Comparative Analysis of Deep Convolution Networks Based Image Super-Resolution Techniques

Renuka Sindge, Maitreyee Dutta, Jagriti Saini

National Institute of Technical Teachers Training and Research, Chandigarh

Corresponding author: Renuka Sindge, Email: renukasindge@gmail.com

Single Image Super-Resolution (SISR) is gaining huge attention in the digital age across various application domains such as surveillance, medical imaging, and agriculture. Numerous SR methods based on deep learning were used by existing researchers to improve image resolution. Literature shows that deep convolutional neural networks (CNNs) perform exceptionally well to handle degraded images. In this study, CNN-based methods from a deep learning environment are compared to reconstruct the High-Resolution (HR) images. Observations show that SRCNN and FSRCNN can achieve considerable image quality after reconstruction; however, performance is limited to small datasets due to shallow network parameters. Furthermore, VDSR and LapSRN were also utilized against heavy datasets due to their huge computational efficiency.

Keywords: Image super-resolution, Single image super-resolution, Convolutional neural network, Deep Learning.

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 1009–1016. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-96

1. Introduction

Image Super-resolution (ISR) is premised on the concept that a High-Resolution (HR) image or image sequence can be created by combining Low-Resolution (LR) noisy images [1]–[5]. Super-Resolution (SR) is an emerging domain with remarkable evolution to address the practical utility and feasibility-related challenges. In computer vision, applications of HR visuals are in demand to achieve more reliable performance. According to it, predicting high-frequency features lost in LR images has received much interest in various domains. HR imagery has been utilized in different situations, including satellite imaging, medical imaging [6], HDTV, surveillance [7], human interpretation, and so on [8]. Various medical, Biometric [9], and surveillance applications are extremely dependent on the authenticity and credibility of the visual content presented in images. The extensive use of these digital images in our daily lives has led to a significant increase in the adoption of simple-to-use SISR methods that have the ability to enhance the overall quality of digital images. The existing studies show evidence for significant improvement in digital imaging applications that require HR visuals. Single image super-resolution (SISR) objectives at constructing an SR image from a single LR image. The main issue of SR is to construct the high-frequency (HF) details that are lost in the LR image. However, the ill-posed problem of SISR [10]creates many possibilities in LR-HR mappings for constructing images.

With the evolution of deep learning (DL) approaches such as neural networks, there has been a lot of growth in image processing applications domains, such as astronomical observation image and video enhancement [11], [12].. The advanced DL methods promise enhanced computational power with the ability to process big data to overcome existing issues in the SR domain. This era witnessed remarkable progress in SR using DL techniques. In the past decade, DL-based SR methods obtained superior performance as compared to the classical methods [3], [8]. And recently, DL-based techniques have made significant progress in the SISR domain as well. Several researchers in the past have worked on SISR field applications, and they marked useful findings. Dong et al. . [13] designed a convolutional neural network (CNN) for SISR to learn the LR-HR mapping, which has achieved reliable performance with the help of an end-to-end approach. With the extension of this pioneer SRCNN method, further Kim et al.[14] improved the network convergence speed with a very deep super-resolution (VDSR) method. This method uses deep layers with a high learning rate and residual learning to ensure high-quality visuals. On the other side, Kim et al. [15] a deep recursive convolutional network (DRCN) proposed a very deep recursive layer with several recursions to control the parameters of the model. Tai et al. [16]proposed a deep recursive residual network (DRRN) that used both residual learning and recursive module to construct the network with fewer parameters. However, in the case of a large upscaling factor, these methods are not much effective. For this reason, Lai et al. [17] proposed a Laplacian pyramid network structure (LapSRN) that reconstructs HR visuals progressively with high scaling factors (4, 8, and more). These methods achieve high-quality visuals with respect to evaluation metrics; Moreover, there are various methods also, which have benefits over computational complexity. For fast speed, Dong et al. [18] introduced the FSRCNN method by inserting a deconvolution layer in the last stage of the network. Jiu et al. [19] proposed dense residual network [12] (RDN) and employ large kernel size deconvolution, which reduces the computational complexity of the network and requires less memory. Further, for SISR, Lan et al. [20] introduced cascading residual network (CARN) with the help of local and global cascading, which requires less network parameters and achieves better results.

