International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 11s, 2025
https://theaspd.com/index.php

Super-Resolution Of Geospatial Images Using Enhanced Gans

Naga Durga Saile K[1],Gaje Gouri Chandana[2],Gunjipalli Chandra Sekhar[3],P V S R Krushik Sarma[4], Rachuri Indhu Sree[5]
[1,2,3,4,5]Department of CSE-AIML & IoT, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering &Technology Hyderabad, Telangana-500090

Abstract
High-resolution imagery is essential for critical applications such as environmental monitoring, urban planning, and disaster response, where accurate details support informed decision-making. However, limitations in available imaging systems for public use and resource constraints often result in low-resolution satellite images lacking the necessary detail. Super-resolution (SR) methods have emerged to address these limitations, with deep learning approaches like Generative Adversarial Networks (GANs) and Transformer-based models offering promising results. This study investigates a GAN-focused SR approach, linking Real-ESRGAN with Transformer-based methods such as SwinIR to obtain higher-resolution usable images. Real-ESRGAN's multi-scale discriminators and Residual-in-Residual Dense Blocks (RRDB) effectively capture complex textures and mitigate noise, making it suitable for high-detail satellite imagery. Our results demonstrate significant improvements in image clarity and overall perceptual quality, supporting applications requiring precise, high-resolution images.
Keywords: *Super-Resolution; Image Enhancement; Geospatial Imagery; GAN; ESRGAN; SwinIR; Transformer Models; Image Quality.*

INTRODUCTION

The very rationale of this research paper, "Super-Resolution of Geospatial Images Using Enhanced GANs," is supported with the need for high-resolution satellite imagery in the applications of environmental monitoring, urban planning, and disaster response whose operations greatly rely on images that are more precise and detailed. From those applications, particularly image quality turns into a matter of fundamental importance to help decide analysis of land cover, following environmental changes, and understanding infrastructure of urban areas. Satellite images lack the necessary resolution due to limitations in capabilities, adverse atmospheric conditions, and scarce resources, thus making it difficult to acquire good visuals. These traditional SR techniques, like bicubic and bilinear interpolation, attempted to achieve an enlarged version with their approximations of the missing pixel values. Though such methods were easy to implement, however they frequently produced outputs that blurred images or failed to preserve rich details of intricate textures. On this count, these initial Super Resolution (SR) techniques could not be applied for applications requiring high resolution images. The development of Convolutional Neural Networks (CNNs) marked a big breakthrough in the field of SR, allowing models like SRCNN to learn mappings from low-to-high resolution images with extensive training on paired datasets. However, despite the leap forward in the use of CNN-based SR, this method suffers from low-frequency detail capture, hence making its applications restricted in the fields of remote sensing and medical imaging, where fine textures and details are very crucial. GANs brought a paradigm shift in SR technology, with GANs employing an architecture of a generator-discriminator network where the generator produces synthetic images at high resolution, while the discriminator aims to distinguish between real and synthetic images. Such an adversarial setup puts pressure on the generator to produce sharper images with more natural textures. Recently, works such as SwinIR have begun to model local and global dependencies within an image using self-attention mechanisms that promise to bring a new revolution in the SR task. Transformer models demonstrated their potential in processing large structural variations characteristic for application such as satellite imagery. This work will leverage the real-ESRGAN along with SWIN IR to form an improved GAN model to fulfil particular high-resolution requirements for satellite imagery, especially at variable terrain and environmental conditions. The rationale of choosing Real-ESRGAN alongside transformer-based approach alternatives, such as SwinIR, is due to the former's better ability to retain high-frequency details; hence, in order to capture complex textures and reduce real-world image degradations like noise and blurring.

Related Work

Traditional SR Methods:

Interpolation and the low-level techniques of super-resolution tried to increase the resolution of an image by filling in missing data based on data from surrounding pixels. Bicubic, bilinear, or other interpolative procedures are rather simple but often produce very blurry and grainy images with very poor detail. Such methods rely on simple mathematical assumptions and cannot reproduce or generate complex textures, useful mostly in high-fidelity applications like geospatial analysis, where fine details and intricate patterns are important[16].

