iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42718

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Enhanced Super-Resolution Using GAN

Chandan Kumar[1], Amzad Choudhary[2], Gurpreet Singh[3], Ms. Deepti Gupta[4]
[1, 2, 3, 4]*Maharaja Agrasen Institute of Technology, New Delhi*

Abstract: ***Super-resolution reconstruction is an increasingly important area in computer vision. To eliminate the problems that super-resolution reconstruction models based on generative adversarial networks are difficult to train and contain artifacts in reconstruction results. besides the breakthroughs in accuracy and speed of single image super-resolution using faster and deeper convolutional neural networks. However, the hallucinated details are often accompanied with unpleasant artifacts. This paper presented ESRGAN model which was also based on generative adversarial networks. To further enhance the visual quality, we thoroughly study three key components of SRGAN – network architecture, adversarial loss and perceptual loss, and improve each of them to derive an Enhanced SRGAN (ESRGAN). In particular, we introduce the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit. Moreover, we borrow the idea from relativistic GAN to let the discriminator predict relative realness instead of the absolute value. Finally, we improve the perceptual loss by using the features before activation, which could provide stronger supervision for brightness consistency and texture recovery. Benefiting from these improvements, the proposed ESRGAN achieves consistently better visual quality with more realistic and natural textures than SRGAN.***

I. INTRODUCTION

With the popularization of the Internet and the development of information technology, the amount of information accepted by humans is growing at an explosive rate. Images, videos and audio are the main carriers of information transmission. Related research has pointed out that the information humans receive through vision accounts for 60%-80% of all media information, so visible images are an important way to obtain information. However, the quality of an image is often restricted by hardware equipment such as imaging systems and the bandwidth during the image transmission process. A low-resolution (LR) image with missing details is eventually presented. The reduction of image resolution will cause a serious decrease in image quality. It will greatly affect people's visual experience and cannot meet the requirements for image quality performance indicators in industrial production. Therefore, how to obtain high-resolution (HR) images has become an urgent issue. Single image super-resolution (SISR), as a fundamental low-level vision problem, has attracted increasing attention in the research community and AI companies. SISR aims at recovering a high-resolution (HR) image from a single low-resolution (LR) one. Since the pioneer work of SRCNN proposed by Dong et al. deep convolutional neural network (CNN) approaches have brought prosperous development. Various network architecture designs and training strategies have continuously improved the SR performance, especially the Peak Signal-to-Noise Ratio (PSNR) value However, these PSNRoriented approaches tend to output over-smoothed results without sufficient high-frequency details, since the PSNR metric fundamentally disagrees with the subjective evaluation of human observers.

Several perceptual-driven methods have been proposed to improve the visual quality of SR results. For instance, perceptual loss is proposed to optimize super-resolution models in a feature space instead of pixel space. Generative adversarial network is introduced to SR to encourage the network to favor solutions that look more like natural images. The semantic image prior is further incorporated to improve recovered texture details. One of the milestones in the way of pursuing visually pleasing results is SRGAN. The basic model is built with residual blocks and optimized using perceptual loss in a GAN framework. With all these techniques, SRGAN significantly improves the overall visual quality of reconstruction over PSNR-oriented methods. However, there still exists a clear gap between SRGAN results and the ground-truth images, as shown in Fig. 1. In this study, we revisit the key components of SRGAN and improve the model in three aspects. First, we improve the network structure by introducing the Residual-in-Residual Dense Block (RDDB), which is of higher capacity and easier to train. We also remove Batch Normalization (BN) layers as in and use residual scaling and smaller initialization to facilitate training a very deep network. Second, we improve the discriminator using Relativistic average GAN (RaGAN), which learns to judge "whether one image is more realistic than the other " rather than "whether one image is real or fake". Our experiments show that this improvement helps the generator recover more realistic texture details. Third, we propose an improved perceptual loss by using the VGG features before activation instead of after activation as in SRGAN. We empirically find that the adjusted perceptual loss provides sharper edges and more visually pleasing results, as will be shown fig 1.

