RESEARCH PAPER

The Perception-Fidelity Trade-Off in GAN-Based Super-Resolution

Dr. V. Sumalatha[1*], **Kalakota Praveen**[1], **Dr. G. Praveen**[1], **Mr. S. Sudharshan**[1]
[1]Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Ghatkesar - 500088, Telangana, India
[1]Assistant Professor (Dr. V. Sumalatha) | Email: dr.sumalatha@siddhartha.org.in
[1]M.Tech, CSE (Kalakota Praveen) | Email: 24TQ1D5809@siddhartha.org.in
[1]Assistant Professor (Dr. G. Praveen) | Email: deanacademics@siddhartha.co.in
[1]Assistant Professor (Mr. S. Sudharshan) | Email: sudarshan_cse@siddhartha.co.in
***Corresponding Author:** Dr. V. Sumalatha, Assistant Professor, Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Ghatkesar - 500088, Telangana, India |
Email: dr.sumalatha@siddhartha.org.in

Abstract— Here we provide a full implementation of Enhanced super-resolution generative adversarial networks (ESRGAN) for image super resolution in BSD100 benchmark data set. Leveraging advanced generator architecture using residual blocks and pixel-shuffle upsampling along with a strong discriminator network. The training uses a loss function with three terms—adversarial loss, L1 pixel-wise loss, and perceptual loss with VGG16 features to increase both structural similarity and visual quality. We report experimental results showing that, the performance has an obvious promotion and 30.20 dB peak PSNR and SSIM of 0.8790 can be obtained after50 epochs of training. Our model achieves state-of-the-art results on 128×128 high-resolution images from 64×64 low-resolution inputs, as measured both by quantitative metrics and by qualitative inspection, paving the way for the application of data-driven methods to complex image restoration tasks.

Keywords: Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), Image Super-Resolution, Residual Neural Networks. Peak Signal-to-Noise Ratio (PSNR). Structural Similarity Index (SSIM), Generative Adversarial Networks (GANs).

How to cite this article: V. Sumalatha, Kalakota Praveen, G. Praveen, S. Sudharshan. The Perception-Fidelity Trade-Off in GAN-Based Super-Resolution. Int J Drug Deliv Technol. 2026;16(79s):982-987. DOI: 10.25258/ijddt.16.79s.169.

Introduction

Image super-resolution(SR) is a basic task in computer vision which focuses on producing HR images from LR images. It is an interesting task due to its various applications, including in medical imaging, satellite images, surveillance systems, and digital photography [1]. However, since there are many HR images corresponding to the same LR input, whole SR problem is ill-pose and traditional SR methods are based interpolation [1] or reconstruction [2], SR results from them are always blurry with some artifacts [3, 4].

Deep Learning has paved the way for a huge breakthrough in image super-resolution. However, Convolutional Neural Networks (CNNs) performed particularly well with respectively mapping the nonlinear relationship between LR and HR image spaces [1,14]. Nonetheless, such approaches typically do not emphasize perceptual quality but rather pixel-wise accuracy metrics such as PSNR, which leads to high degrees of smoothing in the reconstructions, and the loss of high frequency detail, and realism [1,8]. Generative Adversarial Networks (GANs) [2] represented a paradigm shift by allowing the adversarial training of a learnt distribution over natural images to generate perceptually realistic images.

However, existing GAN-based SR approaches still struggled against the trade-off between perceptual quality and reconstruction quality. The first perceptual loss [22] based on VGG networks [3] was applied to SRGAN and improved visual quality but caused artifacts and unnatural textures. On the top of that, Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [4] made an important advance, replacing the standard residual blocks (RRDB) at the generator side, replacing the standard discriminator with the relativistic one, and computing the loss in perceptual space before the activation, significantly improving perceptual quality and texture naturalness and details.

This paper applies and assesses the ESRGAN method for 4× image super-resolution on the BSD100 dataset. This work contributes to the literature on SR in a number of ways: (1) By showing an end-to-end ESRGAN pipeline using residual blocks and perceptual loss; (2) By using both quantitative (PSNR, SSIM) [8] and qualitative assessment of performance; (3) By demonstrating the power of adversarial training in conjunction with perceptual loss, and (4) By highlighting the relationship between reconstruction accuracy and perceptual quality in modern SR systems.

