MDPI

Article

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

Wazir Muhammad[1,*], Zuhaibuddin Bhutto[2,*], Arslan Ansari[3], Mudasir Latif Memon[4], Ramesh Kumar[5], Ayaz Hussain[1], Ali Raza Shah[6] Imdadullah Thaheem[6] and Shamshad Rajper[1]

[1] Department of Electrical Engineering, Balochistan University of Engineering and Technology, Pakistan; wazir.laghari@gmail.com, engr_ayaz@yahoo.com, shamshad.rajpar@gmail.com
[2] Department of Computer System Engineering, Balochistan University of Engineering and Technology, Pakistan; zuhaib_bhutto@hotmail.com
[3] Department of Electronic Engineering, Dawood University of Engineering and Technology, Pakistan; dr.arslanansari@duet.edu.pk
[4] IBA Community College Naushehro Feroze, Sukkur IBA University, Pakistan; mudasirlatif.ccnf@iba-suk.edu.pk
[5] Department of Computer System Engineering, Dawood University of Engineering and Technology, Pakistan; ramesh.kumar@duet.edu.pk
[6] Department of Mechanical Engineering, Balochistan University of Engineering and Technology, Pakistan; razadopasi; @buetk.edu.pk, engrimdad67@gmail.com
* Correspondence: wazir.laghari@gmail.com, zuhaib_bhutto@hotmail.com

Abstract: Deep convolutional neural networks have recently made a breakthrough in the quality of single image super-resolution (SISR) reconstruction. SISR is an efficient approach to improve the quality of spatial resolution in various types of computer vision tasks, such as classification and object detection in case of resolution of camera sensors is limited. Most of the existing convolutional neural network (CNN) based image super-resolution (SR) techniques are used the pre-processing step such as a bicubic interpolation to upscale the low-resolution. Bicubic interpolation technique is not designed for this purpose and introduces the new noises in the model. Furthermore, a deeper network architecture obtained a better result as compared to shallow compared to shallow network architecture, but it introduces the vanish-gradient problem during the training. To resolve these problems, we proposed deep and accurate yet concise network architecture with two types of upscaling layers, i.e., transpose convolution layer, and upsampling layer to reconstructs the HR image. To alleviate the vanishing gradient problem, and reuse all initial features as an input to the subsequent layers, we used ResNet and Inception blocks strategy. The extensive experimental results on image SR using five publicly available test datasets, which shows that 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.

Keywords: super-resolution; deep convolutional neural network; skip connection; inception block.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Image super-resolution plays a vital role in the field of image processing and computer vision-based applications, because high quality or high-resolution (HR) images have more pixel density level and contains more detailed information. This detailed information is applied in various fields of computer vision and image processing tasks, such as image restoration [1], security surveillance [2], object recognition [3], object detection [4], satellite imaging [5], remote sensing imagery [6–8] and medical imaging [9–13]. 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 the similar low-quality LR images by downscaling infinite number of HR images. To overcome these problems, computer vision researchers community are proposed various approaches and can be classified into three classes: interpolation-based methods [14], reconstruction-based methods [15], and learning-based methods [16–23], but here we are present detailed discussion on learning-based methods. 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 [24], bilinear interpolation, and nearest-neighbor interpolation techniques. The reconstruction-based method proposed in [25] and introducing the prior knowledge to reduce the solution space. These 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. Currently, a variety of learning-based approaches have been suggested including the regression-based approaches [26–31] and sparse coding-based approaches [32–40]. Although these learning-based approaches have outstanding performance, but they involve more time-consuming operations of optimization.

Recently, deep convolutional neural networks based approaches [16–18,20,41,42] have been obtained remarkable contributions and significantly increase the progress in the area of image SR tasks, because of their superior capability of the feature representation. The first successful shallow type deep learning-based architecture with three CNN layers followed by two rectified linear units (ReLU) is proposed by Dong *et al.* known as a super-resolution convolutional neural network (SRCNN) [16] 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. To improve the efficiency and speed of SRCNN [16], the same author proposed the faster version known as accelerating the super-resolution convolutional neural network (FSRCNN) [17]. Furthermore, to improve the computational efficiency of the model, Shi *et al.* [42] introduced the Efficient sub-pixel convolutional neural network (ESPCNN) [42]. Traditional deep convolutional neural network methods can optimize the quality of LR images by making the network architecture deeper. The performance of deeper network architectures is best to extract more features information from the image, but these approaches have led to introducing the vanishing gradient problem in the training phase as well as increase the computational cost of the model. In deeper type network architecture when the input of LR image reaches at the end of the last layers after passing many layers, it will disappear and "wash-out" the information.

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.

To address these drawbacks, we suggest the ResNet and Inception based network architecture to extract the fast and accurate feature information for SISR. Our Proposed network architecture mainly includes the three paths/branches parallelly to reconstruct the HR image. These paths are used two types of layers to upscale the LR image, like deconvolution layer, and upsampling layer. The cumulative sum of these three branches is passed by concatenation layer to reconstruct the Super-Resolution HR image. Furthermore, we replace Stochastic Gradient Descent (SGD) optimizer with an ADAM optimizer and ReLU activation function with LeakyReLU [43] to overcome the dead neuron problem and improve the training efficiency of the model.

The main contribution of our proposed method is 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 transpose layer and upsampling layer to reconstruct the high quality of HR image features.
- Traditional deep CNN methods used the ReLU activation function, but our approach used LeakyReLU as an activation function to activate the dead neurons 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 is reported in Section 4, and Section 5, respectively.

2. Related Work

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 played a vital role in the field of image and computer vision-based applications, such as security surveillance videos for face recognition purposes [44], object detection and classification in different scenes [45] especially for small objects [46], astronomical images [47], medical imaging [13], forensics [48] and remote sensing images [49].

The easy and simple technique among them is the interpolation-based technique, such as bilinear, bicubic[24], B-spline kernels, and nearest neighbor interpolation. These approaches take the average pixel point information from the LR image. Interpolation performs better work in the smooth region part of the image, but poor result in the edges of the images and introduce the blurring and jagged ringing effect. Furthermore, reconstruction-based approaches have a more complex process such as random forest [50], neighborhood embedded regression approach [27,51], and sparse coding [32] methods.

A deep learning convolutional neural network-based method for single image SR first time proposed in [16] and has been brought into the deep learning-based era. SRCNN [16] 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 [16] is used as a bicubic upsampled version of the image, which introduces the extra new noises in the model and adds extra computational cost. Inspired by the concept of SRCNN [16] and improved the speed and perceptual quality of the LR image, the same author proposed the concept of Fast Super-Resolution Convolutional Neural Network. The designed network architecture of FSRCNN 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 [17] methods do not use any interpolation technique as a pre-processing step. 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 (ESPCN) [52]. 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 [52] 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 the competitive results as compared earlier approaches.

