**Multi-path Deep CNN with Residual Inception Network for Single Image Super-Resolution**

**Wazir Muhammad**

**ABSTRACT**

Recent research on single-image super-resolution (SISR) using deep convolutional neural networks has made a breakthrough and achieved tremendous performance.

Despite their significant progress, numerous CNNs are limited in practical applications, owing to the requirement of heavy computational cost of the model.

This paper proposes a multi-path network for SISR, known as multi-path deep CNN with residual inception network for single image super-resolution.

In detail, a residual/ResNet block with Inception block support the main framework of the entire network architecture.

In addition, remove the batch normalization layer from the ResNet block and maxpooling layer from the Inception block to further reduce the number of parameters to preventing the over-fitting problem during the training.

Moreover, a conventional ReLU is replaced with Leaky ReLU activation function to speed up the training process.

Specifically, we propose a novel two upscale module, which adopts three paths to upscale the features by jointly using deconvolution and upsampling layers, instead of using single deconvolution layer or upsampling layer alone.

The extensive experimental results on image SR using five publicly available test datasets, which show that the proposed model not only attains the higher score of PSNR/SSIM but also enables faster and more efficient calculations against the existing image SR methods.

For instance, we improved our method in terms of overall PSNR on the SET5 dataset with challenging upscale factor 8x as 1.88 dB over the baseline bicubic method and reduced computational cost in terms of number of parameters 62% by DRCN method.

**Keywords:** image super-resolution, convolutional neural network, deep learning, skip connection, Inception block.

1. **INTRODUCTION**

Image super-resolution plays a vital role in the field of image and computer vision-based applications because the high quality or high-resolution (HR) images have more pixel density level and contains more detailed information.

The detailed information is applied in various fields of computer vision and image processing tasks, such as image restoration [1], security surveillance [2], closed-circuit television surveillance [3] and security systems [4], object recognition [5], object detection [6], satellite imaging [7], remote sensing imagery [8-10], medical imaging [11-15], and atmospheric monitoring [16].

Single image super-resolution (SISR) is a method to reconstruct the visually pleasing high-quality or a high-resolution (HR) output image with rich and clear texture details from the low-quality or degraded version of an input image.

~~However, SISR is a highly ill-posed inverse problem because there are many possible solutions are available and we can recover similar low-quality LR images by downscaling the infinite number of HR images.~~

Traditional SISR methods can be divided into three categories: ~~To overcome these problem~~s, ~~the computer vision researchers community are proposed various approaches and can be classified into three classes:~~ interpolation, -~~based methods~~[17-19], reconstruction, ~~-based methods~~ [20-22], and learning-based methods [23-35].

Implementation of interpolation-based approaches is very simple ~~and easy~~; however, their resultant HR image is prone to blurriness, especially with the large upscale factor.

Furthermore, interpolation-based approaches are limited in applications and suffer from low accuracy.

These approaches are including as bicubic ~~interpolation,~~ bilinear ~~interpolation~~, and nearest-neighbor interpolation techniques.

The reconstruction-based method was specially designed in [36] and introducing prior knowledge to reduce the solution space.

Such types of algorithms can recover details of the sharp edges but rapidly decrease the quality as increase the enlargement factor.

Learning-based approaches or example-based approaches are trying to learn the mappings from millions of co-occurrences of LR to HR example images and then used these learned mapping to reconstruct the desired HR output images.

Recently, deep convolutional neural network-based approaches [37-42] have been obtained significant contributions and ~~significantly~~ increase the progress in the area of image SR tasks, because of their superior capability of the feature representation.

**SRCNN**

First successful shallow type deep learning-based architecture with three CNN layers followed by two rectified linear units (ReLU) is presented by Dong et al. known as a super-resolution convolutional neural network (SRCNN) [37] to solve the SISR problem.

The function of the first CNN layer is used to extract the patches which create the feature mapping information from input images.

The Non-linear mapping is the second layer and its function is to change the feature maps into high dimensional feature vectors.

The function of the final layer is to aggregates the feature maps to reconstruct the HR output image.

**FSRCNN**

To improve the efficiency and speed of SRCNN [37], the same author proposed the faster version known as accelerating the super-resolution convolutional neural network (FSRCNN) [42].

To increase the computational efficiency of the model, Shi et al. [41] introduced the Efficient sub-pixel convolutional neural network (ESPCNN) [41].

**VDSR**

~~Unlike, shallow type network architectures proposed in SRCNN, FSRCNN and ESPCNN.~~  Inspired from the architecture of VGG-net, Kim et al., [39] first time proposed the skip connection based network architecture using the small kernel size of the order (3 × 3), which address the problem of vanishing gradient in the wider and deeper network architectures.

~~in all 20 CNN layers and enlarges the receptive field by increasing the network depth known as VDSR. VDSR [39] extracts the features by global residual learning to ease the training complexity of their network. Moreover Although VDSR [39] has achieved great success, it only extracts single-scale features and ignores the information that is contained in the features at different scales.~~

**LapSRN**

To deal with multi-scale image super-resolution problem, Lai et al. introduced the concept of pyramidal-based network architecture known as the Deep Laplacian pyramid super-resolution network (LapSRN) [43].

This architecture used three sub-branched networks, which can progressively predict the intermediate values of image up to scale factor 8×.

Three basic CNN layers are used to design the whole framework, i.e. the CNN layers, Leaky ReLU [44] layers, and deconvolution layers.

**CARN**

Ahn et al. introduced a new cascading mechanism for the local and global level feature extraction from the multiple layers known as cascading residual network (CARN) [45].

**RCAN**

Inspired from CARN [45], Zhang et al. introduced the concept of residual channel attention network (RCAN) [46].

Although, the deep learning-based image super-resolution research has been greatly improved in the recent decades, but remains a great challenge to capture high-resolution images in some cases such as video security cameras (security surveillance) and human interaction with a computer.

**ResNet architecture**

ResNet architecture first time proposed by He et al. [12] and has been achieved extraordinary performance in the recent decade, due to avoid the vanishing gradient problem during the training.

Though, ResNet still has some challenges and it depends on the Batch Normalization (BN) layer followed ReLU activation function.

BN consumes more training time, because it has requires two times iteration through input data, first for calculating the statistics of batch and second for normalizing its output.

Additionally, the Batch Normalization layer increases the computational cost and more memory consumption. Zhang et al. [47] suggested that BN is not suitable for image super-resolution tasks.

~~Furthermore,~~ Inception blocks borrowed from GoogLeNet, winner of the 2014-ILSVRC competition and main objective of this architecture was to achieve high accuracy with a reduced computational cost [48].

Inception block still faces challenges of max-pooling layer, because it selects the maximum values of the pixel and drops other values of the feature maps.

To address these drawbacks, we suggest Multi-path Deep CNN with Residual Inception Network for Single Image Super-Resolution architecture, namely, MCISIR, which uses the ResNet block without BN layer and Inception block without Maxpooling layer to speed up the feature extraction process as well as reduce the computational complexity of the model.

The extensive quantitative and qualitative evaluations ~~experiments~~ on five benchmark datasets show that our proposed model obtained better perceptual quality as well as reduce the computational cost of the network during the training.  ~~of our proposed model is best as compared to not only achieves higher Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Matrix (SSIM) scores but also enables faster and more efficient calculations against the existing state-of-the-art methods.~~

In summary, in this paper, we establish a novel multi-path deep CNN with residual inception network for single image super-resolution, which yields a noticeable performance in terms of number of parameters, PSNR/SSIM, speed and accuracy.

The main contributions of our proposed method can be summarized as follows:

* Inspired by the ResNet and Inception network architecture, we propose a multi-path deep CNN with Residual and inception network for the SISR method with two upsampling layers to reconstruct the desired HR output images.
* We introduce a new multipath schema to effectively boost the feature representation of the HR image. The multipath schema consists of two layers such as deconvolution layer and upsampling layer to reconstruct the high quality of HR image features.
* Traditional deep CNN methods used the Batch Normalization Layer and MaxPooling Layer followed by ReLU activation function, but our proposed approach removes both Batch Normalization and Maxpooling layers to reduce the computational burden of the model and the conventional ReLU activation function is replaced with LeakyReLU activation function to avoid the vanishing gradient problem during the training.

The remainder section of this paper is organized as follows: Related works are discussed in Section 2, Proposed Network Architecture explained in Section 3. Experimental results and conclusions of the paper are reported in Section 4 and Section 5.

1. **RELATED WORKS**

Single image super-resolution is the key technique to estimate the mapping relationship between low-resolution and high-resolution images.

Recently image super-resolution (SR) has been achieved remarkable attention from the research community.

The main target of image super-resolution is to reconstruct the high quality or high-resolution output image with better perceptual quality and refined details from a given input low quality or low-resolution image.

The image super-resolution is also known as upscaling factor, upsampling process, interpolation, enlargement factor, or zooming process.

Furthermore, image super-resolution has plays dynamic role in the area of digital image processing, machine learning and computer vision-based applications, such as security surveillance videos for face recognition purposes [49], object detection and classification in different scenes [50] especially for small objects [51], astronomical images [52], medical imaging [15], forensics [53] and remote sensing images [54]. ~~Image super-resolution reconstruction methods can be classified into main three categories, i.e., interpolation-based, reconstruction-based, and learning-based. In this paper, we focus on recent deep learning-based image SR methods.~~

* 1. *Deep Learning-Based Image SR*

The rapid developmentof deep convolutional neural networks has made a breakthrough and various methods based on image super-resolution have been proposed by researchers.

**SRCNN**

The pioneering work of image SR is presented by Dong et al. [37], named as SRCNN.

The network architecture of SRCNN [37] consists of three convolutional neural network layers, where each layer is known as feature extraction type layer, non-linear mapping type layer, and reconstruction layer.

