RIMS: Residual-Inception Multiscale Image Super-Resolution Network

Wazir Muhammad[†], Zuhaibuddin Bhutto[††], Jalal Shah[††], Murtaza Hussain Shaikh[†††], Syed Ali Raza Shah[††††], Shah Muhammad Butt[†††††], Salman Masroor[††††] Ayaz Hussain[†]

[†]Department of Electrical Engineering, Balochistan University of Engineering & Technology, Pakistan.
[††]Department of Computer Systems Engineering, Balochistan University of Engineering & Technology, Pakistan.
[†††]Department of Information Systems, Kyungsung University, Busan, South Korea.
[††††]Department of Mechanical Engineering, Balochistan University of Engineering & Technology, Pakistan.
[†††††]Sindh Madressa-tul-Islam University, City Campus, Karachi, Pakistan.

Abstract

The growth of deep learning-based convolutional neural networks (CNNs) for image super-resolution (SR) tasks has improved every day and achieved tremendous performance in recent years. Many deep CNNs based image SR are restricted in practical applications due to their high computational cost, more memory consumption, and more training time. In this paper, we propose a residual-inception multiscale image super-resolution network known as RIMS. Proposed network architecture stacked a 3 CNN layers, 2 skip connection ResNet (SCRB) block and 2 multiscale inception blocks (MSIB) are followed by Leaky ReLU (LReLU) activation function. In addition, shrinking and expanding layers are also used to further reduce the number of parameters while preventing the over-fitting problem during the training. Furthermore, we used a deconvolution layer instead of interpolation to extract the rich features information for reconstruing the high-resolution (HR) output image. The experimental evaluations in terms of both quantitative, as well as qualitative, suggest that the proposed method achieves comparable performance to the existing state-of-the-art methods.

Keywords: *Supper-resolution, Convolutional neural networks, Leaky ReLU, PSNR.*

1. Introduction

One of the most important research challenges in the field of image and computer vision is the single image super-resolution (SISR), which attempts to reconstruct the high quality or high-resolution (HR) output image from their degraded version of low quality or low-resolution (LR) input image. The rapid evaluation of deep convolutional neural network-based image SR achieved remarkable performance as compared to earlier conventional methods. In this regard, Dong et al. first time introduced the shallow type network architecture known as super-resolution convolutional neural network (SRCNN) to recover the HR image [1]. The performance of SRCNN is improved as compared to previous SR approaches, but it increases the computational cost because all CNN layers operations are performed in the high-resolution space. Furthermore, to increase the performance, the same author introduced the new algorithm known as FSRCNN [2]. In FSRCNN authors are replaced the bicubic interpolation with a deconvolutional layer. For reducing the computational cost of the model shrinking, an expanding layer is employed. Aside from the study of shallow CNN methods, there is also a lot of research on deeper network architecture available. Performance of deeper model is better and having a low-computational cost in terms of number of parameters, but it introduces the vanishing gradient problem in the training. Skip connection was originally introduced into VDSR [3] by Kim et al, which is utilized to solve the vanishing gradient problem and aids in the training of deeper and larger networks. The performance of the deeper model is improved than shallow models, but they demand more computational cost and a very slow convergence rate. Additionally, the main issue with deeper network architecture is that the last end layers work as a dead layer because the information is not received properly and more important features are dropped. To report these issues, we proposed a novel architecture known as a residual-inception multiscale image super-resolution network known as RIMS. We used ResNet architecture-based block (SCRB) and Inception-based block (MSIB), which are significantly improved the quantitative as well as qualitative performance and reduced the computational burden of the model.

In summary, we design a new architecture known as single image super-resolution residual-inception multiscale image super-resolution network (RIMS), which produces a satisfactory performance with fewer number parameters as well as takes less CPU processing time. The contributions of this paper can be summarized as follows.

