Performance Analysis of Deep Learning Architectures for Super Resolution

Devi Parvathy Nair, Ann Maria John, E R Amrita Varshini, D Arunima and G Gopakumar*

Department of Computer Science & Engineering, Amrita Vishwa Vidyapeetham, Amritapuri Campus, India

*Corresponding author’s email: gopakumarg@am.amrita.edu

Abstract. Super resolution (SR) being one of the computer vision tasks with increasing applications in modern scenarios, several challenging factors are still prominent despite the numerous breakthroughs achieved in this field in recent years. Introduction of deep convolutional neural networks has brought a booming development to the existing SR techniques tackling many unsolved challenges. As an attempt to perform a relative analysis between currently used methods, this paper explores and establish the capability of Enhanced and Wide super resolution networks. These models are encompassed with improved residual networks with an aim to achieve a higher accuracy with reduced memory usage. The models trained with DIV2K dataset are evaluated using the T91 dataset and found to be showcasing a reliable performance in comparison with other cutting-edge methods devised for super resolution.

1. Introduction

In the domain of image processing, it is always advantageous when the images to be dealt with are of high-resolution (HR) enabling better processing and analysis [1]. HR images are required to enhance the pictorial knowledge so as to have better human observations. Resolution of an image defines its information; hence it is always preferred to have HR images to extract more details from the image. Signal processing methods named super-resolution (SR), have now become a possible way to acquire HR images, given the cost and drawbacks of techniques in enhancing resolution [2]. Image super resolution is the mechanism concerned with using a low resolution (LR) image to retrieve a high-resolution image. The purpose of this is to improve the spatial resolution of images from images of low resolution of the same content. The basic constituent of SR lies in the quality and quantity of additional data that is utilised to recreate HR images [3]. Many studies have proved that LR image is a bicubic downsampled version of corresponding HR image, but in practical applications, effects like blur, noise, etc. can be taken into account for the creation of LR image [4].

Depending on how many LR images are inputted, SR can be of two types: Single Image Super Resolution (SISR) [5] and Multi-Image Super Resolution (MISR) [6]. The objective of SISR is to carry out super resolution using a single LR image [7]. To diminish the mean squared error between the input LR image and its outputted HR image, SISR has drawn attention as an objective function [8]. But this approach has certain disadvantages such that they require information about representation of data in prior [9]. The multi-image approach, which includes non-uniform interpolation method,
iterative back-projection method and standardised SR reconstruction methods, attain the information required with the help of several LR images. However, several LR images cannot be accessible for resolution, hence, in practice, the SISR method is more appropriate than MISR. It is possible to further classify the SISR approach into three categories, which include the interpolation-based [10], reconstruction-based [11], and learning-based methods [12][13]. Interpolation-based methods like bicubic and bilinear interpolation, are the simplest, but multiple specifics are lost, resulting in extremely unfocused corners with unwanted artifacts. Presence of these artifacts were solved using improved interpolation-based methods. Reconstruction-based approaches accomplish super resolution by introducing prior knowledge as constraints to the process of reconstruction of the HR image. Learning-based methods, by producing a visualization form, describe the interrelation of LR and HR images [3].

Recent studies [8][14][15][16] have proved that, when considering structural similarity index method (SSIM) and peak signal-to-noise ratio (PSNR), deep neural networks offer better performance although it showcases limited design optimality. First, the output obtained after reconstructing the model is highly responsive to slight changes in architecture. Second, without understanding and using reciprocal relationships, several existing SR algorithms regard SR of different scale variables as separate challenges [4]. This paper analyses neural network-based models that claims to solve all these problems by introducing ideas of the SRResNet model. Removal of irrelevant modules in the model [4] and training with suitable loss function have outputted better results. Evaluation of these models are performed on benchmark dataset and performances of each model is compared using various quality metrics.

The rest of the paper has been divided into sections with each section covering various fundamental topics. The second section provides a summary on the related works that covers various methods of SISR. The third section details the methodology that has been adopted in the evaluated models. The fourth section details the dataset that has been used to test the models. The fifth section gives a brief summary about the three evaluation metrics that have been used for analysis. The sixth section contains the experimental results that have been obtained by the said models and also provides a detailed analysis of these results by juxtaposing with the existing avant-garde models. The seventh and last section draws a conclusion from the results and also makes some observations about the trends that can be seen in the comparison.