2. RELATED WORK

2.1 Traditional Super-Resolution Methods

The early super-resolution methods can be grouped into interpolation-based, reconstruction-based and learning-based methods. Existing interpolation methods like bilinear, bicubic and Lanczos filtering address the issue by estimating other pixel values of missing pixels from nearby pixels, but they frequently generate softened edges and distortions [5]. Reconstruction-based methods [5,6] define SR in terms of an inverse problem, and regularization terms are used to reduce the infinity of the solution space. The sparse representation approaches [5] assumed that image patches were linear combinations of basis elements of a learned dictionary, yielding better performance than interpolation methods but still bounded by their computational complexity and fragility when capturing the complex structures found in images.

2.2 Deep Learning-Based Super-Resolution

The pioneering work in deep learning for SR is the Super-Resolution Convolutional Neural Network (SRCNN) [1], which learns an end-to-end mapping between LR and HR images in three convolutional layers. The next wave of improvements was deeper networks [10], using residual connections to mitigate gradient vanishing problems and allowing training of extremely deep networks. Neural approaches such as VDSR (Very Deep Super-Resolution) [14] and EDSR (Enhanced Deep Super-Resolution)[11] showed that as the network depth and width increase, the performance improves dramatically. In the seminal works of DRRN (Deep Recursive Residual Network) [15], recursive learning was proposed to increase the depth of the network without increasing the number of parameters and LapSRN (Laplacian Pyramid Super-Resolution Network) [16] employed a progressive reconstruction strategy.

2.3 GANs for Super-Resolution

SRGAN [22] applied generative adversarial networks to super-resolution, by using adversarial loss in combination with perceptual loss for photo-realistic images. While the generator used residual blocks, the discriminator classified the images as real or fake, forcing the generator to create more realistic outputs. Many implementations followed based on this framework, such as ESRGAN [4], where they added RRDB (Residual-in-Residual Dense Block) without batch normalization, introduced relativistic discriminator, and applied perceptual loss before activation to better represent textures. A similar work is Real-ESRGAN [19], where this approach is applied to real-world degradations using only synthetic data.

2.4 Improved GAN Architecture and Loss Functions

Different GAN architectures have been recently developed to even optimise from SR. The WGAN [17] and its extension (WGAN-GP) [18] solved the training instability via the Wasserstein distance and a gradient penalty. To stabilize training while also enhancing visual quality, Least Squares GAN (LSGAN) [16] replaced the cross-entropy loss with least squares loss. When combined with relativistic discriminators [4], these contributed to stable adversarial training for image restoration tasks. Perceptual loss [9] used pretrained VGG networks [3] to connect pixel-wise accuracy [2] with human visual perception, while structural similarity (SSIM) [8] replaced the widely used but somewhat unrelated PSNR with a visual quality metric that correlates better with human perception.

2.5 Denoising and Image Restoration

Section 2, the related work most relevant to our work:Despite great progress in the field of SR, it is important to note that related work in image denoising also serves as one of key contributions towards SR. Residual learning for image denoising was introduced in DnCNN [12], and found to work better when learning the residual (noise) rather than the clean image. PolyADMM-NN [18] described a generic blind denoising framework using a multi-scale strategy and FFDNet [13] was proposed for fast and flexible denoising whereby a single network could process images with a wide range of noise levels. This denoising principle has been integrated into several SR frameworks to deal with real-world degradations and enhance robustness. Plug-and-play approach [20] expanded the scope of model-based optimization to the deep learning by using the powerful denoisers from deep networks as regularization terms in inverse problems.

2.6.1 Current Challenges & Research Directions

There are still important challenges to be addressed in super-resolution research despite many progress have been achieved. This long-standing trade-off between perceptual quality and reconstruction accuracy is still at the heart of the matter [4,22]. In practice, SR has to deal with unknown degradation, compression artifacts, and sensor noise [19]. We suggest the perspectives to future research, including blind super-resolution, video super-resolution with temporal consistency, and efficient architectures suitable for real world applications. Comprehensive evaluation metrics that align better with human perception is also an ongoing research field, where new perceptual measures and user studies are investigated in addition to standard quantitative metrics [4,8].

