SDCN: Synchronized Depthwise Separable Convolutional Neural Network for Single Image Super-Resolution

Wazir Muhammad†, Zuhaibuddin Bhutto††, Syed Ali Raza Shah†††, Jalal Shah††, Murtaza Hussain Shaikh††††, Ayaz Hussain†, Salman Masrour††† Imdadullah Thaheem†††, and Shamshad Ali†

†9Department of Electrical Engineering, Balochistan University of Engineering & Technology, Pakistan.
††Department of Computer Systems Engineering, Balochistan University of Engineering & Technology, Pakistan.
†††Department of Mechanical Engineering, Balochistan University of Engineering & Technology, Pakistan.
††††Department of Information Systems, Kyungsung University, Busan, South Korea.

Abstract

Recently, image super-resolution techniques used in convolutional neural networks (CNN) have led to remarkable performance in the research area of digital image processing applications and computer vision tasks. Convolutional layers stacked on top of each other can design a more complex network architecture, but they also use more memory in terms of the number of parameters and introduce the vanishing gradient problem during training. Furthermore, earlier approaches of single image super-resolution used interpolation technique as a pre-processing stage to upscale the low-resolution image into HR image. The design of these approaches is simple, but not effective and insert the newer unwanted pixels (noises) in the reconstructed HR image. In this paper, authors are proposing a novel single image super-resolution architecture based on synchronized depthwise separable convolution with Dense Skip Connection Block (DSCB). In addition, unlike existing SR methods that only rely on single path, but our proposed method used the synchronizes path for generating the SISR image. Extensive quantitative and qualitative experiments show that our method (SDCN) achieves promising improvements than other state-of-the-art methods.

Keywords: *Image supper-resolution, Deep convolutional neural network, Depthwise Separable convolution, Dense skip connection.*

1. Introduction

Single Image Super-Resolution (SISR) [1] is a key technique to reconstruct the visually pleasing high-quality or high-resolution (HR) output image from its degraded low-quality or low-resolution (LR) input image. Reconstructing the perceptually HR image is almost used in the field of security surveillance [2], medical image enhancement [3], and object recognition [4]. To handle the single image SR problem already many algorithms are proposed by researcher community. SISR approaches have three categories: interpolation approaches, reconstruction methods, and learning-based approaches. Interpolation approaches are including as Bicubic, bilinear, nearest neighbor, B-spline kernels [5], new edge-directed interpolation (NEDI) [6], and Lanczos upsampling [7] methods. Although, the design architecture of these approaches is very simple, but it can introduce the jagged ringing artifacts and added the new noises in the reconstructed HR image. Another category of image SR is the reconstruction-based methods [8-10], including Projection onto Convex Sets [11] and Maximum a Posteriori [12, 13]. These approaches are used to reconstruct the high-quality HR images with sharp edges but loss the high-frequency texture details. Third category is the learning-based approaches including dictionary-based learning methods [14, 15] and example-based methods [1, 16, 17].

With the rapid success of deep learning-based CNN algorithms have been achieved the tremendous performance in image super-resolution (SR). First time Dong et al. [18] introduce the shallow type network with three CNN layers for reconstructing the SISR. To further increase the performance of SRCNN, the same author introduced the new approach known as accelerated super-resolution convolutional neural network (FSRCNN) [17]. In this approach authors replaced the pre-processing step of bicubic interpolation with deconvolution layer. Shi et al. [19] introduced efficient sub-pixel convolution layer for real-time applications. This method is sued for both image as well as videos. First time deeper network architecture proposed by Kim et al. [20]. In this approach authors are used very deep convolutional network (VDSR) with 20 CNN layers. to accelerate the convergence speed of the VDSR residual learning are implemented. Although, the fact that deeper model for single image super-resolution have significant benefits as compared to traditional hand-designed filter methods, but still there are many challenging tasks and more gap is available for researchers.

First, many deep CNN-based approaches are used pre-processing stage to enlarge the LR input image into HR output image by using interpolation method. Second, almost deeper network architecture designed for reconstructing the visually pleasing high-quality HR out image having a more computational cost of the model and take more processing time in real world applications. Moreover, deeper network architectures depend on single serial path for extracting the low, mid as well as high quality features for generating the HR image.