The input of SRCNN [37] is used as a bicubic upsampled version of the image, which introduces the extra new noises in the model and adds extra computational cost.

**FSRCNN**

To address this issue and improved the speed as well as the perceptual quality of the LR image, the same author proposed the concept of a Fast Super-Resolution Convolutional Neural Network [42].

The designed network architecture of FSRCNN [42] is very simple and consists of four CNN layers, namely i.e. feature extraction type layer, shrinking layer, non-linear mapping layer, and deconvolution layer.

FSRCNN [42] methods do not use any interpolation technique as a pre-processing step.

**ESPCN**

Shi et al. proposed a fast super-resolution approach that can operate in real-time images and videos known as a sub-pixel convolutional neural network (ESPCNN) [55].

In traditional SR approaches first upscale the LR image to HR image using bi-cubic interpolation and learn the super-resolution model in HR space, due to this increase the computational cost as well as increase the training time.

ESPCN [55] used an alternate approach to extract the features in the LR space and then used the sub-pixel convolution layer at the final stage to reconstruct the HR image.

ESPCN provides competitive results as compared to earlier approaches.

**VDSR**

Unlike, shallow type network architectures proposed in SRCNN and FSRCNN.

Follow the architecture of VGG-net, Kim et al., [39] introduced the fixed-kernel size of the order (3 × 3) in all 20 CNN layers and enlarges the receptive field by increasing the network depth known as VDSR.

VDSR [39] extracts the features by global residual learning to ease the training complexity of their network.

Although VDSR [39] has achieved great success, it only extracts single-scale features and ignores the information that is contained in the features at different scales.

**DRCN**

DRCN [40] proposed a handless deep CNN architecture recursively to share the depth of the network in terms of network parameters.

**LapSRN**

Pyramidal-based network architecture is known as the Deep Laplacian pyramid super-resolution network (LapSRN) [43].

This architecture used three sub-branched networks that progressively predict the value of image up to enlargement factor 8×.

LapSRN architecture used three types of CNN layers i.e. the convolution layers, Leaky ReLU [44] layers, and deconvolution layers.

**DRRN**

DRRN [56] recursively builds two residual blocks and they handle the pre-processing problem caused by interpolation.

**DnCNN**

Zhang et al. [57] proposed a feed-forward denoising convolutional neural networks architecture known as DnCNN, which is very similar to SRCNN architecture and stacks the convolutional neural network layer side-by-side, followed by batch normalization and ReLU layers.

Although the model reported favorable results, their performance is depending on the accuracy of noise estimation and is computationally expensive due to the use of batch normalization layer after every CNN layer.

* 1. *Residual Skip Connection Based Image SR*

**EDSR**

Lim et al. proposed two even deeper and wider networks: an enhanced deep SR network (EDSR) [58] and a multi-scale deep SR network (MDSR) [58], which both consisted of 1000 convolution layers.

These deep SISR networks improve performance by simply stacking the different blocks.

**CARN**

Ahn et al. proposed a lightweight scenario-based architecture known as cascading residual network (CARN) [45].

The basic design of a CARN [45] architecture is used as a cascading residual block, whose output of each intermediate layer is shifted to the consequent CNN layers.

**SRResNet**

Ledig et al. [59] proposed a residual neural network for SR (SRResNet) with more than 100 layers.

They adopted the generator part of the SRGAN as the model structure and employed the residual connections between layers.

**DRDN**

Musunuri et al. [60] introduced the concept of deep residual dense network architecture for single image super-resolution abbreviated as DRDN.

The network architecture based on the combination of residual and dense blocks with skip connections.

In this architecture authors evaluate qualitative performance with new other matrix, like perception-based image quality evaluation (PIQE) and universal image quality index (UIQI).

* 1. *Multi-Branch Based Image SR*

In contrast to linear/single path with skip-connection-based image super-resolution architecture, the multi-branch-based image SR type architecture obtains a different feature at multi-scales.

The resultant multi-path/multi-scale information is then combined to reconstruct the HR image.

**CMSC**

Cascaded Multi-Scale Cross-network architecture known as CMSC, which is composed of three stages: feature extraction stage, cascaded subnets stage, and reconstruction network stage.

**CNF**

Ren et al. proposed a combination of SRCNN in different layers network known as Context-wise Network Fusion (CNF) model [61].

The resultant output of each SRCNN is passed through a single convolution layer and finally fused as a sum-pooling operation.

**IDN**

The Information Distillation Network, abbreviated as IDN proposed by zheng et.al., [62] and used three blocks named: feature extraction, multiple stacked information distillation, and reconstruction type blocks.

**W.M**

Inspired by GoogLeNet [48], W.M et al. [63]proposed an inception-based multi-path approach to reconstruct the HR image.

In this approach, the author used ResNet block and standard convolution operation replaced with asymmetric convolution operation to reduce the computation complexity of the model.

In recent years, attention mechanism-based models achieved attractive performance in various computer vision tasks, such as image reconstruction [64], natural language processing [65], and also for image super-resolution tasks [66-71].

**DAM**

Wang et al. proposed an attention-based densely connected module, abbreviated as DAM.

The design architecture of the DAM network consists of two parts: the first one is the channel attention module and the second is the dense connection block.

Based on DAM block authors proposed a complete name as Attention-based Densely Connected Network (ADSRNet) for single image super-resolution [72].

**RCAN**

Follow the concept of CARN [45] network architecture, Zhang et al. suggested the idea of residual channel attention network, abbreviated as RCAN [46].

In this framework, authors used residual in residual (RIR) type structure, which consists of different groups of residuals long as well as short skip connections.

**DRLN**

Anwar et al. proposed a Densely Residual Laplacian attention Network, known as (DRLN) to resolve the super-resolution images [73].

**LDCASR**

More recently, Zha et al., proposed a Lightweight Dense Connected Approach with Attention to Single Image Super-Resolution (LDCASR) [74], to resolve the redundant and useless information in the dense network architecture.

Furthermore, the authors used a recursive dense group, which is dependent on Dense Attention Blocks to extract the detailed features for reconstructing the HR image.

The application of DenseNet based architecture also more contribute in the area of image super-resolution, specially SRDenseNet [75], in which authors claim that skip connection mitigates the vanishing gradient problem as well as boost up the training performance.

**MemNet**

A persistent memory type network for image SR is known as MemNet, which is proposed by Tai et al. [76].

The MemNet architecture designed is divided into three stages like SRCNN.

The first stage is the feature extraction stage, which extracts the features information from the original input image.

The second stage is to stack the memory blocks in series wise connection. Final stage is the recursive stage which is same as ResNet type architecture.

The MemNet architecture used the MSE as a loss function.

The total number of six memory blocks is used in the architecture.

The network architecture of single image super-resolution for multiple degradations is abbreviated as SRMD proposed by Zhang et al. [77] with different scale factors.

The authors have also introduced the variant version of SRMD known as SRMDNF.

For all enlargement factors, the number of CNN layers is set to be 12 and the feature maps of each layer are 128.

Three types of operations are performed by each layer, including convolution, ReLU, and batch normalization operations.

**DRFN**

Yang et al. [78] proposed the deep recurrent fusion network of SISR with large factors known as DRFN [78].

It consists of three parts: The first part is called joint feature extraction and upsampling, the second is the recurrent mapping of the image in high-resolution feature space, and the final part is the multi-level fusion reconstruction.

For the training purpose, DRFN [78] used the same training dataset which is used by VDSR [39] with data augmentation in terms of rotation and flipping.

**IKC**

The Iterative Kernel Correction (IKC) method for single image super-resolution proposed by Gu et al. [79], which consists of a super-resolution model, predictor model, and corrector model.

In this approach, the author used the principal component analysis approach to reduce the dimensionality of the kernel.

**MFFRnet**

Jin et al. [80] proposed a new framework known as a multi-level feature fusion recursive network abbreviated as MFFRnet [80] for single image super-resolution without pre-processing any scale of the image.

The network architecture of MFFRnet [80] depends on four basic building blocks: coarse feature extraction, recursive feature extraction, multi-level feature fusion, and reconstruction blocks.

Stacking different shallow type network architecture named as HCNN proposed by Liu et al. [81].

HCNN [81] used three types of functional networks for extraction, reinforcement edges, and image reconstruction.

The edge extraction branch consists of 11 CNN layers with 32 kernels of size 3 × 3.

The edge reinforcement network is used 5 CNN layers with 32 kernels of size 3 × 3.

The final branch is the image reconstruction which has 20 CNN layers with 64 kernels of size 3 × 3.

**SCRSR**

Lin et al. [82] proposed a fast and accurate image SR method known as Split-Concate-Residual Super Resolution (SCRSR) [82].

In this approach, the authors used 58 number of layers and increase the receptive field significantly, because receptive field is proportional to image details.

The overall network architecture divided into four parts: input CNN layer, down sampling type sub-network, upsampling type sub-network, and output CNN layer.

**MIRN**

Qiu et al. [83] suggested the multiple improved residual network abbreviated as MIRN [83] for single image super-resolution.

First, they are designed multiple improved residual blocks in the network architecture and the total number of blocks are eight with upsampling blocks.

Stochastic gradient descent (SGD) algorithm is used to train the MIRN [83] network architecture with an adjustable learning rate.

Inspired by these methods, specially ResNet blocks based architecture, we remove the BN layers from the ResNet architecture and the ReLU activation function is replaced with the LeakyReLU activation function, which can reduce the training time and avoid the vanishing-gradient problem during the training.

Furthermore, we remove the max pooling layers from the inception block to efficiently extract the high-level features and improve the reconstructing as well as the visual pleasing quality of the HR image.

1. **PROPOSED METHOD**

In this section, we describe the motivation and design methodology of our proposed model architecture.