Unlike, shallow type network architectures proposed in SRCNN and FSRCNN. Follow the architecture of VGG-net, Kim *et al.*, [18] 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 [18] extracts the features by global residual learning to ease the training complexity of their network. Although, VDSR [18] had achieved a great success, but it exclusively extracts single-scale features information only and ignores multi-scales features information. Zhang *et al.* [53] 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. Still the model reported favorable results, their performance is depending on the accuracy of noise estimation and computationally expensive due to the use of batch normalization after every CNN layer. DRCN [20] proposed a handless deep CNN architecture recursively to share the depth of the network in terms of network parameters. Pyramidal-based network architecture is known as the Deep Laplacian pyramid super-resolution network (LapSRN) [19]. 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 [43] layers, and deconvolution layers. DRRN [54] recursively builds two residual blocks and they handle the pre-processing problem caused by interpolation. The application of DenseNet in the area of image super-resolution goes to SRDenseNet [55], in which they claim that dense skip connection mitigates the vanishing gradient problem and can boost up the feature propagation. It can achieve better performance as well as faster speed. A persistent memory type network for image SR known is known as MemNet, which is proposed in [56]. The MemNet architecture 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 are used in the architecture. The network architecture of single image super-resolution for multiple degradations abbreviated as SRMD [22] proposed by Zhang *et .al* [22]. 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. Yang *et al.* [57] proposed the deep recurrent fusion network of SISR is known as DRFN [57]. 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 [57] used the same training dataset which is used by VDSR [18] with data augmentation in terms of rotation and flipping. Ahn *et al.* proposed a lightweight scenario-based architecture known as cascading residual network (CARN) [58]. The basic design of a CARN [58] architecture is used as a cascading residual block, whose output of each intermediate layers are shifted to each of the consequent CNN layers. Inspired from CARN [58] Zhang *et al.* introduced the concept of residual channel attention network (RCAN) [23]. They have observed low-frequency content of information and to capture by simple CNN layers. By solving this problem adding multiple long as well as short skip connections for residual dense blocks. Inspired by GoogLeNet [59], W. M *et al.* [60] proposed inception based approach to reconstruct the HR image.

The Iterative Kernel Correction (IKC) method for single image super-resolution proposed by Gu *et al.*, which consists of a super-resolution model, predictor model, corrector model. In this approach, the author used the principal component analysis approach to reduce the dimensionality of the kernel. Jin *et al.* [62] proposed a new framework known as multi-level feature fusion recursive network abbreviated as MF-FRnet [62] for single image super-resolution without pre-processing any scale of the

image. The network architecture of MFFRnet [62] depends on four basic building blocks: coarse feature extraction, recursive feature extraction, multi-level feature fusion, and reconstruction blocks. The three steps based model proposed by Soh *et al.* [63], known as Meta-Transfer Learning for Zero-Shot Super-Resolution abbreviated as (MZSR). These three steps are as: large-scale training step, meta-transfer learning step and meta testing step. To follow the ZSSR [64] architecture, the authors are used 8 simple CNN layers with residual learning in the MZSR network. For meta-training purposes, the authors used DIV2K dataset [65]. Stacking different shallow type network architecture named as HCNN proposed by Liu *et al.* [66]. HCNN [66] 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. Lin *et al.* [67] proposed a fast and accurate image SR method known as Split-Concate-Residual Super Resolution (SCRSR). 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. Qiu *et al.* [68] suggested the multiple improved residual network abbreviated as MIRN for single image super-resolution. First, they are designed a 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 [68] network architecture with an adjustable learning rate.

3. Proposed Method

In this section, we describe the design methodology of our proposed model architecture. Earlier deep learning-based model architectures depends on one branch for restoring the feature texture information and might be discarding the useful LR features information like edges of the image and other high-frequency features information. In this paper we propose three branch network architecture (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.

3.1. Architecture Overview

The main purpose of single image super-resolution is to predict the HR image (I_{HR}) from the corresponding LR image (I_{LR}). Suppose I_{LR} is the low-resolution image followed by an upsampling factor of α to reconstruct the HR image I_{HR}. Furthermore, HR and LR is the pair of the image with color channels C of I_{LR} and I_{HR}, they can be represented in the tensor of the size as $H \times W \times C$ and $\propto H \times \propto W \times \propto C$, respectively. 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. In Branch1, the network architecture takes an original LR image followed by ResNet Block with UpSampling Layer to reconstruct the HR image known as HR_{B1}. Similarly, the same LR image pass through Branch2 followed by ResNet with Inception block followed by Transpose Convolution layer to reconstruct the HR image known as HR_{B2}. Finally, LR image applied on Branch3 followed by ResNet block with UpSampling Layer to reconstruct the HR image known asHR_{B3}. Resultantly output HR images of three branches are concatenate followed by a reconstruction layer to generate the HR output image.

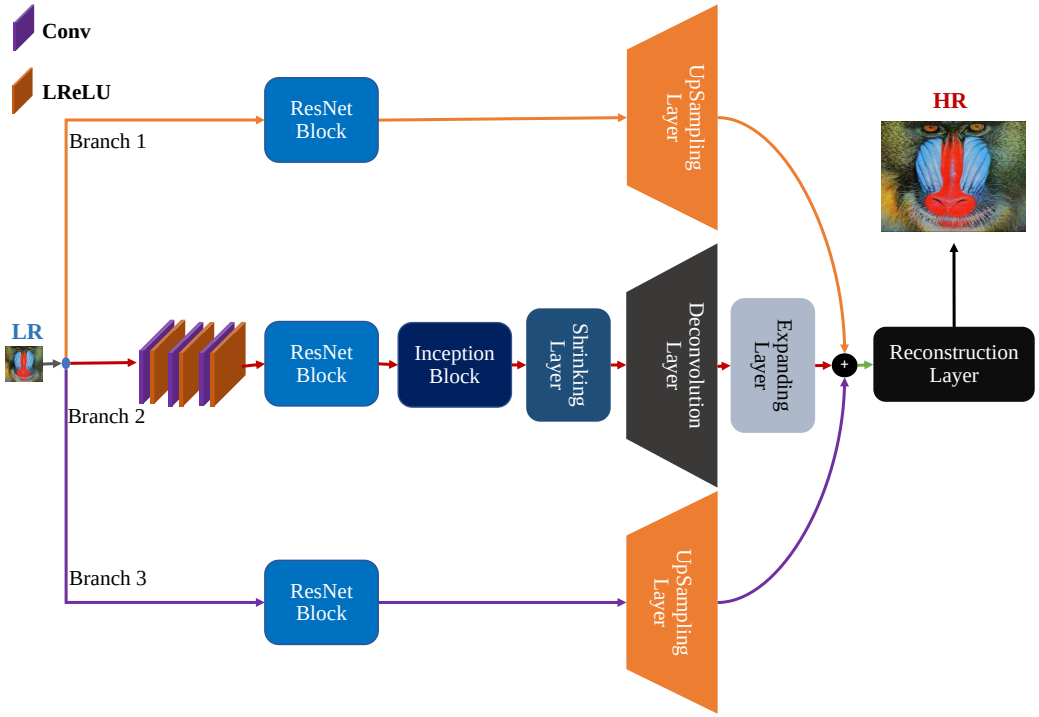


Figure 1. The proposed network architecture of our method with three parallel paths/branches.