2. Related works
Super Resolution, as a notable computer vision task, has been gaining prominence more than ever and therefore, accurate and economical methods of implementing super resolution are essential. Rasti et. al [17] proposed an SR approach established using Iterative Back Projection (IBP) in the registration stage. An imaging model is formed in which a batch of LR images of different geometric transformations is provided as input. Initially a best guess of an HR image is taken which is usually assumed by taking the average of the batch of LR images. These LR images are interpolated and used to register by utilizing IBP to produce a new HR image which is then used to create a new batch of LR images. The iteration of the imaging process equation occurs until the LR images produced correspond to the observed input images which indicates that the best guessed SR image is the correct SR image. If not, then the error image procured by minimising the difference between the set of LR images gets projected in a voxel-by-voxel basis to the best guessed SR image. In order to test the efficiency of the implemented method, it is compared with wavelet zero padding, conventional bicubic interpolation and the method presented by Irani and Peleg [18]. The LR images are 128x128 and the HR images are 256x256 in size. It shows visual comparison between the techniques as well as the SSIM and PSNR values of the resolution enhancement of different images. The results showcase the advantage this SR technique has over others.

Another interesting application and approach of SISR can be observed in the work of Verga et al [19]. It uses SISR to provide an HR image from LR images produced by Computed Tomography (CT) scanners. This technique takes an input of a 3-D image and improves the input image resolution either
by a factor of 2 or by a factor of 4. Two convolutional neural network (CNN) layers are used for rescaling the 3D input volume on $x$, $y$ and $z$ axes and the second stage CNN layer resizes image on one single axis. The first CNN executes 2D super resolution in a slice-by-slice order expanding an input which has been magnified through the $x$ and $y$ axis by a factor $r$, where $r$ is the scale factor and corresponds to the number of slices. Since this layer produces an output containing the same number of slices as the input that has been provided, a second convolutional layer is required to magnify the image further to $r \cdot h \cdot r \cdot w \cdot r \cdot d$ voxels where $w$ is the width, $h$ is the height and $d$ is the depth. This technique can be proven to be much more systematic and efficient from traditional models by comparing the weights generated. By using 2D convolutional filter networks for 3-D super resolution, 578 weights are obtained. This number is comparatively less than the weights produced by a single 3-D CNN which makes this approach superior. The CNN architecture comprises 10 convolutional layers. It also possesses an intermediate sub-pixel convolutional layer to remove bicubic filters. The intermediate sub-pixel convolutional layer divides the CNN layers into two blocks consisting of 6 and 4 layers respectively. The results yielded by the proposed technique are compared with a series of interpolation baselines and the method of Hornik et al [8]. For upscaling factors of 2 and 4, PSNR, SSIM and Information Fidelity Criterion (IFC) values are computed. Compared to Lanczos method [21], by analysing PSNR, SSIM and IFC values, the method is shown to be superior as it shows an increase of 1.31, 0.0180 and 0.43 respectively. Furthermore, the proposed CNN structure provides superior results to Yu et al’s Generative Adversarial Networks (GAN)-based method [22].

Another one of the advanced super resolution methods, which was designed by C. Ledig et al [8], utilizes GAN. This framework aims to reconstruct SR images with high texture detail using upscaling factors of higher numbers and a perceptual loss function. A GAN made up of a generator that uses residual networks and a discriminator utilizing perceptual loss is implemented. The generator network approximates an HR image when given an LR image as input, while using residual blocks designed with 2 batch normalization layers, a parametric rectified linear activation function (ReLU) that operates as the activation function and 2 convolutional layers. The original HR image and resulting SR image are inputted into the discriminator, a network trained to differentiate between the original HR image and SR image. The network consists of 8 convolutional layers, 2 dense layers, a sigmoid function that calculates the probabilities from classifying the inputted images and a leaky ReLU. Although a majority of super resolution models rely on mean squared error (MSE) as the loss function, it has often presented a problem of overly smoothing images and reducing the amount of texture details in the final SR image due to its inability to use the high frequency particulars of the image. The authors of this paper took a different approach by using a perceptual loss function that is obtained from the weighted sum of adversarial loss and content loss. The content loss is obtained from the loss of the VGG layers, and this is added to the adversarial loss which leads the generator to form SR images that become increasingly similar to the original HR images. The PSNR, SSIM, and Mean Opinion Score (MOS) values were determined by taking 4 as the upscaling factor. The results finalized the conclusion that SRGAN was the optimal solution to the super resolution problem, in terms of its MOS ratings.

