LDCSIR: Lightweight Deep CNN-based Approach for Single Image Super-Resolution

Wazir Muhammad†, Murtaza Hussain Shaikh††, Jalal Shah†††, Syed Ali Raza Shah††††, Zuhaibuddin Bhutto†††
Liaquat Ali Lehri††††, Ayaz Hussain†, Salman Masrour††††, Shamshad Ali†, Imdadullah Thaheem†††††
wazir.laghari@gmail.com, zuhaib_bhutto@hotmail.com
†Department of Electrical Engineering, Balochistan University of Engineering & Technology, Pakistan.
††Department of Information Systems, Kyungsung University, Busan, South Korea.
†††Department of Computer Systems Engineering, Balochistan University of Engineering & Technology, Pakistan.
††††Department of Mechanical Engineering, Balochistan University of Engineering & Technology, Pakistan.
†††††Department of Energy System Engineering, Balochistan University of Engineering & Technology, Pakistan.

Abstract

Single image super-resolution (SISR) is an image processing technique, and its main target is to reconstruct the high-quality or high-resolution (HR) image from the low-quality or low-resolution (LR) image. Currently, deep learning-based convolutional neural network (CNN) image super-resolution approaches achieved remarkable improvement over the previous approaches. Furthermore, earlier approaches used hand designed filter to upscale the LR image into HR image. The design architecture of such approaches is easy, but it introduces the extra unwanted pixels in the reconstructed image. To resolve these issues, we propose novel deep learning-based approach known as Lightweight deep CNN-based approach for Single Image Super-Resolution (LDCSIR). In this paper, we propose a new architecture which is inspired by ResNet with Inception blocks, which significantly drop the computational cost of the model and increase the processing time for reconstructing the HR image. Compared with the other state of the art methods, LDCSIR achieves better performance in terms of quantitively (PSNR/SSIM) and qualitatively.

Keywords: *Image super-resolution; Convolutional neural network; PSNR; ResNet block.*

1. Introduction

Single image super-resolution (SISR) is an image enhancement technique to reconstruct the visually pleasing HR image from the original LR input image. Currently, image super-resolution is an attractive research area and widely focused in security surveillance, face recognition and medical image enhancement. Though, SISR is a highly challenging ill-posed inverse problem, as the LR image is a down-sampling version of the HR image. Recently, deep learning-based CNN techniques [1-3] increase the remarkable performance which is lead to significant improvement in the field of image super-resolution tasks. However, earlier design of deep learning-based CNN model connects the layers side by side to increase the size of model. These approaches are very simple, but it introduces the vanishing gradient problem during the training as well as increase the computational cost of the model. Furthermore, existing deep CNN architectures used pre-processing step, such as bicubic interpolation technique to upscale the low-resolution image and to introduce the new noises in the model, because the interpolation technique does not design for this purpose as shown in Figure 1. To solve such issues and further enhance the image quality of existing approaches, we proposed a Lightweight deep CNN-based approach for Single Image Super-Resolution as known as LDCSIR to reconstruct the high perceptual quality of HR image from LR image.

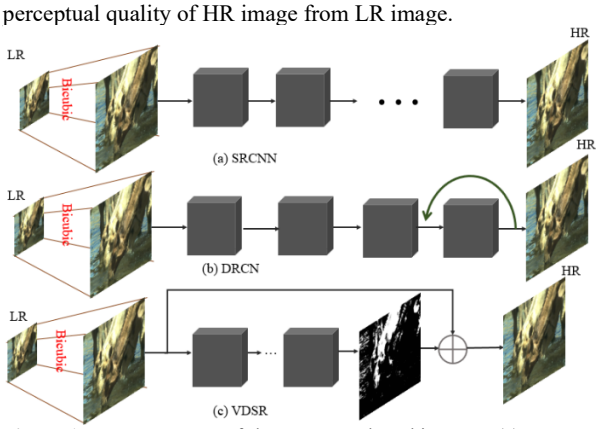


Figure 1. Image SR state-of-the-art network architectures (a) SRCNN, (b) DRCN, and (c) VDSR. All architectures are used pre-processing step to upscale the LR input image into HR output image.

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

- Inspired by Inception-based ResNet architecture, we used ResNet block followed by Inception to extract the multi-scale features information and to reduce the computational cost of the model.
- We replace the conventional Rectified Linear Unit (ReLU) activation function with LeakyReLU to fixes the dying ReLU problem as well as speed up the training process.
- Original Inception block used the maxpooling layer, which are not the best option for image super-resolution, because it extracts the maximum pixel information. In our proposed Inception block remove the maxpooling layer to increase the computational efficiency of the proposed block.