Earlier deep learning-based model architectures depend on single linear path and stacked the CNN layers side by to create the deeper network architecture.

Traditional ResNet and Inception blocks, they increase the computational cost and reduce the perceptual quality of the reconstructed SR image.

The design architecture of single path/branch architecture is simple, but it discards the more useful information like edges of the image and other high-frequency features content.

Additionally, batch normalization and maxpooling layer is not best option for image super-resolution techniques.

To solve, these problems, we propose three branch network architectures (Branch1, Branch2, and Branch3) to enhance the feature information, which is named as multi-path deep CNN with residual inception network for single image super-resolution (MCISIR) as shown in Figure 1.

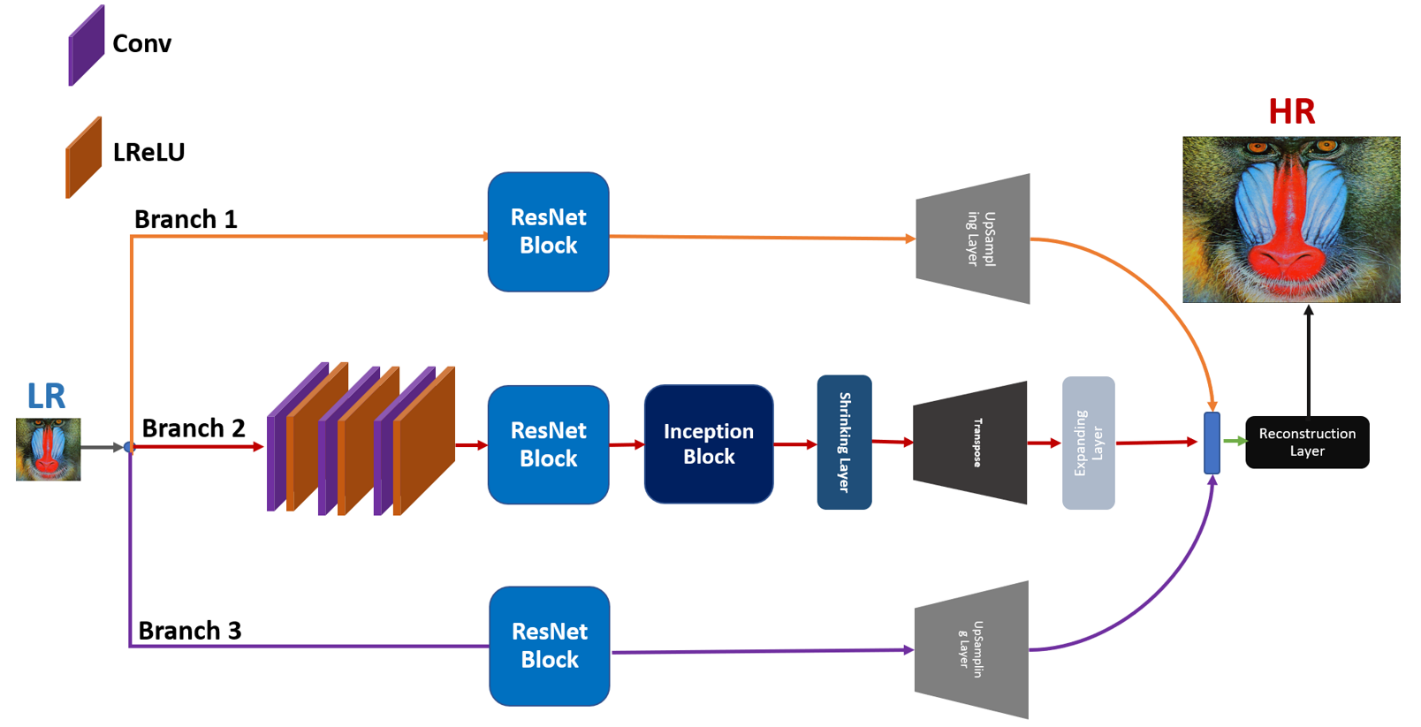


Fig. 1: The proposed network architecture of our method with three parallel paths/branches.

* 1. ***ARCHITECTURE OVERVIEW***

The main purpose of single image super-resolution is to predict the HR image (IHR) from the corresponding LR image (ILR).

Suppose ILR is the low-resolution image followed by an upsampling factor of α to reconstruct the HR image IHR.

Furthermore, HR and LR is the pair of the image with color channels C of ILR and IHR, they can be represented in the tensor of the size as and , respectively.

In this section, we describe the design methodology of our proposed model architecture.

To reconstruct the HR output image, we have proposed a multi-path deep CNN with residual and inception network for single image super-resolution to learn the mapping relationship between the LR and HR images.

The overall network architecture is presented in Figure. 1.

Branch 1 (HRB1) and Branch 3 (HRB3) used only ResNet block with upsampling layer.

Branch 2 (HRB2) used three basic CNN layers to extract the initial low-level features.

The reconstructed basic low-level features are fed to the ResNet blocks followed by Inception block.

For upscaling purpose use the shrinking and expanding layer before and after the deconvolution layer to further reduce the number of model parameters.

In our proposed architecture remove the batch normalization layer from the ResNet block to reduce the memory consumption of GPU, ReLU activation function is replaced with LeakyReLU activation function to avoid the vanishing gradient problem and take out the maxpooling layer from the inception block for best reconstruction of HR image.

Resultantly output HR images of three branches are concatenate followed by a reconstruction layer to generate the HR output image.

* 1. ***FEATURE EXTRACTION***

Following the principle in [84], we used three CNN layers followed by Leaky ReLU [44] of kernel sizes is 3 × 3 with 64 number of channels to reconstruct the feature maps of the main branch (Branch2) feature maps.

The feature maps of these three CNN layers passed through ResNet and Inception blocks to generate the multiscale hierarchical features.

* 1. **Residual Learning Paths**

Earlier approaches are used global residual learning paths with a single CNN layer having a kernel size is bigger than 5 × 5 to extract the low-level features.

The single CNN layer with a bigger kernel size of 5 × 5 is not suitable for low-level feature extraction as well as increases the computational cost of the model.

To overcome this problem, we used a small kernel size of order 3 × 3 followed by upsampled and transposed layer to upscale the LR image.

This type of upsampling strategy improved the accuracy as well as computational efficiency of the model in terms of the number of parameters.

* + 1. ***ResNet Block***

Residual learning [12] is the best way to reduce the computational cost and ease the training complexity.

He et al. [12] first time proposed a ResNet architecture of residual learning for the image classification task.

In [39], Kim et al. proposed a global skip connection to predicting the residual image.

In Figure 2, we compare the building blocks of each network model from the original ResNet [12], SRResNet [59], and our proposed ResNet block.

Original ResNet block as shown in Figure 2(a) used the two layers of convolution, Batch Normalization and ReLU activation after the element-wise addition.

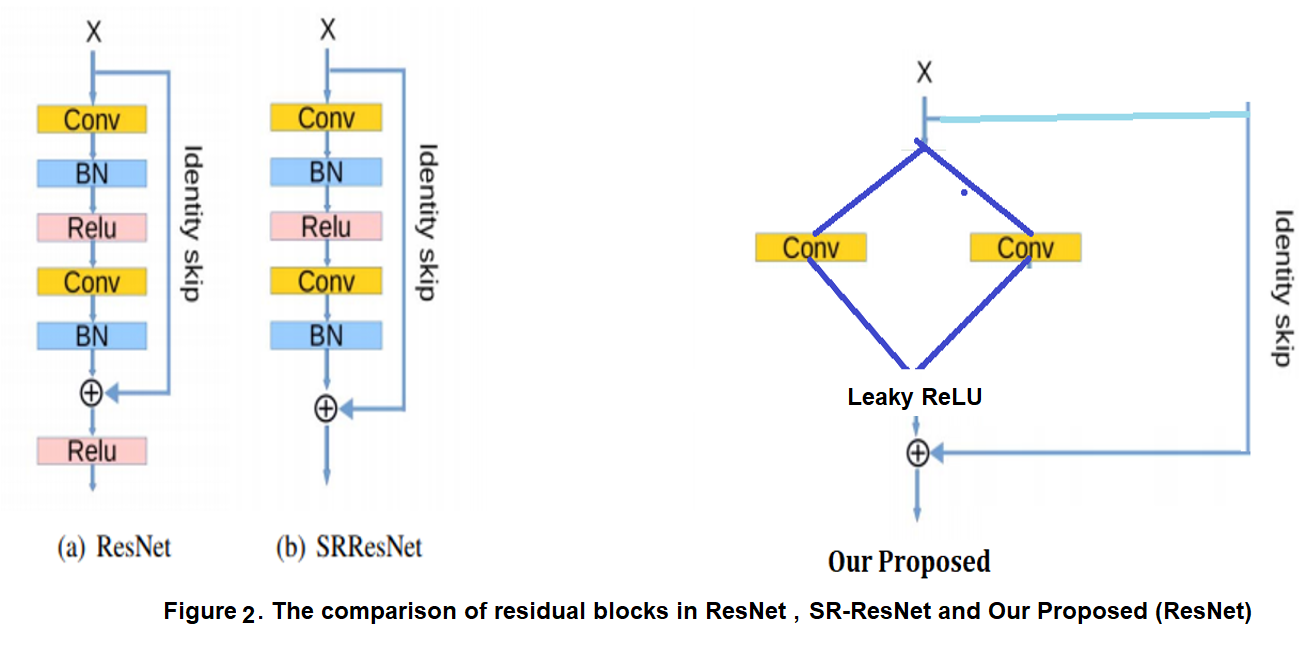
SRResNet [59] block is the modified version of the original ResNet block and removes the ReLU activation layer after the element-wise addition.

