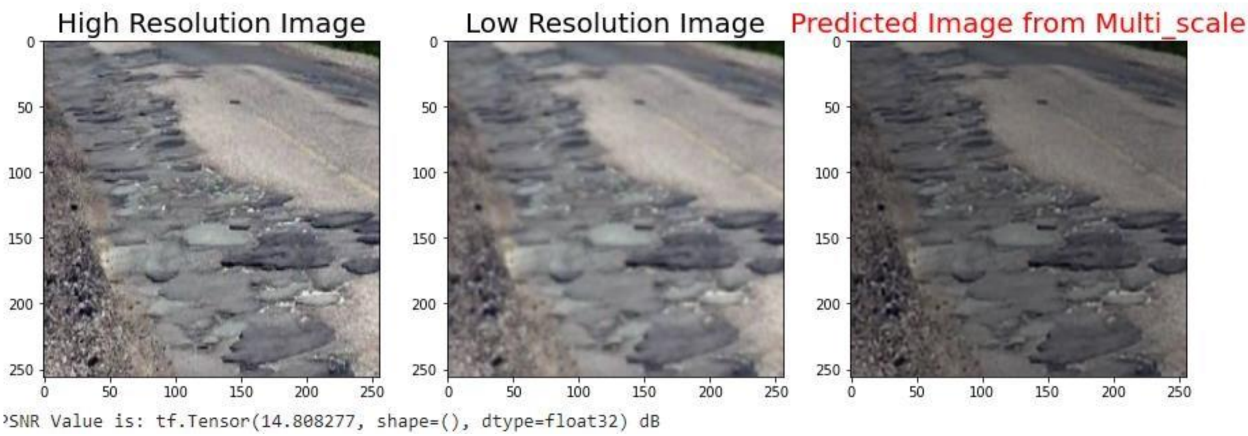

Figure 6-6 Predicted results of Multi Scale Learning

6.4 EDSR

The EDSR model achieves an accuracy of 73.97 % on the training dataset and 69.36 % on the validation dataset after 30 epochs, with a PSNR value of 26.81dB, as stated in the table below.

Table 5 EDSR Results

Training accuracy	Validation accuracy	PSNR
73.97	69.36	26.81 dB

The figures below illustrate that the loss of training and validation datasets decreases with increasing epochs. In terms of accuracy, it began at a minimal level and steadily rose until the third epoch, at which point it increased exponentially until the final epoch. When compared to the total performance of the model, this model provides an excellent performance in terms of accuracy, loss, and PSNR.