In this work, we designed a novel method, known as synchronized depthwise separable CNN for single image super-resolution to learn end-to-end mapping between LR and HR images. Initially, we used the three convolutional neural networks layer (CNN) to extract the different level feature information from the original LR input image. The resultant features are fed into two parallel branches (synchronized) to extract the middle and high-level features information

Manuscript received November 5, 2021
Manuscript revised November 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.11.3

simultaneously. The parallel branches are used five DSCB blocks and concatenate the multi-branch features. We employed the shrinking and expanding layer before and after the deconvolution layer to alleviate the model's training burden and improve its computational efficiency.

The detailed explanation of our proposed method is present in the proposed method section.
The overall main contributions of our paper are summarized as follows:

- We propose a new image SR approach based on a synchronized depthwise separable convolutional neural network for SISR with dense skip connection blocks.
- Authors are introducing the new concept DSCB blocks which are free from memory consumption layers like Batch-Normalization (BN) and Max-pooling layer. Our proposed DSCB block also increase the computational efficiency of the model.
- To resolve the dying ReLU problem, authors are replaced the ReLU activation function with Leaky ReLU. Authors are also drop the bicubic interpolation as a pre-processing step with deconvolution layer for better reconstruction of HR image.

The remaining part of this paper is categorized as follows: Section 2 discuss the information about related works that is extremely important for understand the concept of proposed approach. In Section 3 authors are explain the proposed network architecture in detail. The experimental calculations and quantitative comparison are discussed in the Section 4. Finally, we summarize conclusion in Section 5.

2. Related works

Single image SR algorithms are used to estimate the relationship between low-resolution input image into high-resolution output image. Deep learning-based methods are the best option for feature extraction and image reconstruction. Initially, Dong et al. [18], proposed a shallow type network for single image super-resolution convolutional neural network (SRCNN). In this architecture authors are used the CNN layers followed by Rectified Linear activation function except the last layer. For upscaling purposes used the bicubic interpolation method from LR to HR image. The performance of SRCNN is better as compared to previous approaches, but they have some shortcomings. First, they used interpolation approach for upscaling purpose, which is not designed for this purpose. Second the authors are not extract the features directly from the original input but extract the features from the upscaled version. Thirdly, SRCNN used higher value of kernel size, which loss the more feature extracted information. To handle these types of issues, same author proposed a four layer with one deconvolution layers architecture known as Fast SRCNN [21]. In this approach authors are replaced the bicubic interpolation step with deconvolution layer for upscaling purpose. Similarly, authors are change the ReLU activation function with PReLU activation function. Unlike shallow network type architecture was presented in SRCNN and FSRCNN, Kim et al. [20] proposed the concept of Very Deep Super-Resolution, abbreviated as VDSR. The idea of VDSR is obtained from popular VGG-net architecture with fixed convolutional filters of the size of 3×3. Furthermore, Kim et al. [22] proposed a deeply recursive convolutional neural network (DRCN) for image SR task. Authors used a same layer to 16 times. The quantitative results are improved than pervious methods, but it demands more memory consumption and slow a running time. A pyramidal type network architecture proposed by Lai et al., in [23] known as Deep Laplacian Pyramid Super-Resolution Network abbreviated as LapSRN. LapSRN is used l_1 loss function known as Charbonnier, which can resolve the issue of outliers.
Residual network concept is different form linear network type architecture. The main purpose of ResNet architecture is to avoid the vanishing gradient problem during the training. He et al. [24] first time used this concept in the image classification task. The basic building blocks are used in the complete ResNet architecture as shown in Figure 1 (a). After that, several modified version of ResNet blocks [25, 26] are proposed as shown in Figure 1(b) and (c). Chollet et al., [27] proposed the concept of depthwise separable convolution to reduce the computational cost of the model in terms of number of parameters as well as less number of floating point operations.
A depthwise separable convolution is a type of convolution in which the conventional convolution is factorized into a depthwise convolution and then a pointwise convolution. The depthwise separable convolution process as shown in Figure 2. The chief advantages of depthwise separable convolution is to greatly reduce the number of parameters, channel and regions are processed separately.