For improved performance and numerical stability of the training in SR, we proposed a new design of ResNet Block by removing both batch normalization (BN) layers as suggested by Nah et al. [85], to provide a clean path, because the BN layer is not suitable for the SR task and have a more memory consumption.

Furthermore, in our proposed block original information split into two branches followed two convolution layers parallelly.

The cumulative sum of both convolution layer followed by one common Leaky ReLU Activation function [44].

Leaky ReLU [44] gives better response than ReLU, because it uses a learnable slope parameter instead of a constant slope parameter, which reduces the risk of overfitting in the training.



* + 1. ***Inception Block***

GoogLeNet was the winner of the 2014-ILSVRC competition and the main objective of this architecture was to achieve high accuracy with a reduced computational cost [48].

They introduced the new concept of inception block in CNN, whereby it incorporates multi-scale convolutional transformations using split, transform and merge ideas.

Furthermore, it consists of different parallel convolutional branches with different sizes of the kernel which are then concatenated to increase the width of the network, finally fused the information, respectively.

In the image SR task, most of the earlier approaches used a single kernel size to extract the features for reconstructing the HR image.

However, single kernel size feature extraction is not an efficient way to restore the information completely.

Our proposed block designed is inspired by GoogLeNet [48] architecture to extract the feature information on different kernels to capture better content and structure information from the image.

In our proposed inception block, does not contain the max: pooling layer, because it reduces the ability of the network to learn detailed information, so it is not suitable for image super-resolution tasks.

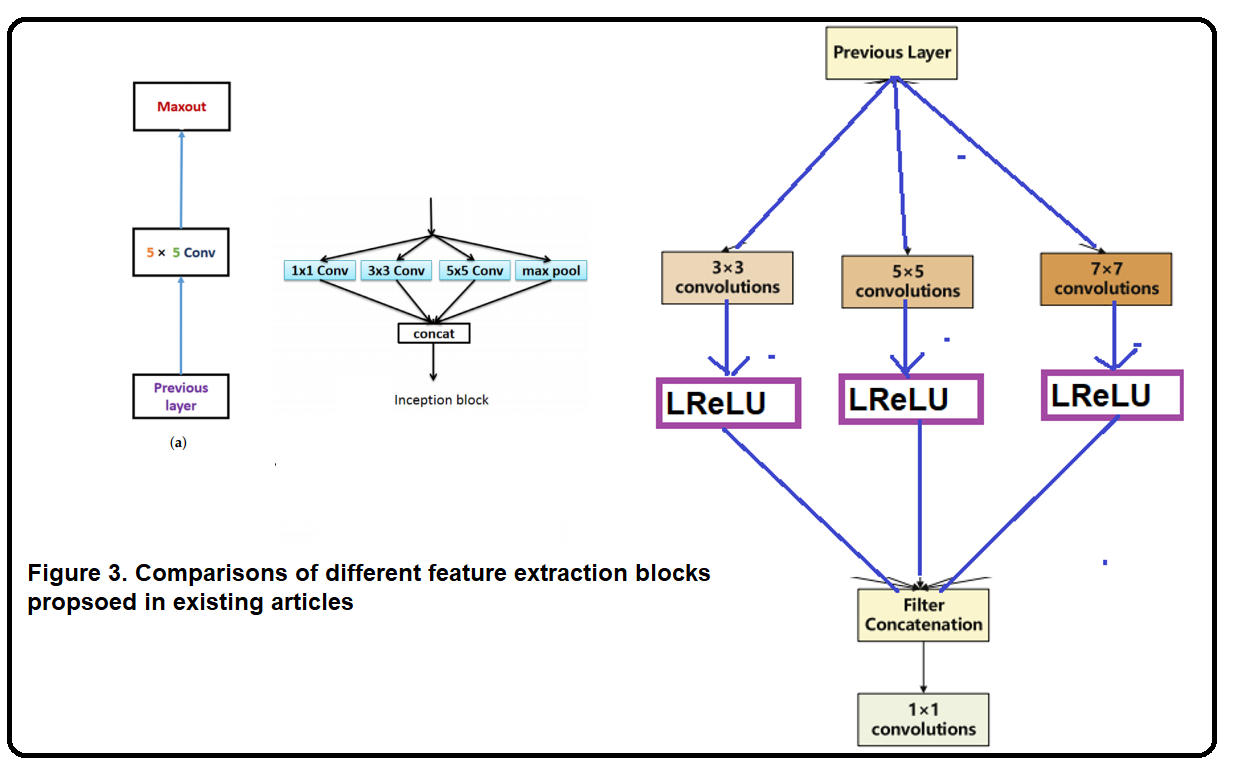


Figure 3 Comparisons of different feature extraction blocks (a) Linear Block, (b) Inception Block, and (c) Our Proposed multi-path Inception Block

Figure 3(a), shows a simple plain network architecture and stacked different CNN layers in a single path.

These types of architecture are used by SRCNN, ESPCN, and FSRCNN SR methods.

The design network architecture of the shallow model is very simple, but it is not suitable for a deeper model architecture and occupies more memory consumption during the training.

Figure 3 (b), shows a conventional inception block to extract the multi-scale feature information.

The problem with this type of block is that it has a higher number of parameters, and so does the higher computational complexity of the model.

Furthermore, these blocks are used the Max pooling layer.

Our proposed block removes the max-pooling layer because the pooling operator considered only the maximum element from the pooling area and ignores other element's information as shown in Figure 3(c).

Proposed block consists of several filters of different sizes. It is used to extract features from the output of the previous layer.

In our network, three scales are utilized, which are 3 × 3, 5 × 5, and 7 × 7 followed by LReLU.

Later, the output of the inception block is mixed in a concatenation layer, and it leads to an increase in the efficiency of the blocks.

* + 1. ***Shrinking Layer***

If a large number of feature maps are directly fed into the deconvolution layer, it will significantly increase the computational cost and size of the model.

The computational complexity and model size will be greatly increased if a large number of feature maps are directly fed into the deconvolution layer.

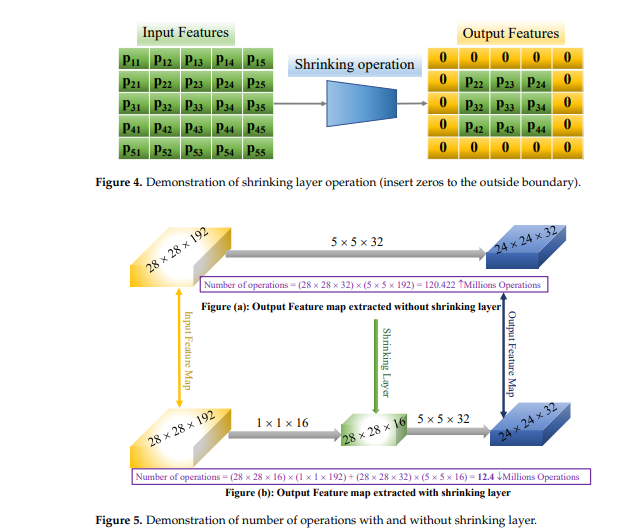
To maintain the model compactness and enhance computational efficiency, we used the bottleneck /shrinking layer, which is a convolution layer having a kernel size of the order 1 × 1 kernel known as bottleneck/shrinking layer [86].

Figure 8 shows the basic operations of shrinking layer to reduce the dimension of the extracted feature maps.

The input feature maps Figure 4 is the order of 5 × 5 and the extracted new feature maps output are the order of 3 × 3, simply insert the zeros to the outside of the boundary.

From computational complexity point of view, we draw a two layers network one is without shrinking layer and other with shrinking layer as shown in Figure 5.

The number of operations in Figure 5(a) is 120.422 Million Operations, very high figure as compared to Figure 9(b) which is 12.4 Million Operation, due to the use of shrinking layer.



* + 1. ***Deconvolution Layer***

Deconvolution layer also called as a [transposed convolutional layer](https://github.com/tensorflow/tensorflow/issues/256#issuecomment-162257789).

The main purpose of this layer is used to upscale the LR image features into HR image features.

The implementation principle of the deconvolution layer as shown in ***Figure 6*** , inspired by [87, 88].

For deconvolution operation the input feature map size is of the order 2 x 2 with kernel size 3 x 3 and reconstructed output is 4 x 4.

In case of convolution operation, the input size is 4 x 4 with kernel is 3 x 3 and reconstructed output size is 2 x 2.

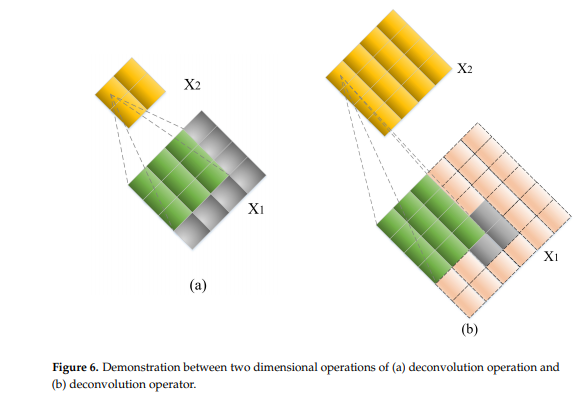
The ***Light blue color*** represented as C1 and ***green color*** represented as reconstructed output as C2.

Furthermore, earlier deep learning-based image super-resolution methods, such as SRCNN [37], VDSR [39], REDNet [89], DRCN [40], and DRRN [56] used an interpolation technique to upscale the input LR image into HR image.

These types of architecture extract the feature information from the interpolated version of the reconstructed image, which introduces the extra new noises in the model and does not achieve better performance as well as increase the computational cost.