The Emergence of CNN-Based SR Models:

The development of Convolutional Neural Networks (CNNs) was an important milestone in the history of SR technology because CNNs enabled models to learn feature mappings between low and high-resolution images, making models better approximations of complex details. The earlier CNN-based models, known as SRCNN[6], paved the path for applying CNNs in SR and depict better quality images. The more advanced architectures of SRCNN led to two more versions: VDSR (Very Deep Super-Resolution)[3] and EDSR (Enhanced Deep Super-Resolution)[3], using deeper layers and residual connections in order to reserve more detailed information and achieve a better quality of images generally.

Though VDSR and EDSR showed impressive progress, they still failed to reproduce realistic textures and fine-grained details, particularly in the satellite imagery, where fine details are of utmost importance for applications such as land cover analysis and infrastructure mapping in cities. CNN-based SR models were decent but frequently failed to reproduce fine grain textures created in complex geospatial images.

Progress by GANs:

The second big advancement in SR was the development of GANs, which have changed the SR paradigm by applying adversarial training to develop sharper and more realistic images. In models such as SRGAN[2], the GAN framework includes a generator that generates high-resolution images and a discriminator that classifies the input as real or generated, thereby forcing the generator to develop sharper, more natural textures. It was able to produce images with very high resolutions by better capturing the high-frequency components than CNN-only models. To improve SRGAN, ESRGAN introduced Residual-in-Residual Dense Blocks to preserve subtle texture better and remove artifacts. Additionally, ESRGAN employed a perceptual loss, calculated using a pre-trained VGG network, which gives more importance to perceptual quality than the pixel-wise accuracy to account for human visual perception[4]. This improvement further made ESRGAN highly effective for applications that need high fidelity in terms of textures, such as satellite imaging, where minute details really matter.

Transformer-Based SR Models:

Some promising alternatives recently have been proposed based on the Transformer architecture. For instance, SwinIR relies on self-attention mechanisms, capturing local and global dependencies across large regions of an image [7]. Transformers may be more capable to attend to spatial relationships across a broader context, and such capabilities would be beneficial when the dependency between pixels is very large in images. SwinIR resorts to the blocks of Swin Transformer to handle different structures effectively, making it as competitive as any other option within the field of SR.

Focus of the Study on Real-ESRGAN:

Comparing against the later comparison, the work demonstrates that real-ESRGAN, which is the more recent GAN model with refinement, is superior to the new challenges that have been set by satellite image SR. Real-ESRGAN can provide fine textures and high-frequency details under the help of multi-scale discriminators and the RRDB architecture so that it is actually viable for real-world degradations, such as noise and blurring in images[13]. This is very important for satellite imagery, where resolution output depends on the extents of downstream applications in mapping, environmental analysis, and all urban planning.

Proposed Method

This work explores the utility of Real-ESRGAN, or Enhanced Super-Resolution Generative Adversarial Networks and a Transformer-based model, which is SwinIR toward enabling high-fidelity image enhancement geospatial image based on specific requirements. It is opted due to its strong performance

in achieving high quality images and addressing complex features from satellite imagery. This research strives to output high resolution toward supporting geospatial downstream applications in mapping, land classification, and environmental monitoring.

Dataset Preparation

The dataset applied for this research is taken from Kaggle [16] which has around 200,000 images of which we used 5,000 images for training, and all of them were carefully prepared to provide a wide array of environmental conditions; hence, it would enable the model to generalize through different geospatial conditions. Such that in arrangement:

High-Resolution Images (Ground Truth): This uses 1000 images as ground truth. These were high-quality images since they are set as the target output meant to evaluate.

Artificially degraded training images 4000: Degradations simulating practical challenges commonly experienced in satellite imagery are added to the images provided for training. The following degradations are applied to the low-resolution images to improve the model's flexibility in dealing with common image degradation seen in satellite data.

Noise Addition and Blurring to simulate atmospheric interference and other environmental impacts, simulating the common challenges encountered in satellite imagery. All images within the dataset have metadata available relating to the type of terrain, weather conditions, and information about acquisition. This metadata would be useful to determine performance through various environmental conditions and ensure the Real-ESRGAN product outputs are correct under multiple geospatial contexts.