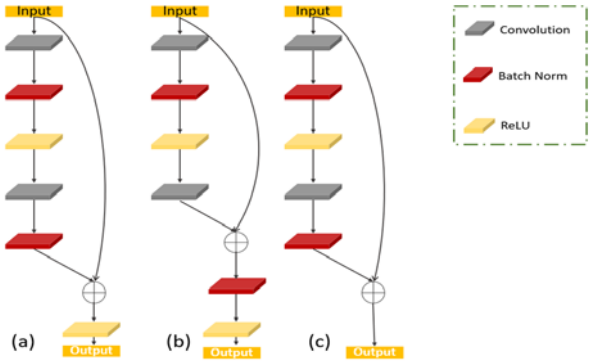

Figure 1. Basic building blocks of ResNet architecture and its derivates used in different SR algorithms.

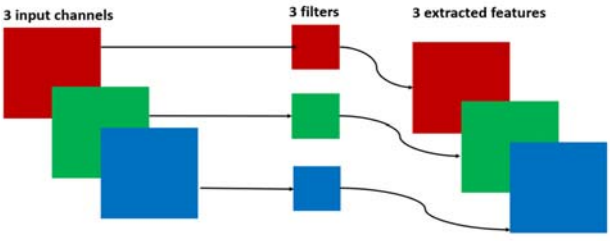

Figure 2. General overview concept of Depthwise Separable Convolution.

3. Proposed Method

In this section we present our proposed SDCN (synchronous depthwise separable convolutional neural network for single image super-resolution) method for reconstructing the high-quality output image from the low-quality input image. Fig 3 shows the overall network architecture comprising on two paths parallelly (also known as synchronized). The proposed architecture split into three phases: feature extraction, non-linear mapping and upsampling phase. Actually, complex serial architecture introduces the vanishing-gradient problem in the training. The main purpose of synchronized based strategy is to stabilize the training process and resolve the issue of vanishing gradient.

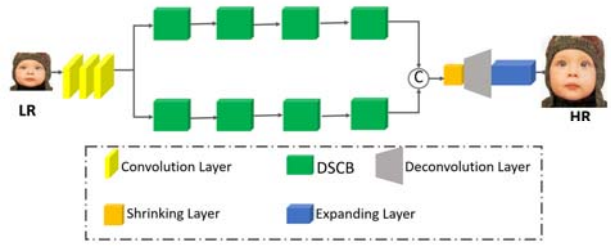

Figure 3. Proposed Network architecture of Synchronized Depthwise Separable Convolutional Neural Network for Single Image Super-Resolution.

3.1 Low-Level Feature Extraction

Earlier approaches are used the extra initial step to enlarge the low-resolution into high-resolution image. In our proposed strategy low-level features are directly extracted from the original LR input image. Our low-level feature extraction stage used three convolutional neural network layers followed LeakyReLU activation function. All feature extraction layers are the order of the 3×3 with 64 feature maps.

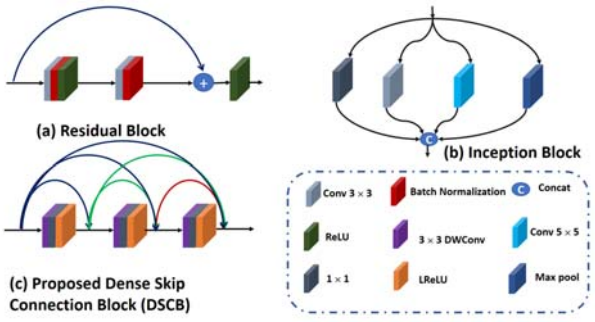

Figure 4. The network architecture of Residual, Inception Proposed Dense Skip Connection Block (DSCB).