Manuscript received December 5, 2021
Manuscript revised December 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.12.63

The remaining section of this paper is structured as follows. Section 2 discuss the literature review of image SR approaches. Our proposed network architecture and its experimental evaluations with existing approaches are explained in Section 3 and 4. The conclusion is explained in Section 5.

2. Related works

The main aim of SISR is to reconstruct the visually pleasing HR output image that has detailed information from the original LR image. The popularity of deep CNN learning-based architecture, many researchers start to implement deep learning approaches to solve the SISR problem. Various approaches have been suggested to resolve the image SR problem, but here we discuss only recent deep learning CNN based approaches in detail.

The first deep CNN based image SR architecture was introduced by Dong et al. [4], known as Super-Resolution Convolutional Neural Network (SRCNN) [4]. SRCNN is the pioneer approach in the field of image SR and significantly improved the performance over all previous methods. In this approach used three CNN layers followed by ReLU activation function to predict the interpolated version of output image. The interpolated version image fed as an input to CNN layers for reconstructing the HR image. The same author Dong et al. [5] realize the adverse effect of bicubic interpolation and to remove the pre-processing step and fed the original LR image to four CNN layers. In SRCNN used pre-upsampling approach, but in the newly proposed architecture used post-upsampling layer (or Deconvolution layer. The performance of FSRCNN [5] are better and have lower computational cost as compared to SRCNN [4] but network capacity is still limited. Kim et al. inspired by the VGG-net architecture used in the implementation of ImageNet classification tasks [6]. Kim et al. introduce the deeper network architecture known as very deep super-resolution network (VDSR) [1]. In this architecture used 20 CNN trainable layers and claimed the improved performance over the SRCNN [4]. Almost, deeper model has the issue of training and introduce the vanishing gradient problem, but VDSR introduce the global residual skip connection with fast convergence rate.

Furthermore, Kim et al. introduced the concept of recursive type architecture known as DRCN (Deeply Recursive Convolutional Network) [2] for the better improvement of image SR tasks. In DRCN used multiple times of CNN layers. the main advantage of DRCN is to use the constant number of network training parameters. The main short coming with DRCN is the training process is too slow. Pyramidal based approach is introduced by Lei et al. [7] for single image super-resolution known as LapSRN. The network design of LapSRN [7] used different levels of pyramids and each pyramidal level is used the deconvolution layer for upscaling purposes. The main issue with LapSRN is that it used the fixed integer scaling enlargement factor, due to this limited the network flexibility. Zhang et al. [8] construct a super-resolution network for various degradations by concatenating a low-resolution image with its degradation mapping type architecture (SRMD). In [9, 10] network designs are very close to SRMD and used a cascade of convolutional layers to extract feature information, followed by the ReLU layer. W.M. et al. [11] suggested multi-scale inception architecture for reconstructing the visually pleasing high quality image. In this approach used the idea of asymmetric-based convolution technique to reduce the computational burden of the model as well as shrink the parameters. In [12] suggested two networks for detailed feature extractions. One network is the capsule image restoration and other is capsule attention. The performance of earlier SISR methods is better, but still some challenging tasks, like computational cost in terms of number of parameters, more memory consumption and required heavy processing time. To resolve these issues, we proposed lightweight architecture to handle these issues.

3. Proposed Methodology

In this section, we clearly discuss the methodology of our proposed network architecture. As similar to the existing SISR methods, proposed network architecture divided into four sections: initial low-level feature extraction used three CNN layers, mid-level feature extraction used three ResNet blocks, for high and multiscale-feature extraction purposes we used one inception block. Finally, we added one deconvolution layer with the support of shrinking and expanding layer before and after to reduce the computational burden during the training as shown in Figure.2.

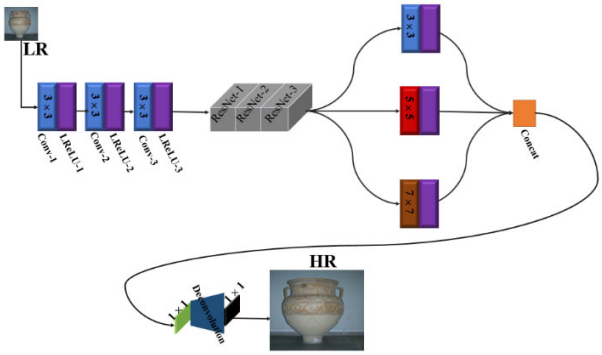


Figure 2. Proposed Lightweight deep CNN-based approach for Single Image Super-Resolution (LDCSIR).

