PLOS ONE

RESEARCH ARTICLE

Multi-scale Xception based depthwise separable convolution for single image super-resolution

Wazir Muhammad[1], **Supavadee Aramvith**[2]*, **Takao Onoye**[3]

1 Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand, **2** Multimedia Data Analytics and Processing Research Unit, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand, **3** Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan

* supavadee.a@chula.ac.th

Citation: Muhammad W, Aramvith S, Onoye T (2021) Multi-scale Xception based depthwise separable convolution for single image super-resolution. PLoS ONE 16(8): e0249278. https://doi.org/10.1371/journal.pone.0249278

Editor: Yan Chai Hum, University Tunku Abdul Rahman, MALAYSIA

Received: October 8, 2020

Accepted: March 15, 2021

Published: August 23, 2021

Data Availability Statement: All relevant data are within the paper.

Funding: The Second Century Fund (C2F), Chulalongkorn University, Bangkok, Thailand.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The main target of Single image super-resolution is to recover high-quality or high-resolution image from degraded version of low-quality or low-resolution image. Recently, deep learning-based approaches have achieved significant performance in image super-resolution tasks. However, existing approaches related with image super-resolution fail to use the features information of low-resolution images as well as do not recover the hierarchical features for the final reconstruction purpose. In this research work, we have proposed a new architecture inspired by ResNet and Xception networks, which enable a significant drop in the number of network parameters and improve the processing speed to obtain the SR results. We are compared our proposed algorithm with existing state-of-the-art algorithms and confirmed the great ability to construct HR images with fine, rich, and sharp texture details as well as edges. The experimental results validate that our proposed approach has robust performance compared to other popular techniques related to accuracy, speed, and visual quality.

1 Introduction

Single image super-resolution (SISR) is more attractive in recovering the high-resolution (HR) output image from a degraded version of a low-resolution (LR) input image generating by a cheaper cost imaging framework within the limited environmental conditions. Recently, SISR, is a very interesting research space in the area of image and computer vision tasks, which is extensively applied in various applications such as; an object detection [1, 2], image segmentation [3, 4] and image classification [5, 6] purposes.

The better performance and higher accuracy of SISR have been encouraged in the area of an image, especially in medical imaging [7–9], face detection and recognition [10, 11], a high-definition television (HDTV) [12], video surveillance [13], satellite imaging [14] and autonomous driving technology [15, 16], where rich details information is greatly desired. Though, image SR is a highly challenging ill-posed inverse problem. Recently, a number of SISR approaches have been discussed to resolve the ill-posed inverse problem. These approaches

can be subdivided into interpolation-based approaches (mostly employed as a pre-processing step to reconstruct the HR image), reconstruction-based approaches, and learning-based approaches. The interpolation-based approaches included as nearest neighbor-based interpolation [17], cubic interpolation [18] and edge guided interpolations. Although, above approaches are simple, and easy to implement, yet they suffer from accuracy shortcomings and are generating the jagged ringing artifacts. Reconstruction-based image super-resolution approaches [19–24] are mostly adopted previous information to narrow-down of the feasible solution which can get the benefit of reconstructing the fine details of edges and suppress the statistical noise effects [25]. However, these methods are time-consuming and rapidly degrading image reconstruction performance on 4× or 8× scale factor enlargements. Learning-based image SR methods are brought into focus by researchers due to outstanding performance and fast computation. Usually, such types of methods are using machine learning approaches to evaluate the relation between a low-resolution and a high-resolution input and output images during the training samples. Chang et al. [26] introduced the concept of neighbor embedding to take the benefit of similar patches generated locally for reconstructing the output of HR image from an input LR image patches. The researchers also used the idea of sparse signal recovery theory [27] and introduced the concept of sparse coding methods [14, 26–30] to solve the SISR problem. Meanwhile, reconstruction based approaches are combined with learning methods to reduce the jagged ringing artifacts and to improve the blurry results [28–31].

Currently, deep neural networks [32–39] provide significantly improved performance and led to dramatic changes in SISR. Furthermore, deep neural network approaches are very fast and accurate, but still, there are some limitations. However, existing deep convolutional neural network model stacked the convolution layer, side by side, to create the deeper network architecture, which leads to increase the computational cost and introduces the vanishing gradient problem during the training. Besides, a bicubic interpolation technique is used in existing deep convolutional neural network approaches as a step of pre-processing to upscale the low-resolution input image and incurs the new noises in the model. For the purpose of solving such issues and improving the quality of the LR image, we proposed a Multi-scale Xception Based Depthwise Separable Convolution for Single Image Super-resolution (MXDSIR) to generate the HR output image from the original LR input image.

In short, our key contributions are three folds across this paper:

- Inspired by the ResNet and Xception networks, we replaced regular convolution blocks with depthwise separable convolution blocks to achieve faster convergence during the period of training and to stop the vanishing gradient problem as well as easing the training complexity.
- The Rectified Linear Unit (ReLU) was replaced with the Parametric Rectified Linear Unit (PReLU) to activate the dead features, due to zero gradients.
- We introduced the new Xception block, which can detect the different image features information for rebuilding the HR image.

The remaining section is structured as follows. Section 2 presents a related work of image SR approaches. Section 3 and 4 explain our proposed method and its experimental results. Section 5 explained the conclusion.

2 Related work

The target of SISR image is to construct the visually pleasing HR output image. The first concrete deep learning-based approach for the SISR problem was suggested by Dong et al. [40]

known as Super-Resolution Convolutional Neural Network (SRCNN) [40] and presented significant improvements over all previous SR methods. SRCNN [40] model used three convolution layers to predict the HR image. Wang et al. [41] introduced the sparse prior deep convolutional neural networks for image SR based approach, named as Sparse Coding Network (SCN) [41]. The performance of SCN [41] is better than SRCNN [40]. The major drawback of SCN [41] is the high computational complexity and also hinders its applications in real-time processing scenarios.

Dong et al. [42] proposed the improved and faster version of SRCNN [41] architecture to accelerate super-resolution image reconstruction, known as Fast Super-Resolution Convolutional Neural Network (FSRCNN) [42]. FSRCNN [42] has a modest network architecture, that depends on four CNN layers and one deconvolution layer for upsampling purposes and using the original input LR images without interpolation techniques. FSRCNN [42] has lower computational complexity and better performance as compared to SRCNN [41] but has a limited network capacity.

A very deep SR network (VDSR) [32] was proposed by Kim et al. [32] who was inspired by the Visual Geometry Group Network (VGG-net) implemented in the ImageNet for classification purpose [5]. VDSR [32] network reported the significant performance improvement over the SRCNN [41] network using the 20 CNN trainable layers. In order to ease the training complexity of a deeper model, they have used the global residual learning with a fast convergence rate. However, VDSR [32] network architecture does not use the actual pixel values but used the interpolated upscaled version of the image, which leads to more memory consumption and heavy computational cost. Kim et al. [33] proposed a Deeply Recursive Convolutional Network for image super-resolution (DRCN) [33] and uses the convolution layers multiple times. The key advantage of DRCN [33] is to fix the number of training parameters, although there are many number of recursions, the main deficiency is to slow the training process. The authors similarly used the skip connection with a recursive manner to optimize model performance. Mao et al. [43] extended the concept of residual type architecture and proposed Residual Encoder-Decoder Networks (RED) [43]. The RED [43] model used residual learning with symmetric convolution operation, which is trained on 30 layers and achieves the best performance. Therefore, such studies replicate the concept of "the Deeper the Better".

Lai et al. [44] proposed a different network architecture for image SR is known as a deep Laplacian Pyramid Super-Resolution Network (LapSRN) [44], to generate the HR image. LapSRN [44] architecture depends on the different levels of the pyramid and each pyramid level is caused by a deconvolution layer as an upsample, but having the problem in scaling factor (fixed integer), which limits the flexibility of the model. Zhang et al. [45] suggested the denoising convolutional neural networks (DnCNNs), to accelerate the improvement of very deep neural network types architectures. DnCNN [45] follows the same architecture as SRCNN [40] and stacked the CNN with batch normalization (BN) layers followed by the ReLU activation function. Although the model provides favorable results, they are computationally expensive due to the use of the batch normalization layer. Zhao et al. [46] proposed a more flexible scaling factor to super-resolved the input LR image named as a gradual upsampling network (GUN) [46]. For Upsampling purposes the GUN [46] network architecture used the bicubic interpolation technique.