Model Architecture

For the model we have used both Real ESRGANs and SWIN models. Of the two models SWIN comes pre trained and training is done only on Real ESRGANs. The architecture of Real-ESRGAN is specifically developed to handle intricate textures and details critical for geospatial applications; it is distinguished from other SR models. It mainly consists of four parts [13]:

Degradation Modeling:

The data given to the model is degraded in two steps. Thus the training data does not require degraded images. The degradation is built such that it simulates various types of noise [13].

U-net Discriminator:

The discriminator part of GAN is built on spectral normalization [13].

Synthetic data training:

The training is done on the synthetic data generated by the degradation model.

Artifact Reduction with Sinc Filters:

Real-ESRGAN incorporates "sinc filters" in the degradation process to simulate and address common artifacts. The sinc filter can be defined as

$$k(i,j) = \frac{\omega_c}{2\pi\sqrt{i^2+j^2}} J_1\left(\omega_c \sqrt{i^2+j^2}\right) \qquad (1)\ [13]$$

where (i, j) is the kernel coordinate; ω_c is the cutoff frequency; and J1 is the first order Bessel function of the first kind.

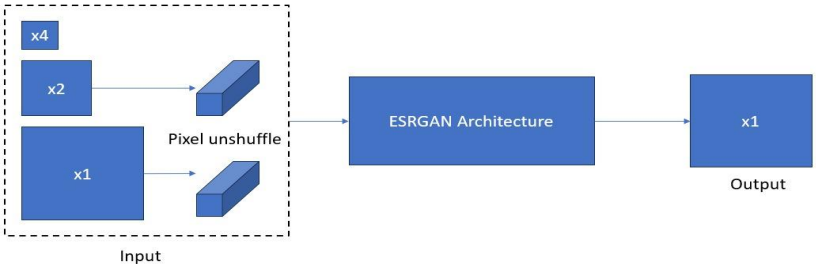

Fig1: Real – ESRGAN block diagram

The network architecture of SwinIR is described in their documentation as following[7]

Shallow Feature Extraction:
A 3x3 convolution extracts basic features from the low-quality input.
Deep Feature Extraction:
Stacked Residual Swin Transformer Blocks (RSTBs) with shifted windows model long-range dependencies and include residual connections for stability.
Image Reconstruction:
Combines shallow and deep features; uses sub-pixel convolution for up sampling or a simple convolution for denoising.
Task-specific Loss:
Uses L1 loss for super-resolution and Charbonnier loss for denoising and artifact reduction