©IJRASET: All Rights are Reserved

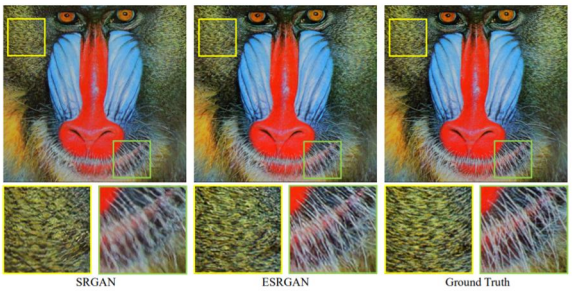


Fig 1.

II. PROPOSED METHODS

Our main aim is to improve the overall perceptual quality for SR. In this section, we first describe our proposed network architecture and then discuss the improvements from the discriminator and perceptual loss. At last, we describe the network interpolation strategy for balancing perceptual quality and PSNR.

A. Network Architecture

In order to further improve the recovered image quality of SRGAN, we mainly make two modifications to the structure of generator G. Firstly, remove all BN layers. secondly, replace the original basic block with the proposed Residual-in-Residual Dense Block (RRDB), which combines multi-level residual network and dense connections as depicted in Fig 2.

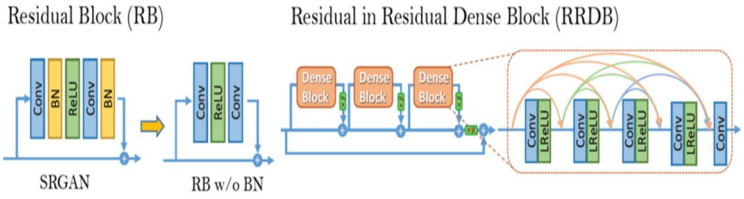


Fig 2

removing BN layers has proven to increase performance and reduce computational complexity in different PSNR-oriented tasks including SR and deblurring. BN layers normalize the features using mean and variance in a batch during training and use estimated mean and variance of the whole training dataset during testing. When the statistics of training and testing datasets differ a lot, BN layers tend to introduce unpleasant artifacts and limit the generalization ability. We empirically observe that BN layers are more likely to bring artifacts when the network is deeper and trained under a GAN framework. These artifacts occasionally appear among iterations and different settings, violating the needs for a stable performance over training. We therefore remove BN layers for stable training and consistent performance. Furthermore, removing BN layers helps to improve generalization ability and to reduce computational complexity and memory usage. We keep the high-level architecture design of SRGAN, and use a novel basic block namely RRDB as depicted in Fig. 2. Based on the observation that more layers and connections could always boost performance the proposed RRDB employs a deeper and more complex structure than the original residual block in SRGAN. Specifically, as shown in Fig. 2 the proposed RRDB has a residual-in-residual structure, where residual learning is used in different levels. A similar network structure is proposed that also applies a multilevel residual network. However, our RRDB differs in that we use dense blocks in the main path as, where the network capacity becomes higher benefiting from the dense connections. In addition to the improved architecture, we also exploit several techniques to facilitate training a very deep network: 1 residual scaling, scaling down the residuals by multiplying a constant between 0 and 1 before adding them to the main path to prevent instability, 2 smaller initialization, as we empirically find residual architecture is easier to train when the initial parameter variance becomes smaller.

B. Relativistic Discriminator

Besides the improved structure of the generator, we also enhance the discriminator based on the Relativistic GAN. Different from the standard discriminator D in SRGAN, which estimates the probability that one input image x is real and natural, a relativistic discriminator tries to predict the probability that a real image is relatively more realistic than a fake one. This modification of the discriminator helps to learn sharper edges and more detailed textures.

©IJRASET: All Rights are Reserved

C. Perceptual Loss

We also develop a more effective perceptual loss Lpercep by constraining on features before activation rather than after activation as practiced in SRGAN. Based on the idea of being closer to perceptual similarity Johnson et al. propose perceptual loss and it is extended in SRGAN. Perceptual loss is previously defined on the activation layers of a pre-trained deep network, where the distance between two activated features is minimized. Contrary to the convention, we propose to use features before the activation layers, which will overcome two drawbacks of the original design. First, the activated features are very sparse, especially after a very deep network.