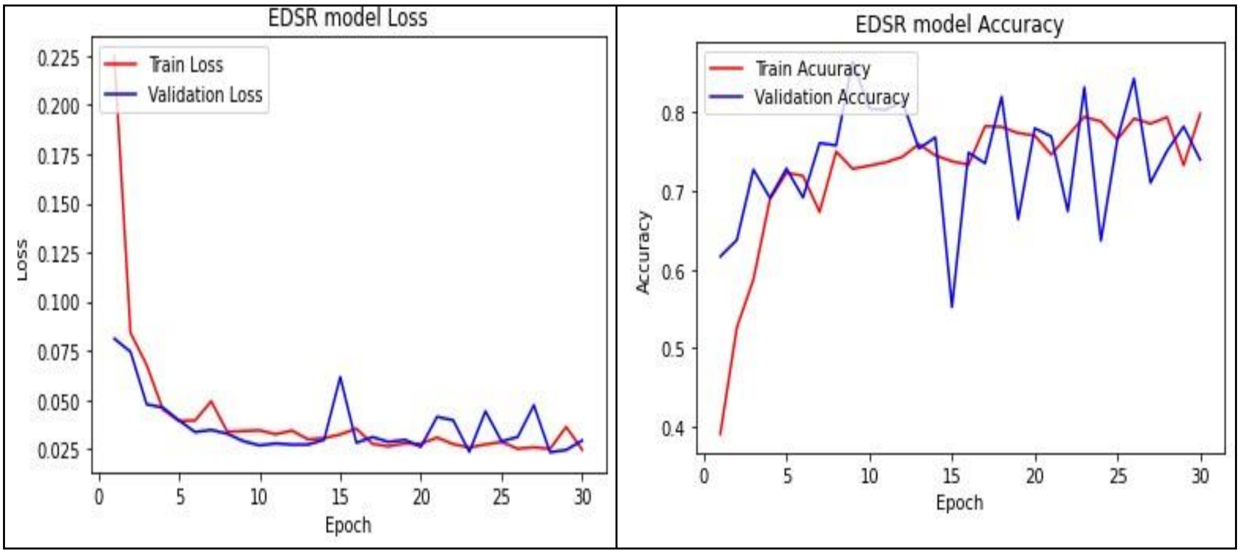

Figure 6-7 graph of accuracy and loss vs. epoch for EDSR model

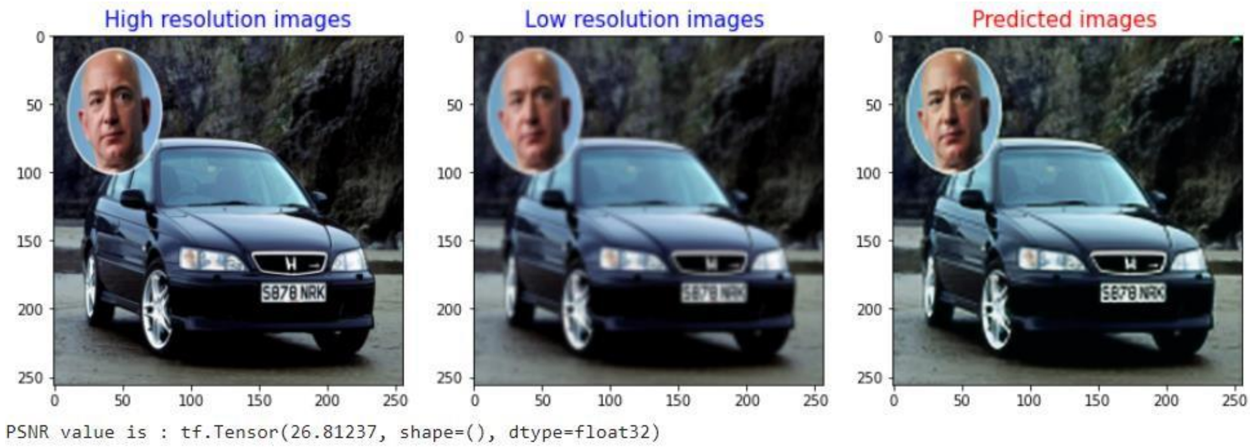

Figure 6-8 Predicted results of EDSR Model

6.5 Discussion

The four models that have been adopted are intended to improve image quality. On a selected dataset, this study presents the quantitative and qualitative assessment findings of four models (CNN, Auto-encoder, EDSR, and Multi scale learning). The accuracy and PSNR criteria are used to evaluate the 130 pictures of the validation dataset. It can be observed from the evolution that the auto-encoder model is working well in terms of accuracy. When it comes to the PSNR value, the EDSR model and the loss value are commendable. The PSNR value of any three predicted images for the four models is shown in the table below. The multi scale learning model achieves the lowest outcome that's because the model's loss is unexpected. Except for multiscale learning, all models have batch size 1. The batch size 1 resulted in model overfitting for multi scale learning when the models were in trial-and-error phase. Below is a table that illustrates how each model measured PSNR for three different images.

Table 6 Comparison of Models

CNN	Multi_scale_learning	Auto-encoder	EDSR
29.52	18.85	34.56	33.45
29.42	14.74	33.72	28.79
27.43	14.87	30.40	28.53

7 Conclusion and Future Work

This research shows how traditional SR approaches could be reformulated into a deep convolutional neural network and comparison of implemented models. The goal of the study was to improve image quality by using single image super resolution models based on convolution neural networks. By tuning the hypermeters of existing models, the CNN, auto encoder, multi scaling learning, and enhanced deep residual network model are built in this research. For that, the number of convolution layers, epochs, batch size, kernel size and kernel number, filter, optimizer, Maxpooling, training speed, classification speed, and accuracy were all considered. A comparatively small dataset is utilized to train the implemented model. According to the evaluation of this paper, the autoencoder model works well in terms of accuracy and PSNR, with an accuracy of roughly 86 % and a PSNR of 34.56 dB. However, the main drawback of this model is the loss value, which is unacceptable when it comes to improving image quality. on the other hand, both multi scale learning and CNN model have higher accuracy with 84% and 86%. However, according to the small dataset size, the EDSR model fits effectively, with a training accuracy of 73 percent and validation accuracy of roughly 70 percent, which is lower than the other three models. Meanwhile, the PSNR value for predicted images is greater than the other models, at 33.45dB.

In the future, this model can be trained on standard benchmark datasets such as "DIV2k," "set4", and "set5" to showcase the performance of model. The implemented CNN and Multiscale learning algorithms have the disadvantage of having a higher loss value. That can be fixed by adjusting the parameter of the model or adding any customized function. The implemented EDSR model makes use of the SRresnet, batch normalization, which require a lot of GPU memory during training time. By eliminating this layer from the network can be reduce the usage of memory. In addition, these models can be tested on real time images rather than just on testing dataset.

Acknowledgement

I am extremely thankful, prof. Vladimir Milosavljevic, for his ongoing guidance in my research project. He was extremely helpful and knowledgeable. He guided me through the documentation process and provided whatever assistance I required. I would also like to thank all of the authors because their contributions in the field of image super resolution benefited me considerably in my research.