- For initial low-level feature extraction purposes, we used 3 CNN layers followed by Leaky ReLU activation.
- Mid and High-level feature extraction, we used SCRB and MSIB blocks. All blocks provide the multiscale features information simultaneously.
- We employed the post-processing strategy for upscaling the LR image into HR image because traditional CNN used pre-processing bicubic interpolation strategy for upscaling.

The remaining section of this paper is divided into different sections. Section 2 discusses the information about related works. In Section 3 and 4, explain the proposed methodology of RIMS architecture, as well as discuss its experimental calculations for reconstructing the HR image. Finally, we present the conclusion

Manuscript received January 5, 2022
Manuscript revised January 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.1.77

in Section 4.

2. Related works

Over the last few years, several research articles have used learning-based algorithms for image super-resolution tasks. Learning-based algorithms, specially CNN-based methods with deep learning techniques, may improve performance over existing traditional SR methods, which rely on hand-designed filters. The main purpose of single image SR is to recreate an HR image from an LR image. Initially, Dong et al. [4], started by using the bicubic interpolating the LR image and then training a CNN to learn a nonlinear mapping from the input image to the HR output. To accelerate the performance of SRCNN, Dong et.al., proposed without pre-processing bicubic interpolation technique known as FSRCNN [2]. The authors used shrinking, expanding, and deconvolution layers for reconstructing the HR output image from the original LR input image. Shi et al. suggested an efficient sub-pixel convolutional neural network (ESPCN) [5]. This architecture replaced both bicubic and deconvolution upscaling into a sub-pixel convolution layer. It is also belonging to shallow type network architecture using three CNN layers followed by the activation function. Kim et al. were the first time to train the deeper model up to a 20-layer network using residual skip connection [3]. The same author proposed another recursive type architecture for image super-resolution known as deeply recursive convolutional network (DRCN) [6], which recursively extends the receptive field while maintaining model capacity. Lai et al. [7] proposed Deep Laplacian Pyramid Networks known as LapSRN. LapSRN is a unique network architecture for accurate and quick SR reconstructing the HR image. This technique contributes two useful features. First, it begins by using cascade learning for residual output and reconstruction results from various scales. Second, instead of using the usual L_2 loss function, LapSRN uses a new Charbonnier loss function.

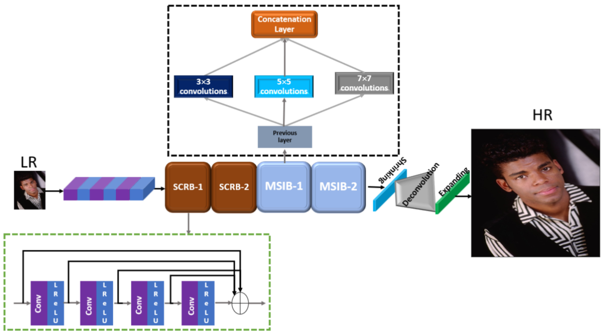


Figure 1. The proposed Network architecture of Residual-Inception Multiscale Image Super-Resolution Network (RIMS).

3. Proposed Methodology

In this section, we will introduce our residual-inception multiscale image super-resolution network known as RIMS in detail. Figure 1 shows the overall architecture of our proposed method and it initially uses three CNN layers for low-level feature extraction. The extracted low-level features are fed into the 2-skip connection ResNet blocks (SCRB) for mid-level feature extraction. The 2 multiscale inception blocks (MSIB) are employed to extract the high-level features information, and all layers are followed by Leaky ReLU (LReLU) activation function. Finally, fed the resultant features information into the deconvolution layer which is connected before and after the bottleneck layer. The bottleneck layer connected before the deconvolution layer is known as the shrinking layer, and after deconvolution is called as expanding layer. The details of each step are discussed as under:

Traditional approaches extract the local features information through the manually hand-designed filter, we extract the features information automatically from the deep learning-based convolutional neural network approach. However, the conventional deep CNN-based image SR method used the bicubic interpolation to upscale the LR image into an HR image. Authors claim that [8] bicubic interpolation is not designed for this purpose and introduces the new noises in the model. Therefore, we used an alternate strategy and extract the features directly from the original LR space domain with the help of CNN layers. For initial low-level feature extraction, we used 3 CNN layers followed by the Leaky ReLU activation function. All 3 CNN layers have a kernel size of the order 3 × 3 with the same padding to preserve the spatial features information.