3. Methodology
This section discusses already implemented methods, Wide Deep Super Resolution (WDSR) and Enhanced Deep Super Resolution (EDSR), which are designed to minimise memory usage and maximize accuracy. These two characteristics enable these models to outperform other modern SR methods [8][23].

3.1. Residual neural network
A residual neural network (ResNet) is a kind of artificial neural network (ANN) that jumps over the layers by skipping connections [24]. Skip connections are mainly applied to prevent the issue of disappearing gradients and this is done by reutilising activations from the preceding layer before its weights are learned by the adjacent layer. The building blocks of ResNet are known as residual blocks.
These blocks are mainly implemented with skip connections of two or more layers along with convolutional layers, batch normalization layers and Rectified Linear Units (ReLU) activation function. Generally, in neural networks, validation accuracy is directly proportional to the number of layers. However, the number of layers added is restricted since adding layers after a particular saturation point degrades the accuracy. As shallow networks solve the problem of degradation, skipping extra layers in dense layers would match the accuracy of the former. In the modified ResNet model used in EDSR and WDSR, batch normalisation layers from the residual blocks are removed so as to eliminate range versatility and reduce memory usage.

3.2. Enhanced deep super resolution
Enhanced deep super resolution (EDSR) model is a popular novel model that has proved its calibre by winning the NTIRE2017 challenge [25]. The resultant HR image produced by the improved EDSR technique showcases refined edge definition and increase in the quality of image [4]. It has a general CNN architecture with a higher number of feature channels rather than the number of layers. This is done in order to achieve maximum capacity of the model while occupying minimum memory. A residual factor of 0.1, which is its default value, is employed to diminish instability of the model during the process of training while increasing the feature channels. The CNN has been constructed with each layer having 64 filters. It is also equipped with a modified residual network. The residual network varies exponentially from its traditional counterpart as it aims to be as effective and occupy very little memory space at the same time. Therefore, the components of the residual network that are deemed unnecessary and thereby being a waste of memory space are removed. Batch normalisation is removed from the residual block of EDSR as it executes feature normalisation and decreases the flexibility of activations by inducing independency between the features. Moreover, it occupies around 40% of the GPU which is almost as much as the memory space occupied by the entire model. It is detrimental to the model’s working and purpose by reducing its efficiency and hindering its ability to train larger models optimally. Batch normalization also reduces covariance shift which tells the quantity by which the hidden unit values shift around. EDSR utilises a sub-pixel convolutional layer which upscales the model by conducting a pixel-shuffle of the input which rearranges the elements to remove bicubic filters [26]. The weights must be initialized by pre-training the model by the upscaling factor of 2 which will accelerate the process of training. This pre-trained network is used to train the models with of 3 and 4 as the upscaling factors, which will lead to faster convergence compared to a model that starts from random initialization.

3.3. Wide deep super resolution
Expanding functions until ReLU in residual blocks allows more knowledge to move from shallow to deeper layers, preserving the deep neural networks’ non-linearity. The key concept of wide deep super resolution (WDSR) leads to explore successful ways to extend features before ReLU, since it is inefficient to simply add more parameters for real-time picture SR scenarios. The simplest approach towards large activation is to minimise the residual identity mapping path characteristics while extending the characteristics before activation. First, implementation of WDSR-A SR residual network, which has a narrow identity mapping track along with wider channels (2 to 4) until each residual block is enabled. Experiments demonstrate that WDSR-A is highly successful in improving SISR accuracy when the expansion ratio is in between 2 and 4, but a ratio above 4 deteriorates accuracy of the model since the pathway of identity mapping becomes too short. Stable channel numbers of identity mapping pathways are maintained to resolve the above shortcomings and inspect more successful ways to extend features, but produces disappointing results. As a last step, an effective linear low-rank convolution that factorises two low-rank convolution kernels with a large convolution kernel was proposed.