3.2. Feature Extraction

Following the principle in [69], we used three CNN layers followed by Leaky ReLU [43] of kernel sizes is 3×3 with 64 number of channels to reconstruct the feature maps of the main branch (Branch2). The feature maps of these three CNN layers passed through ResNet and Inception blocks to generate the multiscale hierarchical features. Similarly, other feature maps are extracted from Branch1 and Branch3.

3.3. Residual Learning Paths

The main target of image super-resolution is the transformation of an image-to-image in which the LR input image is correlate to the desired HR output image.

To extract the low-level features, earlier approaches are used global residual learning paths with a single CNN layer having a kernel size greater than 5×5. 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 files of order 3×3 followed by upsampled and transposed layer to upscalled the LR image. This type of upsampling strategy slightly improved the accuracy as well as computational efficiency of the model in terms of network parameters.

3.3.1. ResNet Block

Residual learning [10] is the best way to increase the computational efficiency and ease the training complexity. He et al. [10] first time proposed a ResNet architecture of residual learning for the image classification task. In [18], 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 block [10], SRResNet block [70], and our proposed ResNet block. Figure 2 (a) shows the Original ResNet block and it uses the two convolution layers, two BN layers, one ReLU activation function used before and after the element-wise addition with Identity Skip connection. SRResNet [70] 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 BN layers proposed by Nah et al. [71], to provide the clean path, because 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 CNN layers followed by one common activation function (Leaky ReLU). Leaky ReLU [43] 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.

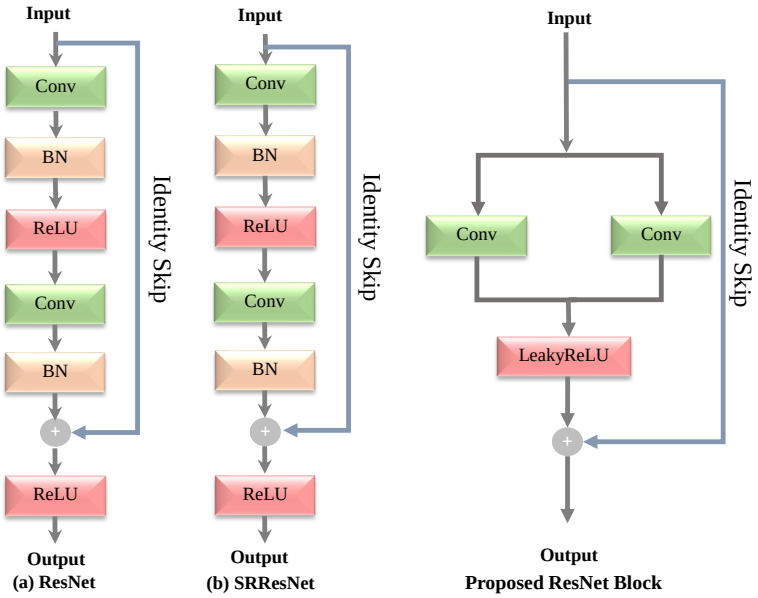


Figure 2. The comparision of residual blocks in ResNet, SRResNet and our proposed (ResNet).

3.3.2. Inception Block

The winner of the 2014-ILSVRC competition was GoogLeNet, whose major goal was to attain excellent accuracy at a low computational cost [59]. Inception block is a new concept in the CNN architecture that integrates multi-scale convolutional transformations using split, transform, and merge principles. Furthermore, it consists of different parallel paths with various 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 previous techniques extracted the features for reconstructing the HR image using a single size kernel. However, single size kernel is not an efficient way to restoring the feature information completely. Our proposed block designed is inspired by GoogLeNet [59] architecture to extract the feature information on different kernels to capture better content and structure information from the image. Moreover, the max: pooling layer is not included in our inception block because it decreases the network's ability to learn the detailed information and not suitable for SR tasks.

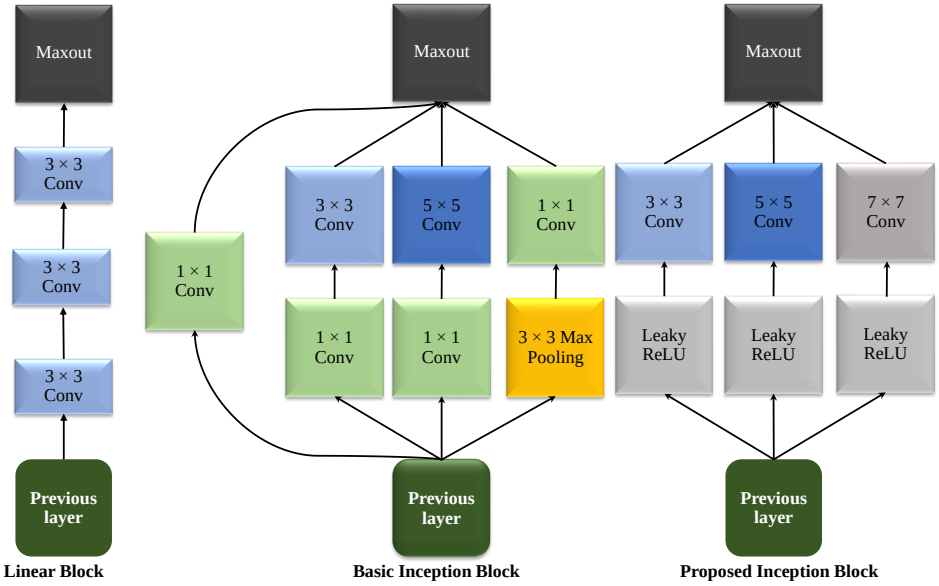


Figure 3. Comparison of different types of blocks used for feature extraction (a) Linear/Single path type block, (b) Inception-based Multi path block, and (c)Our proposed Multi path Inception block.

In 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 deeper model architecture and occupy more memory consumption during the training. Figure 3(b), shows a conventional inception block to extract the multi-scale feature information. The drawback in such type of designed blocks is that it contains more parameters, which means the model is more computationally expensive. 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 others element's information as shown in Figure 3(c). Our proposed block consists of several filters of different sizes. It extracts the features from the previous layer's output. In our proposed inception block is used three types of kernel sizes having the order of 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.

3.3.3. Shrinking Layer

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 [72].