Inspired by [9], a skip connection technique is used to extract features information of the first, second, third, and fourth layers concatenated together. All CNN layers are followed by the Leaky ReLU activation function to avoid the dying ReLU problem. The resultant features are fed into the next block, which is the multiscale inception blocks (MSIB) to extract the multiscale features information smoothly. The concept of this block is borrowed from GoogLeNet [10] as shown in Figure 2 (a). The original inception block used 3 types of convolution layers with kernel size is 1 × 1, 3 × 3, 5 × 5, and one max pooling layer to extract the multiscale features information. In our proposed MSIB block consists of two stages and is connected by a concatenation layer as shown in Figure 2(b). Furthermore, our proposed block removes the max poling layer because it is a type of candidate selection and not preserve the spatial information.

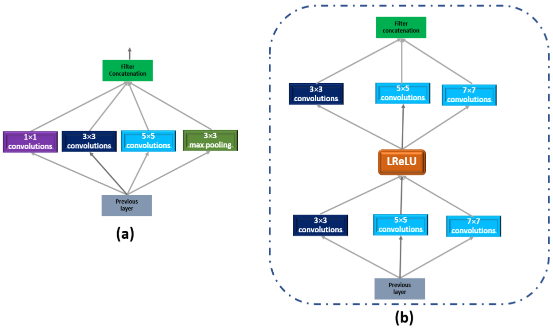


Figure 2. Inception module, naive version [11] (a), and our proposed MSIB (b)

Upscaling plays a vital role during the reconstruction of HR images. Earlier approaches use different ways to upscale the LR image into HR, such as bicubic, bilinear, and nearest-neighbor interpolation. These hands designed technique is not suitable for converting the LR image into an HR image, because these approaches are not designed for image upscaling purposes. To resolve such types of problems, we employed the deconvolution layer to upscale the LR into HR features. Additionally, deconvolution operation performs the inverse of the convolution operation. To maintain the computational complexity and processing speed of the model, we add two bottleneck layers before and after the deconvolution layer. The first bottleneck layer work as a shrinking layer and the later layer is known as expanding layer. Finally, reconstruct the visually pleasing high-quality HR output image.

4. Experimental Results

method are discussed in this section. Furthermore, we also present the quantitative as well as qualitative comparison with the existing state-of-the-art SR method on enlargement scale factor 2× and 4×. Our model has trained on the combination of Yang 91 [12] and BSDS200 [13] image datasets. Additionally, to avoid the overfitting problem, the authors used the data augmentation technique in terms of flipping, rotation, and scaling. The low-resolution images were generated using the bicubic built-in function in MATLAB in the scale factor 2× and 4×. The Adam optimizer is used with an initial learning rate is 0.0003 including 32 as a mini-batch size. Our model fully converges on 200 epochs. For training purposes, we run our code on NVIDIA GPU RTX 2070. Keras TensorFlow library is used for designing the model architecture. We evaluated the performance of our proposed model on publicly available benchmark test datasets such as SET14 [14], BSDS100 [13], and URBAN100 [15]. The most used technique to measure the perceptual quality of the image is the peak-signal-to-noise ratio (PSNR). Higher PSNR means the reconstructed image has more visually pleasing and vice versa. PSNR can be easily explained by mean squared error (MSE).