The authors of this study compared and analyzed the performance of four DL-based methods for SISR application. Among these methods, SRCNN was the primary method used for constructing the SISR. With some additional layers to the basic SRCNN method, the FSRCNN method was developed, which enhanced the quality of the image. However, both methods utilized a shallow network that is unable to process heavy datasets. Therefore, this study further analyzed the performance of a large dataset using VDSR and LapSRN methods. Both methods were effective in a large dataset with increased quality of the image. For experimental work, this study used Set5, Set14, and Urban100 for testing and T91, BSD200, and ImageNet for the training purpose. For assessment of network performance, authored preferred using structure similarity index (SSIM) and peak-signal-noise ratio (PSNR) metrics. The main goal of this study is to conduct a comparative analysis of DL-based SISR methods with a detailed comparison of reconstruction results.

2. Methods

Deep CNN became more popular due to its simple structure that gives better reconstruction results. Various factors are the reason of importance in CNN to use: (i) Powerful training implementation with the help of GPU [21] ,(ii) Use of the Rectified Linear Unit (ReLU) that leads to faster convergence [22],and (iii) the easy access to the huge dataset for large models [23]. Among various state-of-art methods, this study analyzes and compares the following popular SR methods of the DL-based approach.

A. Super-Resolution Convolutional Neural Network (SRCNN)

In ISR reconstruction development, the SRCNN method is one of the widely preferred methods for image reconstruction needs [13]. This method uses bicubic interpolation for upsampling the inputs. It has a simple structure with three layers, namely patch extraction, non-linear mapping, and lastly, reconstruction (Figure 1). The patch representation layer extracts the feature from the bicubic interpolation. Then, these extracted high-dimensional features were forwarded to the next layer for non-linear mapping. Lastly, the output from the non-linear mapping layer was used to reconstruct the HR image. This method uses filters to optimize end-to-end mapping, and with the help of various filter sizes set in non-linear mapping, it can utilize the maximum information.

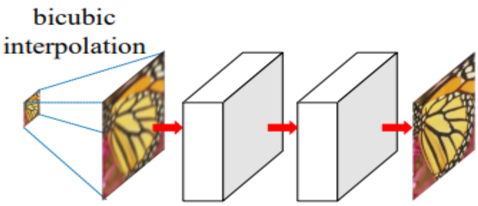

Fig 1. Network architectures of SRCNN [13]

B. Fast Super-Resolution Convolutional Neural Network (FSRCNN)

Dong et al. [18] recommend adding more convolutional layers to SRCNN for non-linear mapping. However, more convolutional layers increase the computational time. Therefore, to overcome this issue, FSRCNN was developed. This method consists of five main slices, given in Figure 2. These were feature extraction, non-linear mapping, expanding, shrinking, and image reconstruction. The presence of a shrinking layer and expanding layer in FSRCNN architecture is the main difference between FSRCNN and SRCNN. The dimension of extracted features was reduced by the shrinking layers from the preceding layer. In the meantime, the expanding layer expands the output feature of the non-linear layer mapping, which works exactly in reverse to the layer shrinking. Lastly, the FSRCNN method utilized deconvolution for upsampling.

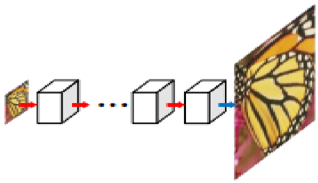

Fig 2. Network architectures of FSRCNN[18]

C. Very Deep Super Resolution (VDSR)

This method was introduced by Kim et al. [14] with additional mapping layers in comparison to SCRNN. VDSR method in this comparative analysis was trained and tested on depth range from 5 to 20. Observations show considerable improvement in performance with higher depth. The network performance improved rapidly with the additional layers. Residual learning has been utilized by this VDSR method for mapping input and output features. Through a skip connection, the residual learning further added output features to the interpolated features. Combining the low-level features and the high-level features with the help of the skip connection increases the performance of the model. The network learns the residual error present in input and output to overcome the SRCNN problem, which learns the HR image directly. The network convergence speed is accelerated via residual learning. As a result, the best outcome is achieved in the shortest amount of time. The main advantage of using residual learning was enhanced convergence speed in comparison to SRCNN. VDSR network design is shown in Figure 3.

