

Image Super Resolution for Satellite Imagery

Bella Wahmilyana Asril¹, Hogan Eighfansyah Susilo², Fadjrianah^{3*}

¹ Institut Teknologi Del, Department of Information Technology, Faculty of Vocational Studies, Sitoluama, Toba, 22381, INDONESIA

² National Research and Innovation Agency (BRIN), Research Centre for Satellite Technology, Aviation and Space Research Organization, Rancabungur, Bogor, 16310, INDONESIA

³ Institut Teknologi Sumatera, Department of Telecommunication Engineering, Faculty of Industrial Technology Jalan Terusan Ryacudu, Way Hui, South Lampung, 35365, INDONESIA

*Corresponding Author: fadjrianah@tt.itera.ac.id

Article Info

Received: 20 August 2025
Revised: 25 September 2025
Accepted: 30 September 2025
Available online: 30 September 2025

Keywords

ESRGAN, satellite imagery, super resolution

Abstract

Super resolution (SR) techniques play a critical role in enhancing the quality of satellite imagery by reconstructing high resolution (HR) images from low resolution (LR) inputs. This study investigates the application of the Enhanced Super Resolution Generative Adversarial Network (ESRGAN) to improve the spatial resolution of satellite images. The ESRGAN framework integrates a Residual-in-Residual Dense Block (RRDB) generator, a Relativistic average GAN (RaGAN) discriminator, and an improved perceptual loss function to achieve sharper textures and more realistic structures. The 4× Satellite Image Super Resolution dataset, consisting of paired LR (2 m/pixel) and HR (0.5 m/pixel) images, was employed. Quantitative evaluation was conducted using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), with Lanczos5 interpolation adopted as a baseline for comparison. Experimental results on ten test images show that ESRGAN consistently outperforms Lanczos5, achieving an average PSNR of 25.976 dB and an average SSIM of 0.750, compared to 25.179 dB and 0.698, respectively, for Lanczos5. Qualitative assessments further confirm that ESRGAN effectively restores high frequency details, enhances fine textures, and preserves object boundaries more accurately than the interpolation method.

1. Introduction

High resolution (HR) satellite imagery is essential for mapping, land-use monitoring, and urban analytics because it enables reliable interpretation of fine spatial details such as road networks, buildings, and parcel boundaries. However, sensor limitations, high acquisition costs, and the substantial bandwidth required for

transmission mean that many workflows rely on low resolution (LR) imagery that is inadequate for operational analysis. Super-resolution (SR) addresses this gap by reconstructing spatial detail to improve the readability of edges, repeating patterns, and semantically important structures.

Classical interpolation methods such as bicubic often yield blurred edges and unnatural textures, which diminishes the imagery's operational value. Recent advances in deep learning for single-image super-resolution (SISR) have changed this landscape: SRGAN introduced an adversarial formulation with perceptual losses that prioritize visual realism rather than merely maximizing PSNR [1]. EDSR further refined residual architectures by removing superfluous modules, delivering consistent quantitative gains across standard SISR benchmarks [2]. Within remote sensing, edge-aware GAN-based methods have been shown to sharpen linear structures and man-made objects in satellite scenes [3].

Following EDSR, the Residual Channel Attention Network (RCAN) introduced a residual-in-residual architecture with channel attention that adaptively emphasizes informative feature channels, delivering consistent gains in single-image SR and establishing a strong post-EDSR baseline [4]. More recently, Transformer-based architectures such as SwinIR leverage shifted-window self-attention to capture long-range dependencies, improving the recovery of both local details and global context in image restoration including single-image super-resolution for satellite data [5]. Building on this foundation, this study evaluates ESRGAN for satellite imagery to assess how well the model sharpens edges and textures relevant to operational mapping and image analytics, and to discuss implications for perceptual quality and practical utility in remote-sensing workflows.

2. Literature Review

Super-resolution (SR) has been widely studied in image and video processing, with the goal of reconstructing high-resolution (HR) images from low-resolution (LR) inputs. Approaches are generally categorized into multi-image super-resolution (MISR) and single-image super-resolution (SISR). MISR exploits multiple LR observations of the same scene for fusion and reconstruction, but this requires well-aligned images that are often unavailable in satellite imaging. In contrast, SISR aims to recover HR content from a single LR input and has therefore become the dominant focus in remote sensing applications [6].

The most basic SISR approaches rely on interpolation techniques, such as nearest-neighbor, bilinear, bicubic, spline, and Lanczos methods [7]. These methods are computationally efficient but operate solely on pixel intensities without considering image semantics. Consequently, they often produce blurred textures, ringing artifacts, and poorly defined edges.