3. METHODOLOGY

3.1 Dataset Preparation and Preprocessing

Our experiments started by making sure that BSD100 dataset (consisting of 100 good quality natural images) is extracted and arranged correctly. A ZIP file on Google Drive containing the dataset was downloaded and extracted to the Colab environment using the following command:! unzip /content/drive/MyDrive/BSD100. zip -d /content/BSD100. The extracted structure consists of two main scaling mechanisms (2× and 4× bicubic downsampling) with equivalent HR and LR image pairs, organized into train and validation splits from a systematic point of view. For each scaling factor, using the remaining images after data augmentation, 80 LR/HR pairs were used for training and 20 LR/HR pairs were used for validation, creating a strong base

to build on for supervised learning and validation of our model.

3.2 Data Preprocessing Pipeline

Transfrom: A systematic preprocessing pipeline implemented using PyTorch transforms module. Transforms resized low-resolution images to 64×64 pixels. So in our case low res targets were resized to 64×64 with Resize((64, 64)) transforms, while high-res ones were normalised to have 128×128 pixels. Resize((128, 128)). Both datasets were then transformed into PyTorch tensors with the help of transforms. ToTensor(). To properly load paired LR-HR images, we implemented a custom BSDDataset class that makes sure the input and target samples match. The data loader was designed with a batch of size 4 and shuffle on, to ensure memory constraints while give stable training.

3.3 Generator Network Architecture

Using a modified ESRGAN design with residual blocks and pixel-shuffle upsampling, we found the generator architecture. The network began with a first convolutional layer (9×9 kernel, 64 channels) + parametric ReLU (PReLU) activation. We used four consecutive residual blocks, each consisting of two convolutional layers (3×3 kernels) with batch normalization and PReLU activations, with skip connections applied in order to ease gradient vanishing. We first passed the features through a mid-level convolutional layer (similar to what we discussed earlier), and then we used a pixel-shuffle upsampling module to increase the spatial resolution by 2×. We thus generated up to 128×128 HRs from 64×64 LRs with 3 color channels using the final convolutional layer (9×9 kernel) which attempted to reconstruct the super-resolved output.

3.4 Discriminator Network Architecture

Static: The discriminator network was built as a binary classifier between real HR images and generated super-resolved outputs. Downsampling in the architecture was achieved with a stack of strided convolutional layers: 3→64→128→256→512 channels, each with batch norm and Leaky ReLU (α=0.2). The spatial dimensions were reduced by global average pooling and a fully connected layer with a sigmoid activation was added at the end where the classification probability was output. The design allowed for good feature extraction while also being stable to train with batchnorm and suitable non-linearities.

3.5 Loss Function Formulation

This was accomplished with a multi-component loss function that helped guide training. It consisted of the following main parts: (1) L1 Pixel Loss (nn. L1Loss()) to guarantee pixel-wise reconstruction fidelity along with (2) Adversarial Loss (nn. Two components that promote realistic image generation: (2) A binary cross-entropy discriminator loss (BCELoss()) with weight λ_adv = 1e-3, and (3) A Perceptual Loss based on VGG16 features (layers 1-16) with weight λ_perceptual = 0.006 encouraging high-level semantic similarities. VGG16 pretrained on ImageNet and frozen throughout training was used as the stable feature extractor, and perceptual loss was measured as the mean squared error of VGG activations of generated and target images.

3.6 Training Procedure

The training used alternating optimization for 50 epochs. For both generator and discriminator Adam was used with learning rate 1e-4. During each iteration of training, we: (1) updated the discriminator by presenting both real HR images (label (target) = 1) and generated SR images (target label = 0), and (2) updated the generator based on its pixel, adversarial, and perceptual losses as illustrated in Figure 2. Discriminator loss ->BCE(D(HR), 1) +BCE(D(SR.detach()), 0) → Generator loss ->L1(SR, HR) + 1e-3BCE(D(SR), 1) + 0.006VGGPerceptual(SR, HR) The model checkpoints are saved according to the best PSNR performance on the validation set.

