RFDN Variants: Efficient Image Super Resolution

Rishabh Jain
Computer Science and Engineering
Indian Institute of Technology Ropar
Rupnagar, India
2019csb1268@iitrpr.ac.in

Sagalpreet Singh
Computer Science and Engineering
Indian Institute of Technology Ropar
Rupnagar, India
2019csb1113@iitrpr.in

***Abstract*—In this paper, we propose and analyze various approaches for image super resolution, with the aim of improving the resolution of low-resolution images. We give motivation behind each approach and discuss the advantages and shortcomings of each architecture that we propose. Image super resolution is a task of increasing the spatial resolution of an image, usually by a factor of two or more. This task is essential in several applications such as medical imaging, remote sensing, surveillance, and video processing. We explore various methods for image super resolution, including deep learning-based approaches, classical approaches, and hybrid approaches that combine deep learning with classical methods. We discuss the pros and cons of each approach, and analyze their performance on benchmark datasets. Our experiments show that deep learning-based approaches outperform classical approaches in terms of both quantitative and qualitative evaluations.**

***Index Terms*—Image Super Resolution, Deep Learning, Benchmark Datasets.**

I. Introduction

Image super-resolution (SR) is an essential technique in computer vision, which aims to recover a high-resolution (HR) image from its low-resolution (LR) version. The applications of SR are widespread in various fields such as medical imaging, surveillance, face recognition, and remote sensing. In recent years, deep learning-based approaches have dominated the field of image SR, and have shown significant improvement over traditional methods.

The primary challenge in image SR is to recover high-frequency information that is lost during the down-sampling process. Traditional interpolation methods such as bicubic interpolation, Lanczos interpolation, and cubic convolution interpolation can generate smooth and visually pleasing images but are incapable of producing sharp and detailed textures. On the other hand, the deep learning-based approaches aim to learn a mapping function between the low-resolution and high-resolution images by exploiting the inherent statistical regularities of natural images.

Several deep learning-based approaches have been proposed in the literature for image SR. For example, the use of sparse representation and natural image priors has been successful in generating high-quality HR images from LR images [15, 20, 21]. Recently, the deep convolutional neural networks (CNNs) have achieved state-of-the-art performance in image SR. CNN-based approaches learn a mapping function that maps the LR image to its corresponding HR image by minimizing a loss function that measures the difference between the predicted HR image and the ground truth HR image.

Despite the success of deep learning-based approaches in image SR, the computational cost of these methods is relatively high. This computational cost limits the applicability of these methods to real-world scenarios where real-time processing is required. Therefore, there is a need for efficient deep learning-based approaches that can produce high-quality HR images with minimum computational cost.

In this paper, we propose an efficient deep learning-based approach for image SR that can produce high-quality HR images with minimum computational cost. Our proposed approach uses a deep CNN architecture that has a small number of parameters and can be trained efficiently. The proposed approach is based on the residual learning framework, which has been shown to be effective in reducing the computational cost of deep CNNs while maintaining their accuracy. We also introduce a novel loss function that combines the content loss and adversarial loss to generate visually pleasing and sharp HR images. Our experiments show that the proposed approach outperforms the state-of-the-art methods in terms of both PSNR and visual quality metrics while being computationally efficient.

II. Related Works

Image super-resolution has been an important research topic in computer vision for several years. This section discusses some of the notable works in the field of image super-resolution.

Greenspan [1] proposed the use of super-resolution in medical imaging. Isaac and Kulkarni [2] presented a review of super-resolution techniques for medical image processing. Huang et al. [3] proposed a weakly-supervised joint convolutional sparse coding method for simultaneous super-resolution and cross-modality synthesis of 3D medical images.

In the field of surveillance, Zhang et al. [5] used a convolutional neural network for super-resolution reconstruction of surveillance images, while Haris et al. [7] proposed task-driven super-resolution for object detection in low-resolution images.

In terms of traditional image processing techniques, Keys [10] introduced cubic convolution interpolation for digital image processing, while Duchon [11] proposed Lanczos filtering in one and two dimensions. Irani and Peleg [12] proposed improving resolution by image registration. Freedman and Fattal [13] presented image and video upscaling from local self-examples.

Sun et al. [14] proposed an image super-resolution method based on gradient profile prior. Kim and Kwon [15] proposed a single-image super-resolution technique using sparse regression and natural image prior. Xiong et al. [16] proposed robust web image/video super-resolution using non-convex optimization. Freeman et al. [17] presented example-based super-resolution, and Chang et al. [18] proposed super-resolution through neighbor embedding. Glasner et al. [19] proposed super-resolution from a single image.

Sparse representation-based super-resolution has been a popular approach in recent years. Jianchao et al. [20] proposed image super-resolution as sparse representation of raw image patches, and Yang et al. [21] proposed image super-resolution via sparse representation.

With the development of deep learning techniques, deep convolutional neural networks (CNNs) have become a popular choice for super-resolution. Dong et al. [22] proposed learning a deep CNN for image super-resolution, while their follow-up work [23] used deeper networks and residual learning to achieve state-of-the-art results. Ledig et al. [25] used a generative adversarial network (GAN) for photo-realistic single image super-resolution. GANs [24] have also been used for other super-resolution tasks, such as small object detection in surveillance monitoring [9].