Tai et al. [47] introduced the idea of the deep recursive residual network (DRRN) [47] with 52 CNN layers. The authors introduced a stable training process for a deeper network with parallel architecture. Ledig et al. [34] employ a deep residual connection with 16 blocks using skip-connection to recover the upscaled version of the image. Lim et al. [48] proposed a method to develop deep SR architecture to increase the training efficiency of a model by eliminating the BN layers and their method to win the NTIRE2017 SR challenge [49]. Meanwhile,

Tai et al. [50] suggested the deepest model, known as a persistent memory Network for image restoration purposes (MemNet) [50], in which multiple memory blocks are stacked to obtain persistent memory. Yamanaka et al. [51] presented a combined architecture of skip connection layers and parallelized CNN layers for development of a deep learning-based architecture for SISR and used mainly two networks, the first network is utilized for extracting the features of different levels and the second is the image reconstruction type network. This model is shallower than VDSR [32].

Han et al. [52] proposed the idea of Dual-State Recurrent Network (DSRN) [52], which exchanges the information from LR to the HR state. At each state, they update the signal information and then transmit to the HR state. Li et al. [53] used an adaptive feature detection process to obtain the features fusion at different scales, named as a multi-scale residual network [53]. This approach used the complete hierarchical type of feature information to reconstruct an accurate image super-resolution. Ahn et al. [54] proposed scale-specific upsampling type modules with multiple shortcut connections to learn residuals in LR feature space and to handle the multi-scale information with appropriate specific pathways. Zhang et al. [55] took a concatenated version of the low-resolution image with its degradation mapping type architecture named as super-resolution network for multiple degradations (SRMD) [55].

Wang et al. [56] introduced a dilated CNN network to enhance a receptive field without increasing the size of the kernel. The relative size of the receptive field increases in the case of shallow network type architecture. In dilated convolutional network for SR (DCNSR) [56] uses 12 layers to extract the contextual information efficiently. In [57], the authors proposed End-to-End Image SR via Deep and Shallow (EEDS) [57] CNN architecture and to replace the bicubic interpolation upsampling with the transposed upsampling layer. The HR image is obtained from deep as well as shallow branch simultaneously. Yang et al. [58] suggested a deep recurrent fusion network (DRFN) [58] for image super-resolution, which used the transposed convolution layer with large scale factors. Su et al. [59] proposed a novel type structure, that consists of several sub-networks for reconstructing the HR image progressively. In each subnetwork, the input shall be utilized with the LR feature map and transposed convolution output will be fused with residuals to get the finer one. Wang et al. [60] solves the problem of single image SR using Heaviside Function with iterative refinement. The authors used the binary classification of images to reconstruct the HR image.

Hung et al. [61] proposed a super-sampling network (SSNet) [61] type architecture, which used depthwise separable convolution for image SR. In this architecture a number of parameters as well multiple operations can be significantly reduced by depthwise separable convolution technique. Barzegar et al. [62] introduced a small architecture to prevent the training problem in the deeper model. The design of a DetailNet architecture in such a way, that LR image information can be increased by any approach, then pass through main architecture to boost the perceptual quality of LR image. Hsu et al. [63] inspired by the capsule neural network to extract more potential features information for image SR. In this work authors designed two networks Capsule Image Restoration Neural Network and the Capsule Attention and Reconstruction Neural Network (CARNN) [63] for image SR. The CARNN [63] network generates super-resolution features information efficiently. Liu et al. [64] proposed a new hierarchical convolutional neural network (HCNN) [64] architecture for SR purpose and to learn the features information at different stages. In this approach, the authors have used a three-step hierarchical process, which depends on the extraction of the edge branch, a branch of edge reinforcement, and the SR image reconstruction branch. Muhammad et al. [65] proposed multi-scale inception based super-resolution using a deep learning approach (MSISRD) [65] for image reconstruction. In this approach, the authors used the concept of asymmetric

convolution operation to enhance the computational efficiency of the model and finally used the inception block to reconstruct the multiscale feature information for image SR.

Tian et al. [66] resolve the problem of instability during the training and proposed the new network architecture known as Coarse-to-fine CNN for SISR (CFSRCNN). The proposed network architecture consists of feature extraction, enhancement, construction and refinement of blocks to learn the robust image super-resolution model. The stacked feature extraction blocks are used to learn the short as well as long path features, and then finally fuses the learnable features by expending the effect of a shallow to deeper network to enhance the representing of the features.

Qiu et al. [67] proposed the method of multiple improved residual network (MIRN) image SR network architecture. In this network architecture deep residual network with different levels of skip connection is used to resolve the lack of correlation between the information of adjacent CNN layers. Stochastic gradient descent method (SGD) is used to train the MIRN network architecture. Lan et al. [68] proposed the new dense lightweight network architecture known as fast and lightweight network for SISR. This method addresses the problem of feature extraction and feature correlation learning.

The deep CNN based image SR network architectures used an excessive amount of CNN layers and parameters. Usually, used high computational cost and more memory consumption for training a SR model. To resolve these problems Tian et al. [69] proposed the lightweight enhanced super-resolution based SRCNN known as (LESRCNN). In this approach authors are used the three types of successive blocks as an information extraction, enhancement, and reconstruction block with information refinement block.

3 Proposed method

In this section, we have discussed comprehensive details regarding our proposed network architecture for image SR based on ResNet and Xception blocks. Like the existing SISR methods, our proposed method is classified into five stages namely feature extraction, shrinking, upsampling, expanding, and multi-scale reconstruction, as shown in Fig 1.

3.1 Feature extraction

This part is similar to the previous methods but different from the input image. However, majority of the previous deep convolutional neural network type SISR approaches extract the features information from a bicubic interpolated upsampled version of the HR image. It is important to note that the bicubic interpolation technique damages vital information of LR image and introduces new noise in the model [57, 70]. In contrast, we have used an alternative strategy in our proposed model for extracting the features information directly from the LR image without using interpolation techniques.

Our initial feature extraction stage consists of one convolution layer and two ResNet Blocks with skip connection followed by Parametric Rectified Linear Unit (PReLU) [71] activation function. The said stage extracts the low, middle, and high-level features of information simultaneously. Inspired by VDSR [32], we have used one convolution layer of filter size 3 × 3 with 64 number of filters accompanied by the Parametric Rectified Linear Unit (PReLU) [71]. Mathematically, the convolution layer can be explained as:

$$F_l(Y) = PReLU(W_l * F_{l-1}(Y) + B_l, \quad (1)$$

where F_l denoted the resultant output features map, B_l denoted the biases of l^{th} layer.

$$F_l = max(0, W_l * F_{l-1} + b_l), \quad (2)$$

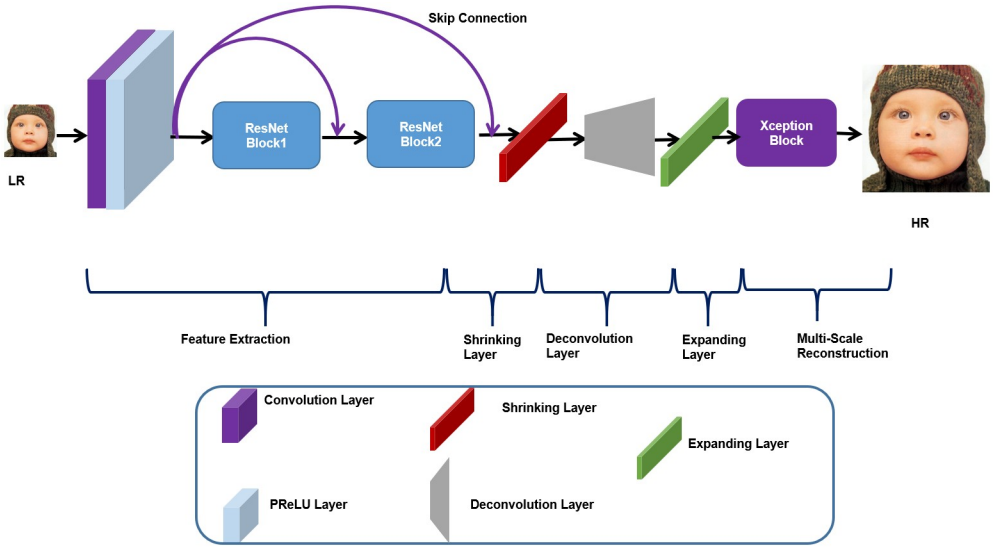

Fig 1. Proposed network architecture of Xception based single image super-resolution reconstruction.