A breakthrough came with deep learning-based methods. Dong et al. [8] introduced SRCNN (Super-Resolution Convolutional Neural Network), the first end-to-end trainable CNN for image SR, which significantly outperformed classical interpolation. Building on this, Ledig et al. [1] developed SRGAN, introducing adversarial and perceptual losses to emphasize natural textures over pixel similarity. Lim et al. [2] later proposed EDSR (Enhanced Deep Residual Networks), simplifying residual structures and achieving state-of-the-art performance on standard benchmarks.

Further advancements were demonstrated by RCA-GAN by Cai et al. [9], which introduced a Residual Channel Attention Block within a GAN structure, enabling the model to adaptively emphasize critical high-frequency features and generate more natural textures and finer details. Recent studies, including Asril et al. [10], have applied ESRGAN experimentally, confirming its potential in practical settings.

In the remote sensing domain, GAN-based approaches have demonstrated particular promise. Lavreniuk et al. [11] applied GANs for satellite data super-resolution, showing improvements in feature reconstruction. Similarly, Jiang et al. [3] introduced edge-enhanced GANs that effectively preserved structural elements such as roads and parcel boundaries. More broadly, reviews such as Xu et al. [12] highlight the growing role of deep learning-based SR in Earth observation, with demonstrated benefits for land-use monitoring, urban mapping, and environmental analysis.

Despite these advances, challenges remain regarding model generalization across different satellite sensors, spectral ranges, and acquisition conditions. Many SR models are trained on natural image datasets

and applied in fields such as medicine [13] and agriculture [14], which differ significantly from satellite imagery in terms of structure and noise characteristics. This motivates the present study, which evaluates ESRGAN on a dedicated satellite dataset and benchmarks it against classical interpolation, with the aim of improving both perceptual quality and analytical value in operational applications.

3. Materials and Methods

3.1 Materials

The dataset used in this study is the 4x Satellite Image Super Resolution dataset [15], which consists of paired high resolution (HR) and low resolution (LR) satellite images specifically designed for 4× super-resolution tasks. The HR images have a spatial resolution of 0.5 meters per pixel, serving as the ground truth, while the LR images have a resolution of 2 meters per pixel, corresponding to the input data for upscaling. All image pairs are geographically aligned, ensuring pixel-to-pixel correspondence between HR and LR images. This dataset provides a suitable benchmark for evaluating the proposed super resolution method due to its high-quality alignment and clearly defined resolution scaling.

3.2 Methods

This study employs the Enhanced Super Resolution Generative Adversarial Network (ESRGAN) as the primary method for generating high-resolution satellite images from low resolution inputs. ESRGAN is a deep learning-based approach that enhances image details, enabling the reconstruction of sharper textures and more realistic structures [16]. Fig. 1. illustrates the block diagram of the process, from the low resolution input to the high resolution output.

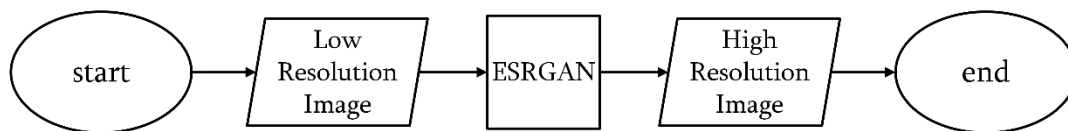


Fig. 1 Workflow of the super resolution experiment using ESRGAN

The Enhanced Super Resolution Generative Adversarial Network (ESRGAN) is an advancement over SRGAN aimed at narrowing the gap between generated super-resolved images and the ground truth in terms of sharpness and texture detail. The improvements are made in three main areas. First, the generator architecture replaces the SRGAN's basic block with a Residual-in-Residual Dense Block (RRDB), which offers higher capacity and training stability, removes Batch Normalization (BN) to reduce artifacts and improve generalization, and applies residual scaling with smaller weight initialization to enable stable training of very deep networks. Second, the discriminator is enhanced by adopting the Relativistic average GAN (RaGAN) via the Relativistic average Discriminator (RaD), which judges image realism in a relative sense rather than absolute, helping the generator to produce more realistic textures. Third, the perceptual loss is improved by using VGG features before activation instead of after, which has been shown to yield sharper edges and more visually pleasing results compared to the original SRGAN approach [17].