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

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

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

$$PSNR = 20\log_{10}(MAX_I) - 10\log_{10}(MSE) \qquad (3)$$

Similarly, another perceptual quality metric is the structural similarity index (SSIM), which quantifies the image quality degradation due to losses or compression. SSIM depends on main three factors such as structure, luminance, and contrast. We recorded the execution time of all the compared algorithms using the same workstation with a 3.40 GHz Intel i7 CPU to assess their effectiveness (16 GB RAM) as shown in Figures 3 and 4.

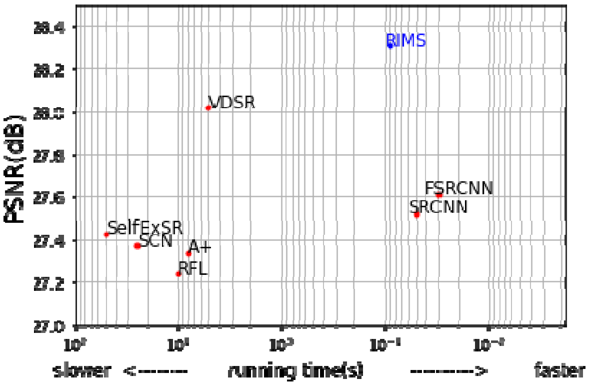


Figure 3. Trade-off performance in terms of PSNR and running time on SET14 test dataset with enlargement factor 4×.

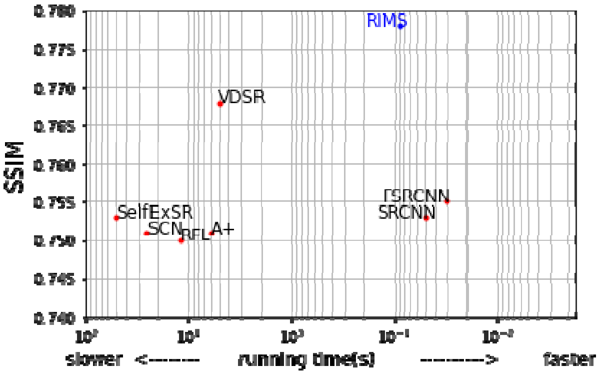


Figure 4. Trade-off performance in terms of SSIM and running time on SET14 test dataset with enlargement factor 4×.

The quantitative comparison in terms of PSNR and SSIM of our model with state-of-the-art methods including Bicubic, A+, RFL, SelfExSR, SRCNN, FSRCNN, and VDSR as shown in Table 1 and 2. Our model achieves the best performance in terms of PSNR/SSIM on all test datasets, such as SET14, BSDS100, and URBAN100. Furthermore, quantitively comparison results are presented in Figures 5 and 6, that our method has a higher ranking than other methods.

Table 1. Experimental evaluation in terms of PSNR of our proposed method with other image SR methods with scale factor 2×, and 4×. First-best values are indicated in red color with bold and second-best in blue colors with an underline.

Algorithms	*Factor*	**SET14** **PSNR**	**BSDS100** **PSNR**	**URBAN100** **PSNR**
Bicubic	2×	30.25	29.57	26.89
A+ [14]	2×	32.32	31.24	29.25
RFL [16]	2×	32.29	31.18	29.14
SelfExSR [15]	2×	32.24	31.20	29.55
SRCNN [4]	2×	32.51	31.38	29.53
FSRCNN [2]	2×	32.66	31.53	29.88
SCN [17]	2×	32.35	31.26	29.52
VDSR [3]	2×	33.05	31.90	30.77

RIMS (ours)	2×	**33.09**	**31.92**	**30.79**
Bicubic	4×	26.01	25.97	23.15
A+ [14]	4×	27.34	26.83	24.34
RFL [16]	4×	27.24	26.76	24.20
SelfExSR [15]	4×	27.41	26.84	24.83
SRCNN [4]	4×	27.52	26.91	24.53
FSRCNN [2]	4×	27.61	26.98	24.62
SCN [17]	4×	27.39	26.88	24.52
VDSR [3]	4×	28.02	27.29	25.18
RIMS (ours)	4×	**28.31**	**27.35**	**25.23**

Table 2. Experimental evaluation of our proposed method in terms of SSIM with other image SR methods with scale factor 2×, and 4×. First-best values are indicated in red color with bold and second-best in blue colors with an underline.