https://doi.org/10.1371/journal.pone.0249278.g001

where W_l are the weights of the filter and b_l are the biases of the l^th layers, respectively. The output of the features map is denoted by F_l and "$*$" represents the convolution operation. The W_l supports $n_l \times f_l \times f_l$ number of parameters, where, f_l indicates the filter size, n_l represents number of filters. The CNN layer and ResNet blocks have the same sizes of $3 \times 3 \times c$ of kernels which generate the "c" features map, where "c" represents 64 number of channels.

3.1.1 PReLU. Earlier approaches used the convolution layers or blocks which were followed by the rectified Linear Unit (ReLU), like SRCNN [40] and VDSR [32]. These types of models have a fair response, but results are still not satisfactory, because, in most of the cases ReLU has a constant gradient. Whereas, in the proposed model, we have used the Parametric Rectified Linear Unit (PReLU) [71], which not only resolves the problem of constant gradient but also has a relatively faster speed of convergence during the training. Mathematically, PReLU [71] activation function can be explained as:

$$PReLU(x_i) = max(x_i, 0) + a_i min(0, x_i), \tag{3}$$

where x_i is the activation function of i^{th} layer input image, and the negative coefficient part of PReLU is denoted by a_i, where a_i parameter is used as ReLU for zero value and PReLU for learnable purpose. The main purpose of PReLU is used to avoid the "dead features", which is produced by zero gradients in the ReLU activation function. The resultant output feature maps using PReLU activation function can be written as:

$$F_l(Y) = PReLU(W_l * F_{l-1}(Y) + B_l, \tag{4}$$

where F_l denoted the resultant output features map, B_l denoted the biases of l^{th} layer.

3.1.2 Feature extraction blocks. The layer stacked, side by side, increases the network depth but reduces the transmission of information to the final layers [72]. Resultantly, the vanishing gradient problem arises in the model and the computational cost of the model is increased. He et al. [73] proposed the ResNet blocks to resolve the above-said problems. The ResNet blocks, these days, are extensively used in the deep learning type SISR image to reconstruct the HR image. Furthermore, the deeper ResNet architecture has a superior performance

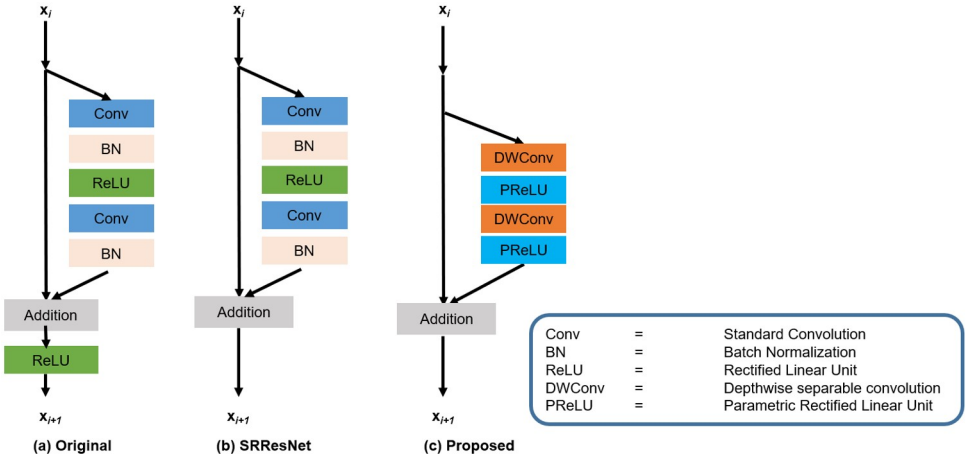

Fig 2. Comparison diagram of different ResNet blocks with the proposed ResNet block. (a) Original ResNet block. (b) SRResNet without final ReLU activation function. (c) Our Proposed ResNet block that removes the BN and replaces the regular convolution and ReLU activation function with depthwise separable convolution followed by the PReLU activation function.

https://doi.org/10.1371/journal.pone.0249278.g002

and is effectively used in the field of image SR [34, 48]. In our proposed method, we have used different residual skip connections which make fast training convergence and reduce the complexity of the model. In Fig 2; we have shown the comparison diagrams of the original residual skip [74] connection, SRResNet [34], and our proposed ResNet block.

The architecture of the ResNet block as expressed in Fig 2(a); uses a direct path and skip connection by way of transmitting the features information and the summed up resultant information followed by the ReLU activation function. SRResNet [34] block as indicated in Fig 2(b); uses the alternative strategy to remove the ReLU activation function and provides a simple and clear path from one block to another. Fig 2(c); shows our proposed ResNet block, which eliminates the Batch Normalization (BN) [74] layers for improving the efficiency of the Graphics Processing Unit (GPU) memory card and enhances the computational efficiency of the model. Furthermore, we replace the operation of regular convolution with depthwise separable convolution followed by point wise convolution and ReLU activation function with PReLU. The PReLU is used to avoid the problem of vanishing gradient and to reduce the training complexity as well as enhances the efficiency of the block. For the middle and the high-level feature extraction, we applied 2 ResNet blocks, each block consists of two 3×3 depthwise separable convolution kernels with 64 filters followed by PReLU nonlinearity.

3.2 Shrinking layer

If more features are directly applied to the transpose convolution layer, it will led to increase in computational cost as well as in size of the model. However, we have employed a one CNN layer as a shrinking layer before the deconvolution layer. This type of arrangement has also been observed in the latest convolutional neural network architectures for computer vision applications. Authors, proposed in [57, 65, 75] are using a shrinking layer for increasing the computational efficiency of the model.

3.3 Deconvolution layer

Researchers have suggested in [40, 57, 76] that the purpose of upscaling the LR image resolution before the initial layer is to increase the computational cost and damage critical

information due to the fact that the processing speed is directly dependent on resolution of the image. Furthermore, the use of upscaled techniques before the initial layer does not provide additional information, however, introduces the jagged ringing artifacts in the SR image. We propose for generating the high-resolution image directly from the actual low-resolution feature domain to handle these types of problems. For this purpose, we have applied the deconvolution layer as an upscaling operation before the Xception block. The size of the deconvolution layer is 16×16 of stride that is equal to enlargement factors.

3.4 Expanding layer

The expanding layer performs the inverse operation of a shrinking layer and produces the HR image more accurately. Furthermore, if the HR image is directly reconstructed from LR features, the final reconstruction quality of the image will be poor. Therefore, after the deconvolution layer, we are applying the expanding layer to recover the original feature's information smoothly.