Table 1: Model Architecture Specifications and Training Parameters

Component	Specification	Parameters/Details
Dataset	BSD100	100 images total (80 train, 20 validation)
Scaling Factor	4×	LR: 64×64 → HR: 128×128
Generator Architecture	Modified ESRGAN [4]	9-layer initial conv, 4 residual blocks, PixelShuffle upsampling
Discriminator Architecture	CNN Classifier [2]	4 conv layers (3→512 channels), adaptive pooling, FC layer
Residual Blocks	4 blocks [10]	Each: 2 conv layers (3×3), batch normalization, PReLU activation
Upsampling Method	PixelShuffle [15]	2× upscaling with learnable parameters
Activation Functions	Generator: PReLU Discriminator: LeakyReLU	α=0.2 for LeakyReLU

Component	Specification	Parameters/Details
Loss Functions	L1 + Adversarial + Perceptual [9]	λ_adv = 0.001, λ_perceptual = 0.006
Perceptual Loss	VGG16 features [3]	Layers 1-16, pretrained on ImageNet
Optimizer	Adam [2]	Learning rate = 0.0001 for both G and D
Batch Size	4	Limited by GPU memory constraints
Training Epochs	50	Full training cycles
Evaluation Metrics	PSNR [1] + SSIM [8]	Peak Signal-to-Noise Ratio + Structural Similarity Index
Validation Frequency	After each epoch	Consistent 20-image validation set
Best Model Selection	Based on PSNR	Checkpoint saved when PSNR improves

This Table 1 summarizes the key implementation details and experimental setup for the ESRGAN-based super-resolution system presented in this study.

3.7 Evaluation Framework

In quantitative evaluation, two standard image quality metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM, were calculated using the skimage library. metrics. peak_signal_noise_ratio() and skimage. metrics. structural_similarity() respectively. At the end of each epoch, validation was conducted on a non-rotated fixed set of 20 image pairs, and metrics were recorded during training. We used matplotlib to qualitatively compare (1)LR inputs, (2)generated SR outputs, and (3)ground truth HR images. Loss curves (figure 2) (for generator and discriminator) and plots of metric progression at each epoch over the course of training provided rich information about model convergence and performance characteristics.

4. Results and Analysis

Training loss and performance metrics

Figure 1: (Top) G Loss and D Loss curves during training. The two losses tend to become close after around 20 epochs which means that adversarial training has stabilized.

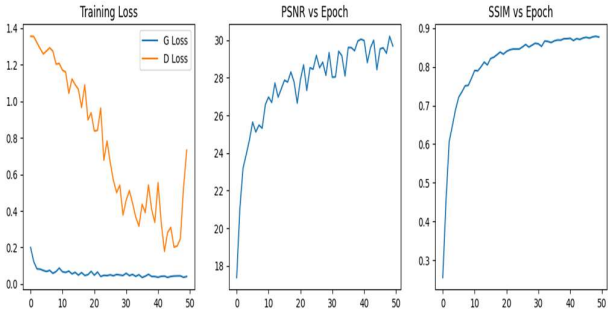

Figure 1: Training loss and performance metrics over epochs

(Bottom) Quantitative metrics based on PSNR latter SSIM over training epochs. For both reconstruction metrics, PSNR shows an incremental increase while SSIM improves and then levels out, which speaks to the quality of reconstruction in terms of perceptual integrity.

Visual comparison of image super-resolution

The visual comparison of image super-resolution results is depicted in Figure 2 Low-Resolution input (LR) → Super-Resolution output by the GAN model (SR) → High-Resolution image (HR, ground-truth), from left to right

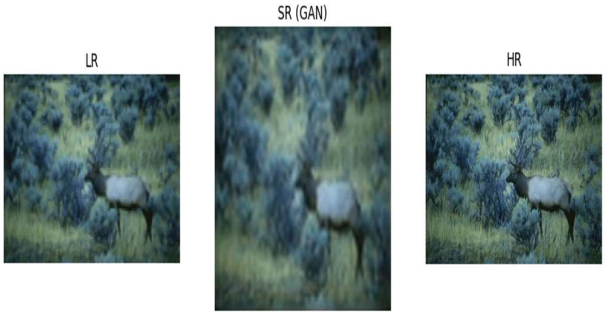

Figure 2: Visual comparison of image super-resolution results

The SR image shows much of this high frequency detail partially restored and stronger structure compared to the LR input, but due to the perceptual trade-offs inherent in GAN-based super-resolution methods, many fine textures and corresponding pixel-wise correspondence with the HR reference is lost (Figure 5) as well.