In our proposed network architecture original LR input image is fed to three convolutional neural network layers to extract the low-level features. All three CNN layers are followed by LeakyReLU non-linear activation function. The LeakyReLU is the derived version of ReLU activation function [13]. The main advantages of using LeakyReLU is to resolve the issue of dying ReLU, because in deeper network architecture neurons are died and it creates the vanishing gradient problem. Before each LeakyReLU activation function, used CNN layers of the order of 32 number of filters with kernel size is 3×3. The design of deeper model architecture introduces the vanishing gradient problem [14]. Previous published work about design of deeper network just add the layer side by side and less information reached on the final layers [15]. He et al. [16] first time proposed the ResNet blocks to resolve the issue of vanishing gradient

during the training and extensively used in the field of image processing specially image super-resolution tasks

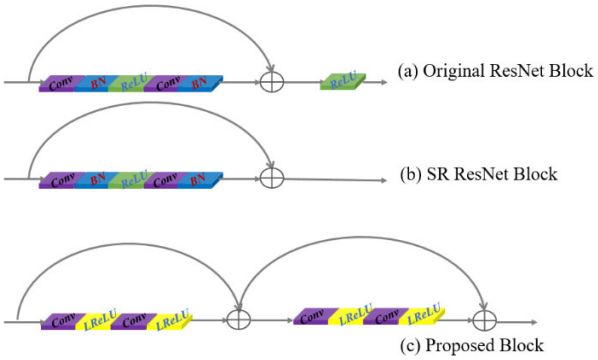
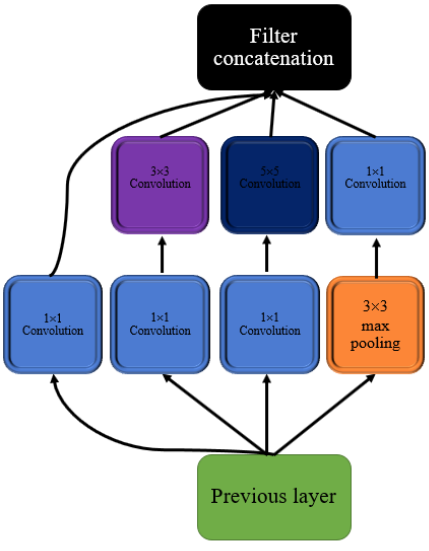


Figure 3. Shows a relationship between different ResNet blocks with our newly suggested ResNet block. (a) Original ResNet block with 2 CNN, Batch Normalization (BN) and ReLU activation layers. (b) shows the derived version of ResNet block known as SRResNet. (c) Our ResNet block used two blocks with local skip connections.

Figure 3(a) is the original ResNet block architecture proposed by He et al. [16] with convolution, batch normalization and ReLU activation layers with skip connection. The SRResNet block is derived from the original ResNet block and only remove the one ReLU activation layer as shown in Figure 3(b). In our proposed ResNet block is totally different from both blocks as shown in Figure 3(c). Proposed block used two ResNet block with locally skip connections and remove the Batch Normalization and ReLU activation layers to provide the clean path to network architecture for improving the efficiency of GPU as well as reduce the computational cost of the model. For middle feature extraction purpose inserted three ResNet blocks, each block consists of two 3x3 convolution kernel with 32 number of filters followed by LeakyReLU activation function.



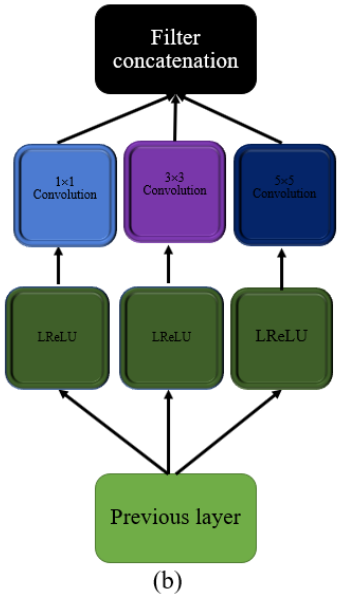


Figure 4. Comparison of original Inception block with our proposed Inception block.