Therefore, recent works [24, 63, 89] have introduced the operation of deconvolution layers to learn the upscaling filters and also extract the features detailed of the LR image efficiently.

We added the deconvolution layer at the end of the network because our whole feature extraction process was performed in the LR space.



***Figure 6:*** Demonstration between 2-dimensional operations of (a) Deconvolution operation and (b) Convolution Operation.

* + 1. ***Expanding Layer***

The function of expanding layer is the inverse operation of shrinking layer.

This operation greatly improves the reconstructed quality of the HR image.

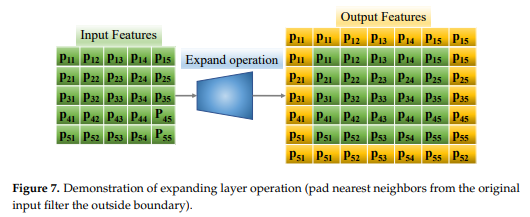
If we reconstruct the HR image directly from the LR features, the restoration quality will be poor.

Generally, shrinking layer reduce the dimension of the 64-channel input into 4 features outputs for upsampling purposes.

After performing the upsampling operation now we recover back original 64 feature map again from the 4-channel input feature map.

For this purpose, we used the expanding layer of kernel size 1×1 followed by LeakyReLU to increase the nonlinearity function.

Furthermore, for detailed explanation as shown in Figure 10, the input feature map is the size of 5 × 5 and pass through the expand operation of convolution layer of the kernel size is 1 × 1 to reconstruct the output feature maps of the order 7 × 7 just padding the nearest neighbor pixels on the outside of the boundaries.



* + 1. ***UpSampling Layer***

To enhance the computational efficiency and reduce memory consumption, we used weight free layer known as UpSampling Layer followed by LReLU activation.

UpSampling layer upscales the features extracted from Branch 1 and Branch 3 through ResNet block followed by common LeakyReLU activation function.

UpSampling layer kernel size depends on the scale factor.

* 1. ***CONCATENATION LAYER***

Earlier approaches [37, 39] uses only a single path to extract the feature information for reconstructing the HR output image.

These types of network architectures are very simple, but they cannot extract the feature information completely and later end layers face severe problems and, in some cases, it works as dead layers.

To resolve said problems we extract the features information from different routes/branches and concatenate it via the concatenation layer.

* 1. ***RECONSTRUCTION LAYER***

In our proposed model the resultant feature maps are used to generate the HR images via the reconstruction layer.

The reconstruction layer is a convolution layer with a kernel size of 3 × 3.

1. **Experiments**
   1. **Training and Testing Datasets**

In our proposed method we combined two datasets of different color images, which are 200 images obtained from BSD200 [90] datasets and 91 images from Yang et al. [35]for training purposes.

The dataset is split using k-fold cross-validation approach and 80% for training and 20% for testing.

To improve the quality of available data for training the model, we used data augmentation techniques such as flipping, rotation, and cropping.

For creating the training and testing datasets in coding we used Keras built in function “image\_dataset\_from\_directory” having main parameters required are crop\_size, upscale\_factor, input\_size and batch\_size.

After that, we rescale the images in the range of [0,1].

Increasing the training efficiency of the model, we convert RGB color image space into the YUV color space.

For the input of low-resolution image data, we crop the image and retrieve the Y channel (luminance) and resize using Bicubic area method obtained from the Pillow, which is the Python imaging library.

In our training model, we only consider on luminance channel in the YUV color space, because humans are more sensitive with luminance change.

During the training we also used callbacks function to monitor our training process with Early Stopping function having patience value is 10.

In the testing phase, five standard publicly available benchmark test datasets, Set5 [91], Set14 [92], BSDS100 [93] , Urban100 [94] and Manga109 [95] are used.

The number of images in said datasets is 5, 14, 100, 100, and 109 images, respectively.

Each of five benchmark datasets exhibit different characteristics. Set5 [91], Set14 [92] and BSDS100 [93] consists of natural scenes.

Urban100 [94] contains challenging urban scene images with details in various frequency bands.

Finally, Manga109 [95] also known as Japanese comic images is the class of multimodal type of artwork, which is collected from Japanese Manga.

* 1. **Implementation Details**

Training the deep CNN architecture we used the Adam optimizer [96] rather than Stochastic Gradient Descent because SGD is extremely time-consuming.

The initial learning rate is set to be 0.0001.

The experimental setup was performed on Windows 10 operating system.

The deep-learning framework used included (Keras, Tensorflow, OpenCV and MATLAB R2017a), CUDA Version 10.2, Python 3.7, and an NVIDIA GeForce RTX 2070 GPU.

During the training process, curves of training loss and test loss initially decrease rapidly, but after some epochs loss decrease gradually as number of epochs greater than 50 as sown in Figure 9.

In Figure 10, the accuracy of test data set has been increased as the training epoch improved.

We will observed that best results can be obtained by increasing the number of epochs and providing a longer training time.

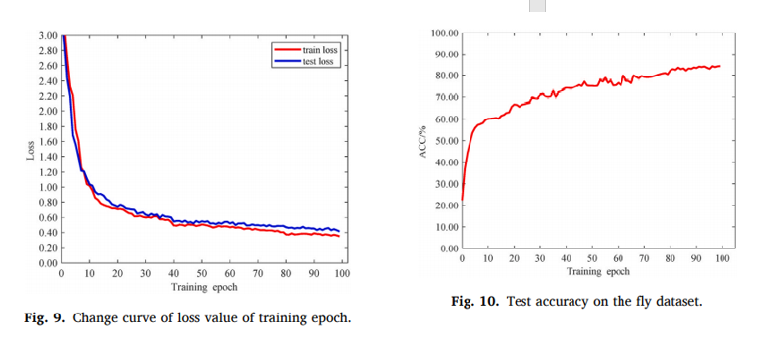


Figure: Traing and Testing Loss

* 1. **Comparison with other State-of-the-Art Methods**

The image quality matrix of PSNR / SSIM is the most popular reference quality metric, which is widely used in the image SR tasks, and they directly apply to the intensity of the image.

We evaluate the performance of our proposed model MCISIR on the five benchmark test datasets with challenging enlargement scale factors 4× and 8×.

For quantitative comparison point of view, we selected thirteen different state-of-the-art algorithms, along with the baseline.

As seen from Table 1, our method achieves, on average better PSNR/SSIM than other existing state-of-the-art methods.

Furthermore, our model can improvement overall PSNR on SET5 dataset with challenging upscale factor 8× as 1.88 dB, 0.75 dB, 0.90 dB, 0.79 dB, 0.69 dB, 0.53 dB, 0.95 dB, 0.68 dB, 0.35 dB, 0.35 dB, 0.14 dB, 0.10 dB, and 0.12 dB, Bicubic, A+, RFL, SelfExSR, SCN, ESPCN, SRCNN, FSRCNN, VDSR, DRCN, LapSRN, DRRN, and MemNet, respectively.

Furthermore, the performance of the image super-resolution model also correlates with the network depth.

The deeper model performed better than the shallow model proposed by Kim et al. [39].

However, the deeper model has more parameters as a compared shallow model. Table 2 shows that our proposed model has a smaller number of parameters as compared to VDSR, DRCN, LapSRN, and MemNet, due to the multi-branch approach.

In this approach, we used a combination of ResNet with Inception block followed by Leaky ReLU learning strategy, which greatly reduces the computational cost in terms of model parameters.