3.2 Proposed Dense Skip Connection Block

Figure 4 shows the comparison of different blocks used in the field of deep learning by researcher community to increase the quality of LR image. Earlier deep learning-based model used the traditional approach to stack the layer side by side and the fed the information of each layer to the next layer. The design of conventional approach is very simple, but the later end layer suffers the problem and information cannot receive properly and sometimes overfitting problems are introduced during the training. To overcome this issue He et al. [24] proposed the ResNet architecture and stacked the residual blocks side by side to learn the residual feature with shortcut skip connections as shown in Figure 4(a). ResNet blocks support the high-speed training and avoiding the vanishing gradient problem during the training. Inception block is borrowed from Inception modules [28] which are used in deep convolutional neural network to reduce the computational cost and deeper through dimensionality reduction with stacked 1×1 convolution as shown in Figure 4(b). The main basic purpose of this block is to solve the computational cost problem as well as reduce the chances of overfitting. The solution is very simple to take the multiple sizes of kernels within single CNN block rather than stacking the different layer of kernels sequentially. Figure 4(c) is the proposed dense skip connection block (DSCB), that is totally different from both previous blocks as shown in Figure 4(a) and 4(b). In our proposed block remove the two layers such as Batch Normalization (BN) [29] layer and Max-pooling layer. Earlier works [25, 30, 31] suggested that Batch Normalization layer is not best choice for image SR tasks and reduce the accuracy of the model as well as increase the more memory consumption. However, it is quantitatively evaluated that batch normalization layer introduce the delays in the accuracy of image SR. Thus, recent research on single image SR network architectures [25, 30, 31] avoid the use of batch normalization layer. The max pooling layer is also not suitable choice for image reconstruction, because it cannot differentiate whether the relevant feature in one of the rows is available and it forgets the order of features in which they occur. Proposed DSCB block replaces the standard convolution operation with depthwise separable convolution operation followed by point-wise layer with leaky rectified linear (Leaky ReLU). DSCB block used the dense global as well as local skip connections that combines hierarchical feature maps to extract the richer feature representations. As the feature maps of the previous convolution layer are concatenated with those of the current convolution layer within a dense block, it requires more memory capacity to store massive feature maps and network parameters.

3.3 Deconvolution Layer

Deconvolution layer is to perform the inverse operation of convolution and main function is to increase the original LR image into HR image. Instead of using hand-designed filters approach, we used the deep learning-based approach. To reduce the training time and memory cost, we used two bottlenecks 1×1 layer. First convolution layer is the shrinking layer and other named as expanding layer to recover the original number of features after deconvolution layer.

4. Experiments

The quantitative and qualitative evaluations are present to validate the performance of our proposed method. Initially, we discussed the training and testing datasets for image SR. Next, we present the implementation details of our method and finally, evaluate the performance comparison with other methods.

4.1 Quantitative Evaluation Matrix

There are many quantitative evaluation matrix available to evaluate the quality of reconstructed HR out image from the original LR input image. For testing purposes, we used two quality matrix is PSNR/SSIM. The quantitative value of PSNR/SSIM is higher it indicates the better reconstruction of HR image. Generally, value of PSNR is calculated by MSE:

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

Mathematically, Peak signal-to-noise ratio (PSNR) is expressed as:

$$PSNR = 20.\log_{10}\left(\frac{MAX_1}{\sqrt{MSE}}\right), \qquad (2)$$

where the value of I and K are the representation of two images having size of the order $m \times n$ and measured in decibel (dB). Another quantitative evaluation matrix is the structural similarity index (SSIM).
Mathematically, SSIM is calculated as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma\mu_2^2 + C_2)}, \quad (3)$$

Where μ_x describe as an average value of x, μ_y represented as an average value of y. The value of variance and covariance of x are represented as σ_{xy}.

4.2 Model Computational Complexity Evaluation.

The model computational complexity validates in terms of number of parameters versus PSNR as shown in Figure 5. For evaluation purpose we used the SET5 test dataset on enlargement factor $4\times$. The proposed model has about 70% fewer parameters than LapSRN and 86% fewer parameters than DRCN. As compared to the Bicubic and SRCNN, the proposed SDCN improvement of about 3.20 and 1.13 dB on the challenging dataset SET5 with enlargement factor $4\times$ SR.

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.

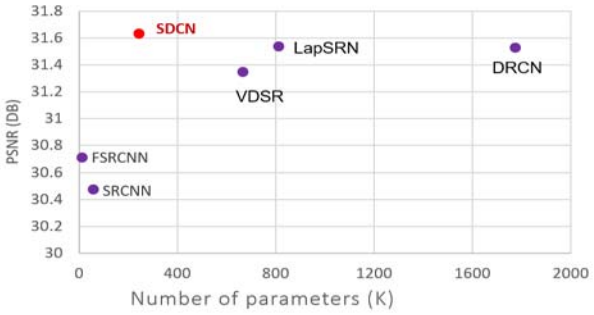

Figure 5. Computational cost comparison in terms of the model parameters (K) and PSNR on image SR dataset SET5 enlargement factor 4×.

Table 1. Quantitative comparisons of different single image SR methods using factor $4\times$ and $8\times$. Numerical value bold with red-colors represent the best value as compared to blue-colors with underline shows the second-best.