3.5 Multi-scale reconstruction

3.5.1 Depthwise separable convolution. Originally, depthwise separable convolution was proposed by Sifre [77] and was applied for image classification purposes. Factorizing a convolution operation is a form of depthwise separable convolution in which it converts regular convolution operation into a depthwise separable convolution operation followed by a pointwise convolution operation. The separable convolution operation performs a single filter per channel input and finally combines the linear input channels. The convolution process substitutes a factorized convolution layer with two layers; one is used for space filter, and the other is used for combining purposes. Thus, the depthwise separable convolution will sufficiently lessen both the number of parameters and size of the model. The regular type of convolution kernel takes three parameters such as; height (h), width (w), and input channel (c_{in}) of an input feature map (I). The resultant convolution layer ($h \times w \times c_{in}$) is applied as $K \times K \times c_{in} \times c_{out}$, where c_{out}, is the number of output channels. The depthwise separable convolution depends on two convolution operations: depthwise separable convolution operation and pointwise convolution operation. Mathematically, the depthwise separable convolution operation can be written as:

$$G(y, x, j) = \sum_{u=1}^{k} \sum_{v=1}^{k} K(u, v, j) \times I(y + u - 1, x + v, j), \tag{5}$$

where K represents the kernels of depthwise separable convolution operation of size $K \times K \times c_{in}$. The n^{th} filter in the kernel K is applied on the n^{th} number of channels in the input feature map of I to reconstruct the G output feature map. While reconstructing new features, we apply the pointwise convolution. Mathematically, the pointwise convolution can be written as:

$$O(y, x, l) = \sum_{j=1}^{c_{in}} G(x, y, j) \times P(j, l), \tag{6}$$

where the size of the kernel of pointwise convolution operation is $1 \times 1 \times c_{in} \times c_{out}$.

3.5.2 Xception block. In the final phase, we have employed a multi-scale Xception block that stands for a multi-scale Extreme version of Inception block, which is adopted from GoogLeNet [78] with a modified depthwise separable convolution better than Inception v-3 [79]. Multi-scale Xception block is used to choose the correct kernel size, as kernel size performs a pivotal role in model design, training procedure, and multi-scale reconstruction purposes. The

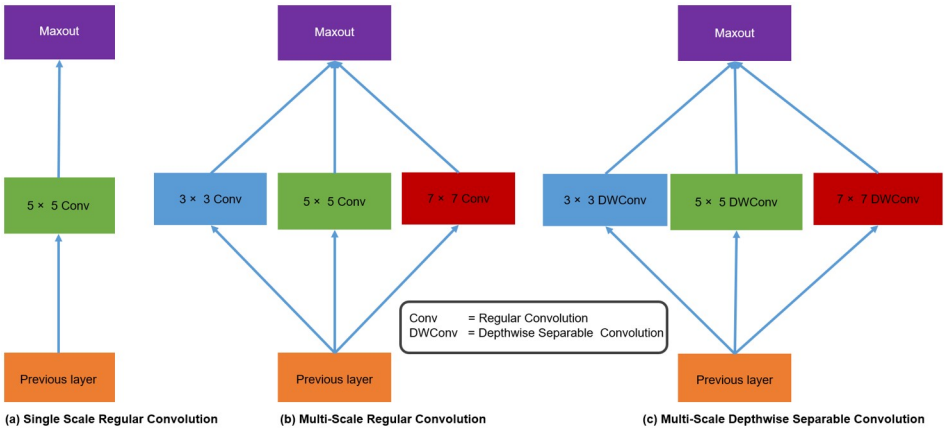

Fig 3. Comparison of a single scale, multi-scale regular and depthwise separable convolution blocks. (a) Single Scale Regular Convolution (b), Multi-Scale Regular Convolution, and (c) Multi-Scale Depthwise Separable Convolution (Our proposed).

https://doi.org/10.1371/journal.pone.0249278.g003

larger size of the kernel is more suitable, when the features information is distributed globally, whereas the smaller size of the kernel is better, when features information is distributed locally. The Xception architecture employs this concept and includes more depthwise separable convolution on kernels of various sizes. Fig 3(a); shows a single scale regular convolution plain type of architectures, in which several convolution layers are stacked in a single straight-line path. Such type of architectures are implemented by a well-known image super-resolution methods, like SRCNN [40] and FSRCNN [42]. These types of architectures are easy in implementation, however, deeper network architecture has more memory consumption and enhances the network depth of the model. Fig 3(b); uses the regular convolution type inception block to extract the multi-scale feature information efficiently. Fig 3(c); shows our proposed block of multi-scale depthwise separable convolution. The proposed Xception block consists of different depthwise separable convolution kernel sizes, like 3×3, 5×5, and 7×7 followed by pointwise convolution with PReLU activation function, to reconstruct the SR image.

4 Experimental results

In this section, initially, we discuss the selection procedure of training and testing datasets with hyper-parameters. The training as well as testing datasets were downloaded from Kaggle website [80]. Afterwards, we have evaluated the quantitative as well as the qualitative performance in terms of PSNR/SSIM [81] and perceptual vision quality on five test datasets which are publicly available. Finally, we have compared the computational cost and processing speed of our proposed model in terms of PSNR versus the running time and network depth (number of K parameters).

4.1 Training datasets

The various sizes of the image datasets have been available for the training purposes to train the model for single image super-resolution. Yang et al. [23] and the Berkeley Segmentation Dataset (BSD300) [82] are commonly used image datasets, because these datasets are used by well-known SR methods, like VDSR [32], DRCN [33] and LapSRN [44] for the training purpose. In order to enhance the training dataset, data augmentation technique has been applied in terms of rotation and flipping. All the experimental evaluations were done on the original

image and for data manipulation purposes, we used a programming language python 3.7.9, deep learning Keras 2.1.5 library supported back-end as Tensor Flow and PyTorch version 1.6.0. Various types of loss functions were also available to evaluate model performance. Deep learning-based CNN SR architecture has mostly used the mean square error (MSE) as the loss function. So, we have also used similar type of loss with our proposed method. Mathematically, the loss function may be calculated as:

$$L(\theta) = \frac{1}{m}\sum_{i=1}^{m} F((Y_i;\theta) - X_i)^2, \tag{7}$$

where $F(Y_i, \theta)$ is the recovered output image, X_i is the high-quality HR images, Y_i is corresponding the low quality image, and the number of small size batches is the m in the training. In the training phase, we have used an adaptive momentum estimation optimizer (Adam) [83] having a 0.0004 initial learning rate with mini-batch size of 16. The process of training takes 200 epochs to converge the model properly. We train our model on a NVIDIA GeForce RTX2070 GPU, having 2.6 GHz Ci7-9750H CPU with 16 GB RAM under the Windows 10 operating system's environment.

4.2 Testing datasets

We have assessed the performance of proposed network architecture on five standard datasets. The Set5 [84] dataset comprises of five images having different sizes like 228 × 228 and 512 × 512 pixels. The Set14 [85] images consist of different sizes of fourteen images. BSD100 [82] test dataset depends on 100 different natural scenes of images. Urban100 [86] is the challenging test image dataset having different frequency bands with detailed information. Manga109 [87] test image dataset depends on different comic type images with fine structures.

4.3 Implementation details

Under the Windows 10 operating system environment, our proposed approach was trained and tested with NVIDIA GeForce RTX2070 GPU with 16 GB RAM. We have trained our model on the scale enhancement factor of 2×, 4×, and 8× in Keras 2.1.5, PyTorch 1.6.0 and MATLAB 2018a framework.

4.4 Comparison with other state-of-the-art-methods

We compare the performance of our MXDSIR SR method with ground-truth HR image, including baseline method (Bicubic interpolation) and twelve other state-of-the-art methods are A+ [88], RFL [89], SelfExSR [86], SCN [41], SRCNN [40], FSRCNN [42], VDSR [32], DRCN [33], LapSRN [44], DRRN [47], MemNet [50], and MSISRD [65] by both objective PSNR/SSIM [81] and subjective measures. The summary of quantitative evaluation performed on five benchmark datasets as shown in Table 1. We can observe from Table 1, that our model achieves the best quantitative results in terms of PSNR/SSIM on enlargement factor 2× and 8×. The maximum and minimum range of the average PSNR improvement on scale factor 2× is 0.13dB to 4.12dB. Similarly, we also used another quality matrix to evaluate the performance of our proposed model is the SSIM. The minimum and maximum average range of SSIM improvement on scale factor 2× are in the range of 0.001 to 0.05. In the enlargement factor 4×, our model achieves the second-best performance as compared to other existing methods, though DRRN [47] and MSISRD [65] are the most comparable, but these models incur a higher computational complexity as they have more model parameters. Finally, our minimum and maximum improvement on challenging enlargement factor 8×, our range of the

Table 1. Presents benchmark results of the average value of PSNR/SSIM [81] for enlargement factor 2×, 4×, and 8× on Set5 [84], Set14 [85], BSD100 [82], Urban100 [86], and Manga109 [87] test datasets. Bold indicated results with red colors are the **best** values. The underlined results with blue color are second − best values.