3.3.4. Deconvolution Layer

Earlier deep convolutional neural network based image super-resolution approaches used an interpolation technique to upscale the input LR image into HR image, such as SRCNN [16], VDSR [18], REDNet [73], DRCN [20], and DRRN [54]. These types of architecture extract the features information from the interpolated version of reconstructed image, which introduces the extra new noises in the model and not achieve better performance as well as increase the computational cost. Therefore, recent works [17,60,74] 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 performed in the LR space.

3.3.5. Expanding Layer

The function shrinking layer, which comes before the deconvolution layer, reduces the dimension of the 64 channel input into 4 feature outputs for upsampling purposes. To recover the original 64 feature map back from the 4-channel input feature map, we used expanding layer of kernel size is 1×1 followed by LReLU to increase the nonlinearity function.

3.3.6. 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.

3.4. Concatenation Layer

Earlier approaches [16,18] uses only the 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. So, to resolve said problems we extract the features information from different routes/branches and concatenate it via the concatenation layer.

3.5. Reconstruction Layer

In our proposed model the resultant feature maps are used to reconstruct the high quality or high-resolution images via a reconstruction layer. The reconstruction layer is a basically type of CNN layer having a kernel size of the order 3×3.

4. Experiments

4.1. Datasets

In our proposed method, we combine two datasets of different color images, which are 200 images obtained from BSD200 [75] datasets and 91 images from Yang et al. [32] for training purposes. In the testing phase, we used five standard publicly available test datasets, including Set5 [76], Set14 [77], BSDS100 [78], Urban100 [79] and Manga109 [80]. The number of images in Set5 is 5, Set14 is14, BSDS100 is 100, Urban100 is 100, and Manga109 is 109, respectively. Each of the five benchmark test datasets has its own set of characteristics. Natural scenes can be found Set5 [76], Set14 [77] and BSDS100 [78]. The images in Urban100 [79] are challenging images with details in a variety of frequency bands. Finally, Manga109 [80] also known as Japanese comic images is the class of multimodal type of artwork, which is collected from Japanese Manga.

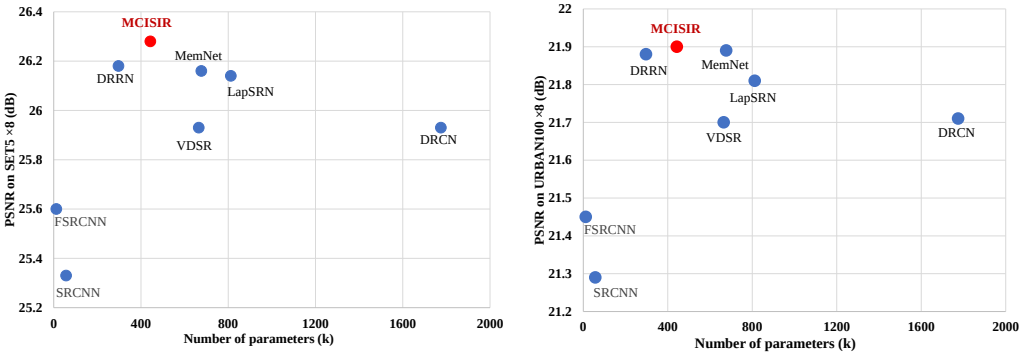
4.2. Training Settings

The LR images are generated by using the Bicubic kernel of challenging enlargement scale factor $4\times$ and $8\times$. To improve the quality of available data for training the model, we used data augmentation techniques such as flipping, rotation, and cropping. Training the deep CNN architecture we used the Adam optimizer [81] 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, and OpenCV), CUDA Version 10.2, Python 3.7, and an NVIDIA GeForce RTX 2070 GPU.

4.3. Comparisons with current existing state-of-the-art approaches

The PSNR/SSIM image quality matrix is the most generally used as a reference quality metric in the field of image SR, because they are directly related with the intensity of the image. Our proposed model evaluates on five publicly available benchmark test datasets with challenging enlargement factor 4× and 8×. For quantitative comparison point of view, we have used thirteen different image super-resolution methods with the baseline method. The quantitative results of selected methods and our method as shown in the Table 1. Our proposed approach achieves the better PSNR/SSIM on average scale than other image SR methods. Furthermore, our model can improvement overall PSNR on SET5 dataset with challenging upscale factor 8x as 1.88, 0.75, 0.90, 0.79, 0.69, 0.53, 0.95, 0.68, 0.35, 0.35, 0.14, 0.10, and 0.12 dB's, Bicubic, A+, RFL, SelfExSR, SCN, ESPCN, SRCNN, FSRCNN, VDSR, DRCN, LapSRN, DRRN, and MemNet, respectively.

The performance of the image super-resolution model also correlates with the network depth. The performance of deeper model is better as compared to shallow model proposed by Kim et al. [18]. However, the deeper model has more parameters than shallow model. Table 2 presents the existing image SR algorithms in terms of number of filters, network depth (number of layers), network parameters and type of loss functions. Our proposed method has significantly reduced the number of parameters as well as network depth on same number filters 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.



(a) Figure **(b)** Figure

Figure 4. Performance comparisons in terms of PSNR versus network model parameters on SET5 and URBAN100 test datasets with enlargement factor 8×.

Figure 4, depicts the tradeoff between the model performance(PSNR) versus size of the model (Number of parameters). Both results are performed on the Set5 and URBAN100 datasets for challenging enlargement 8× scale factor. From quantitative evaluations, we clearly observed that our model achieves outperforms than current state-of-the-art approaches. For example, our MCISIR achieves much better performance than VDSR, MemNet, LapSRN, and DRCN on scale factor 8×, with the size of network parameters is decreased by 33%, 35%, 45%, and 62%, respectively. In Figure 5, and 6, noticed that our proposed model achieved better PSNR/SSIM on all public test datasets at challenging scale factor 8×.

Table 1. Quantitative comparison of PSNR/SSIM of recent image super-resolution methods on challenging enlargement factor 4× and 8×. The first best result is indicated by bold with red color and the second-best result is described by blue color.

Method	Scale	#Parameters ↓	SET5 PSNR↑/SSIM↑	SET14 PSNR↑/SSIM↑	BSDS100 PSNR↑/SSIM↑	URBAN100 PSNR↑/SSIM↑	MANGA109 PSNR↑/SSIM↑	Average PSNR↑/SSIM↑
Bicubic [24]	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+ [28]	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 [50]	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 [79]	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 [41]	4×	42k	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 [42]	4×	20k	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 [16]	4×	57k	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 [17]	4×	12k	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 [18]	4×	665k	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 [20]	4×	1,775k	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 [19]	4×	812k	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 [54]	4×	297k	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 [56]	4×	677k	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×	443k	**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 [24]	8×	-/-	24.40/0.658	23.10/0.566	23.67/0.548	20.74/0.516	21.47/0.649	22.68/0.587
A+ [28]	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 [50]	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 [79]	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 [41]	8×	42k	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 [42]	8×	20k	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 [16]	8×	57k	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 [17]	8×	12k	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 [18]	8×	665k	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 [20]	8×	1,775k	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 [19]	8×	812k	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 [54]	8×	297k	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 [56]	8×	677k	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×	443k	**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. Comparison of existing well-known Image super-resolution methods in terms of network parameters, number of layers (depth), number of filters and type of loss function.