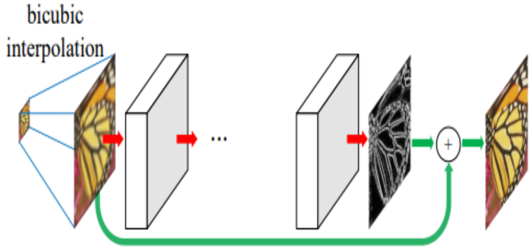

Fig 3. Network architectures of VDSR[14]

D. Laplacian Pyramid Super-Resolution Network (LapSRN)

In contrast to one-step upsampling, like SRCNN, the LapSRN network reconstructs the HR image residuals at multiple pyramid levels progressively, with a specific scale factor (i.e., 2, 4, 8). Without the use of bicubic, it extracts the features directly from the LR input image, which causes less computational cost. Laplacian Pyramid has been used for decades. It has two branches- first is feature extraction branch and second is the image reconstruction branch (Figure 4). Feature extraction upsamples the features where network features learn the complex mappings of the lower level with the higher levels. In the image reconstruction process, the upsampled images are combined with the set of predicted residual images from the feature extraction branch and they further generate an HR output image.

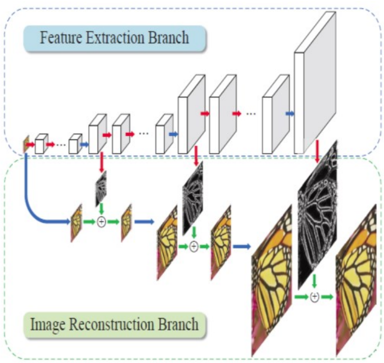

Fig 4. Network architectures of LapSRN [17]

Among all these four CNN-based SR methods, LapSRN uses a progressive reconstruction of the image. In contrast, the remaining methods use direct reconstruction. Furthermore, SRCNN and VDSR methods use bicubic interpolation for upsampling. The network architectures of these methods have a different number of layers. Each method uses the L2 loss function except LapSRN, which uses charbonnier loss. The comparative analysis of all four methods is provided in Table 1.

E. Experimental Details

The CNN-based DL methods were used in this experimental work, such as the pioneer method of CNN, i.e., SRCNN, FSRCNN with additional layers to the primary SRCNN. Experimental results show that both works fine on low scaling factors and small datasets due to shallow networks. Hence, the authors expanded the experimental work to VDSR with 20 layers and LapSRN with ten layers to observe and improve the performance with a large scaling factor, i.e., 4×, and a huge dataset. In order to validate the given network performance, the authors' preferred using common training and testing datasets for fair comparative analysis. In this study, SRCNN and FSRCNN used ImageNet [21], T91, and General100 datasets. However, for VDSR and LapSRN, large datasets like BSD200 and T91 were used. For the testing purpose, the authors used three benchmark datasets, including Set5 [24], Set14 [25], and Urban100 dataset, which consists of urban images [26]. All methods are supervised by pixel-wise loss function with Adam optimizer and experiments are conducted using Python.

Table 1. Comparisons of CNN based SR algorithms

SR Methods	Layers	Network input	Residual learning	Loss function	Reconstruction
SRCNN	3	LR+bicubic	No	L2	Direct
FSRCNN	8	LR	No	L2	Direct
VDSR	20	LR+bicubic	Yes	L2	Direct
LapSRN	27	LR	Yes	Charbonnier	Progressive

3. Results and Discussions

This section presents the results of the already explained CNN-based SR algorithms for reconstruction of SR image with respect to their performance on different training and test datasets. Table II shows quantitative comparisons for scaling factors 2×, 3×, and 4×. This experimental work does not give the result on scaling factors 2× and 3× for SRCNN and FSRCNN methods due to relatively shallow network attributes. Observations state that VDSR and LapSRN perform better against existing methods on most datasets. Very less difference was observed in SRCNN and FSRCNN methods in terms of their output. FSRCNN acceleratees its performance significantly, i.e., about n^2 times faster than SRCNN, where n is the scaling factor. However, VDSR and LapSRN methods have a good convergence speed as compared to the SRCNN method. Also, VDSR used extremely high learning rates; it is generally10^4 times higher when compared to SRCNN). LapSRN is faster than other methods except FSRCNN. The texture of the image was better in VDSR due to the inclusion of residual learning, which was not in the SRCNN method.