Algorithms	*Factor*	**SET14 SSIM**	**BSDS100 SSIM**	**URBAN100 SSIM**
Bicubic	2×	0.870	0.844	0.841
A+ [14]	2×	0.906	0.887	0.895
RFL [16]	2×	0.905	0.885	0.891
SelfExSR [15]	2×	0.904	0.887	0.898
SRCNN [4]	2×	0.908	0.889	0.896
FSRCNN [2]	2×	0.909	0.892	0.902
SCN [17]	2×	0.905	0.885	0.897
VDSR [3]	2×	0.913	0.896	0.914
RIMS (ours)	2×	**0.920**	**0.896**	**0.915**
Bicubic	4×	0.704	0.670	0.660
A+ [14]	4×	0.751	0.711	0.721
RFL [16]	4×	0.747	0.708	0.712
SelfExSR [15]	4×	0.753	0.713	0.740
SRCNN [4]	4×	0.753	0.712	0.725
FSRCNN [2]	4×	0.755	0.715	0.728
SCN [17]	4×	0.751	0.711	0.726
VDSR [3]	4×	0.768	0.726	0.754
RIMS (ours)	4×	**0.778**	**0.731**	**0.758**

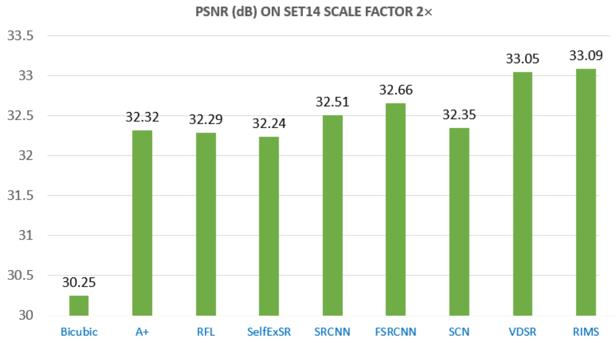


Figure 5. Quantitative evaluation of PSNR on SET14 dataset enlargement factor 2× with other state-of-the-art methods. Our proposed method obtained the highest PSNR as compared to other methods.

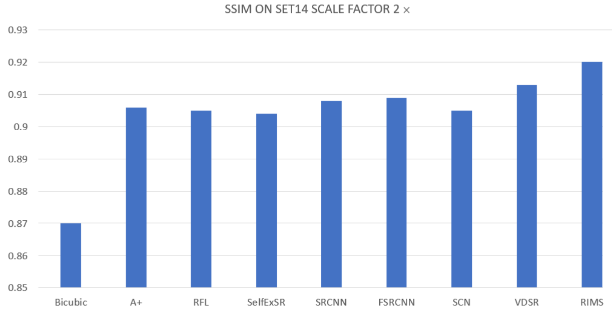
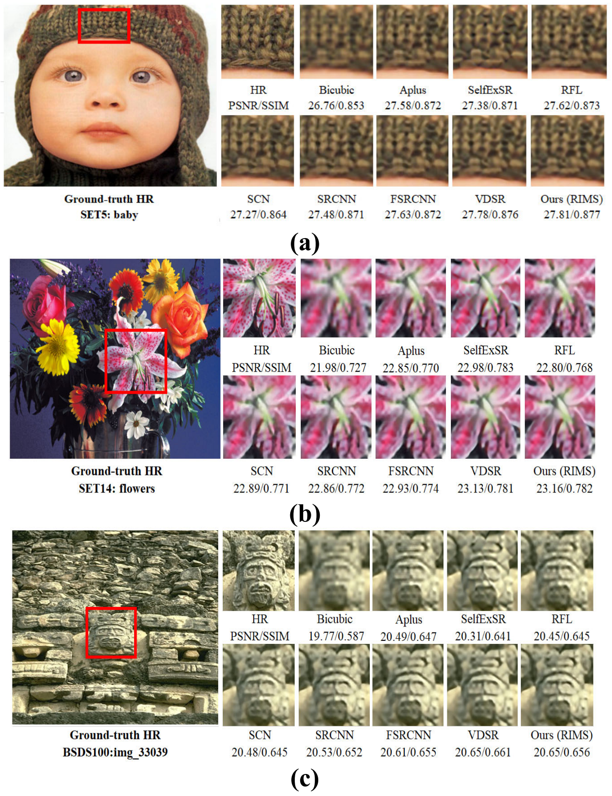
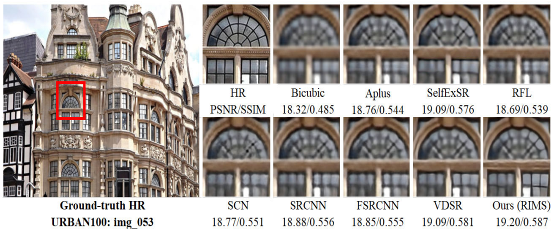