The ideas of wider activation and linear low-rank convolution together constitute the WDSR-B SR residual network [27]. In this implementation, expansion of channel numbers is executed before applying ReLU activation function after the convolution layer. The number of channels on the identity
mapping path and in each of the residual block is cut down to improve the architecture without an escalation in the total number of parameters. As batch normalisation reduces the accuracy of the model during training, weight normalisation is introduced to train the model with a higher learning rate resulting in better performance and convergence. In this paper, analysis of WDSR-B SR residual network is performed.

4. Dataset
To analyse the models [4][27] implemented for super resolution technique and to compare the performance of each model with the latest methods [8][23][28], certain datasets are used. The analysed models were trained using the DIV2K dataset [29], which is a recently developed, high standard dataset that finds applications mainly in image restoration tasks. The dataset provides a diverse set of images ranging from people and objects in their lives, to various environments such as cities, forests, or oceans, and different living organisms like animals and flowers. It comprises 800 training, 100 validation and 100 testing images of LR and SR categories. The neural network models explored in this paper are trained using this DIV2K dataset as it possesses good number of high-resolution images with diverse visual elements. To confirm its interoperability across multiple datasets and to compare the performance of the models, T91 dataset has been chosen for testing. This dataset, which is widely used in super resolution, have a set of 91 images of size 264x204 with diverse visual features. Though its use as a training dataset for a neural network model is limited due to its small size, it is large and versatile enough to evaluate a trained model for super resolution.

5. Image quality assessment metrics
Different deep learning-based SR algorithms can be correlated with the help of image quality assessment methods. Image quality assessment, which determines the typical property of an image, can be used to identify the deterioration and enhancement in the quality of the images after performing certain operations. The quality of an image is generally calculated with respect to the input image known as the reference image. It gives human perception related to the image [30]. After completing the process, some information given by the image features may get disfigured. Hence, the need to assess the quality of images by the human view perception is necessary [31][32]. The majority of researchers in image processing conduct both subjective and objective assessment. Subjective assessments are fast, but not cost-effective. Various elements of the images under analysis are included in the objective assessment, and several techniques and measurements are developed. The most common quantitative measurements used in the image processing field are the PSNR and MSE [33]. In certain cases, without taking structural details into account, quantitative measurements will not be sufficient to determine the quality of the picture. Researchers have therefore developed SSIM that has become very important and effective. SSIM deduces a mean value of structural similarity between the source and target images which has been normalised [34].

5.1. Mean Squared Error (MSE)
MSE is an extensively used image quality metric. Low values which approximate zero indicate high similarity between the source and target images. The MSE is calculated between two images as the average of squared error of corresponding pixel intensities. MSE between two images \( g(x, y) \) and \( g'(x, y) \) is described as [34]

\[
MSE = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} [g'(x, y) - g(x, y)]^2
\]

5.2. Peak Signal-to-Noise Ratio (PSNR)
PSNR is calculated as the fraction of the power of the maximal signal to that of the distorting noise. This value indicates the quality of image description. PSNR is measured with respect to decibel scale logarithmic terms since the signals have an outspread range [35]. PSNR can be expressed as
where \( \text{PSNR} \) is the highest value in the image data. PSNR is an excellent metric for the analysis of results obtained after super resolution. There is an inverse correlation between PSNR and MSE. As quality of an image increases, PSNR increases, and it indirectly indicates that MSE is less. With MSE approaching zero, the value of PSNR approaches infinity; this means that an image of good quality is given by a higher PSNR value.