Method	Scale	Para	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM	Average PSNR/SSIM
Bicubic	2×	-/-	33.69/0.931	30.25/0.870	29.57/0.844	26.89/0.841	30.86/0.936	30.52/0.884
A+ [88]	2×	-/-	36.60/0.955	32.32/0.906	31.24/0.887	29.25/0.895	35.37/0.968	32.96/0.922
RFL [89]	2×	-/-	36.59/0.954	32.29/0.905	31.18/0.885	29.14/0.891	35.12/ 0.966	32.86/0.920
SelfExSR [86]	2×	-/-	36.60/0.955	32.24/0.904	31.20/0.887	29.55/0.898	35.82/0.969	33.08/0.923
SCN [41]	2×	42	36.58/0.954	32.35/0.905	31.26/0.885	29.52/0.897	35.51/0.967	33.04/0.922
SRCNN [40]	2×	57	36.72/0.955	32.51/0.908	31.38/0.889	29.53/0.896	35.76/0.968	33.18/0.923
FSRCNN [42]	2×	12	37.05/0.956	32.66/0.909	31.53/0.892	29.88/0.902	36.67/0.971	33.56/0.926
VDSR [32]	2×	665	37.53/0.959	33.05/0.913	31.90/0.896	30.77/0.914	37.22/0.975	34.09/0.931
DRCN [33]	2×	1775	37.63/0.959	33.06/0.912	31.85/0.895	30.76/0.914	37.63/0.974	34.19/0.931
LapSRN [44]	2×	812	37.52/0.959	33.08/0.913	31.80/ 0.895	30.41/0.910	37.27/0.974	34.02/0.930
DRRN [47]	2×	297	37.74/0.959	33.23/0.914	32.05/**0.897**	31.23/**0.919**	37.92/**0.976**	34.43/0.933
MemNet [50]	2×	677	37.78/0.959	33.28/0.914	32.08/**0.897**	31.31/**0.919**	37.72/0.974	34.43/0.933
MSISRD [65]	2×	240	37.80/**0.960**	33.84/**0.920**	32.09/0.895	31.10/0.913	37.70/0.975	34.51/0.933
MXDSIR	2×	222	**37.93**/0.959	**33.87**/**0.920**	**32.12**/**0.897**	**31.33**/0.918	**37.93**/**0.976**	**34.64**/**0.934**
Bicubic	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+ [88]	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 [89]	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 [86]	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×	42	30.41/0.863	27.39/0.751	26.88/0.711	24.52/0.726	27.39/0.857	27.32/0.782
SRCNN [40]	4×	57	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	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 [32]	4×	665	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 [33]	4×	1775	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 [44]	4×	812	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 [47]	4×	297	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 [50]	4×	677	31.74/**0.889**	28.26/**0.772**	27.40/**0.728**	25.50/0.763	29.42/0.894	28.46/0.809
MSISRD [65]	4×	240	31.62/0.886	28.51/0.771	27.33/0.727	25.42/0.757	**31.61**/0.891	**28.90**/0.806
MXDSIR	4×	222	**32.37**/0.888	**28.63**/**0.772**	**27.45**/**0.728**	**25.54**/0.763	30.21/0.895	28.84/0.809
Bicubic	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+ [88]	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 [89]	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 [86]	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×	42	25.59/0.706	24.02/0.603	24.30/0.573	21.52/0.560	22.68/0.701	23.62/0.629
SRCNN [40]	8×	57	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	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 [32]	8×	665	25.93/0.724	24.26/0.614	24.49/0.583	21.70/0.571	23.16/0.725	23.91/0.643
LapSRN [44]	8×	812	26.15/0.738	24.35/0.620	24.54/0.586	21.81/0.581	23.39/0.735	24.05/0.652
MemNet [50]	8×	677	26.16/**0.741**	24.38/0.619	24.58/0.584	21.89/**0.582**	23.56/0.738	24.11/0.653
DRCN [33]	8×	1775	25.93/0.723	24.25/0.614	24.49/0.582	21.71/0.571	23.20/0.724	23.92/0.643
MSISRD [65]	8×	240	26.26/0.737	24.38/0.621	24.73/0.586	22.53/**0.582**	23.50/0.738	24.28/0.653
MXDSIR	8×	222	**26.31**/0.740	**24.42**/**0.622**	**24.77**/**0.587**	**22.91**/**0.582**	**23.63**/**0.739**	**24.41**/**0.654**

https://doi.org/10.1371/journal.pone.0249278.t001

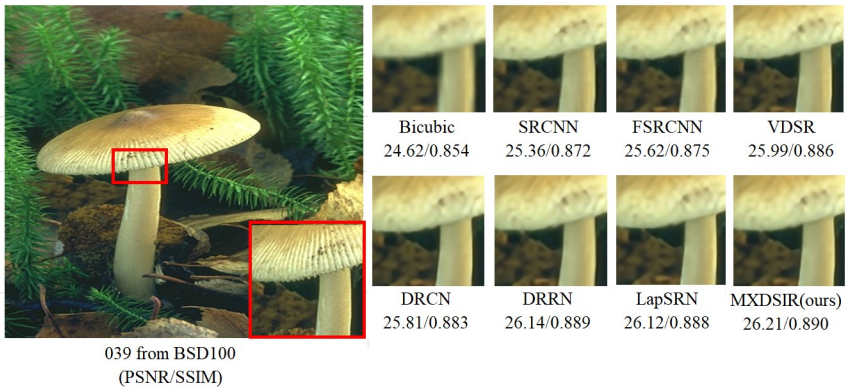

Fig 4. Visual performance of images with 4× enlargement factor of image 039 from BSD100 dataset.

https://doi.org/10.1371/journal.pone.0249278.g004

improvement in terms of average PSNR is 0.13dB to 1.73dB. Similarly, our model achieves minimum and maximum average SSIM improvement is 0.003 to 0.082.

Apart from the quantitative comparison, the qualitative performance of our method and existing state-of-the-art methods are shown in Figs 4–8, were obtained from Huang [86] (https://github.com/jbhuang0604/SelfExSR) and [90] PLOS ONE Journal (https://doi.org/10.1371/journal.pone.0241313.g007).

From these images clearly observed that the baseline bicubic method cannot reconstruct any extra details information, but introduce the new noises in the image as well as more blurry results especially on enlargement scale factor 4× and 8×. The deep learning based image super-resolution approach, like SRCNN [40], FSRCNN [42] and VDSR [32] can produce, in some cases, fair reconstruction details from the original LR input image, but still results in blurry image contours due to their model designed in linear fashion (stacked layer side by side). In case of LapSRN [44] as well as family of deeper model, results are fair, but miss some edges and lines, because deeper model only relies on the single scale kernel. As we compare existing deeper model for image SR, our model achieves noticeable improvement in terms of perceptual quality, due to multiscale kernel used in the Xception block. The noticeable improvement observed in Fig 6; especially "080" image from Urban100 has excessive amount of artifacts, but our method produces sharper boundaries and richer textures with less amount of artifacts. Similar artifacts also observed on the image Figs 7 and 8 respectively.