Method	*Scale*	SET5 PSNR/SSIM	Set14 PSNR/SSIM	BSDS100 PSNR/SSIM
Bicubic	4×	28.43/0.811	26.01/0.704	25.97/0.670
A+ [32]	4×	30.32/0.860	27.34/0.751	26.83/0.711
RFL [33]	4×	30.17/0.855	27.24/0.747	26.76/0.708
SelfExSR	4×	30.34/0.862	27.41/0.753	26.84/0.713
SRCNN	4×	30.50/0.863	27.52/0.753	26.91/0.712
FSRCNN	4×	30.72/0.866	27.61/0.755	26.98/0.715
SCN	4×	30.41/0.863	27.39/0.751	26.88/0.711
VDSR	4×	31.35/0.883	28.02/0.768	27.29/0.726
DRCN	4×	31.54/0.884	28.03/0.768	27.24/0.725
LapSRN	4×	31.54/0.885	28.19/0.772	27.32/0.727
SDCN (ours)	4×	**31.63/0.886**	**28.24/0.773**	**27.34/0.729**
Bicubic	8×	24.40/0.658	23.10/0.566	23.67/0.548
A+ [32]	8×	25.53/0.693	23.89/0.595	24.21/0.569
RFL [33]	8×	25.38/0.679	23.79/0.587	24.13/0.563

SelfExSR	8×	25.49/0.703	23.92/0.601	24.19/0.568
SRCNN	8×	25.33/0.690	23.76/0.591	24.13/0.566
FSRCNN	8×	25.60/0.697	24.00/0.599	24.31/0.572
SCN	8×	25.59/0.706	24.02/0.603	24.30/0.573
VDSR	8×	25.93/0.724	24.26/0.614	24.49/0.583
DRCN	8×	25.93/0.723	24.25/0.614	24.49/0.582
LapSRN	8×	26.15/0.738	24.35/0.620	24.54/0.586
SDCN (ours)	8×	**26.23/0.744**	**24.39/0.625**	**24.58/0.587**

Furthermore, our proposed method evaluates visually as shown in Figure 5 and 6 with enlargement factor $4\times$. In Figure 6 authors used the image of zebra obtained from Set14 and the reconstructed results of our proposed method visually pleasing as compared to baseline method Bicubic. Similarly, the perceptual quality of Figure 7 image the generated result by Bicubic is blurry and jagged artifacts, but our proposed method is better reconstructed HR image.

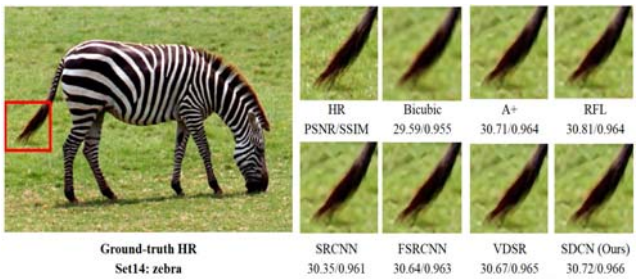

Figure 6. Perceptual quality comparisons of SDCN approach with other super-resolution approaches on sale factor 4×.

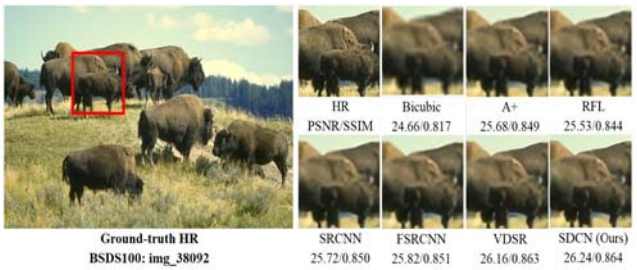

Figure 7. Perceptual quality comparisons of SDCN approach with other super-resolution approaches on sale factor 4×.

5. Conclusion

In this paper authors are proposed synchronized depthwise separable convolutional neural network for single image super-resolution (SDCN). The main architecture extracts the detailed features from the two-path followed by dense skip connection block (DSCB). In our proposed approach authors are changed the operation of standard convolution with depthwise separable CNN operation to reduce the computational cost. Furthermore, to resolve the issue of vanishing-gradient appear in training in deeper network architecture due to dying ReLU activation function, authors are replaced with LeakyReLU activation function. Experimental results confirm that SDCN obtained the favorable performance against existing deep learning-based SR methods.