It was also observed that SRCNN and FSRCNN methods were easy to train due to a limited number of parameters. However, FSRCNN generates better PSNR performance as compared to SRCNN. In comparison to both shallow networks, VDSR and LapSRN approaches reconstructed substantially superior image quality performance. However, because of the deeper network property, they are difficult to train. For comparing model performance, this study used two common and popular evaluation metrics: PSNR and SSIM. Figure 5 shows the average PSNR values of all these four SR methods with respect to scaling factor 4×.

Table 2. Evaluation of CNN-based SR algorithms in terms of PSNR/SSIM for different scaling factors

SR Methods	**Scale**	**Set5**	**Set14**	**Urban100**
		PSNR/SSIM	**PSNR/SSIM**	**PSNR/SSIM**
SRCNN	2	36.66/0.954	32.45/0.906	-
FSRCNN		37.00/0.958	32.63/0.908	-
VDSR		**37.53** / 0.958	32.97 / **0.913**	**30.77/ .914**
LapSRN		37.52 / **0.959**	**33.08/ .913**	30.41 / 0.910
SRCNN	3	32.75/0.909	29.30/0.8215	-
FSRCNN		33.16/0.914	29.43/0.824	-
VDSR		33.66/0.921	29.77/0.831	**27.14**/0.827
LapSRN		**33.82/0.922**	**29.87/0.832**	27.07/**0.828**
SRCNN	4	30.49 / 0.862	27.61 / 0.754	24.53 / 0.724
FSRCNN		30.71 / 0.865	27.70 / 0.756	24.61 / 0.727
VDSR		31.35 / 0.882	28.03 / 0.770	25.18 / 0.753
LapSRCN		**31.54/ .885**	**28.19/ 0.772**	**25.21/ 0.756**

The analysis provided in the comparison Table 2 shows that the SRCNN and FSRCNN methods give better results on low scaling factors with a smaller number of input images. Moreover, both methods have less computational complexity with a minimum number of layers. In comparison, VDSR and LapSRN methods give reliable results on high as well as low scaling factors with all types of benchmark datasets. However, these two methods have more computational complexity as compared to SRCNN and FSRCNN due to a greater number of layers.

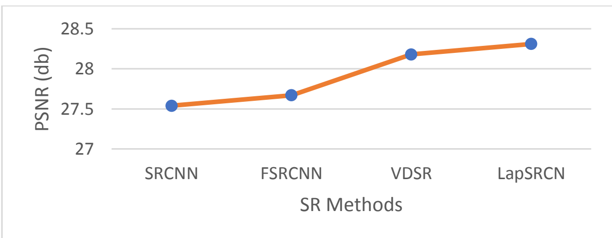

Fig 5. PSNR average performance of SR method with scale factor 4×.

4. Conclusions

In today's digital environment, ISR has received a lot of interest from several application sectors. The progress of DL motivated to utilize CNN-based ISR to accomplish the best network model with less computation complexity. This paper evaluated and analyzed four SR approaches in order to achieve HR output images. The first pioneer CNN-based method was SRCNN. Other SR methods such as FSRCNN, VDSR, and LapSRN were also evaluated to obtain the HR visuals. This study observed that SRCNN and FSRCNN methods give the best result with less input and low scaling factor and have less computational complexity with a smaller number of layers. In contrast, VDSR and LapSRN methods give the best performance in the case of high scaling factors with large benchmark datasets. The comparative analysis reveals that the DL-based SR approaches show significant results for obtaining HR images by processing LR images as input.

However, there are several parameters associated with these DL-based methods that are still to be analyzed to achieve enhanced outcomes for real-life applications. This comparative analysis will be helpful for upcoming researchers to expand the performance of the ISR methods in terms of output image quality. Furthermore, future researchers also need to work on specific application domains such as surveillance and medical imaging to address the image resolution-related challenges.

References

[1] T. Angeland and R. Colomo-Palacios, "Deep Learning for Single Image Super-Resolution: A Systematic Mapping Study," p. 26, 2019.

[2] S. Anwar, S. Khan, and N. Barnes, "A Deep Journey into Super-resolution: A survey," *arXiv:1904.07523 [cs]*, Mar. 2020, Accessed: Aug. 31, 2021. [Online]. Available: http://arxiv.org/abs/1904.07523

[3] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A comprehensive review of deep learning-based single image super-resolution," *PeerJ Computer Science*, vol. 7, p. e621, Jul. 2021, doi: 10.7717/peerj-cs.621.

[4] R. A. A. Al-falluji, A. A.-H. Youssif, and S. K. Guirguis, "Single Image Super Resolution Algorithms: A Survey and Evaluation," vol. 6, no. 9, p. 8, 2017.