In the generator, RRDB integrates multi-level residual learning with dense connections, leveraging the benefits of deeper residual structures to enhance feature representation. Removing BN not only reduces computational complexity but also prevents artifacts that tend to occur in deep networks under the GAN framework. On the discriminator side, RaGAN allows the model to benefit from gradients derived from both real and generated data simultaneously, improving the learning of texture details and edge sharpness [18]. Additionally, the pre-activation-based perceptual loss addresses the shortcomings of conventional post-activation loss, which often results in highly sparse activations and inconsistent brightness compared to the ground truth. Through this combination of architectural enhancements, loss function refinements, and

training strategies, ESRGAN consistently outperforms state-of-the-art methods in both image sharpness and fine visual detail.

As a classical baseline, Lanczos5 interpolation is used for comparison with the proposed ESRGAN method. Lanczos interpolation is a resampling method used to increase or decrease image resolution while preserving detail and clarity. It employs the Lanczos kernel, a windowed sinc function designed to produce smooth and consistent interpolations. Applied systematically across both image rows and columns, Lanczos interpolation enables accurate upscaling and downscaling with high smoothness, making it a valuable technique for improving the quality of resampled images [19 – 20].

3.3 Evaluation Metrics

To objectively assess the performance of the super-resolution method, a quantitative evaluation metric was employed to compare the reconstructed images with the corresponding high-resolution ground truth. In this study, the Peak Signal-to-Noise Ratio (PSNR) was used, as it is widely adopted in image quality assessment for super-resolution tasks. PSNR measures the fidelity of the reconstructed image by quantifying the ratio between the maximum possible signal power and the noise introduced during reconstruction. A higher PSNR value indicates better reconstruction quality [21 – 22]. The PSNR, measured in decibels (dB), can be calculated using the following formula:

$$PSNR = 10 \log_{10} \left(\frac{A^2}{\frac{1}{n} \sum_{i=1}^n (Y(i) - \hat{Y}(i))^2} \right) \quad (1)$$

For an 8-bit image, A is typically 255 and represents the maximum possible pixel value in the image, determined by the number of bits used in the pixel representation. n denotes the total number of pixels in the image, $Y(i)$ is the value of the i -th pixel in the ground truth image, and $\hat{Y}(i)$ is the value of the i -th pixel in the reconstructed image.

On the other hand, in contrast to PSNR which operates on absolute pixel-wise differences, the Structural Similarity Index Measure (SSIM) is designed to model the human visual system's sensitivity to structural information in images. SSIM is a widely used perceptual metric for assessing image quality by measuring the degree of similarity between two images. Its value ranges from 0 to 1, with higher values indicating lower distortion. An SSIM score of 1 indicates that the two images are perfectly identical [23].

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

Mathematically, the Structural Similarity Index Measure (SSIM) between two image patches x and y is defined as shown in Eq. (2). In this formulation, μ_x and μ_y denote the mean intensity values of images x and y , representing luminance. The terms σ_x^2 and σ_y^2 correspond to the variances of x and y , which capture contrast information, while σ_{xy} represents the covariance between the two images, reflecting their structural similarity. To guarantee numerical stability, particularly when the denominators are close to zero, two small constants, c_1 and c_2 are incorporated into the equation.

4. Results and Discussion

This section presents and discusses the experimental results obtained from applying the super-resolution method to the satellite image dataset. The quantitative evaluation uses the Peak Signal-to-Noise Ratio (PSNR) to measure the fidelity of the reconstructed images against the high-resolution ground truth. The results are analyzed to identify performance trends across different test images. In this research, we used ten different images, and the PSNR results are presented in Table 1. For comparison, Lanczos5 interpolation is also evaluated as a classical baseline to provide a quantitative reference against the ESRGAN results.

Table 1 PSNR Evaluation for 10 Images in Super-Resolution

Image Number	PSNR (dB)	
	ESRGAN	LANCZOS5
1	25.90	24.79
2	26.30	25.59
3	26.64	26.38
4	25.63	24.74
5	27.37	26.18
6	25.16	24.03
7	25.62	25.19
8	25.46	24.93
9	26.49	25.70
10	25.19	24.26

Based on Table 1, the ESRGAN method consistently outperforms the Lanczos5 method across all test images. The average PSNR value for ESRGAN is 25.976 dB, while Lanczos5 achieves an average of 25.179 dB. This indicates that the reconstructed images produced by ESRGAN are closer to the high-resolution ground truth. The largest improvement is observed in image number 5, with a difference of 1.19 dB, while the smallest improvement is found in image number 3, with a difference of only 0.26 dB. These quantitative results demonstrate that ESRGAN excels in preserving detail and sharpness in satellite images compared to classical resampling methods such as Lanczos5. This consistent improvement shows that deep learning-based approaches can deliver better results than conventional interpolation.