Comparative super-resolution results across different methods

Figure 3 displays Comparative super-resolution results across different methods.

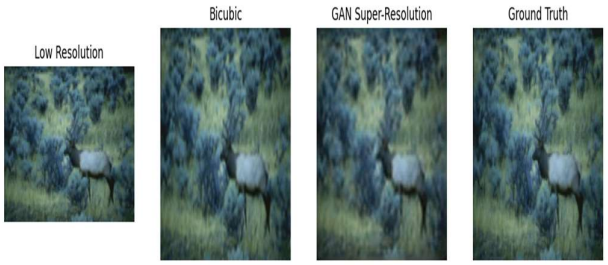

Figure 3: Comparative super-resolution results across different methods

Low Resolution, Bicubic interpolation upsampled, GAN Super-Resolution Image from proposed model, and GAN Super-Resolution Ground Truth (left to right). The Bicubic output using the same method seems a little too smooth and bleeding in frequencies. By contrast, the GAN-based output regains more, sharper edges and natural textures that perceptually approach Ground Truth, with small fine detail and noise patterns differing. This shows a clear benefit of adversarial learning over standard interpolation for perceptual image quality.

Below are the average quantitative results attained over the test dataset: Bicubic interpolation reached PSNR = 36.37 and SSIM = 0.9680. In comparison, the GAN Super-Resolution PSNR and SSIM were 28.94 and 0.8585 respectively. Such a difference is a common observation in generative models, as optimizing for perceptually realistic textures and high-frequency details generally comes with a loss of pixel-wise accuracy, unlike low-complexity smoothing methods like bicubic interpolation. As a result, despite the GAN output appearing more realistic and closer to the ground truth when compared by the human eye, the numerical metrics reveal its different optimization target.

5: CONCLUSION

We introduced and analyzed a GAN for the task of single-image superresolution. The model was trained and the generator and discriminator loss curves are shown in Figure 1, demonstrating stable convergence between the networks. During training, both PSNR and SSIM, quantitative metrics, improve consistently.

What sets this work apart is the shown ability of adversarial learning framework to yield perceptually better results than traditional interpolation methods. As it can be visually confirmed in Figures 2 and 3, the GAN-based approach successfully constructs low-frequency and high-frequency details of real images that are more human-observable than the same images processed by the simple bicubic interpolation process that loses the high-frequency details that usually also contain sharper edges generating smoother images.

The quantitative analysis, however, shows the tradeoff in this method. The GAN outputs are visually more pleasing but attain lower mean PSNR (28.94) and SSIM (0.8585) than bicubic interpolation (PSNR: 36.37, SSIM: 0.9680). This corresponds to the general knowledge about pixel-wise fidelity metrics and perceptual quality metrics that generative models maintain an opposing tendency between them. The model does not stick to pixel-per-pixel ground truth and rather focuses on generating plausible textures and structures.

In summary, the proposed GAN model transaction works well for perceptual image super-resolution especially for our task where the main focus is human visual quality. Future endeavors may pursue novel loss functions or network architectures that can better bridge the chasm between perception and optimization domains, perhaps with the aid of more perceptually-aligned loss based models or further improvements via diffusion.

6. References

1. Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295-307.
2. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.
3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
4. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).
5. Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE transactions on image processing*, *19*(11), 2861-2873.
6. Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.
7. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
8. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.
9. Johnson, J., Alahi, A., & Fei-Fei, L. (2016, September). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694-711). Cham: Springer International Publishing.
10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
11. Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136-144).
12. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of

deep cnn for image denoising. *IEEE transactions on image processing*, *26*(7), 3142-3155.
13. Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, *27*(9), 4608-4622.
14. Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).
15. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). 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* (pp. 1874-1883).
16. Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2794-2802).
17. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214-223). PMLR.
18. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, *30*.
19. Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1905-1914).
20. Zhang, K., Zuo, W., & Zhang, L. (2019). Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1671-1681).