[5] S. Dubey, Prof. P. Sahu, Prof. S. Bazal, and Department of ECE, GGITS, Jabalpur, Madhya Pradesh, India, "Literature Review on Single Image Super Resolution," *IJTSRD*, vol. Volume-2, no. Issue-5, pp. 2485–2490, Aug. 2018, doi: 10.31142/ijtsrd18339.

[6] J. Jurek, M. Kociński, A. Materka, M. Elgalal, and A. Majos, "CNN-based superresolution reconstruction of 3D MR images using thick-slice scans," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 111–125, Jan. 2020, doi: 10.1016/j.bbe.2019.10.003.

[7] P. Shamsolmoali, M. Zareapoor, D. K. Jain, V. K. Jain, and J. Yang, "Deep convolution network for surveillance records super-resolution," *Multimed Tools Appl*, vol. 78, no. 17, pp. 23815–23829, Sep. 2019, doi: 10.1007/s11042-018-5915-7.

[8] J. Li, Z. Pei, and T. Zeng, "From Beginner to Master: A Survey for Deep Learning-based Single-Image Super-Resolution," *arXiv:2109.14335 [cs, eess]*, Sep. 2021, Accessed: Feb. 24, 2022. [Online]. Available: http://arxiv.org/abs/2109.14335

[9] K. Nguyen, C. Fookes, S. Sridharan, M. Tistarelli, and M. Nixon, "Super-resolution for biometrics: A comprehensive survey," *Pattern Recognition*, vol. 78, pp. 23–42, Jun. 2018, doi: 10.1016/j.patcog.2018.01.002.

[10] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1122–1131. doi: 10.1109/CVPRW.2017.150.

[11] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[12] V. Bellandi, A. Chehri, S. Cuomo, and G. Jeon, "Special issue on deep learning for emerging big multimedia super-resolution," *Multimedia Systems*, vol. 27, no. 4, pp. 581–587, Aug. 2021, doi: 10.1007/s00530-021-00814-5.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *arXiv:1501.00092 [cs]*, Jul. 2015, Accessed: Jan. 17, 2022. [Online]. Available: http://arxiv.org/abs/1501.00092

[14] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.

[15] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," *arXiv:1511.04491 [cs]*, Nov. 2016, Accessed: Jan. 24, 2022. [Online]. Available: http://arxiv.org/abs/1511.04491

[16] Y. Tai, J. Yang, and X. Liu, “Image Super-Resolution via Deep Recursive Residual Network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 2790–2798. doi: 10.1109/CVPR.2017.298.
[17] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution,” *arXiv:1704.03915 [cs]*, Oct. 2017, Accessed: Jan. 24, 2022. [Online]. Available: http://arxiv.org/abs/1704.03915
[18] C. Dong, C. C. Loy, and X. Tang, “Accelerating the Super-Resolution Convolutional Neural Network,” *arXiv:1608.00367 [cs]*, Aug. 2016, Accessed: Aug. 30, 2021. [Online]. Available: http://arxiv.org/abs/1608.00367
[19] J. Xu, Y. Chae, B. Stenger, and A. Datta, “Dense Bynet: Residual Dense Network for Image Super Resolution,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 71–75. doi: 10.1109/ICIP.2018.8451696.
[20] R. Lan *et al.*, “Cascading and Enhanced Residual Networks for Accurate Single-Image Super-Resolution,” *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 115–125, Jan. 2021, doi: 10.1109/TCYB.2019.2952710.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012, vol. 25. Accessed: Apr. 01, 2022. [Online]. Available: https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
[22] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Madison, WI, USA, Jun. 2010, pp. 807–814.
[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, “ImageNet: a Large-Scale Hierarchical Image Database,” Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
[24] M. Bevilacqua, A. Roumy, C. Guillemot, and M. A. Morel, “Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding,” in *Procedings of the British Machine Vision Conference 2012*, Surrey, 2012, p. 135.1-135.10. doi: 10.5244/C.26.135.
[25] R. Zeyde, M. Elad, and M. Protter, “On Single Image Scale-Up Using Sparse-Representations,” in *Curves and Surfaces*, vol. 6920, J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 711–730. doi: 10.1007/978-3-642-27413-8_47.
[26] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5197–5206. doi: 10.1109/CVPR.2015.7299156.