5.3. Structural Similarity Index Method (SSIM)

SSIM is a common metric of image quality which measures the similitude of the target and source images [36]. The image’s visual quality depends on three parameters: correlation, luminance, and contrast. The general image quality can be influenced by changes in any of these quality parameters, or combinations of them. Therefore, the consistency of the image can be quantified by calculating the possible parameter alterations. This method is known as SSIM, which is defined within the specified range. A value of 0 as the SSIM index indicates the maximal difference between the images compared, and as it reaches 1, there is an increase in the resemblance of the images [37]. The SSIM for source and target images \( m \) and \( n \) respectively is defined as

\[
SSIM(m, n) = \left[ \frac{l(m, n)^\alpha \cdot c(m, n)^\beta \cdot s(m, n)^\gamma}{\sigma_m + \sigma_n} \right]
\]

where \( \alpha > 0, \beta > 0 \) and \( \gamma > 0 \) governs the relative importance of the terms in the equation. \( l \), \( c \) and \( s \) define the luminance, contrast and structure of the images, which are calculated as

\[
l(m, n) = \frac{2\mu_m\mu_n + c_1}{\mu_m^2 + \mu_n^2 + c_1}
\]

\[
c(m, n) = \frac{2\sigma_m\sigma_n + c_2}{\sigma_m^2 + \sigma_n^2 + c_2}
\]

\[
s(m, n) = \frac{\sigma_{mn} + c_3}{\sigma_m + \sigma_n + c_3}
\]

Here, \( \mu_m, \mu_n \) are average values of pixel intensities of \( m \) and \( n \) respectively, and \( \sigma_m, \sigma_n \) are their variances. \( \sigma_{mn} \) denotes the covariance of \( m \) and \( n \). In addition, \( c_1 = (k_1L)^2 \), \( c_2 = (k_2L)^2 \) and \( c_3 = 0.5c_2 \), where \( k_1 = 0.01 \) and \( k_2 = 0.03 \) by default and \( L \) defines the dynamic range of image pixels.

6. Results

The studied model architectures [4][27] have been compared with the following models using the PSNR, MSE, and SSIM values when tested with the T91 dataset.

6.1. SRGAN-GAN

SRGAN-GAN trains a generator from a GAN to take an LR image and contrive an SR image, then feed it along with the HR image into the discriminator. The discriminator identifies the differences and trains the model to produce high-resolution SR images utilizing a perceptual loss calculated from the adversarial loss and content loss [8].

6.2. SRGAN-PRE

Rather than training the generator from a GAN, like in SRGAN-GAN, this model is first pretrained using a pixel-wise loss function, then further optimised using a perceptual loss function [8].

6.3. Bicubic Interpolation

During geometric transformation, interpolation is used for estimation of data at unknown points as there may be discrepancies during pixel transformation due to indirect coincidence of these pixels. Bicubic Interpolation is a non-adaptive algorithm used to create a double resolution image from a two-
dimension source. This method surveys the 4*4 region about the known pixels totalling it to 16 pixels, to determine the output.

6.4. Super Resolution Convolutional Neural Network (SRCNN)
SRCNN [28] is a three CNN layer network: the first layer intervolves the image with a set of basis filters to carry out patch extraction, the second layer involves non-linear mapping which requires mapping of a one-dimensional LR image vector to a two-dimensional HR image vector, and the last layer reconstructs the image.

![Comparison of SR images obtained from various methods with its LR image](image)

**Figure 1.** Comparison of SR images obtained from various methods with its LR image

6.5. Fast Super Resolution Convolutional Neural Network (FSRCNN)
In order to provide acceleration to the SRCNN, the model [23] uses a convolutional layer in place of interpolation, thereby mapping the learning directly to the HR from the LR image. Along with a
deconvolutional layer at the end of the model, as well as usage of smaller filters, FSRCNN applies a shrinkage before mapping and an expansion after it on the input features.