D. Network Interpolation

To remove unpleasant noise in GAN-based methods while maintaining a good perceptual quality, we propose a flexible and effective strategy – network interpolation. Specifically, we first train a PSNR-oriented network G-PSNR and then obtain a GAN-based network G-GAN by fine-tuning.

The proposed network interpolation enjoys two merits. First, the interpolated model is able to produce meaningful results for any feasible α without introducing artifacts. Second, we can continuously balance perceptual quality and fidelity without re-training the model. We also explore alternative methods to balance the effects of PSNR-oriented and GAN-based methods. For instance, one can directly interpolate their output images (pixel by pixel) rather than the network parameters. However, such an approach fails to achieve a good trade-off between noise and blur, i.e. the interpolated image is either too blurry or noisy with artifacts. Another method is to tune the weights of content loss and adversarial loss. But this approach requires tuning loss weights and fine-tuning the network, and thus it is too costly to achieve continuous control of the image style. We compare the effects of network interpolation and image interpolation strategies in balancing the results of a PSNR-oriented model and GAN-based method. We apply simple linear interpolation on both the schemes. The pure GAN-based method produces sharp edges and richer textures but with some unpleasant artifacts, while the pure PSNRoriented method outputs cartoon-style blurry images. By employing network interpolation, unpleasing artifacts are reduced while the textures are maintained. By contrast, image interpolation fails to remove these artifacts effectively.

E. Training Details

Following SRGAN, all experiments are performed with a scaling factor of 4×4 between LR and HR images. The experimental platform we use is NVIDIA GeForceMX150, Intel (R) Core (TM) i7-8550U CPU@2.00GHz, 8 GB RAM, the compilation software we use are pycharm2017 and and the pytorch deep learning toolbox is used to build and train the network. . This paper uses the DIV2K dataset, which consists of 800 training images, 100 validation images and 100 testing images. We augment the training data with random horizontal flips and 90 rotations. We perform experiments on three widely used benchmark datasets Set5, Set14 and BSD100. All experiments are performed with a scale factor of 4×4 between low- and high-resolution images. The mini-batch size is set to 16. The spatial size of the cropped HR patch is 128 × 128. We observe that training a deeper network benefits from a larger patch size, since an enlarged receptive field helps to capture more semantic information. However, it costs more training time and consumes more computing resources. This phenomenon is also observed in PSNR-oriented methods.

III. RESULTS

Our proposed ESRGAN outperforms previous approaches in both sharpness and details. For instance, ESRGAN can produce sharper and more natural baboon's whiskers and grass textures than PSNR-oriented methods, which tend to generate blurry results, and than previous GAN-based methods, whose textures are unnatural and contain unpleasing noise. ESRGAN is capable of generating more detailed structures in building while other methods either fail to produce enough details (SRGAN) or add undesired textures. Moreover, previous GAN-based methods sometimes introduce unpleasant artifacts, e.g., SRGAN adds wrinkles to the face. Our ESRGAN gets rid of these artifacts and produces natural results.