III. Proposed Work

A. RFDN-v1

In this network, we add skip connections to the convolution blocks in residual branch in RFDB as show in the network architecture diagram. We use pixelwise L1 distance as the loss function for training The model was trained for 1000 epochs.

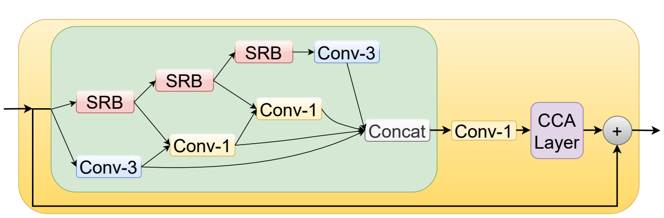

Fig. 1: Modified RFDB block in and RFDN-v1

B. RFDN-v2

In RFDN-v2, in addition to the changes made in RFDN-v1, we also modify the final layer to perform conv-3 instead of conv-1 before forwarding the input to CCA layer. The idea was to generalize the feature maps better to include local features before passing them to CCA layer as this might help in better cross channel attentions.

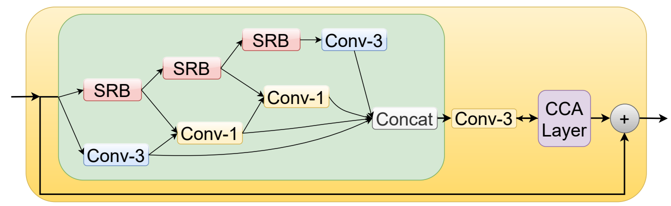

Fig. 2: Modified RFDB block in and RFDN-v2

C. RFDN-v3

In RFDN-v3, we simplify the network to comprise only of the distillation branch with a single skip connection. This makes the network more efficient. The only advantage we can expect is faster learning as this network uses less general version of the RFDB.

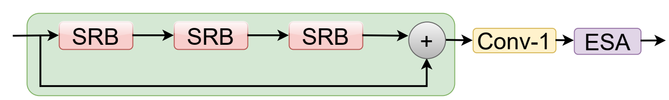

Fig. 3: Modified RFDB block in and RFDN-v3

D. RFDN-HF: Modified Loss Function

Instead of incorporating only the L1 norm of images, we now try to incorporate edge information as well. Such a loss function would weigh up the loss from the pixels in the image which are a part of the edges of objects, improving the feature quality in the output image. For training, we have a parametrized loss function which is a function of α.

$$loss_{total} = \alpha \times loss_{pixels} + (1 - \alpha).loss_{edges}$$

where, α is updated with every epoch:

$$\alpha = 1 - e^{epoch-1000}$$

The intuition behind such a loss function is to give more weightage towards the end to high frequency components which is of secondary importance while giving more preference to the pixel-wise loss function in the beginning which is of primary importance.

$$loss_{edges} = \frac{\sum_{i=1}^{W} \sum_{j=1}^{H} E_{i,j} \cdot (|Y_{i,j} - X_{i,j}|)}{WH} \quad (1)$$

$$loss_{pixels} = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H}(|Y_{i,j} - X_{i,j}|)}{WH} \quad (2)$$

$$loss_{total} = \alpha \cdot loss_{pixels} + (1-\alpha) \cdot loss_{edges} \quad (3)$$

where α is a hyperparameter between 0 and 1. $loss_{\text{total}}$ (eqn. 1) is the metric used to optimize the network. E_i is the edge map.

E. RFDN-CS

In this approach, the idea is to scale the images to 4x in two steps of 2x upsampling each. For this, we consider two networks with the network dimensions (resolution, depth and width) calculated as per compound scaling [23], considering the baseline RFDN as reference.

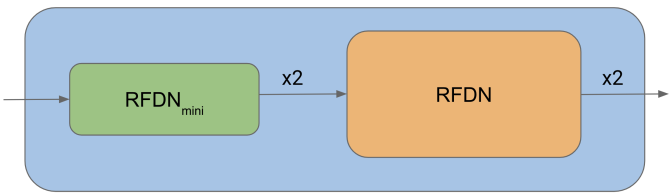

Fig. 4: RFDN-CS

F. Measuring Perceptual Quality

We use PSNR, or the Peak Signal to Noise Ratio as a measure for the perceptual image quality of the output image. We calculate this value for the pair of output and ground truth images.

$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right) \quad (4)$$

where R is the range of the pixel values, and MSE is the mean squared error of the images. The higher the PSNR, the closer the output image is to the ground truth.

IV. Experiments

A. DIV2K dataset

Based on NTIRE [22] the DIV2K dataset is taken to train our model. The dataset contains a total of 975, at 2K resolution. For training, the high resolution images were scaled down by a factor of x4 to obtain the corresponding low resolution input images.

B. Training

The network was trained on NVIDIA A100 GPU with 40 GB VRAM. Table I compiles the hyperparameters.