**TABLE 1. Comparison of Average PSNR/SSIM of different image super-resolution models on scale factor 4× and 8×. The result bold with the red color indicated the first best and underline with blue color is the second-best performance of the model.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **Scale** | **#Parameters ↓** | **SET5 [91]**  **PSNR↑/SSIM↑** | **SET14 [92]**  **PSNR↑/SSIM↑** | **BSDS100 [93]**  **PSNR↑/SSIM↑** | **URBAN100 [94]**  **PSNR↑/SSIM↑** | **MANGA109 [95]**  **PSNR↑/SSIM↑** | **Average**  **PSNR ↑/SSIM ↑** |
| Bicubic [19] | 4× | -/- | 28.43 / 0.811 | 26.01 / 0.704 | 25.97 / 0.670 | 23.15 / 0.660 | 24.93 / 0.790 | 25.70 / 0.727 |
| A+ [97] | 4× | -/- | 30.32 / 0.860 | 27.34 / 0.751 | 26.83 / 0.711 | 24.34 / 0.721 | 27.03 / 0.851 | 27.17 / 0.779 |
| RFL [98] | 4× | -/- | 30.17 / 0.855 | 27.24 / 0.747 | 26.76 / 0.708 | 24.20 / 0.712 | 26.80 / 0.841 | 27.03 / 0.773 |
| SelfExSR [94] | 4× | -/- | 30.34 / 0.862 | 27.41 / 0.753 | 26.84 / 0.713 | 24.83 / 0.740 | 27.83 / 0.866 | 27.45 / 0.787 |
| SCN [38] | 4× | 42 k | 30.41 / 0.863 | 27.39 / 0.751 | 26.88 / 0.711 | 24.52 / 0.726 | 27.39 / 0.857 | 27.32 / 0.782 |
| ESPCN [41] | 4× | 20 k | 29.21 / 0.851 | 26.40 / 0.744 | 25.50 / 0.696 | 24.02 / 0.726 | 23.55 / 0.795 | 25.74 / 0.762 |
| SRCNN [37] | 4× | 57 k | 30.50 / 0.863 | 27.52 / 0.753 | 26.91 / 0.712 | 24.53 / 0.725 | 27.66 / 0.859 | 27.42 / 0.782 |
| FSRCNN [42] | 4× | 12 k | 30.72 / 0.866 | 27.61 / 0.755 | 26.98 / 0.715 | 24.62 / 0.728 | 27.90 / 0.861 | 27.57 / 0.785 |
| VDSR [39] | 4× | 665 k | 31.35 / 0.883 | 28.02 / 0.768 | 27.29 / 0.726 | 25.18 / 0.754 | 28.83 / 0.887 | 28.13 / 0.804 |
| DRCN [40] | 4× | 1775 k | 31.54 / 0.884 | 28.03 / 0.768 | 27.24 / 0.725 | 25.14 / 0.752 | 28.98 / 0.887 | 28.19 / 0.803 |
| LapSRN [43] | 4× | 812 k | 31.54 / 0.885 | 28.19 / **0.772** | 27.32 / 0.727 | 25.21 / 0.756 | 29.09 / 0.890 | 28.27 / 0.806 |
| DRRN [56] | 4× | 297 k | 31.68 / 0.888 | 28.21 / **0.772** | 27.38 / 0.728 | 25.44 / **0.764** | 29.46 / **0.896** | 28.43 / **0.810** |
| MemNet [76] | 4× | 677 k | 31.74 / **0.889** | 28.26 / **0.772** | 27.40 / 0.728 | 25.50 / 0.763 | 29.42 / 0.894 | 28.46 / 0.809 |
| MCISIR (Our) | 4× | 443 k | **31.77** / **0.889** | **28.29** / **0.772** | **27.43** / **0.729** | **25.54** / **0.764** | **29.48** / **0.896** | **28.50** / **0.810** |
| Bicubic [19] | 8× | -/- | 24.40 / 0.658 | 23.10 / 0.566 | 23.67 / 0.548 | 20.74 / 0.516 | 21.47 / 0.650 | 22.68 / 0.588 |
| A+ [97] | 8× | -/- | 25.53 / 0.693 | 23.89 / 0.595 | 24.21 / 0.569 | 21.37 / 0.546 | 22.39 / 0.681 | 23.48 / 0.617 |
| RFL [98] | 8× | -/- | 25.38 / 0.679 | 23.79 / 0.587 | 24.13 / 0.563 | 21.27 / 0.536 | 22.28 / 0.669 | 23.37/ 0.607 |
| SelfExSR [94] | 8× | -/- | 25.49 / 0.703 | 23.92 / 0.601 | 24.19 / 0.568 | 21.81 / 0.577 | 22.99 / 0.719 | 23.68 / 0.634 |
| SCN [38] | 8× | 42 k | 25.59 / 0.706 | 24.02 / 0.603 | 24.30 / 0.573 | 21.52 / 0.560 | 22.68 / 0.701 | 23.62 / 0.629 |
| ESPCN [41] | 8× | 20 k | 25.75 / 0.673 | 24.21 / 0.510 | 24.37 / 0.527 | 21.59 / 0.542 | 22.83 / 0.671 | 23.75 / 0.585 |
| SRCNN [37] | 8× | 57 k | 25.33 / 0.690 | 23.76 / 0.591 | 24.13 / 0.566 | 21.29 / 0.544 | 22.46 / 0.695 | 23.39 / 0.617 |
| FSRCNN [42] | 8× | 12 k | 25.60 / 0.697 | 24.00 / 0.599 | 24.31 / 0.572 | 21.45 / 0.550 | 22.72 / 0.692 | 23.62 / 0.622 |
| VDSR [39] | 8× | 665 k | 25.93 / 0.724 | 24.26 / 0.614 | 24.49 / 0.583 | 21.70 / 0.571 | 23.16 / 0.725 | 23.91 / 0.643 |
| DRCN [40] | 8× | 1775 k | 25.93 / 0.723 | 24.25 / 0.614 | 24.49 / 0.582 | 21.71 / 0.571 | 23.20 / 0.724 | 23.92 / 0.643 |
| LapSRN [43] | 8× | 812 k | 26.14 / 0.738 | 24.35 / 0.620 | 24.54 / 0.586 | 21.81 / 0.581 | 23.39 / 0.735 | 24.05 / 0.652 |
| DRRN [56] | 8× | 297 k | 26.18 / 0.738 | 24.42 / 0.622 | 24.59 / 0.587 | 21.88 / 0.583 | 23.60 / 0.742 | 24.13 / 0.654 |
| MemNet [76] | 8× | 677 k | 26.16 / 0.741 | 24.38 / 0.619 | 24.58 / 0.584 | 21.89 / 0.582 | 23.56 / 0.738 | 24.11 / 0.653 |
| MCISIR (Our) | 8× | 443 k | **26.28** / **0.743** | **24.93** / 0.625 | **24.61** / **0.589** | **21.90** / **0.584** | **23.62** / **0.745** | **24.27** / **0.657** |

**TABLE 2. Some CNN-Based SR Algorithms are compared. The depth stands for how many numbers of layers are in the network.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Method** | **Network input** | **# Parameters (K)** | **Depth** | **Filters** | **Reconstruction** | **Loss function** |
| SRCNN [37] | LR + bicubic | 57 | 3 | 64 | Direct | *l2* |
| VDSR [39] | LR + bicubic | 665 | 20 | 64 | Direct | *l2* |
| DRCN [40] | LR + bicubic | 1775 | 20 | 256 | Direct | *l2* |
| LapSRN [43] | LR | 812 | 27 | 64 | Progressive | *l1* |
| MemNet [76] | bicubic | 677 | 80 | 64 | Direct | *l2* |
| MCISIR (Our) | LR | 443 k | 26 | 64 | Direct | *l2* |

Chart, scatter chart

Description automatically generated

Figure 4: Performance comparisons in terms of PSNR versus Number of model parameters on SET5 and URBAN100 enlargement factor 8×.

The tradeoff between the performance and size of the model is shown in Figure 4. Both results are performed on the Set5 and URBAN100 datasets for challenging enlargement 8× scale factor.

We can observe that our MCISIR model outperforms the existing state-of-the-art methods.

For example, our MCISIR achieves much better performance than VDSR, MemNet, LapSRN, and DRCN on scale factor 8×, with the number of parameters is being reduced by 33%, 35%, 45%, and 62%, respectively.

In Figures 5 and 6, noticed that our proposed model achieved better PSNR/SSIM on all public test datasets at challenging scale factor 8×.

Chart

Description automatically generated

Figure 5 Peak signal to noise ratio versus different algorithms on enlargement scale factor 8×.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Figure 6 Plot the PSNR/SSIM of all publicly available test image datasets versus different algorithms on enlargement scale factor 8×.

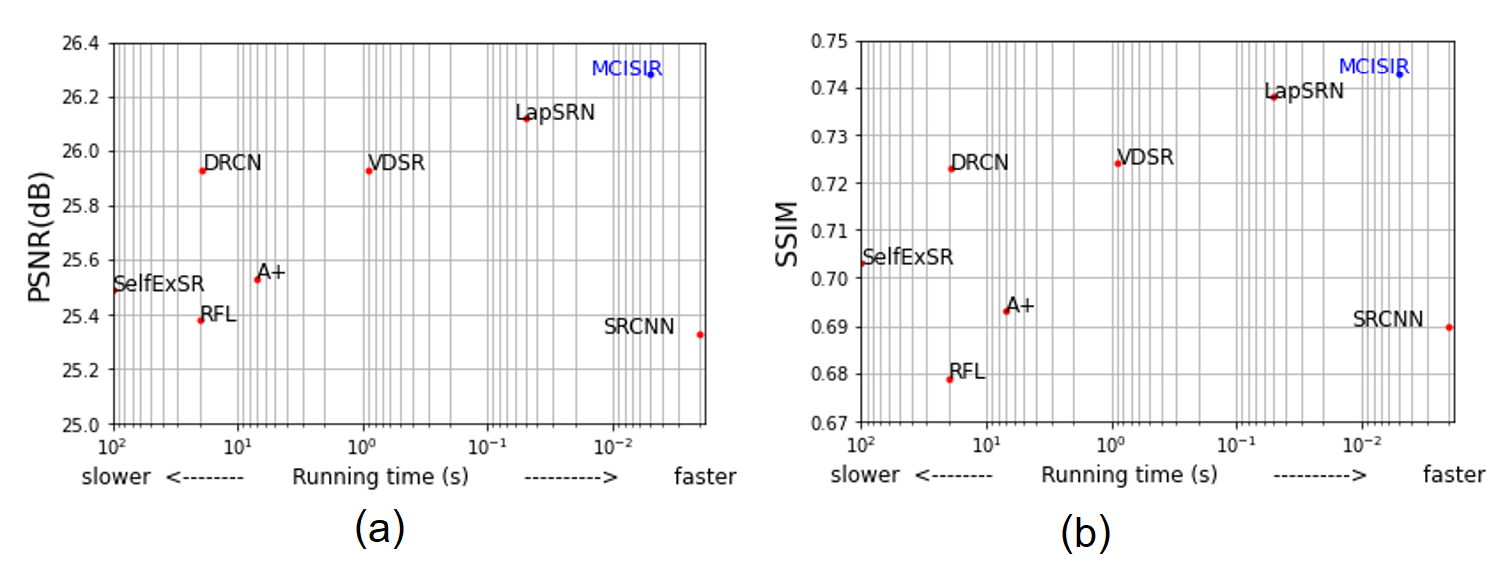


Figure 6: performance comparison in terms of execution time versus PSNR / SSIM on Set5 dataset with enlargement scale factor 8x.

Finally, we compare the performance of running time with the other existing state-of-the-art methods.

We evaluate the performance on Set5 dataset. Quantitative average value of PSNR/SSIM of our proposed method higher and processing time is at near to faster level.

To further evaluate the perceptual quality of our proposed model with recent state-of-the-art methods as shown in Figures 7 and 8.

In Figure 7, we present the visual comparison performance of different approaches on baboon image obtained from publicly available dataset Set14 with enlargement factor 8×.