In addition to PSNR, the Structural Similarity Index Measure (SSIM) was employed to further assess the perceptual quality of the reconstructed images. Table 2 summarizes the SSIM values obtained for ESRGAN and Lanczos5 across ten test images. Similar to the PSNR results, ESRGAN consistently achieves higher SSIM scores than Lanczos5. The average SSIM for ESRGAN is 0.750, compared to 0.698 for Lanczos5, indicating that ESRGAN is more effective in preserving structural information and visual similarity to the ground truth. These results reinforce the conclusion that ESRGAN not only reduces distortion but also maintains perceptual fidelity better than the classical interpolation method.

Table 2 SSIM Evaluation for 10 Images in Super-Resolution

Image Number	SSIM	
	ESRGAN	LANCZOS5
1	0.762	0.699
2	0.760	0.714
3	0.718	0.680
4	0.740	0.681
5	0.767	0.717
6	0.772	0.696
7	0.726	0.690
8	0.730	0.685
9	0.759	0.714
10	0.765	0.701

The visual results in this section present a qualitative comparison between the low-resolution inputs and the corresponding super resolution outputs produced by ESRGAN. From the ten test images used in this study, three representative examples, namely images 1, 3, and 5, are selected for display. These examples highlight the visual improvements achieved by the super-resolution method, particularly in enhancing fine details and restoring sharper textures in satellite imagery.

The images in Fig. 2, Fig. 3, and Fig. 4 show a qualitative comparison between the low-resolution inputs, the super-resolved outputs produced by ESRGAN, and the results obtained using Lanczos5 interpolation. Compared to the low-resolution images, both methods improve clarity, but ESRGAN produces sharper edges, clearer building outlines, and more distinguishable fine details. In contrast, Lanczos5 provides smoother but less detailed reconstructions, highlighting the advantage of deep learning-based approaches over classical interpolation. These visual enhancements demonstrate the method's effectiveness of reconstructing high-frequency details that are lost in low resolution images and align with the quantitative PSNR results, further confirming the effectiveness of ESRGAN in enhancing satellite imagery.



Fig. 2 Image 1 (a) Low resolution; (b) Lanczos5; (c) ESRGAN



Fig. 3 Image 3 (a) Low resolution; (b) Lanczos5; (c) Super resolution



Fig. 4 Image 5 (a) Low resolution; (b) Lanczos5; (c) Super resolution

5. Conclusion

This study demonstrated the effectiveness of the Enhanced Super Resolution Generative Adversarial Network (ESRGAN) in improving the quality of low-resolution satellite imagery. Quantitative evaluation using Peak Signal-to-Noise Ratio (PSNR) showed that ESRGAN consistently outperformed the classical Lanczos5 interpolation method across all test images. The largest PSNR gain of 1.19 dB was observed for image 5, while the smallest gain of 0.26 dB occurred for image 3. Similarly, ESRGAN achieved higher Structural Similarity Index Measure (SSIM) scores, with the most significant improvement of 0.076 observed for image 6 and the smallest improvement of 0.036 for image 7. Qualitative analysis further confirmed these findings, where ESRGAN outputs exhibited sharper edges, clearer object boundaries, and more distinguishable fine details in urban and natural features. These results indicate that deep learning-based super resolution methods, such as ESRGAN, are highly effective in reconstructing high-frequency details and enhancing the visual quality of satellite imagery compared to conventional interpolation techniques.

For future work, several directions can be explored to further strengthen the evaluation of super-resolution methods. First, a closer inspection of specific features such as roads, rooftops, and vegetation could provide more detailed insights into the visual improvements achieved by ESRGAN. Second, in addition to PSNR and SSIM, the Learned Perceptual Image Patch Similarity (LPIPS) metric may be incorporated to better capture perceptual quality from a human-visual perspective. Finally, further comparisons with other deep learning-based super-resolution architectures, such as RCAN, SwinIR, or EDSR, could provide a more comprehensive benchmark.

Acknowledgement

The authors would like to thank the Department of Information Technology, Institut Teknologi Del, the Department of Telecommunication Engineering, Institut Teknologi Sumatera, and the National Research and Innovation Agency (BRIN) for their support.

Author Contribution

The authors confirm contribution to the paper as follows: study conception and design: Fadjrianah; data collection: Hogan Eighfansyah Susilo; analysis and interpretation of results: Bella Wahmilyana Asril; draft manuscript preparation: Bella Wahmilyana Asril, Fadjrianah, Hogan Eighfansyah Susilo. All authors reviewed the results and approved the final version of the manuscript.