Method	**Input**	**No: of Filters**	**No: of layers**	**#Network Parameters(k)**	**Reconstruction**	**Loss Function**
SRCNN [16]	LR+bicubic	64	3	57	Direct	ℓ_2
VDSR [18]	LR+bicubic	64	20	665	Direct	ℓ_2
DRCN [20]	LR+bicubic	256	20	1,775	Direct	ℓ_2
LapSRN [19]	LR	64	27	812	Progressive	ℓ_1
MemNet [56]	bicubic	64	80	677	Direct	ℓ_2
MCISIR (Our)	LR	64	26	443	Direct	ℓ_2

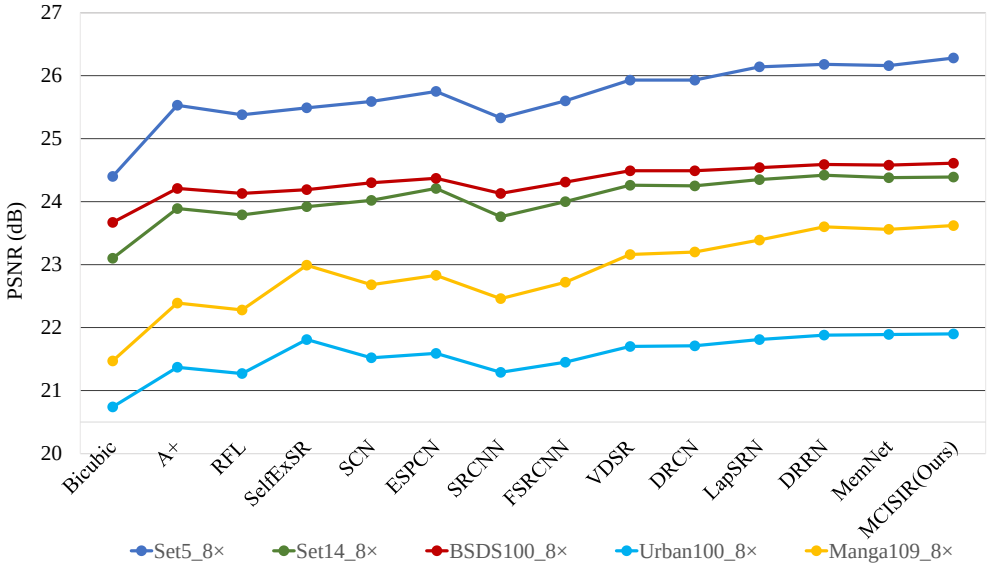


Figure 5. Peak signal to noise ratio versus different algorithms on enlargement scale factor 8×

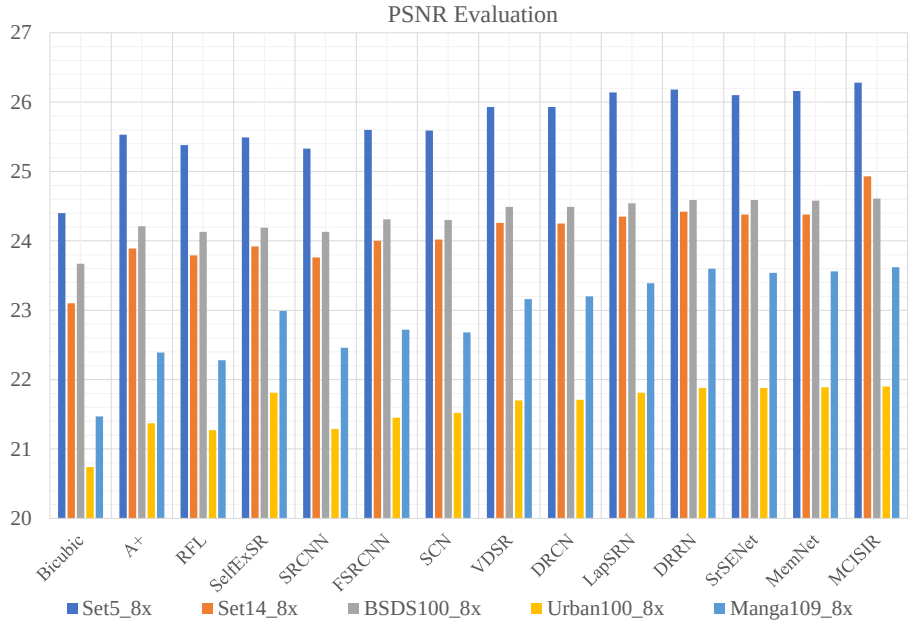


Figure 6. Plot the PSNR of all publicly available test image datasets versus different algorithms on enlargement scale factor 8×

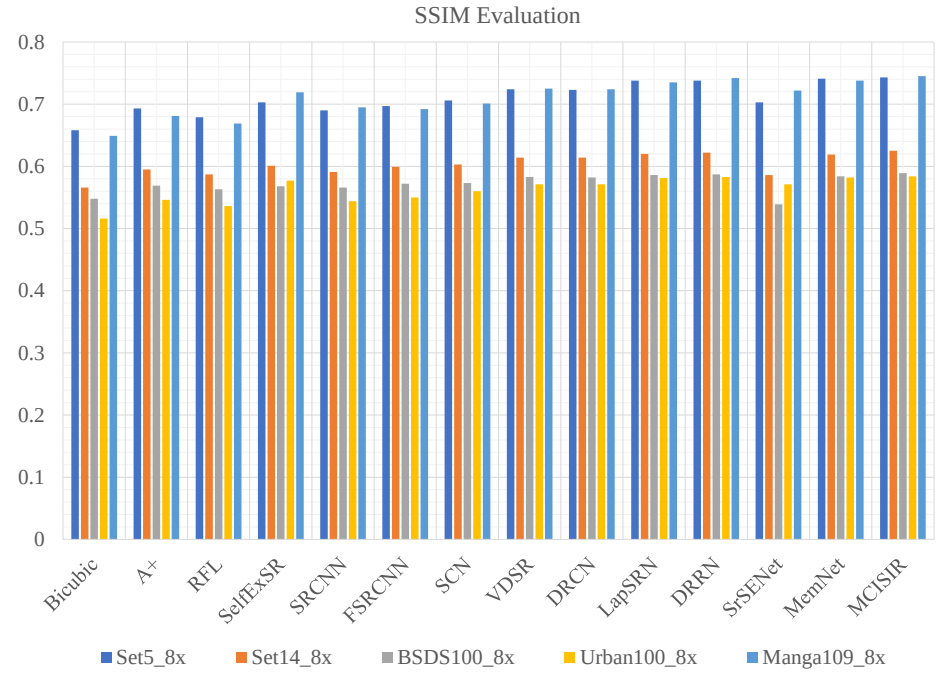


Figure 7. Plot the SSIM of all publicly available test image datasets versus different algorithms on enlargement scale factor 8×