References

1. Freeman, W.T., E.C. Pasztor, and O.T. Carmichael, *Learning low-level vision.* International journal of computer vision, 2000. **40**(1): p. 25-47.
2. Shamsolmoali, P., et al., *Deep convolution network for surveillance records super-resolution.* Multimedia Tools and Applications, 2019. **78**(17): p. 23815-23829.
3. Isaac, J.S. and Kulkarni, R. *Super resolution techniques for medical image processing.* in *2015 International Conference on Technologies for Sustainable Development (ICTSD)*. 2015. IEEE.
4. Yang, X., et al., *Long-distance object recognition with image super resolution: A comparative study.* IEEE Access, 2018. **6**: p. 13429-13438.
5. Unser, M., A. Aldroubi, and M. Eden, *Fast B-spline transforms for continuous image representation and interpolation.* IEEE Transactions on pattern analysis and machine intelligence, 1991. **13**(3): p. 277-285.
6. Li, X. and M.T. Orchard, *New edge-directed interpolation.* IEEE transactions on image processing, 2001. **10**(10): p. 1521-1527.
7. Duchon, C.E., *Lanczos filtering in one and two dimensions.* Journal of Applied Meteorology and Climatology, 1979. **18**(8): p. 1016-1022.
8. Marquina, A. and S.J. Osher, *Image super-resolution by TV-regularization and Bregman iteration.* Journal of Scientific Computing, 2008. **37**(3): p. 367-382.
9. Sun, J., Z. Xu, and H.-Y. Shum. *Image super-resolution using gradient profile prior*. in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008. IEEE.
10. Tai, Y.-W., et al. *Super resolution using edge prior and single image detail synthesis*. in *2010 IEEE computer society conference on computer vision and pattern recognition*. 2010. IEEE.
11. Stark, H. and P. Oskoui, *High-resolution image recovery from image-plane arrays, using convex projections.* JOSA A, 1989. **6**(11): p. 1715-1726.
12. Schultz, R.R. and R.L. Stevenson, *Extraction of high-resolution frames from video sequences.* IEEE transactions on image processing, 1996. **5**(6): p. 996-1011.
13. Schultz, R.R. and R.L. Stevenson. *Improved definition video frame enhancement.* in *1995 International Conference on Acoustics, Speech, and Signal Processing*. 1995. IEEE.
14. Jia, K., X. Wang, and X. Tang, *Image transformation*

based on learning dictionaries across image spaces. IEEE transactions on pattern analysis and machine intelligence, 2012. **35**(2): p. 367-380.

15. Bevilacqua, M., et al., *Low-complexity single-image super-resolution based on nonnegative neighbor embedding.* 2012.
16. Freeman, W.T., T.R. Jones, and E.C. Pasztor, *Example-based super-resolution.* IEEE Computer graphics and Applications, 2002. **22**(2): p. 56-65.
17. Chang, H., D.-Y. Yeung, and Y. Xiong. *Super-resolution through neighbor embedding*. in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* 2004. IEEE.
18. Dong, C., et al. *Learning a deep convolutional network for image super-resolution*. in *European conference on computer vision*. 2014. Springer.
19. Shi, W., et al. *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
20. Kim, J., J.K. Lee, and K.M. Lee. *Accurate image super-resolution using very deep convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
21. Dong, C., C.C. Loy, and X. Tang. *Accelerating the super-resolution convolutional neural network*. in *European conference on computer vision*. 2016. Springer.
22. Kim, J., J.K. Lee, and K.M. Lee. *Deeply-recursive convolutional network for image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
23. 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.
24. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
25. Lim, B., et al. *Enhanced deep residual networks for single image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.
26. Ahn, N., B. Kang, and K.-A. Sohn. *Fast, accurate, and lightweight super-resolution with cascading residual network*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
27. Chollet, F. *Xception: Deep Learning with Depthwise Separable Convolutions*. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
28. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
29. Ioffe, S. and C. Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. in *International conference on machine learning*. 2015. PMLR.
30. Zhang, Y., et al. *Residual dense network for image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
31. Fan, Y., et al. *Balanced two-stage residual networks for image super-resolution*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017.
32. Timofte, R., V. De Smet, and L. Van Gool. *A+: Adjusted anchored neighborhood regression for fast super-resolution*. in *Asian conference on computer vision*. 2014. Springer.
33. Schulter, S., C. Leistner, and H. Bischof. *Fast and accurate image upscaling with super-resolution forests*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.