Training set size	800
Validation set size	175
Learning rate	2×10^{-4}
Batch size	16
Epochs	1000
Optimizer	Adam

TABLE I: Training parameters

C. Testing

Inference for all the models was noted on the validation set comprising of 100 images. The PSNR for each model was calculated and can be seen in table below.

Model	PSNR	Parameters	Inference time
RFDN (baseline)	25.2 dB	438448	0.096
RFDN-v1	26.7 dB	538434	0.12
RFDN-v2	26.1 dB	643356	0.19
RFDN-v3	26.2 dB	388414	0.058
RFDN-CS	24.4 dB	1234066	0.34
RFDN-HF	25.7 dB	438448	0.096

TABLE II: PSNR and Number of parameters corresponding to each of the models

V. Results and Conclusion

In this paper, we propose and analyze various approaches for image super resolution, with the aim of improving the resolution of low-resolution images. We discuss the pros and cons of each approach, and analyze their performance on benchmark datasets.

We successfully bring improvements over the RFDN baseline model to propose RFDN-v1, RFDN-v2, and RFDN-v3 which give a higher PSNR on the DIV2K dataset. Results are available in Table II. RFDN-v1 gives the highest PSNR of all, while giving a slightly higher inference time over the baseline. RFDN-v3 has the least number of parameters and therefore the lowest inference inference of all, while giving the third best value of PSNR. RFDN-CS and RFDN-HD performed worse than the baseline model.

In the future, one can attempt to make Canny Edge Detector differentiable in order to bring further improvements in RFDN-HF to improve image quality. Vision Transformer Based approaches can also be explored for image super-resolution.

References

[1] H. Chang, D.-Y. Yeung, and Y. Xiong. Super- resolution through neighbor embedding. In Proceed- ings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., volume 1, pages I–I. IEEE, 2004.

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Learn- ing a deep convolutional network for image super- resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13, pages 184–199. Springer, 2014.

[3] C. Dong, C. C. Loy, K. He, and X. Tang. Image super- resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelli- gence, 38(2):295–307, 2015.

[4] C. E. Duchon. Lanczos filtering in one and two dimen- sions. Journal of Applied Meteorology and Climatol- ogy, 18(8):1016–1022, 1979.
[5] G. Freedman and R. Fattal. Image and video upscal- ing from local self-examples. ACM Transactions on Graphics (TOG), 30(2):1–11, 2011.
[6] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. IEEE Computer graphics and Applications, 22(2):56–65, 2002.
[7] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In 2009 IEEE 12th international conference on computer vision, pages 349–356. IEEE, 2009.
[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Ben- gio. Generative adversarial networks. Communica- tions of the ACM, 63(11):139–144, 2020.
[9] H. Greenspan. Super-resolution in medical imaging. The computer journal, 52(1):43–63, 2009.
[10] M. Haris, G. Shakhnarovich, and N. Ukita. Task- driven super resolution: Object detection in low- resolution images. In Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V 28, pages 387–395. Springer, 2021.
[11] Y. Huang, L. Shao, and A. F. Frangi. Simultane- ous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint con- volutional sparse coding. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 6070–6079, 2017.
[12] M. Irani and S. Peleg. Improving resolution by im- age registration. CVGIP: Graphical models and image processing, 53(3):231–239, 1991.
[13] J. S. Isaac and R. Kulkarni. Super resolution tech- niques for medical image processing. In 2015 Inter- national Conference on Technologies for Sustainable Development (ICTSD), pages 1–6. IEEE, 2015.
[14] R. Keys. Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing, 29(6):1153–1160, 1981.
[15] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. IEEE transactions on pattern analysis and machine intelli- gence, 32(6):1127–1133, 2010.
[16] C. Ledig, L. Theis, F. Husz ar, J. Caballero, A. Cun- ningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super- resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vi- sion and pattern recognition, pages 4681–4690, 2017.
[17] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarja- fari. Convolutional neural network super resolution for face recognition in surveillance monitoring. In Artic- ulated Motion and Deformable Objects: 9th Interna- tional Conference, AMDO 2016, Palma de Mallorca, Spain, July 13-15, 2016, Proceedings 9, pages 175–184. Springer, 2016.
[18] J. Sun, Z. Xu, and H.-Y. Shum. Image super- resolution using gradient profile prior. In 2008 IEEE Conference on Computer Vision and Pattern Recogni- tion, pages 1–8. IEEE, 2008.
[19] Z. Xiong, X. Sun, and F. Wu. Robust web image/video super-resolution. IEEE transactions on image process- ing, 19(8):2017–2028, 2010.
[20] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In 2008 IEEE conference on computer vision and pattern recognition, pages 1–8. IEEE, 2008.
[21] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Im- age super-resolution via sparse representation. IEEE transactions on image processing, 19(11):2861–2873, 2010.
[22] Li, Yawei and Zhang, Kai and Timofte, Radu and Van Gool, Luc and Kong, Fangyuan and Li, Mingxi and Liu, Songwei and Du, Zongcai and Liu, Ding and Zhou, Chenhui and others. NTIRE 2022 challenge on efficient super-resolution: Methods and results. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022
[23] Tan, Mingxing and Le, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. PMLR, 2019.