To further evaluate the performance of our proposed model, we present a visual perceptual quality assessment compared to the existing 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 butterfly and baboon image obtained from publicly available dataset Set5 and Set14 with enlargement factor 8×. Upscaled region of the image indicated by 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 Baseline method, but still, results are largely blurry. In our proposed model reconstruct the 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. Similarly, the perceptual quality of the reconstructed images is evaluated on another two challenging datasets of Urban100 and BSDS100. Image_100 obtained from Urban100 dataset and image_098 is obtained from BSDS100. In Figure 9, the image of lizard (image_098) clearly observed that our proposed method reconstructs the better patch result as compared to Bicubic, SelfExSR, SRCN and FSRCNN methods. Likewise, VDSR, DRCN, DRRN and LapSRN reconstructed results are fairly acceptable, but our method has strong strip of lizard as compared to others.

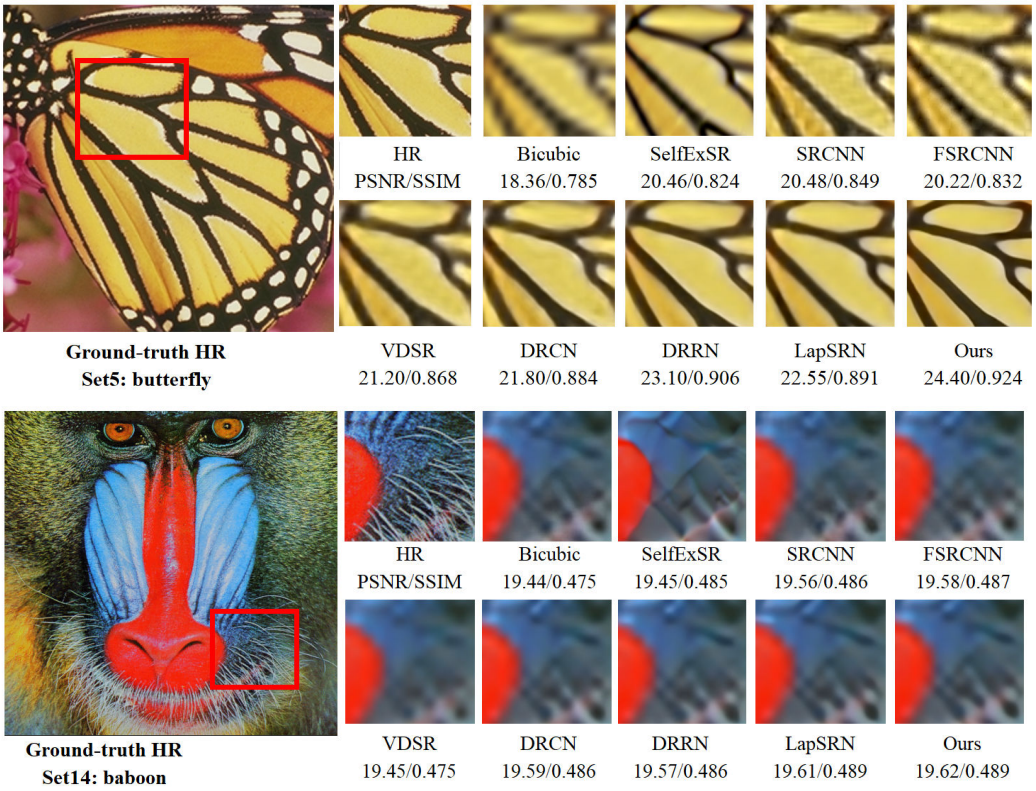


Figure 8. Presents the visual quality performance comparison for 8× image SR methods on Set5 and Set14 datasets.

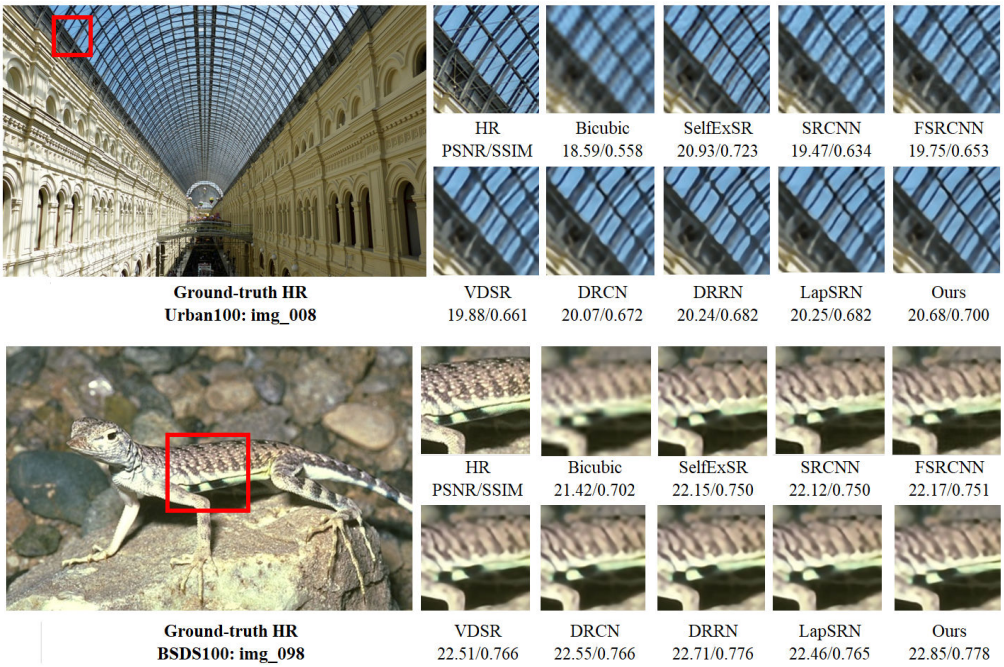


Figure 9. Presents the visual quality performance comparison for 4× image SR methods on Urban100 and BSDS100 datasets.

5. Conclusions

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 reconstruc-

tion 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) applies 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.