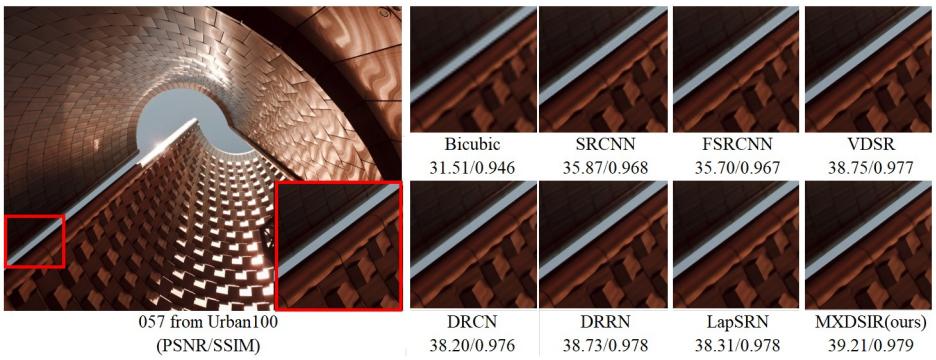

Fig 5. Visual performance of images with 4× enlargement factor of image 057 from Urban100 dataset.

https://doi.org/10.1371/journal.pone.0249278.g005

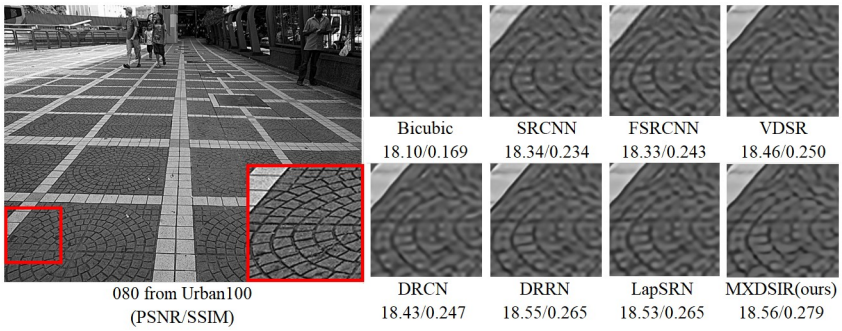

Fig 6. Visual performance of images with 8× enlargement factor of image 080 from Urban100 dataset.

https://doi.org/10.1371/journal.pone.0249278.g006

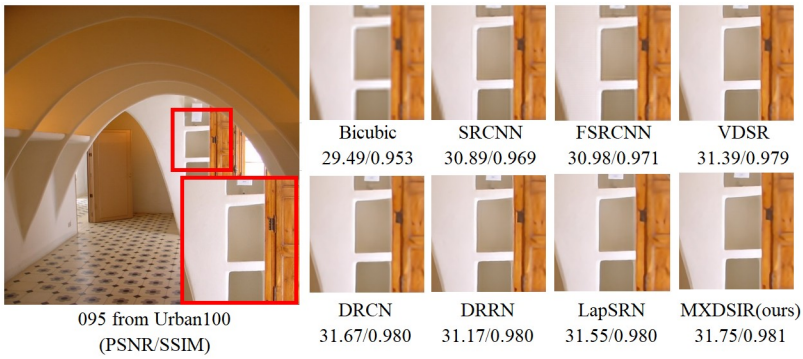

Fig 7. Visual performance of images with 8× enlargement factor of image 095 from Urban100 dataset.

https://doi.org/10.1371/journal.pone.0249278.g007

In summary, our proposed method can achieve better quality improvement measured by PSNR, SSIM index, and visual image quality comparison compared to other methods. In the following sections, our proposed architecture provides a favorable trade-off in terms of computational cost and visual quality improvement.

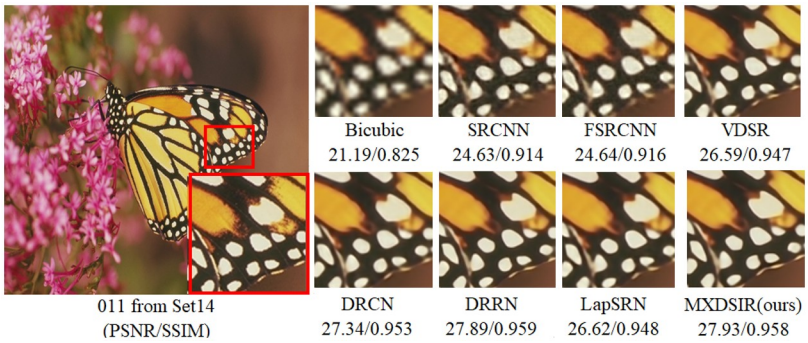

Fig 8. Visual performance of images with 8× enlargement factor of image 011 from Set14 dataset.

https://doi.org/10.1371/journal.pone.0249278.g008

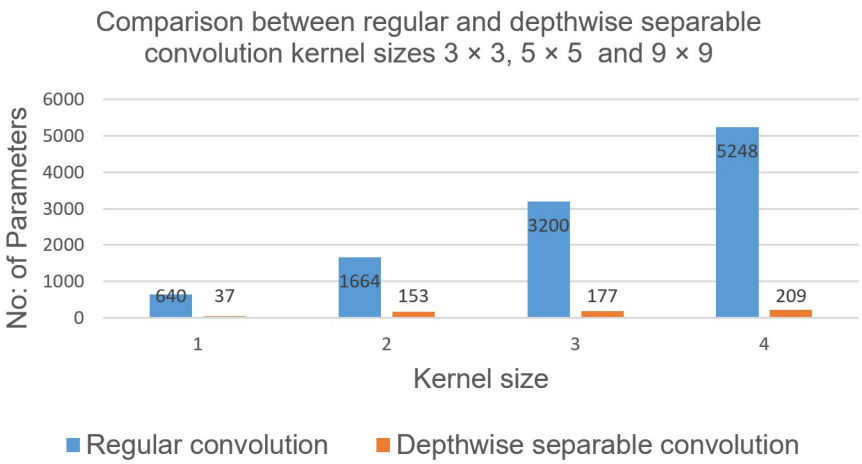

Fig 9. Complexity comparison between the regular convolution kernel versus the depthwise separable convolution kernel.

https://doi.org/10.1371/journal.pone.0249278.g009

4.5 Performance comparison in terms of the kernel size

The size as well as the type of the convolution kernel plays a key role in terms of the model size and computational cost. In Fig 9; we have selected the two different convolution kernels, one is regular convolution kernel and the other is a depthwise separable convolution kernel, with the same 64 number of feature maps. Performance of our proposed depthwise separable convolution kernel is more computationally efficient as compared to the regular convolution kernel.

4.6 Comparison in terms of the number of the model parameters

We have presented the complexity of the model related to network depth (number of parameters) versus PSNR [81] as shown in Fig 10. By using the depthwise separable convolution layer, our proposed model decreases the number of parameters as compared to other publicly available methods. Our MXDSIR method has parameters about 66% less than the VDSR [32], 87%

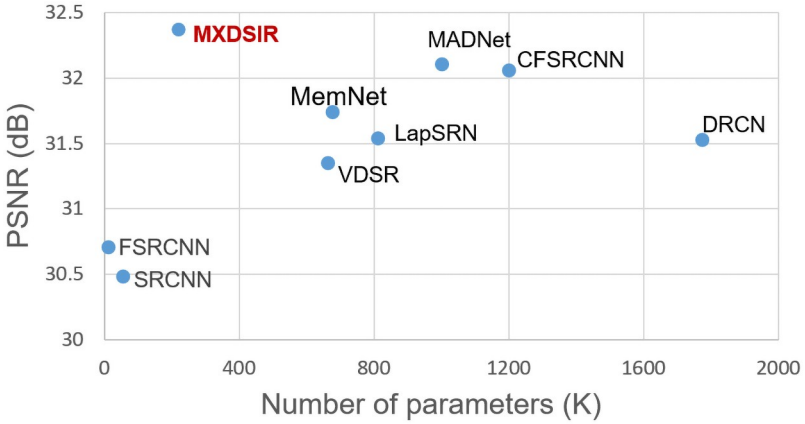

Fig 10. The performance comparison measurement on PSNR [81] versus the depth of the network (number of parameters). The performance results on the Set5 [84] dataset with scale factor 4×.