References

- [1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4681–4690.
- [2] Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 136–144.
- [3] Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., & Jiang, J. (2019). Edge-enhanced GAN for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8), 5799–5812.
- [4] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. *Proceedings of the European Conference on Computer Vision (ECCV)*, 294–310
- [5] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image restoration using Swin Transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 1833–1844
- [6] Yang, J., & Huang, T. (2017). Image super-resolution: Historical overview and future challenges. In S. Chaudhuri & T. Huang (Eds.), *Super-resolution imaging* (pp. 1–34). CRC Press.
- [7] Wang, X. (2022). Interpolation and sharpening for image upsampling. *2022 2nd International Conference on Computer Graphics, Image and Virtualization (ICCGIV)*, 73–77. <https://doi.org/10.1109/ICCGIV57403.2022.00020>
- [8] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *European Conference on Computer Vision (ECCV)* (pp. 184–199). Springer.
- [9] Cai, J., Meng, Z., & Ho, C. M. (2020). Residual channel attention generative adversarial network for image super-resolution and noise reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 454–455).
- [10] Asril, B. W., Hamid, E. Y., Satyawan, A. S., & Prini, S. U. (2023). Image super resolution using ESRGAN: A preliminary experimental study. *2023 International Conference on Robotics, Automation, and Mechatronics Technology (ICRAMET)*. <https://doi.org/10.1109/ICRAMET60171.2023.10366740>
- [11] Lavreniuk, M., Kussul, N., Shelestov, A., Lavrenyuk, A., & Shumilo, L. (2022). Super-resolution approach for the satellite data based on the generative adversarial networks. *IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium*, 1095–1098. <https://doi.org/10.1109/IGARSS46834.2022.9884460>
- [12] Xu, X., Li, J., & Sun, X. (2021). Remote sensing image super-resolution: A review. *Remote Sensing*, 13(9), 1615.
- [13] Bing, X., Zhang, W., Zheng, L., & Zhang, Y. (2019). Medical image super resolution using improved generative adversarial networks. *IEEE Access*, 7, 145030–145038.
- [14] Han, H., Feng, Z., Du, W., Guo, S., Wang, P., & Xu, T. (2024). Remote sensing image classification based on multi-spectral cross-sensor super-resolution combined with texture features: a case study in the Liaohe planting area. *IEEE Access*, 12, 16830–16843.
- [15] Tudela, C. (n.d.). *4× Satellite image super-resolution* [Data set]. Kaggle. Retrieved August 2025, from <https://www.kaggle.com/datasets/cristobaltudela/4x-satellite-image-super-resolution>
- [16] Musunuri, Y. R., Kim, C., Kwon, O. S., & Kung, S. Y. (2024). Object detection using ESRGAN with a sequential transfer learning on remote sensing embedded systems. *IEEE Access*.

- [17] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... Loy, C. C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 0–0.
- [18] Liu, J., Cao, J., Zhao, L., You, J., & Li, H. (2024). Super-Resolution Reconstruction of Seismic Images Based on Deep Residual Channel Attention Mechanism. *IEEE Access*.
- [19] Bituin, R. C., & Antonio, R. (2024). Ensemble model of Lanczos and bicubic interpolation with neural network and resampling for image enhancement. *Proceedings of the 2024 7th International Conference on Software Engineering and Information Management (ICSEIM)*, 110–115.
- [20] Kumar, A., Islam, T., Ma, J., Kashiyama, T., Sekimoto, Y., & Mattmann, C. (2023). WindSR: Improving spatial resolution of satellite wind speed through super-resolution. *IEEE Access*, 11, 69486-69494.
- [21] Tseng, S. M., & Chen, Y. F. (2018). Average PSNR optimized cross layer user grouping and resource allocation for uplink MU-MIMO OFDMA video communications. *IEEE Access*, 6, 50559-50571.
- [22] Zhang, J., Shi, W., Zhou, L., Gong, R., Wang, L., & Zhou, H. (2019). A low-power and high-PSNR unified DCT/IDCT architecture based on EARC and enhanced scale factor approximation. *IEEE Access*, 7, 165684-165691.
- [23] Panigrahy, A. K., Maniyath, S. R., Sathiyarayanan, M., Dholvan, M., Ramaswamy, T., Hanumanthakari, S., ... & Swain, R. (2024). A faster and robust artificial neural network based image encryption technique with improved SSIM. *IEEE Access*, 12, 10818-10833.