A visual interpretation, as given in Fig. 1, clearly depicts an improved edge definition in case of HR images produced by EDSR and WDSR. On the other hand, by comparing the values given in Table 1-3, we can observe a reasonable range of PSNR scores for EDSR and WDSR ranging higher than those of the other methods analysed, with the best performance being 46.946 dB for EDSR and 46.936 dB in the case of WDSR, showing the finer quality of the SR image reconstructed using the studied models. The MSE scores show a drastically lower range of values for the models we analysed in a range of 3-48 with the worst being 48.892, while the other established methods had MSE values soar up to 310.640 in the worst-case scenario. In the case of the SSIM scores, although the EDSR model performed the best among all other methods, WDSR turned out to showcase slightly worse performance than two of the other models with the worst SSIM of WDSR being 0.95744. In the paper [27] by Yu et. al, WDSR outperformed EDSR when tested with DIV2K dataset [29]. This is assumed to be due to the varied test dataset selection. An overall analysis of the values obtained for the different image quality assessment metrics - PSNR, MSE and SSIM - shows a clear upper hand in the case of the examined models.

### Table 1. PSNR scores obtained for T91 dataset

| Image | EDSR | WDSR | SRGAN-GAN [8] | SRGAN-PRE [8] | Bicubic Interpolation | SRCNN [28] | FSRCNN [23] |
|-------|------|------|--------------|--------------|-----------------------|------------|-------------|
| t1    | 39.769 | 39.633 | 32.557       | 39.246       | 32.809                | 33.318     | 33.524      |
| t2    | 37.072 | 36.977 | 32.026       | 36.777       | 34.006                | 35.206     | 35.723      |
| t3    | 46.296 | 46.521 | 34.818       | 45.223       | 37.838                | 39.260     | 39.730      |
| t4    | 36.373 | 36.402 | 31.672       | 35.973       | 31.577                | 31.334     | 31.490      |
| t5    | 36.100 | 36.009 | 27.979       | 35.624       | 31.577                | 31.943     | 32.176      |
| t6    | 40.482 | 40.136 | 32.414       | 39.724       | 37.313                | 34.872     | 35.255      |
| t8    | 46.946 | 46.936 | 32.995       | 46.025       | 39.760                | 38.741     | 39.026      |
| t58   |       |       |              |              |                      |            |             |

### Table 2. MSE scores obtained for T91 dataset

| Image | EDSR | WDSR | SRGAN-GAN | SRGAN-PRE | Bicubic Interpolation | SRCNN | FSRCNN |
|-------|------|------|-----------|-----------|-----------------------|------|-------|
| t1    | 20.570 | 21.225 | 108.265   | 23.202    | 102.143               | 90.863 | 86.653 |
| t2    | 38.723 | 39.125 | 122.337   | 40.965    | 77.538                | 58.823 | 52.219 |
| t3    | 4.576  | 4.345  | 64.316    | 5.859     | 32.090                | 23.130 | 20.757 |
| t4    | 44.967 | 44.663 | 132.735   | 49.304    | 135.650               | 143.472 | 138.417 |
| t5    | 47.885 | 48.892 | 310.640   | 53.428    | 151.774               | 124.703 | 118.172 |
| t6    | 17.456 | 18.904 | 111.889   | 20.782    | 36.209                | 63.520  | 58.167  |
| t8    | 3.940  | 3.949  | 97.876    | 4.871     | 20.613                | 26.065  | 24.410  |
| t58   |       |       |           |           |                       |        |       |

### Table 3. SSIM scores obtained for T91 dataset

| Image | EDSR | WDSR | SRGAN-GAN | SRGAN-PRE | Bicubic Interpolation | SRCNN | FSRCNN |
|-------|------|------|-----------|-----------|-----------------------|------|-------|
| t1    | 0.99273 | 0.97536 | 0.98484   | 0.99044   | 0.92298               | 0.94788 | 0.95015 |
| t2    | 0.99254 | 0.97360 | 0.98140   | 0.99158   | 0.94121               | 0.97078 | 0.97614 |
| t3    | 0.99418 | 0.97759 | 0.97214   | 0.99259   | 0.96138               | 0.97925 | 0.98173 |
| t4    | 0.98668 | 0.95744 | 0.97738   | 0.98584   | 0.89036               | 0.91644 | 0.92055 |
| t5    | 0.99131 | 0.96857 | 0.98473   | 0.99057   | 0.91935               | 0.96267 | 0.96659 |
| t6    | 0.99450 | 0.98751 | 0.98570   | 0.99382   | 0.98027               | 0.97674 | 0.98096 |
| t8    | 0.99557 | 0.99348 | 0.98439   | 0.99480   | 0.97448               | 0.97505 | 0.97685 |
| t58   |       |       |           |           |                       |        |       |