References

Gerchberg, R.W., 1974. Super-resolution through error energy reduction. *Optica Acta: International Journal of Optics*, *21*(9), pp.709-720.

Tsai, R.Y. and Huang, T.S., 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on pattern analysis and machine intelligence*, (1), pp.13-27.

Bashir, S.M.A., Wang, Y., Khan, M. and Niu, Y., 2021. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science*, *7*, p.e621.

Qin, B. and Li, D., 2020. Identifying facemask-wearing condition using image superresolution with classification network to prevent COVID-19. *Sensors*, *20*(18), p.5236.

Nasrollahi, K. and Moeslund, T.B., 2014. Super-resolution: a comprehensive survey. *Machine vision and applications*, *25*(6), pp.1423-1468.
C. Y. Yang, C. Ma and M. H. Yang, "Single-image super-resolution: A benchmark", *Proc. Eur. Conf. Comput. Vis.*, pp. 372-386, 2014.

Zhao, H., Gallo, O., Frosio, I. and Kautz, J., 2015. Loss functions for neural networks for image processing. *arXiv preprint arXiv:1511.08861*.

Wang, Z., Chen, J. and Hoi, S.C., 2020. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*.

Dong et al. (2014) Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image superresolution. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer Vision–ECCV 2014. Lecture Notes in Computer Science.* Vol. 8692. Cham: Springer; 2014

Zeng, K., Yu, J., Wang, R., Li, C. and Tao, D., 2015. Coupled deep autoencoder for single image super-resolution. *IEEE transactions on cybernetics*, *47*(1), pp.27-37.

Dong, W., Zhang, L., Shi, G. and Li, X., 2012. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, *22*(4), pp.1620-1630.

Du, X., Qu, X., He, Y. and Guo, D., 2018. Single image super-resolution based on multiscale competitive convolutional neural network. *Sensors*, *18*(3), p.789.

Szegedy, C., Ioffe, S. and Vanhoucke, V., Inception-ResNet and the impact of residual connections on learning [J]. *arXiv preprint arXiv:1602.07261*.

Cao, F. and Chen, B., 2019. New architecture of deep recursive convolution networks for super-resolution. *Knowledge-Based Systems*, *178*, pp.98-110.

Li, S., Fan, R., Lei, G., Yue, G. and Hou, C., 2018. A two-channel convolutional neural network for image super-resolution. *Neurocomputing*, *275*, pp.267-277.

Peyrard, C., 2017. *Single image super-resolution based on neural networks for text and face recognition* (Doctoral dissertation, Université de Lyon)

Albluwi, F., Krylov, V.A. and Dahyot, R., 2018, September. Image deblurring and superresolution using deep convolutional neural networks. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

Yang, X., Li, X., Li, Z. and Zhou, D., 2021. Image super-resolution based on deep neural network of multiple attention mechanism. *Journal of Visual Communication and Image Representation*, *75*, p.103019.

Li, S., Zhao, M., Fang, Z., Zhang, Y. and Li, H., 2020. Image Super-Resolution Using Lightweight Multiscale Residual Dense Network. *International Journal of Optics*, *2020*.

Hu, Y., Gao, X., Li, J., Huang, Y. and Wang, H., 2021. Single image super-resolution with multi-scale information cross-fusion network. *Signal Processing*, *179*, p.107831.

Lim, B., Son, S., Kim, H., Nah, S. and Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136-144).

Zhang, Y., He, X., Jing, M., Fan, Y. and Zeng, X., 2019. Enhanced Recursive Residual Network for Single Image Super-Resolution. In *2019 IEEE 13th International Conference on ASIC (ASICON)* (pp. 1-4). IEEE.

Kim, S., Jun, D., Kim, B.G., Lee, H. and Rhee, E., 2021. Single image super-resolution method using cnn-based lightweight neural networks. *Applied Sciences*, *11*(3), p.1092.
Kim, P., 2017. Convolutional neural network. In *MATLAB deep learning* (pp. 121-147). Apress, Berkeley, CA.

Huang, D., Huang, W., Yuan, Z., Lin, Y., Zhang, J. and Zheng, L., 2018. Image superresolution algorithm based on an improved sparse autoencoder. *Information*, *9*(1), p.11
Wang, D., Guo, Q., Song, Y., Gao, S. and Li, Y., 2019. Application of multiscale learning neural network based on CNN in bearing fault diagnosis. *Journal of Signal Processing Systems*, *91*(10), pp.1205-1217.

Duan, M., Zhang, Y., Li, H., Wang, Y., Fang, J., Wang, J. and Zhao, Y., 2021, September. Learning a Deep ResNet for SAR Image Super-Resolution. In *2021 SAR in Big Data Era (BIGSARDATA)* (pp. 1-3). IEEE

Kune, R., Konugurthi, P., Agarwal, A., Rao, C. R. and Buyya, R. (2016). The anatomy of big data computing, *Softw., Pract. Exper.* 46(1): 79–105.