Author Contributions: Conceptualization, Wazir and Zuhaibuddin; methodology, Wazir and Zuhaibuddin; investigation, Arsalan and Ramesh; writing—original draft preparation, Wazir, Zuhaibuddin and Ayaz; writing—review and editing, Ali Raza, Mudasir and Imdadullah; supervision, Wazir, Zuhaibuddin and Shamshad; All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: No conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BN	Batch normalization
CARN	Cascading Residual Network
CNN	Convolutional Neural Network
DnCNN	Denoising Convolutional Neural Network
DRCN	Deeply-recursive convolutional network
DRFN	Deep Recurrent Fusion Network
DRRN	Deep Recursive Residual Network
ESPCNN	Efficient sub-pixel convolutional neural network
FSRCNN	Fast Super-Resolution Convolutional Neural Network
HCNN	Hierarchical convolutional neural network
IKC	Iterative Kernel Correction
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LapSRN	Laplacian Pyramid Super-Resolution Network
LR	Low-Resolution
LReLU	Leaky ReLU
MemNet	Memory network
MFFRnet	Multi-level feature fusion recursive network
MIRN	Multiple improved residual networks
MSE	Mean squared error
MZSR	Meta-Transfer Learning for Zero-Shot Super-Resolution
PSNR	Peak Signal-to-Noise Ratio
RCAN	Residual channel attention networks
RED-Net	Residual encoding–decoding convolutional neural network
ReLU	Rectified linear units
SCRSR	Split-Concate-Residual Super Resolution
SGD	Stochastic Gradient Descent
SISR	Single Image Super-Resolution
SRCNN	Super-resolution convolutional neural network
SRDenseNet	Super-Resolution Dense Network
SRMD	Super-Resolution Network for Multiple Degradations
SSIM	Structural Similarity Index Matrix
VDSR	Very Deep Super Resolution
VGG-Net	Visual Geometry Group Net Architecture
ZSSR	Zero-Shot SR

References

1. Wang, X.; Yu, K.; Dong, C.; Loy, C.C, Recovering realistic texture in image super-resolution by deep spatial feature transform, Proceedings of the IEEE conference on computer vision and pattern recognition, Location of Conference, Country, June 2018.
2. Zhang, L.; Zhang, H.; Shen, H.; Li, P, A super-resolution reconstruction algorithm for surveillance images. *Signal Processing* **2010**, *90*, 619–626.
3. Sajjadi, M.S.; Scholkopf, B,; Hirsch, M. EnhanceNet: Single image super-resolution through automated texture synthesis, Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 Oct. 2017.
4. Hou, Q.; Cheng, M.; Hu, X.; Borji, A.; Tu, Z., Torr, P., Deeply supervised salient object detection with short connections, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017.
5. Thornton, M.W.; Atkinson, P.M.; Holland, D.A. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. *International Journal of Remote Sensing* **2006**, *27*, 473–491.
6. Jiang, K.; Wang, Z.; Yi, P.; Jiang, J.; Xiao, J.; Yao, Y. Deep Distillation Recursive Network for Remote Sensing Imagery Super-Resolution. *Remote Sens.* **2018**, *10*, 1700.
7. Jiang, K.; Wang, Z.; Yi, P. A progressively enhanced network for video satellite imagery superresolution. *IEEE Signal Processing Letters*, **2018**, *25*, 1630–1634.
8. Jiang, K.; Wang, Z.; Yi, P.; Guangcheng W.; Lu, T.; Jiang, J. Edge-enhanced GAN for remote sensing image superresolution. *IEEE Transactions on Geoscience and Remote Sensing*, **2019**, *8*, 5799–5812.
9. Li, K.; Zhu, Y.; Yang, J.; Jiang, J. Video super-resolution using an adaptive superpixel-guided auto-regressive mode. *Pattern Recognition*, **2016**, *51*, 59–71.
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016.
11. Luo, W.; Zhang, Y.; Feizi, A.; *et al*. Pixel super-resolution using wavelength scanning. *Light Science & Applications*, **2016**, *5*, e16060.
12. Zhou, F.; Yang, W.; Liao, Q. Interpolation-based image super-resolution using multisurface fitting. *IEEE Transactions on Image Processing*, **2012**, *21*, 3312–3318.
13. Hayit, G. Super-resolution in medical imaging. *The Computer Journal*, **2009**, *52*, 43–63.
14. Zhang, L.' and Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Transactions on Image Processing* , **2006**, *15*, 2226–2238.
15. Zhang, K.; Gao, X.; Tao, D.; Li, D. Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression. *IEEE Transactions on Image Processing* , **2012**, *21*, 4544–4556.
16. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2016**, *38*, 295–307.
17. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network, Proceedings of the European Conference on Computer Vision (ECCV), Sept. 2016.
18. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016.
19. Lai, W.; Huang, J.; Ahuja, N.; Yang, M. Deep laplacian pyramid networks for fast and accurate super-resolution, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017.
20. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016.
21. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017.
22. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018.
23. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks, Proceedings of the European Conference on Computer Vision (ECCV), Oct. 2018.
24. Keys, R. Cubic convolution interpolation for digital image processin. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **2018**, *29*, 1153–1160.
25. Sun, S.; Xu, Z.; Shum, H,Y. Image super-resolution using gradient profile prior, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, June 2008.
26. Wang, H.; Gao, X.; Zhang, K.; Li, J.; Single image super-resolution using Gaussian process regression, Conference on Computer Vision and Pattern Recognition (CVPR), **2011**.
27. Timofte, R.; Smet, V.D; Gool, L,V. Image super-resolution using gradient profile prior, Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 2013.
28. Timofte, R.; Smet, V.D; Gool, L,V. A+: Adjusted anchored neighborhood regression for fast super-resolution, Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, Singapore, 2014.
29. Karl, S.N; and Nguyen, T.Q. Image super-resolution using support vector regression. *IEEE Transactions on Image Processing*, **2007**, *16*, 1596–1610.
30. Kim, I.M; and Kwon, Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2010**, *32*, 1127–1133.