Figure 6. Quantitative evaluation of SSIM on SET14 dataset enlargement factor 2× with other state-of-the-art methods. Our proposed method obtained the highest SSIM as compared to other methods.

Evaluation from the perceptual quality point of view is shown in Figure 7. In Figure 7, we used baby image obtained from SET5, flowers image obtained from SET14, and the other two images are from BSDS100 and URBAN100. Perceptual quality results of Bicubic and SRCNN are blurry and not clear view, but VDSR and our (RIMS) obtained visually pleasing as compared to the baseline method.





(d)

Figure 7. Qualitative as well as quantitative comparisons of our (RIMS) approach with other state-of-the-art image SR approaches on sale factor 4×.

5. Conclusion

In this paper, we propose a residual-inception multiscale image super-resolution network known as RIMS. The proposed architecture stacked 3 CNN layers, skip connection ResNet (SCRB) block and multiscale inception blocks (MSIB) are followed by Leaky ReLU (LReLU). For increasing the computational efficiency, the authors used shrinking and expanding layers before and after the deconvolution layer. Additionally, earlier approaches are feed interpolated versions of HR images into a convolutional neural network for extracting the low, mid, and high-level features. Such methodologies improved the performance, but it introduces extra new noises in the model and increase the computational burden during the training. In our approach used a deconvolution layer at the later end to extract the features information efficiently. The quantitative and qualitative calculations suggest that our proposed approach achieves comparable performance to the other image SR methods. As future work, we plan to extend our work with the xception model, which is also introduced the GoogLeNet and its performance is better as compared to all previous approaches

References

1. Dong, C., et al., *Image super-resolution using deep convolutional networks.* IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(2): p. 295-307.
2. Dong, C., C.C. Loy, and X. Tang. *Accelerating the super-resolution convolutional neural network*. in *European conference on computer vision*. 2016. Springer.
3. 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.
4. Dong, C., et al. *Learning a deep convolutional network for image super-resolution*. in *European conference on computer vision*. 2014. Springer.
5. 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.
6. 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.
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. Wang, Y., et al., *End-to-end image super-resolution via deep and shallow convolutional networks.* IEEE Access, 2019. **7**: p. 31959-31970.
9. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
10. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
11. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
12. Yang, J., et al., *Image super-resolution via sparse representation.* IEEE transactions on image processing, 2010. **19**(11): p. 2861-2873.
13. 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.
14. 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.
15. Huang, J.-B., A. Singh, and N. Ahuja. *Single image super-resolution from transformed self-exemplars*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
16. 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.
17. Wang, Z., et al. *Deep networks for image super-resolution with sparse prior*. in *Proceedings of the IEEE international conference on computer vision*. 2015.