Upscaled region of the image indicated by a rectangle with red color, where high chances of texture expectation.

In the case of the Bicubic interpolation technique, hair present on the baboon beard fails to resolve the textures and generated a highly blurred output.

The VDSR, DRCN, DRRN, and LapSRN approaches produce better texture results as compared to the Baseline method, but still, results are largely blurry.

In our proposed model reconstruct the detailed texture details around the beard hair of the baboon with any prominent artifacts.

Similar effects are observed in Figure 8.

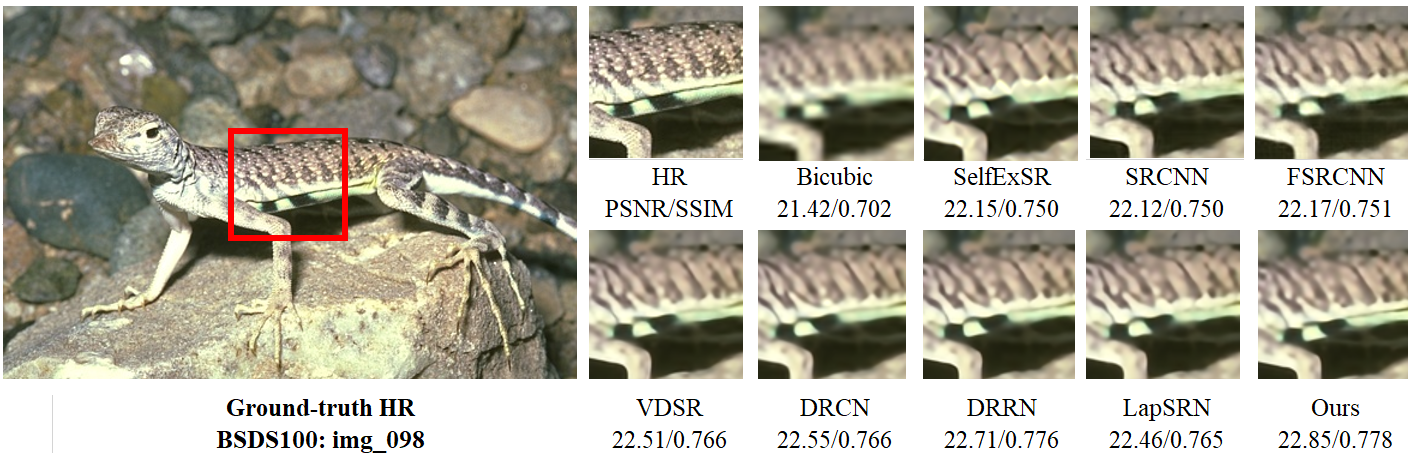
From a comparison point of view, our method reduces the effects of edge bending and reconstructs the high-frequency details efficiently.

This is because of the multi-path arrangement of network architecture to reconstruct the HR image. The above results are verifying the superior performance of our MCISIR especially with fine texture details of the reconstructed image patch.





Figure 7. Presents the visual quality performance comparison for 8× image SR methods on Set14 and Set5 datasets.



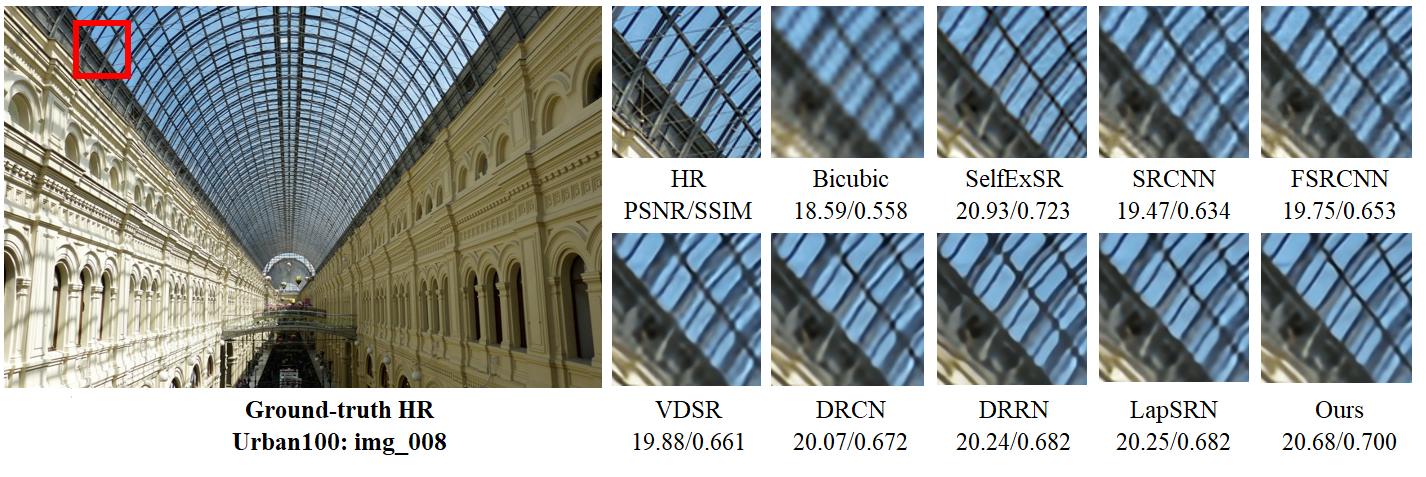


Figure 8. Presents the visual quality performance comparison for 4× image SR methods on BSDS100 and Urban100 datasets.

**CONCLUSION**

In this paper, we proposed a novel deep learning-based CNN model called multi-path deep CNN with Residual and Inception Network for Single Image Super-Resolution. In our proposed network model predicts the result of image super-resolution reconstruction through three branches. Branch 1 and 3 pass the original input LR image through the ResNet block and upscale the resultant features by the up-sampling layer. The second branch (Branch 2) use the original input image on two different ResNet and Inception blocks upscaled by deconvolution layer. The resultant output is finally combined to reconstruct a high-resolution image. This alternate strategy of deeper network model is to further reduce the computational complexity and to avoid the vanishing gradient problem during the training. The experimental result of image super-resolution reconstruction shows that our proposed model has better reconstruction performance with a similar or smaller number of parameters than other state-of-the-art deep learning-based image super-resolution algorithms. Although, our model obtained promising results on SR with the enlargement scale factor of 8× to reconstruct the HR images, but still exhibits some limitations, such as computational cost, speed, and visual perception. To address these limitations, in our future work, we will apply lightweight convolution operations, such as Octave convolution and Grouped convolution-based ResNet block with Inception module, that will help to reduce computational cost and improve the perceptual quality of LR images.

**REFERENCES**

1. Wang, X., et al. *Recovering realistic texture in image super-resolution by deep spatial feature transform*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

2. Zhang, L., et al., *A super-resolution reconstruction algorithm for surveillance images.* 2010. **90**(3): p. 848-859.

3. Turchini, F., et al., *Deep learning based surveillance system for open critical areas.* Inventions, 2018. **3**(4): p. 69.

4. Hazra, D. and Y.-C. Byun, *Upsampling Real-Time, Low-Resolution CCTV Videos Using Generative Adversarial Networks.* Electronics, 2020. **9**(8): p. 1312.

5. Sajjadi, M.S., B. Scholkopf, and M. Hirsch. *Enhancenet: Single image super-resolution through automated texture synthesis*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

6. Hou, Q., et al. *Deeply supervised salient object detection with short connections*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

7. Thornton, M.W., P.M. Atkinson, and D.J.I.J.o.R.S. Holland, *Sub‐pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super‐resolution pixel‐swapping.* 2006. **27**(3): p. 473-491.

8. Jiang, K., et al., *Deep distillation recursive network for remote sensing imagery super-resolution.* 2018. **10**(11): p. 1700.

9. Jiang, K., et al., *A progressively enhanced network for video satellite imagery superresolution.* 2018. **25**(11): p. 1630-1634.

10. Jiang, K., et al., *Edge-enhanced GAN for remote sensing image superresolution.* 2019. **57**(8): p. 5799-5812.

11. Li, K., et al., *Video super-resolution using an adaptive superpixel-guided auto-regressive model.* 2016. **51**: p. 59-71.

12. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

13. Luo, W., et al., *Pixel super-resolution using wavelength scanning.* 2016. **5**(4): p. e16060-e16060.

14. Zhou, F., W. Yang, and Q.J.I.T.o.I.P. Liao, *Interpolation-based image super-resolution using multisurface fitting.* 2012. **21**(7): p. 3312-3318.

15. Greenspan, H.J.T.c.j., *Super-resolution in medical imaging.* 2009. **52**(1): p. 43-63.

16. Dudczyk, J., *A method of feature selection in the aspect of specific identification of radar signals.* Bulletin of the Polish Academy of Sciences. Technical Sciences, 2017. **65**(1).

17. Zhang, L. and X. Wu, *Wu, X.: An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Transactions on Image Processing 15(8), 2226-2238.* IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 2006. **15**: p. 2226-38.

18. Aràndiga, F., *A nonlinear algorithm for monotone piecewise bicubic interpolation.* Applied Mathematics and Computation, 2015. **272**.

19. Keys, R., *Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Process.* Acoustics, Speech and Signal Processing, IEEE Transactions on, 1982. **29**: p. 1153-1160.

20. Zhang, K., et al., *Single image super-resolution with non-local means and steering kernel regression.* 2012. **21**(11): p. 4544-4556.

21. Stark, H. and P. Oskoui, *High-resolution image recovery from image-plane arrays, using convex projections.* JOSA A, 1989. **6**(11): p. 1715-1726.

22. Yang, X., et al., *An improved iterative back projection algorithm based on ringing artifacts suppression.* Neurocomputing, 2015. **162**: p. 171-179.