7. Conclusion
Recent studies have proved that SR techniques rooted on deep neural networks outperform other existing methods in terms of image super resolution. EDSR and WDSR, which use deep neural networks as base, are explored in this paper and are compared with the traditional SR methods to judge its efficiency and precedence. Experimental analysis of the aforementioned SR methods is conducted using pretrained weights (trained on DIV2K dataset) by testing on the T91 dataset. The models’ performances are analysed by contrasting and comparing them with existing deep neural network-based models like SRGAN, FSRCNN, and so on. The image quality assessment metrics used for comparison are MSE, SSIM and PSNR. The results table shows both EDSR and WDSR produce higher PSNR scores, lower MSE scores, and higher SSIM scores in comparison to those of the other super resolution methods observed. However, in case of WDSR, SRGAN-PRE and SRGAN-GAN shows better performance in terms of its SSIM scores. By comparing the derived results, we can subjectively and quantitatively assess that EDSR and WDSR outperform other top-level super resolution methods. Moreover, the modifications in the residual blocks of the models have proven to be advantageous as they induce superior accuracy with minimum memory usage. We believe the work presented helps the readers to visualize the versatility of the EDSR and WDSR models, thereby, encouraging the creation of enhanced SR models in the future.

References
[1] Isaac J S and Kulkarni R 2015 Super resolution techniques for medical image processing 2015 International Conference on Technologies for Sustainable Development (ICTSD) (Mumbai, India: IEEE) pp 1–6
[2] Yang J and Huang T 2017 Image super-resolution: Historical overview and future challenges Super-Resolution Imaging 2013–34
[3] Pan Z, Yu J, Huang H, Hu S, Zhang A, Ma H and Sun W 2013 Super-Resolution Based on Compressive Sensing and Structural Self-Similarity for Remote Sensing Images IEEE Trans. Geosci. Remote Sensing 51 4864–76
[4] Lim B, Son S, Kim H, Nah S and Lee K M 2017 Enhanced Deep Residual Networks for Single Image Super-Resolution 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Honolulu, HI, USA: IEEE) pp 1132–40
[5] Glasner D, Bagon S and Irani M 2009 Super-resolution from a single image 2009 IEEE 12th International Conference on Computer Vision (Kyoto, Japan: IEEE)
[6] Kawulok M, Benecki P, Piechaczek S, Hrynczenko K, Kostrzewa D and Nalepa J 2020 Deep Learning for Multiple-Image Super-Resolution IEEE Geosci. Remote Sensing Lett. 17 1062–6
[7] John N, Viswanath A, Sowmya V and Soman K P 2016 Analysis of Various Color Space Models on Effective Single Image Super Resolution Intelligent Systems Technologies and Applications 384, ed S Berretti, S M Thampi and P R Srivastava (Cham: Springer International Publishing) pp 529–40
[8] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z and Shi W 2017 Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI: IEEE) pp 105–14
[9] Chen Y, Xie Y, Zhou Z, Shi F, Christodoulo A G and Li D 2018 Brain MRI super resolution using 3D deep densely connected neural networks 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (Washington, DC: IEEE) pp 739–42
[10] Fei Zhou, Wenming Yang, and Qingmin Liao 2012 Interpolation-Based Image Super-Resolution Using Multisurface Fitting IEEE Trans. on Image Process. 21 3312–8
[11] Ida, Takashi, Matsumoto, and K. Isogawa. "Reconstruction-Based Super-resolution using Self-congruency of Images.” The Institute of Electronics, Information and Communication Engineers Technical Report, CS2007-52, IE2007-135 (Dec. 2007) (2007): 135-140.
[12] Kim C, Choi K, Lee H, Hwang K and Jong B R 2010 Robust learning-based super-resolution
2010 IEEE International Conference on Image Processing (Hong Kong, China: IEEE)