IV. CONCLUSION

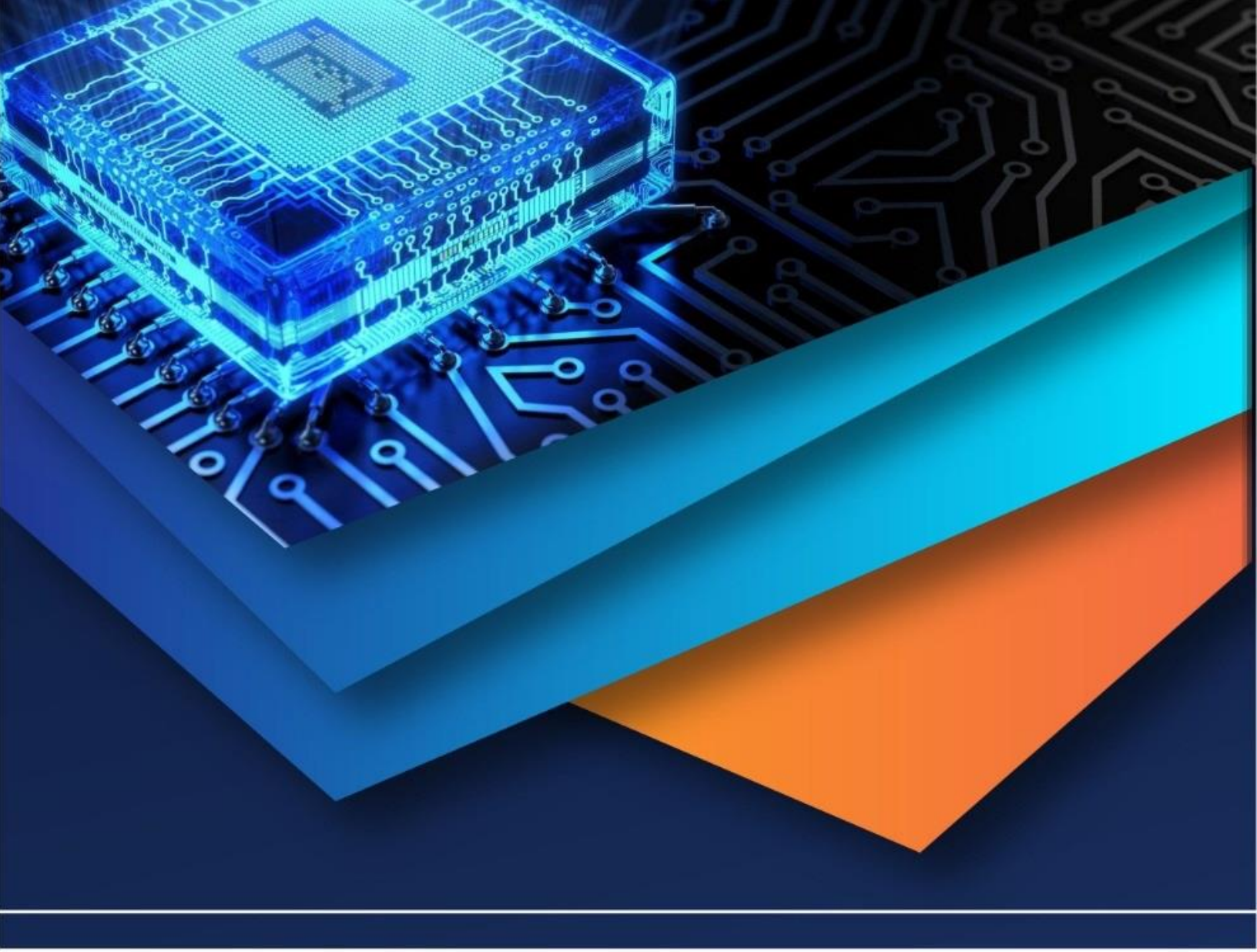
We have presented an ESRGAN model that achieves consistently better perceptual quality than previous SR methods. We have formulated a novel architecture containing several RDDB blocks without BN layers. In addition, useful techniques including residual scaling and smaller initialization are employed to facilitate the training of the proposed deep model. We have also introduced the use of relativistic GAN as the discriminator, which learns to judge whether one image is more realistic than another, guiding the generator to recover more detailed textures. Moreover, we have enhanced the perceptual loss by using the features before activation, which offer stronger supervision and thus restore more accurate brightness and realistic textures.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPRW (2017)
[2] Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017)
[3] Bell, S., Upchurch, P., Snavely, N., Bala, K.: Material recognition in the wild with the materials in context database. In: CVPR (2015)
[4] Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC. BMVA press (2012)
[5] Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: 2018 pirm challenge on perceptual image super-resolution. arXiv preprint arXiv:1809.07517 (2018)
[6] Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: CVPR (2017)
[7] Bruna, J., Sprechmann, P., LeCun, Y.: Super-resolution with deep convolutional sufficient statistics. In: ICLR (2015)
[8] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV (2014)
[9] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI 38(2), 295–307 (2016)
[10] Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: NIPS (2015).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)