Finally, for high and multiscale-feature extraction purpose, used a multi-scale Inception block which is adopted from GoogLeNet [17]. The main purposes of using multi-scale inception is to select the proper kernel size. The kernel size plays a vital role in any network architecture. The feature extraction information is distributed globally, we prefer the larger kernel size. The extracted information distributed locally smaller kernel size is more suitable. The inception architecture followed this concept as shown in Figure 4(a). The original inception block used unbalanced layers, due to uses of four 1 × 1 CNN layer. The other main issue with architecture is to uses of maxpooling layer, because maxpooling is not the best choice for image SR tasks. Maxpooling layer only select the maximum value of pixels and drop other values, due to this most prominent information is loosed and reconstructed image introduces the ringing and blurry HR output image. In Figure 4(b), removing the maxpooling layer as well as balance the distribution of filters like 3 x 3, 5 x 5, and 7 x 7 followed by LeakyReLU activation function, to reconstruct the SR image.

For upscaling the LR feature into HR feature we used an alternate technique and cannot directly send high level features to deconvolution layer for maintaining the computational burden of the model. It has been proved in an earlier study [18] that a 1 x 1 convolution layer used as a shrinking or bottleneck layer to reduce the size of the feature maps. However, we employ a 1 x 1 CNN layer before the deconvolution layer. Conventional based deep CNN model used pre-processing step to upscale the LR image into HR image. These types of arrangements are not suitable for deeper network architectures and introduced the extra new noises in the model. Furthermore, we lose the benefit of deep learning approach and it is training takes more time. For this purpose, we used one deconvolution layer as an enlargement /upscaling factor. The size of deconvolution layer is 3 x 3, padding is same, and stride is equal to upscaling factor. After the deconvolution layer, used expanding layer, which performs an inverse operation of a shrinking to reconstruct the visually

pleasing high-quality HR image.

4. Experimental Results

In the experimental section, first of all, we explain the training procedure of datasets with different model of hyper-parameters. Afterward, we evaluate the performance in terms of peak signal-to-Noise ratio (PSNR)/SSIM on publicly three benchmark test datasets. Finally, we evaluate the computational complexity in terms of PSNR versus Number of parameters.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \quad (1)$$

$$PSNR = 10\times\log_{10}\left(\frac{MAX_I^2}{\sqrt{MSE}}\right),$$

$$PSNR = 20\times\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

$$PSNR = 20\times\log_{10} MAX_I - 10\times\log_{10}(MSE) \quad (2)$$

4.1 Training and Testing Datasets

There has been the numerous datasets are available on image super-resolution for training purpose. The most commonly image SRtraining datasets are Yang et al. [19] and the Berkeley Segmentation Dataset (BSDS300) [20]. For loss function, we used mean squared error (MSE). We used the Adam optimizer during the training with initial learning rate is 0.0001 having batch size is 32. All experimental results were conducted on an Ubuntu 18.04 operating system with NVIDIA Titan Xp GPU, having 3.5 GHz Intel i7-5960× CPU with 16 GB RAM. We trained our model on the luminous channel for increasing the training speed. We evaluate the experimental results on three benchmark test datasets. The Set5 [21] having five number of images with different height and width such as 228 × 228 and 512 × 512. Set14 [22] uses the png format of total number 14 images with different sizes. BSDS100 [20] dataset uses the total 100 number of different natural images.

4.2 Quantitative Comparison in Terms of Network Depth.

We compare the network depth in term of number of k parameters versus PSNR of model as shown in Figure 5. By using the strategy of ResNet and Inception, our proposed LDCSIR model has less number of parameters and achieve better PSNR then existing other publicly available methods.

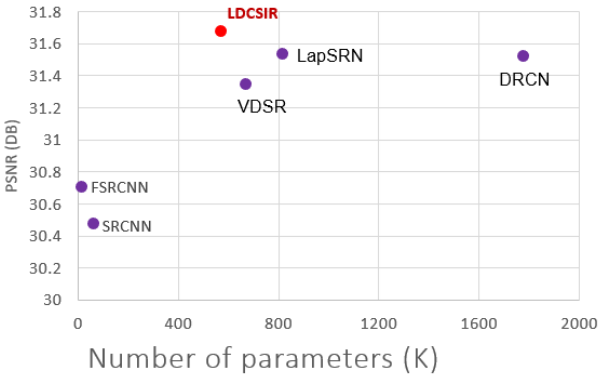


Figure 5. The performance measure on PSNR versus the number of network parameters. The performance is assessed with scale factor 4x on the Set5 dataset.