31. Deng, C.; Xu, J.; Zhang, K.; Tao, D.; Li, X. Similarity constraints-based structured output regression machine: An approach to image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, **2016**, *27*, 2472–2485.
32. Yang, J.; Wright, J.; Thomas, S.H.; Ma, Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, **2010**, *11*, 2861–2873.
33. Dong, W.; Zhang, L.; Shi, G.; Li, X. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, **2013**, *22*, 1620–1630.
34. Yang, W.; Tian, Y.; Zhou, F.; Liao, Q.; Chen, H.; Zheng, C. Consistent coding scheme for single-image super-resolution via independent dictionaries. *IEEE Transactions on Multimedia*, **2016**, *3*, 313–325.
35. Li, J.; Gong, W.; Li, W. Dual-sparsity regularized sparse representation for single image super-resolution. *Information Sciences*, **2015**, *298*, 257–273.
36. Gong, W.; Tang, Y.; Chen, X.; Qiane, Y.; Weigong, L. Combining edge difference with nonlocal self-similarity constraints for single image super-resolution. *Neurocomputing*, **2017**, *249*, 157–170.
37. Chen, C.L.P; Liu, L.; Chen, L.; Tang, Y.Y.; Zhou, Y. Weighted couple sparse representation with classified regularization for impulse noise removal. *IEEE Transactions on Image Processing*, **2015**, *24*, 4014–4026.
38. Liu, L.; Chen, L.; Chen, C.L.P.; Tang, Y.Y.; Pun, C.M Weighted Joint Sparse Representation for Removing Mixed Noise in Image. *IEEE Transactions on Cybernetics*, **2017**, *47*, 600–611.
39. Zhang, Y.; Shi, F.; Cheng, J.; Li, W.; Yap, P.T.; Shen, D. Longitudinally Guided Super-Resolution of Neonatal Brain Magnetic Resonance Images. *IEEE Transactions on Cybernetics*, **2019**, *49*, 662–674.
40. Jiang, J.; Yu, Y.; Tang, J.; Ma, A. Aizawa,; Aizawa. K. Context-patch face hallucination based on thresholding localityconstrained representation and reproducing learning. *IEEE Transactions on Cybernetics*, **2018**, *50*, 324–337.
41. Wang, Z.; Liu, D.; Yang, J.; Han, W.; Huang. T. Deep networks for image super-resolution with sparse prior, Proceedings of the IEEE international conference on computer vision (ACCV), 370–378, 2015.
42. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang Z. Real-time single image and video superresolution using an efficient sub-pixel convolutional neural network, Proceedings of the IEEE conference on computer vision and pattern recognition, 1874–1883, 2016.
43. Mass, A.L.; Hannum, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models, Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013.
44. Mudunuri, S.P.; and Biswas, S. Low resolution face recognition across variations in pose and illumination. *IEEE transactions on pattern analysis and machine intelligence*, **2015**, *38*, 1034–1040.
45. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, **2015**, *38*, 142–158.
46. Bai, Y.; Zhang, Y.; Ding, M.; Ghanem, B. Sod-mtgan: Small object detection via multi-task generative adversarial network, Proceedings of the European Conference on Computer Vision (ECCV), 206–221, 2018.
47. Lobanov, A.P. Resolution limits in astronomical images. *arXiv preprint*, astro-ph/0503225.
48. Swaminathan, A.; Wu, M.; Liu, K.R. Digital image forensics via intrinsic fingerprints. *IEEE transactions on information forensics and security*, **2008**, *1*, 101–117.
49. Lillesand, T.; Kiefer, R.W.; Chipman, J. Remote sensing and image interpretation. *John Wiley & Sons*, **2015**.
50. Schulter, S.; Leistner, C.; Bischof, H. Fast and accurate image upscaling with super-resolution forests, Proceedings of the IEEE conference on computer vision and pattern recognition, 3791–3799, 2015.
51. Chang, H.; Yeung, D.Y.; Xiong, Y. Super-resolution through neighbor embedding, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 2004.
52. Shi, W.; Caballero, J.; Theis, L.; Huszar, F.; Aitken, A.; Ledig, C.; Wang, Z. Is the deconvolution layer the same as a convolutional layer?. *arXiv preprint*, arXiv:1609.07009.
53. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, **2017**, *26*, 3142–3155.
54. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network, Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, USA. 2017.
55. Tong, T.; Li, G. Liu, X.; Gao, Q. image super-resolution using dense skip connections, Proceedings of the IEEE international conference on computer vision, Venice, Italy, 4799–4807, 2017.
56. Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration, Proceedings of the IEEE international conference on computer vision, Venice, Italy, 4539–4547, 2017.
57. Yang, X.; Mei, H.; Zhang, J.; Xu, K.; Yin, B.; Zhang, Q.; Wei, X. Drfn: Deep recurrent fusion network for single-image superresolution with large factors. *IEEE Transactions on Multimedia*, **2018**, *21*, 328–337.
58. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and lightweight super-resolution with cascading residual network, Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 252–268, 2018.
59. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA., 1–9, 2015.
60. Muhammad, W.; and Aramvith, S. Multi-scale inception based super-resolution using deep learning approach, *Electronics*, **2019**, *8*, 892.

61. Gu, J.; Lu, H.; Zuo, W.; Dong, C. Blind super-resolution with iterative kernel correction, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA., 1604–1613, 2019.
62. Jin, X.; Xiong, Q.; Xiong, C.; Li, Z.; Gao, Z. Single image superresolution with multi-level feature fusion recursive network. *Neurocomputing*, **2019**, *370*, 166–173.
63. Soh, J.W.; Cho, S.; Cho, N.I. Meta-transfer learning for zeroshot super-resolution, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA., 3516–3525, 2020.
64. Shocher, A.; Cohen, N.; Irani, M. "zero-shot" super-resolution using deep internal learning, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA., 3118–3126, 2018.
65. Agustsson, E.; and Timofte, R. Ntire 2017 challenge on single image super-resolution: Dataset and study, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 126–135, 2017.
66. Liu, B.; Boudaoud, D.A. Effective image super resolution via hierarchical convolutional neural network. *Neurocomputing*, **2020**, *374*, 109–116.
67. Lin, D.; Xu, G.; Xu, W.; Wang, Y.; Sun, X.; Fu, K. Scrsr: An efficient recursive convolutional neural network for fast and accurate image super-resolution. *Neurocomputing*, **2020**, *398*, 399–407.
68. Qiu, D.; Zheng, L.; Zhu, J.; Huang, D. Multiple improved residual networks for medical image super-resolution. *Future Generation Computer Systems*, **2021**, *116*, 200–208.
69. Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*, arXiv:1409.1556.
70. Ledig, C.; Theis, L.; *et al*., Photorealistic single image super-resolution using a generative adversarial network, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA., 4681–4690, 2017.
71. Nah, S.; Kim, T.H.; Lee, K.M. Deep multi-scale convolutional neural network for dynamic scene deblurring, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA., 2017.
72. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA., 2016.
73. Mao, X.; Shen, C.; Yang, Y.B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems*, **2016**, 2802–2810.
74. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation, Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015.
75. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, **2010**, *33*, 898–916.
76. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M. L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding, Proceedings of the 23rd British Machine Vision Conference (BMVC), 2012.
77. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations, Proceedings of the International conference on curves and surfaces, Avignon, France, 2010.
78. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, Proceedings of the 8th IEEE International Conference on Computer Vision (ICCVs), Vancouver, BC, Canada, 2001.
79. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars, Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 5197–5206, 2015.
80. Matsui, Y.; Ito, K.; Aramaki, Y.; Fujimoto, A.; Ogawa, T.; Yamasaki, T.; Aizawa, K. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, **2017**, *76*, 21811–21838.
81. Kingma, D.P.; and Adam, J.L.B. A method for stochastic optimization, *Computing Research Repository (CoPR)*, **2015**,arXiv:1412.6980.