https://doi.org/10.1371/journal.pone.0249278.g010

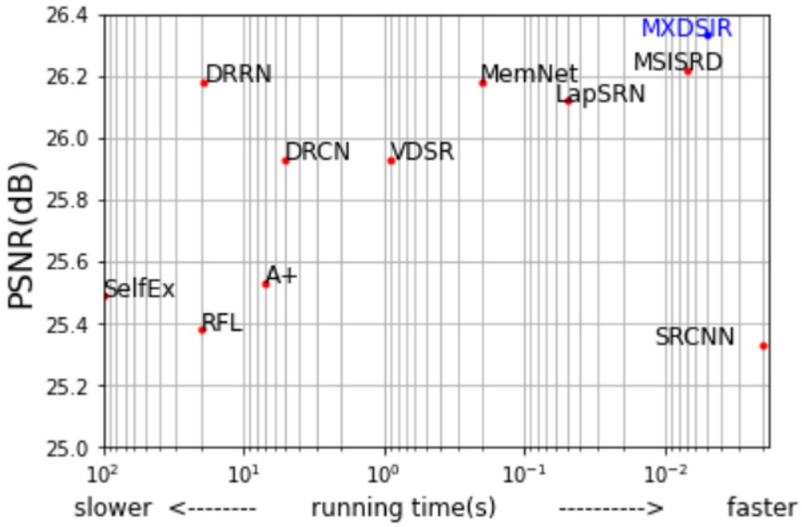

Fig 11. Quantitative comparison between the PSNR [81] performance vs. runtime on Set5 [84] scale 8× enlargement.

https://doi.org/10.1371/journal.pone.0249278.g011

less than the DRCN [33], 72% less than the LapSRN [44], 67% less than the MemNet [50], 74% less than MADNet [68] and 81% less than CFSRCNN [66].

4.7 Quantitative comparison in terms of run time versus PSNR

In this part, as shown in Fig 11; we have evaluated our method in terms of running or execution time versus PSNR [81]. As for the execution of time performance is concerned, we have used the public access codes given by the authors to evaluate the state-of-the-art methods with 2.6 GHz Ci7-9750H CPU 16GB RAM. The comparative analysis between the execution of time and performance on the Set5 [84] dataset for 8× SR reveals that our method is 0.16 dB higher than LapSRN [44] on PSNR [81] and, approximately, 10 times faster than LapSRN [44].

5 Conclusion

In this paper, we have presented fast and computationally efficient Xception based residual CNN network architecture for image SR to extract the features information locally as well as globally from the input LR image, and to generate the HR output image. The proposed network architecture used the two ResNet blocks and three Xception block, which is adopted from the ResNet and GoogLeNet to recover several features during the extraction and reconstruction stages. The proposed technique ensured that the network shows fast convergence speed and low computational cost, by replacing the interpolation technique with the learned transposed convolution layer and regular convolution operation with the depthwise separable convolution. Furthermore, our network architecture is relatively simple and well designed for images and computer vision tasks. Extensive experimental results on different image datasets not only provides satisfactory results on the performance of image SR quantitatively but also have favorable results in terms of complexity and provided visual pleasing quality as compare to the existing state-of-the-art SR methods.

Author Contributions

Conceptualization: Wazir Muhammad, Supavadee Aramvith.

Data curation: Wazir Muhammad.

Formal analysis: Takao Onoye.

Funding acquisition: Supavadee Aramvith.

Investigation: Supavadee Aramvith, Takao Onoye.

Methodology: Wazir Muhammad.

Project administration: Supavadee Aramvith.

Software: Wazir Muhammad.

Supervision: Supavadee Aramvith, Takao Onoye.

Validation: Supavadee Aramvith.

Visualization: Wazir Muhammad.

Writing – original draft: Wazir Muhammad.

Writing – review & editing: Supavadee Aramvith, Takao Onoye.

References

1. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence. 2016; 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031 PMID: 27295650
2. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 779–788.
3. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2961–2969.
4. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017; 39 (12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615 PMID: 28060704
5. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012; 25:1097–1105.
6. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.
7. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging. 2016; 35(5):1299–1312. https://doi.org/10.1109/TMI.2016.2535302 PMID: 26978662
8. Peled S, Yeshurun Y. Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2001; 45(1):29–35. https://doi.org/10.1002/1522-2594(200101)45:1%3C29::AID-MRM1005%3E3.0.CO;2-Z PMID: 11146482
9. Shi W, Caballero J, Ledig C, Zhuang X, Bai W, Bhatia K, et al. Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In: International conference on medical image computing and computer-assisted intervention. Springer; 2013. p. 9–16.
10. Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 815–823.
11. Gunturk BK, Batur AU, Altunbasak Y, Hayes MH, Mersereau RM. Eigenface-domain super-resolution for face recognition. IEEE transactions on image processing. 2003; 12(5):597–606. https://doi.org/10.1109/TIP.2003.811513 PMID: 18237935
12. Goto T, Fukuoka T, Nagashima F, Hirano S, Sakurai M. Super-resolution System for 4K-HDTV. In: 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 4453–4458.

13. Zhang L, Zhang H, Shen H, Li P. A super-resolution reconstruction algorithm for surveillance images. Signal Processing. 2010; 90(3):848–859. https://doi.org/10.1016/j.sigpro.2009.09.002

14. Aplin P, Atkinson PM, Curran PJ. Fine spatial resolution simulated satellite sensor imagery for land cover mapping in the United Kingdom. Remote sensing of Environment. 1999; 68(3):206–216. https://doi.org/10.1016/S0034-4257(98)00112-6

15. Uçar A, Demir Y, Güzeliş C. Object recognition and detection with deep learning for autonomous driving applications. Simulation. 2017; 93(9):759–769. https://doi.org/10.1177/0037549717709932

16. Pelliccione P, Knauss E, Heldal R, Ågren SM, Mallozzi P, Alminger A, et al. Automotive architecture framework: The experience of volvo cars. Journal of systems architecture. 2017; 77:83–100. https://doi.org/10.1016/j.sysarc.2017.02.005

17. Zhao L, Qi W, Li SZ, Yang SQ, Zhang HJ. Content-based retrieval of video shot using the-improved nearest feature line method. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221). vol. 3. IEEE; 2001. p. 1625–1628.

18. Keys R. Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing. 1981; 29(6):1153–1160. https://doi.org/10.1109/TASSP.1981.1163711

19. Dai S, Han M, Xu W, Wu Y, Gong Y, Katsaggelos AK. Softcuts: a soft edge smoothness prior for color image super-resolution. IEEE Transactions on Image Processing. 2009; 18(5):969–981. https://doi.org/10.1109/TIP.2009.2012908 PMID: 19342335

20. Sun J, Xu Z, Shum HY. Image super-resolution using gradient profile prior. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2008. p. 1–8.

21. Yan Q, Xu Y, Yang X, Nguyen TQ. Single image superresolution based on gradient profile sharpness. IEEE Transactions on Image Processing. 2015; 24(10):3187–3202. https://doi.org/10.1109/TIP.2015.2414877 PMID: 25807567

22. Freeman WT, Jones TR, Pasztor EC. Example-based super-resolution. IEEE Computer graphics and Applications. 2002; 22(2):56–65. https://doi.org/10.1109/38.988747

23. Yang J, Wright J, Huang TS, Ma Y. Image super-resolution via sparse representation. IEEE transactions on image processing. 2010; 19(11):2861–2873. https://doi.org/10.1109/TIP.2010.2050625 PMID: 20483687

24. Kim KI, Kwon Y. Example-based learning for single-image super-resolution. In: Joint Pattern Recognition Symposium. Springer; 2008. p. 456–465.

25. Chan TM, Zhang J, Pu J, Huang H. Neighbor embedding based super-resolution algorithm through edge detection and feature selection. Pattern Recognition Letters. 2009; 30(5):494–502. https://doi.org/10.1016/j.patrec.2008.11.008

26. Chang H, Yeung DY, Xiong Y. Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 1. IEEE; 2004. p. I–I.

27. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing. 2006; 54(11):4311–4322. https://doi.org/10.1109/TSP.2006.881199

28. Zhang K, Tao D, Gao X, Li X, Li J. Coarse-to-fine learning for single-image super-resolution. IEEE transactions on neural networks and learning systems. 2016; 28(5):1109–1122. https://doi.org/10.1109/TNNLS.2015.2511069 PMID: 26915133

29. Yu J, Gao X, Tao D, Li X, Zhang K. A unified learning framework for single image super-resolution. IEEE Transactions on Neural networks and Learning systems. 2013; 25(4):780–792.