23. Dong, C., *Image super-resolution using deep convolutional networks.* IEEE Trans. Pattern Anal, 2015. **38**: p. 295-307.

24. Dong, C., C.C. Loy, and X. Tang, *Accelerating the Super-Resolution Convolutional Neural Network*. Vol. 9906. 2016. 391-407.

25. Kim, J., J. Lee, and K. Lee, *Accurate Image Super-Resolution Using Very Deep Convolutional Networks*. 2016. 1646-1654.

26. Lai, W.-S., et al., *Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution.* 2017.

27. Kim, J., J. Lee, and K. Lee, *Deeply-Recursive Convolutional Network for Image Super-Resolution*. 2016. 1637-1645.

28. Lim, B., et al., *Enhanced Deep Residual Networks for Single Image Super-Resolution.* 2017.

29. Zhang, Y., et al., *Residual Dense Network for Image Super-Resolution.* 2018.

30. Zhang, Y., et al., *Image Super-Resolution Using Very Deep Residual Channel Attention Networks*. 2018.

31. Chang, H., D.-Y. Yeung, and Y. Xiong, *Super-resolution through neighbor embedding*. Vol. 1. 2004. I-275.

32. Timofte, R., V. De, and L. Van Gool, *Anchored Neighborhood Regression for Fast Example-Based Super-Resolution*. 2013. 1920-1927.

33. Jiang, J., et al., *Ensemble Super-Resolution with A Reference Dataset*. 2019.

34. Chen, C., *Single Image Super-Resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means.* IEEE Transactions on Multimedia, 2016.

35. Yang, J., J. Wright, and L. Yu, *Image Super-Resolution Via Sparse Representation.* Image Processing, IEEE Transactions on, 2010. **19**: p. 2861-2873.

36. Sun, J., Z. Xu, and H.-Y. Shum. *Image super-resolution using gradient profile prior*. in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008. IEEE.

37. Dong, C., et al., *Image super-resolution using deep convolutional networks.* 2015. **38**(2): p. 295-307.

38. Wang, Z., et al. *Deep networks for image super-resolution with sparse prior*. in *Proceedings of the IEEE international conference on computer vision*. 2015.

39. Kim, J., J. Kwon Lee, and K. Mu Lee. *Accurate image super-resolution using very deep convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

40. Kim, J., J. Kwon Lee, and K. Mu Lee. *Deeply-recursive convolutional network for image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

41. Shi, W., et al. *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

42. Dong, C., C.C. Loy, and X. Tang. *Accelerating the super-resolution convolutional neural network*. in *European conference on computer vision*. 2016. Springer.

43. Lai, W.-S., et al. *Deep laplacian pyramid networks for fast and accurate super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

44. Maas, A.L., A.Y. Hannun, and A.Y. Ng. *Rectifier nonlinearities improve neural network acoustic models*. in *Proc. icml*. 2013. Citeseer.

45. Ahn, N., B. Kang, and K.-A. Sohn. *Fast, accurate, and lightweight super-resolution with cascading residual network*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

46. Zhang, Y., et al. *Image super-resolution using very deep residual channel attention networks*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

47. Zhang, W., et al., *Learning to predict high-quality edge maps for room layout estimation.* IEEE Transactions on Multimedia, 2016. **19**(5): p. 935-943.

48. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

49. Mudunuri, S.P. and S. Biswas, *Low resolution face recognition across variations in pose and illumination.* IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(5): p. 1034-1040.

50. Girshick, R., et al., *Region-based convolutional networks for accurate object detection and segmentation.* 2015. **38**(1): p. 142-158.

51. Bai, Y., et al. *Sod-mtgan: Small object detection via multi-task generative adversarial network*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

52. Lobanov, A.P.J.a.p.a.-p., *Resolution limits in astronomical images.* 2005.

53. Swaminathan, A., et al., *Digital image forensics via intrinsic fingerprints.* 2008. **3**(1): p. 101-117.

54. Lillesand, T., R.W. Kiefer, and J. Chipman, *Remote sensing and image interpretation*. 2015: John Wiley & Sons.

55. Shi, W., et al., *Is the deconvolution layer the same as a convolutional layer?* 2016.

56. Tai, Y., J. Yang, and X. Liu. *Image super-resolution via deep recursive residual network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

57. Zhang, K., et al., *Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising.* 2017. **26**(7): p. 3142-3155.

58. Lim, B., et al. *Enhanced deep residual networks for single image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.

59. Ledig, C., et al. *Photo-realistic single image super-resolution using a generative adversarial network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

60. Musunuri, Y. and O.-S. Kwon, *Deep Residual Dense Network for Single Image Super-Resolution.* Electronics, 2021. **10**: p. 555.

61. Ren, H., M. El-Khamy, and J. Lee. *Image super resolution based on fusing multiple convolution neural networks*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017.

62. Hui, Z., X. Wang, and X. Gao. *Fast and accurate single image super-resolution via information distillation network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

63. Muhammad, W. and S. Aramvith, *Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach.* Electronics, 2019. **8**: p. 892.

64. Liu, C., et al., *Food and ingredient joint learning for fine-grained recognition.* IEEE Transactions on Circuits and Systems for Video Technology, 2020.

65. Bahdanau, D., K. Cho, and Y. Bengio, *Neural machine translation by jointly learning to align and translate.* arXiv preprint arXiv:1409.0473, 2014.

66. Wang, F., H. Hu, and C. Shen, *BAM: A Lightweight and Efficient Balanced Attention Mechanism for Single Image Super Resolution.* arXiv preprint arXiv:2104.07566, 2021.

67. Hang, Y., et al. *Attention Cube Network for Image Restoration*. in *Proceedings of the 28th ACM International Conference on Multimedia*. 2020.

68. Xu, Q., et al., *CSA-MSO3DCNN: Multiscale octave 3D CNN with channel and spatial attention for hyperspectral image classification.* Remote Sensing, 2020. **12**(1): p. 188.

69. Wang, X., et al., *Dual residual attention module network for single image super resolution.* Neurocomputing, 2019. **364**: p. 269-279.

70. Xing, X. and D. Zhang, *Image Super-Resolution Using Aggregated Residual Transformation Networks With Spatial Attention.* IEEE Access, 2019. **7**: p. 92572-92585.

71. Salvetti, F., et al., *Multi-image super resolution of remotely sensed images using residual attention deep neural networks.* Remote Sensing, 2020. **12**(14): p. 2207.

72. Wang, Z., et al., *Single image super-resolution with attention-based densely connected module.* Neurocomputing, 2021. **453**: p. 876-884.

73. Anwar, S. and N. Barnes, *Densely residual laplacian super-resolution.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

74. Zha, L., et al., *A Lightweight Dense Connected Approach with Attention on Single Image Super-Resolution.* Electronics, 2021. **10**(11): p. 1234.

75. Tong, T., et al. *Image super-resolution using dense skip connections*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

76. Tai, Y., et al. *Memnet: A persistent memory network for image restoration*. in *Proceedings of the IEEE international conference on computer vision*. 2017.

77. Zhang, K., W. Zuo, and L. Zhang, *Learning a Single Convolutional Super-Resolution Network for Multiple Degradations.* Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.

78. Yang, X., et al., *DRFN: Deep Recurrent Fusion Network for Single-Image Super-Resolution With Large Factors.* 2019. **21**(2): p. 328-337.

79. Gu, J., et al. *Blind super-resolution with iterative kernel correction*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

80. Jin, X., et al., *Single image super-resolution with multi-level feature fusion recursive network.* 2019. **370**: p. 166-173.

81. Liu, B. and D.J.N. Ait-Boudaoud, *Effective image super resolution via hierarchical convolutional neural network.* 2020. **374**: p. 109-116.

82. Lin, D., et al., *SCRSR: An efficient recursive convolutional neural network for fast and accurate image super-resolution.* 2020. **398**: p. 399-407.

83. Qiu, D., et al., *Multiple improved residual networks for medical image super-resolution.* 2021. **116**: p. 200-208.

84. Simonyan, K. and A.J.a.p.a. Zisserman, *Very deep convolutional networks for large-scale image recognition.* 2014.

85. Nah, S., T. Hyun Kim, and K. Mu Lee. *Deep multi-scale convolutional neural network for dynamic scene deblurring*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

86. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

87. Dumoulin, V. and F. Visin, *A guide to convolution arithmetic for deep learning.* 2016.

88. Xiankai, L., et al., *Learning Deconvolutional Network for Object Tracking.* IEEE Access, 2018. **PP**: p. 1-1.

89. Mao, X.-J., C. Shen, and Y.-B. Yang, *Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections.* 2016.

90. Arbelaez, P., et al., *Contour detection and hierarchical image segmentation.* 2010. **33**(5): p. 898-916.

91. Bevilacqua, M., et al., *Low-complexity single-image super-resolution based on nonnegative neighbor embedding.* 2012.

92. Zeyde, R., M. Elad, and M. Protter. *On single image scale-up using sparse-representations*. in *International conference on curves and surfaces*. 2010. Springer.

93. Martin, D., et al. *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*. in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. 2001. IEEE.

94. Huang, J.-B., A. Singh, and N. Ahuja. *Single image super-resolution from transformed self-exemplars*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

95. Matsui, Y., et al., *Sketch-based manga retrieval using manga109 dataset.* 2017. **76**(20): p. 21811-21838.

96. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* 2014.

97. Timofte, R., V. De Smet, and L. Van Gool. *A+: Adjusted anchored neighborhood regression for fast super-resolution*. in *Asian Conference on Computer Vision*. 2014. Springer.

98. Schulter, S., C. Leistner, and H. Bischof. *Fast and accurate image upscaling with super-resolution forests*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.