[13] Thapa D, Raahemifar K, Bobier W R and Lakshminarayanan V 2014 Comparison of super-resolution algorithms applied to retinal images J. Biomed. Opt. 19 056002

[14] Kim J, Lee J K and Lee K M 2016 Accurate Image Super-Resolution Using Very Deep Convolutional Networks 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV, USA: IEEE) pp 1646–54

[15] Kim J, Lee J K and Lee K M 2016 Deeply-Recursive Convolutional Network for Image Super-Resolution 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV, USA: IEEE) pp 1637–45

[16] Jayanarayan A, Sowmya V and Soman K P 2020 Remote Sensing Image Super-Resolution Using Residual Dense Network Soft Computing and Signal Processing 1118, ed V S Reddy, V K Prasad, J Wang and K T V Reddy (Singapore: Springer Singapore) pp 721–9

[17] Rasti P, Demirel H and Anbarjafari G 2014 Iterative back projection based image resolution enhancement 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP) (Zanjan, Iran: IEEE)

[18] Irani M and Peleg S 1991 Improving resolution by image registration CVGIP: Graphical Models and Image Processing 53 231–9

[19] Georgescu M-I, Ionescu R T and Verga N 2020 Convolutional Neural Networks With Intermediate Loss for 3D Super-Resolution of CT and MRI Scans IEEE Access 8 49112–24

[20] Hornik K 1991 Approximation capabilities of multilayer feedforward networks Neural Networks 4 251–7

[21] Duchon C E 1979 Lanczos Filtering in One and Two Dimensions Journal of Applied Meteorology and Climatology 18 1016–22

[22] You C et al. 2020 CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE) IEEE Trans. Med. Imaging 39 188–203

[23] Dong C, Loy C C and Tang X 2016 Accelerating the Super-Resolution Convolutional Neural Network arXiv:1608.00367 [cs]

[24] Gouda N and Amudha J 2020 Skin Cancer Classification using ResNet 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) (Greater Noida, India: IEEE) pp 536–41

[25] Timofte R et al. 2017 NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Honolulu, HI, USA: IEEE) pp 1110–21

[26] Aitken A, Ledig C, Theis L, Caballero J, Wang Z and Shi W 2017 Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize arXiv:1707.02937 [cs]

[27] Yu J, Fan Y, Yang J, Xu N, Wang Z, Wang X and Huang T 2018 Wide Activation for Efficient and Accurate Image Super-Resolution arXiv:1808.08718 [cs]

[28] Li Y, Hu J, Zhao X, Xie W and Li J 2017 Hyperspectral image super-resolution using deep convolutional neural network Neurocomputing 266 29–41

[29] Agustsson E and Timofte R 2017 NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Honolulu, HI, USA: IEEE) pp 1122–31

[30] Oszust M 2016 Full-Reference Image Quality Assessment with Linear Combination of Genetically Selected Quality Measures ed Y Shi PLoS ONE 11 e0158333

[31] Wang Z and Simoncelli E P 2005 Reduced-reference image quality assessment using a wavelet-domain natural image statistic model Electronic Imaging 2005 ed B E Rogowitz, T N Pappas and S J Daly (San Jose, CA) p 149

[32] Lahouliou A, Larabi M C, Beghdadi A, Viennet E and Bouridane A 2016 Knowledge-based Taxonomic Scheme for Full-Reference Objective Image Quality Measurement Models Journal of Imaging Science and Technology 60 604061–6040615
[33] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image Quality Assessment: From Error Visibility to Structural Similarity IEEE Trans. on Image Process. 13 600–12
[34] Sara U, Akter M and Uddin M S 2019 Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study JCC 07 8–18
[35] Deshpande R G, Ragha L L and Sharma S K 2018 Video Quality Assessment through PSNR Estimation for Different Compression Standards IJECS 11 918
[36] John A M, Khanna K, Prasad R R and Pillai L G 2020 A Review on Application of Fourier Transform in Image Restoration 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (Palladam, India: IEEE) pp 389–97
[37] Hor’e A and Ziou D 2010 Image Quality Metrics: PSNR vs. SSIM 2010 20th International Conference on Pattern Recognition (Istanbul, Turkey: IEEE)