4. 3 Comparison with state-of-the-art methods

The quantitative evaluation of proposed algorithm with 10 existing state-of-the-art methods including Bicubic, A+, RFL, SelfExSR, SRCNN, FSRCNN, SCN, VDSR, DRCN and LapSRN were experimentally compared with our proposed method. Table 1 present the PSNR (dB) / SSIM comparison result with the existing deep CNN-based image SR methods on three main benchmark datasets SET5, SET14 and BSDS100 with challenging factor $4\times$ and $8\times$. Table 1 clearly shows our method achieves improved performance as compared other SR methods.

Figure 6 clearly shows the perceptual quality performance for scale 4× enlargements image SR on butterfly image, ppt3 image, img_108005, imge_16077 obtained from SET5, SET14 and BSDS100 datasets. The Bicubic, SRCNN and FSRCNN results appear blurry with a lack of clarity on high-frequency details. Although, image enhancement on enlargement factor 4× is a challenging task, yet our method reconstructs the details of texture and suppresses the artifacts correctly.

5. Conclusion

In this paper, we have proposed lightweight deep CNN-based approach for single image super-resolution known as LDCSIR to obtain the low, mid and high features directly from an original LR input image. The proposed method is inspired by ResNet with Inception block to reduce the number of network depth in terms of number of k parameters and produce the multiscale features information during the feature extraction and reconstruction processes. The proposed LDCSIR demonstrates the low-computational cost, due to replacing the bicubic upscaling technique with deconvolution layer and ReLU activation function with LeakyReLU. Extensive experimental results on different images are drawn from there representative image datasets. Resultantly, our proposed method obtained the satisfactory performance quantitatively as well as qualitatively as compared to the other state-of-the-art SR methods.

Table 1. evaluations of existing deep convolutional neural network SR methods with our proposed method. The results are reported on the average value of PSNR and SSIM with enlargement factor 4×. Bold and red color results are considered as the best performance. The blue color with underline considered as a second-best performance.

Method	No. of Parameters	Scale	SET5 PSNR/SSIM	Set14 PSNR/SSIM	BSDS100 PSNR/SSIM
Bicubic	-/-	4×	28.43 / 0.811	26.01 / 0.704	25.97 / 0.670
		8×	24.40 / 0.658	23.10 / 0.566	23.67 / 0.548
A+	-/-	4×	30.32 / 0.860	27.34 / 0.751	26.83 / 0.711
		8×	25.53/ 0.693	23.89 / 0.595	24.21 / 0.569
SelfExSR	-/-	4×	30.34 / 0.862	27.41 / 0.753	26.84 / 0.713
		8×	25.49 / 0.703	23.92 / 0.601	24.19 / 0.568
RFL	-/-	4×	30.17 / 0.855	27.24 / 0.747	26.76 / 0.708
		8×	25.38 / 0.679	23.79 / 0.587	24.13 / 0.563
SCN	42	4×	30.41 / 0.863	27.39 / 0.751	26.88 / 0.711
		8×	25.59 / 0.706	24.02 / 0.603	24.30 / 0.573
SRCNN	57	4×	30.50 / 0.863	27.52 / 0.753	26.91 / 0.712
		8×	25.33 / 0.690	23.76 / 0.591	24.13 / 0.566
FSRCNN	12	4×	30.72 / 0.866	27.61 / 0.755	26.98 / 0.715
		8×	25.60 / 0.697	24.00 / 0.599	24.31 / 0.572
VDSR	665	4×	31.35 / 0.883	28.02 / 0.768	27.29 / 0.726
		8×	25.93 / 0.724	24.26 / 0.614	24.49 / 0.583
DRCN	1770	4×	31.54 / 0.884	28.03 / 0.768	27.24 / 0.725
		8×	25.93 / 0.723	24.25 / 0.614	24.49 / 0.582
LapSRN	812	4×	31.54 / **0.885**	28.19 / **0.772**	27.32 / **0.727**
		8×	26.15 / 0.738	24.35 / **0.620**	24.54 / 0.586
Proposed (LDCSIR)	570	4×	**31.62** / **0.885**	**28.21** / 0.771	**27.36** / 0.726
		8×	**26.33** / **0.739**	**24.45** / **0.620**	**24.57** / **0.587**