30. Deng C, Xu J, Zhang K, Tao D, Gao X, Li X. Similarity constraints-based structured output regression machine: An approach to image super-resolution. IEEE transactions on neural networks and learning systems. 2015; 27(12):2472–2485. https://doi.org/10.1109/TNNLS.2015.2468069 PMID: 26357410

31. 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; 18(3):313–325. https://doi.org/10.1109/TMM.2016.2515997

32. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1646–1654.

33. Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1637–1645.

34. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, 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. p. 4681–4690.

35. Muhammad W, Ullah I, Ashfaq M. An Introduction to Deep Convolutional Neural Networks With Keras. In: Machine Learning and Deep Learning in Real-Time Applications. IGI Global; 2020. p. 231–272.

36. Jiang K, Wang Z, Yi P, Jiang J. Hierarchical dense recursive network for image super-resolution. Pattern Recognition. 2020; 107:107475. https://doi.org/10.1016/j.patcog.2020.107475

37. Zhang D, Shao J, Liang Z, Gao L, Shen HT. Large Factor Image Super-Resolution with Cascaded Convolutional Neural Networks. IEEE Transactions on Multimedia. 2020.

38. Zhu L, Zhan S, Zhang H. Stacked U-shape networks with channel-wise attention for image super-resolution. Neurocomputing. 2019; 345:58–66. https://doi.org/10.1016/j.neucom.2018.12.077

39. Li Z, Li Q, Wu W, Yang J, Li Z, Yang X. Deep recursive up-down sampling networks for single image super-resolution. Neurocomputing. 2020; 398:377–388. https://doi.org/10.1016/j.neucom.2019.04.004

40. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence. 2015; 38(2):295–307. https://doi.org/10.1109/TPAMI.2015.2439281

41. Wang Z, Liu D, Yang J, Han W, Huang T. Deep networks for image super-resolution with sparse prior. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 370–378.

42. Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer; 2016. p. 391–407.

43. Mao XJ, Shen C, Yang YB. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. arXiv preprint arXiv:160309056. 2016.

44. Lai WS, Huang JB, Ahuja N, Yang MH. Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 624–632.

45. 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(7):3142–3155. https://doi.org/10.1109/TIP.2017.2662206 PMID: 28166495

46. Zhao Y, Li G, Xie W, Jia W, Min H, Liu X. GUN: Gradual upsampling network for single image super-resolution. IEEE Access. 2018; 6:39363–39374. https://doi.org/10.1109/ACCESS.2018.2855127

47. Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 3147–3155.

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

49. Timofte R, Agustsson E, Van Gool L, Yang MH, Zhang L. Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2017. p. 114–125.

50. Tai Y, Yang J, Liu X, Xu C. Memnet: A persistent memory network for image restoration. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 4539–4547.

51. Yamanaka J, Kuwashima S, Kurita T. Fast and accurate image super resolution by deep CNN with skip connection and network in network. In: International Conference on Neural Information Processing. Springer; 2017. p. 217–225.

52. Han W, Chang S, Liu D, Yu M, Witbrock M, Huang TS. Image super-resolution via dual-state recurrent networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 1654–1663.

53. Li J, Fang F, Mei K, Zhang G. Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 517–532.

54. Ahn N, Kang B, Sohn KA. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 252–268.

55. Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 3262–3271.

56. Wang R, Gong M, Tao D. Receptive field size versus model depth for single image super-resolution. IEEE Transactions on Image Processing. 2019; 29:1669–1682. https://doi.org/10.1109/TIP.2019.2941327

57. Wang Y, Wang L, Wang H, Li P. End-to-end image super-resolution via deep and shallow convolutional networks. IEEE Access. 2019; 7:31959–31970. https://doi.org/10.1109/ACCESS.2019.2903582

58. Yang X, Mei H, Zhang J, Xu K, Yin B, Zhang Q, et al. DRFN: Deep recurrent fusion network for single-image super-resolution with large factors. IEEE Transactions on Multimedia. 2018; 21(2):328–337. https://doi.org/10.1109/TMM.2018.2863602

59. Su M, Lai S, Chai Z, Wei X, Liu Y. Hierarchical Recursive Network for Single Image Super Resolution. In: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE; 2019. p. 595–598.

60. Wang XY, Huang TZ, Deng LJ. Single image super-resolution based on approximated Heaviside functions and iterative refinement. Plos one. 2018; 13(1):e0182240. https://doi.org/10.1371/journal.pone.0182240 PMID: 29329298

61. Hung KW, Zhang Z, Jiang J. Real-time image super-resolution using recursive depthwise separable convolution network. IEEE Access. 2019; 7:99804–99816. https://doi.org/10.1109/ACCESS.2019.2929223

62. Barzegar S, Sharifi A, Manthouri M. Super-resolution using lightweight detailnet network. Multimedia Tools and Applications. 2020; 79(1):1119–1136. https://doi.org/10.1007/s11042-019-08218-4

63. Hsu JT, Kuo CH, Chen DW. Image super-resolution using capsule neural networks. IEEE Access. 2020; 8:9751–9759. https://doi.org/10.1109/ACCESS.2020.2964292

64. Liu B, Ait-Boudaoud D. Effective image super resolution via hierarchical convolutional neural network. Neurocomputing. 2020; 374:109–116. https://doi.org/10.1016/j.neucom.2019.09.035

65. Muhammad W, Aramvith S. Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach. Electronics. 2019; 8(8):892. https://doi.org/10.3390/electronics8080892

66. Tian C, Xu Y, Zuo W, Zhang B, Fei L, Lin CW. Coarse-to-fine CNN for image super-resolution. IEEE Transactions on Multimedia. 2020.

67. 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. https://doi.org/10.1016/j.future.2020.11.001

68. Lan R, Sun L, Liu Z, Lu H, Pang C, Luo X. MADNet: a fast and lightweight network for single-image super resolution. IEEE transactions on cybernetics. 2020.

69. Tian C, Zhuge R, Wu Z, Xu Y, Zuo W, Chen C, et al. Lightweight image super-resolution with enhanced CNN. Knowledge-Based Systems. 2020; 205:106235. https://doi.org/10.1016/j.knosys.2020.106235

70. Li S, Fan R, Lei G, Yue G, Hou C. A two-channel convolutional neural network for image super-resolution. Neurocomputing. 2018; 275:267–277. https://doi.org/10.1016/j.neucom.2017.08.041

71. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 1026–1034.

72. Hui Z, Wang X, Gao X. Fast and accurate single image super-resolution via information distillation network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 723–731.

73. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.

74. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR; 2015. p. 448–456.

75. Lin M, Chen Q, Yan S. Network in network. arXiv preprint arXiv:13124400. 2013.

76. Chen Y, Pock T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE transactions on pattern analysis and machine intelligence. 2016; 39(6):1256–1272. https://doi.org/10.1109/TPAMI.2016.2596743 PMID: 27529868

77. Sifre L, Mallat S. Rigid-motion scattering for texture classification. arXiv preprint arXiv:14031687. 2014.

78. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–1258.

79. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 2818–2826.

80. Kaggle Datasets. https://www.kaggle.com. Accessed: December of; 2020.

81. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing. 2004; 13(4):600–612. https://doi.org/10.1109/TIP.2003.819861 PMID: 15376593

82. Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2010; 33(5):898–916. https://doi.org/10.1109/TPAMI.2010.161

83. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

84. Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel ML. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

85. Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: International conference on curves and surfaces. Springer; 2010. p. 711–730.

86. Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 5197–5206.

87. Matsui Y, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T, et al. Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications. 2017; 76(20):21811–21838. https://doi.org/10.1007/s11042-016-4020-z

88. Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Asian conference on computer vision. Springer; 2014. p. 111–126.

89. Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3791–3799.

90. Xiong Zhengqiang, Lin, et al. Single image super-resolution via Image Quality Assessment-Guided Deep Learning Network. In: PloS one; 2020. https://doi.org/10.1371/journal.pone.0241313 PMID: 33119656