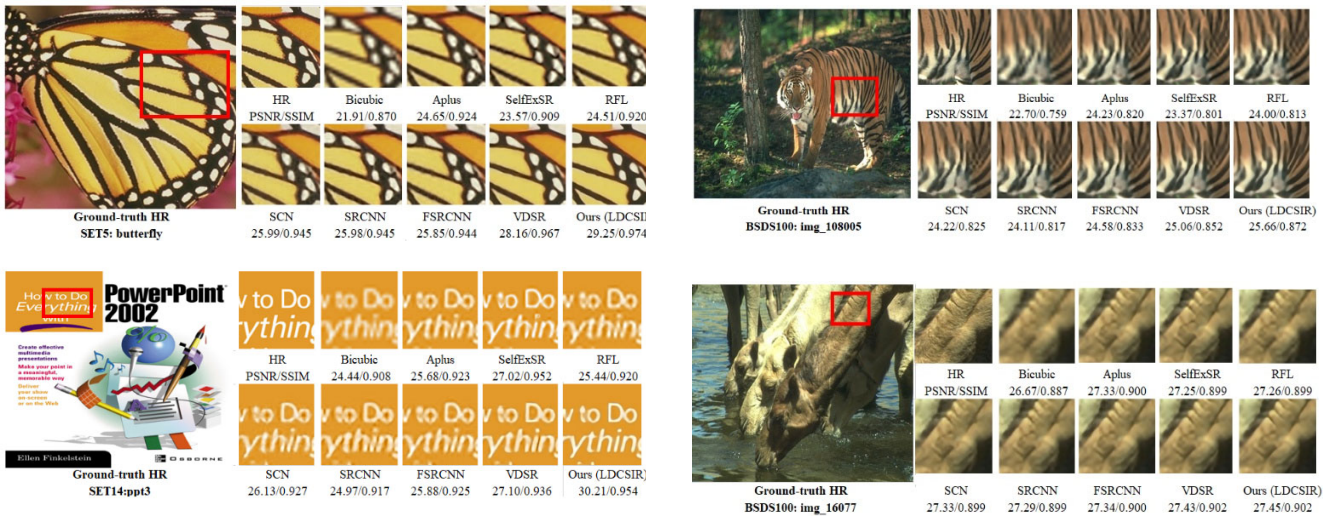


Figure 6. Perceptual quality performance of different images of Set5, Set14 and BSDS100.

References

1. Kim, J., J. Kwon Lee, and K. Mu Lee. *Accurate image super-resolution using very deep convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
2. Kim, J., J. Kwon Lee, and K. Mu Lee. *Deeply-recursive convolutional network for image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
3. Ledig, C., et al. *Photo-realistic single image super-resolution using a generative adversarial network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
4. Dong, C., et al., *Image super-resolution using deep convolutional networks.* 2015. **38**(2): p. 295-307.
5. Dong, C., C.C. Loy, and X. Tang. *Accelerating the super-resolution convolutional neural network*. in *European conference on computer vision*. 2016. Springer.
6. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
7. Lai, W.-S., et al. *Deep laplacian pyramid networks for fast and accurate super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
8. Zhang, K., W. Zuo, and L. Zhang. *Learning a single convolutional super-resolution network for multiple degradations*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
9. Dong, C., et al., *Image Super-Resolution Using Deep Convolutional Networks.* 2016. **38**(2): p. 295-307.
10. Zhang, K., et al. *Learning deep CNN denoiser prior for image restoration*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
11. Muhammad, W. and S.J.E. Aramvith, *Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach.* 2019. **8**(8): p. 892.
12. Hsu, J.-T., C.-H. Kuo, and D.-W.J.I.A. Chen, *Image Super-Resolution Using Capsule Neural Networks.* 2020. **8**: p. 9751-9759.
13. Nair, V. and G.E. Hinton. *Rectified linear units improve restricted boltzmann machines*. in *Icml*. 2010.
14. Simonyan, K. and A.J.a.p.a. Zisserman, *Very deep convolutional networks for large-scale image recognition.* 2014.
15. Hui, Z., X. Wang, and X. Gao. *Fast and accurate single image super-resolution via information distillation network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
16. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
17. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
18. Lin, M., Q. Chen, and S. Yan, *Network in network.* arXiv preprint arXiv:1312.4400, 2013.
19. Yang, C.-Y., C. Ma, and M.-H. Yang. *Single-image super-resolution: A benchmark*. in *European conference on computer vision*. 2014. Springer.
20. Martin, D., et al. *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*. in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. 2001. IEEE.
21. Bevilacqua, M., et al., *Low-complexity single-image super-resolution based on nonnegative neighbor embedding.* 2012.
22. Zeyde, R., M. Elad, and M. Protter. *On single image scale-up using sparse-representations*. in *International conference on curves and surfaces*. 2010. Springer.