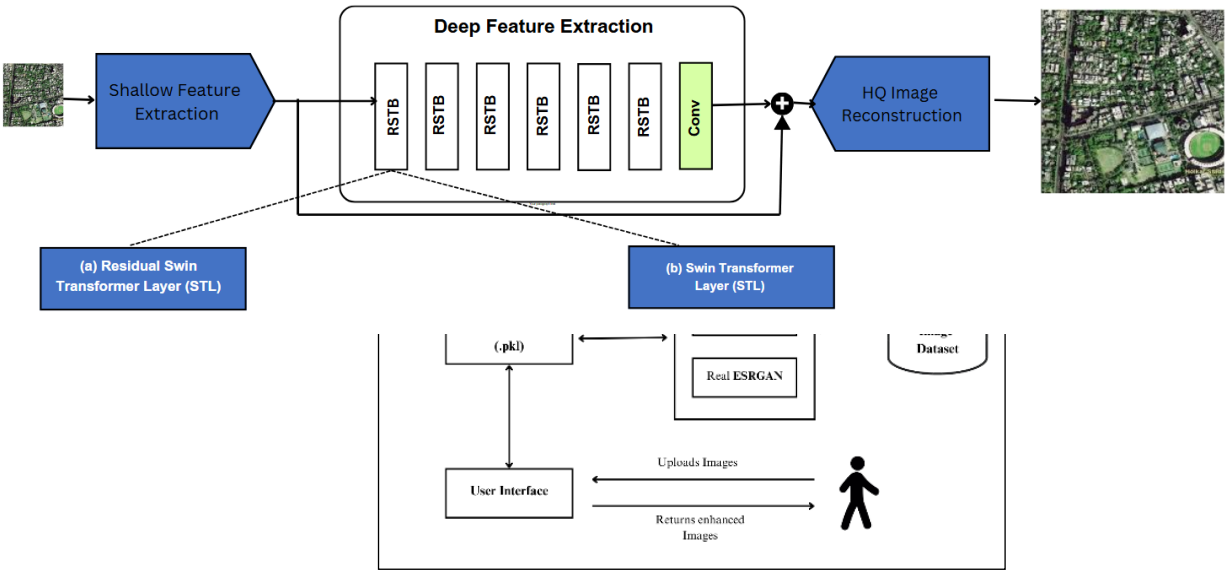

Figure2: SwinIR Architecture
Using outputs from both the models we intend to build a system where the strong points of these models can be used. Once the models are trained the outputs are overlayed to preserve most of the features. The input images can be taken from the user interface.
Training Procedure
We trained Real-ESRGAN using the procedure listed in the project. Following the steps, we reduced the number of epochs and training data due to limitations in computing power. The training data consisted of 5000 satellite images on 2000 epochs. The original model was trained on animated images and we trained it on satellite images. The training took approximately 6 hours on a GPU of size 8 gigabytes. The training could be faster if a more powerful computer was used as the procedure requires lot of computing power.
Inference Process
We have inferred the models using python scripts [13]. The code sourced included these scripts. Inference can be done by downloading pre trained models. We used a pre trained model for SwinIR [7] and used the trained model for Real – ESRGAN. For the resultant image we overlayed the outputs given by both the models. The output from the Real – ESRGAN model can be used inherently. The output from the other model requires some preprocessing before being used to produce the resultant image. The output from the SwinIR model have a yellowish tint due to the nature of the input images from the dataset, thus we adjusted the intensity value of blue to reduce this yellowness. One of the output images transparencies is reduced to 0.5 and overlayed on the other. Since SwinIR produces an output of varying size we resize the image and shift it slightly such that the features align perfectly.
We used an image of moderate quality as an input for the first test case. The output is an image with increased resolution and more sharper features. An image of lower quality is used in the second test case. The output is an image with better quality and the noise in the image is reduced. These scripts are

connected to a user interface built with Streamlit framework in python. Once an image is uploaded the script is run in the terminal. The UI can only take one image input at a time and can be further improved to take multiple inputs.

Evaluation Metrics

There are several evaluation metrics using which we can compare the input and resultant images. The traditional techniques used to compare the quality of the images require both the images to be of same size. With our solution the images are of different size. There are two ways evaluate the output. With the first method we can resize the input image using traditional resizing techniques and compare the output from the models. This method might be inconsistent as all the output images are not the same size.

With the second method we can train deep learning models to evaluate the image. This can take a lot of time and computing power. Due to the lack of resources we stick to visual comparison of images.

Comparison of Images

Image 1: moderate resolution image

The image has been upscaled to a factor of 4. There is minimal loss of features on visual observation. Though the image has a light-yellow tint. Minute details are also preserved which can be observed.

Image 2: Resultant Image

Results

The models have produced an enhanced image which is 4 times the size of the original image. The features and quality are preserved on a visual basis. There is a major enhancement in images that have a lot of noise and are distorted. The models could allow us to build a larger dataset for other purposes.

Challenges: Using two models required a lot of computing power the time taken to enhance each image. The time taken to receive results is around 10 minutes per image. We are using computer with limited computing power and graphics processing power.

CONCLUSION

This paper highlights the contributions of Real-ESRGAN and SwinIR in enhancing the quality of satellite imagery. These models Were versatile and adaptable, making them suitable not only for satellite images but also for training across various other domains of images. Both models have delivered good results when applied to satellite imagery. Real-ESRGAN employs a GAN-based approach, which is particularly effective in generating high-quality image enhancements. SwinIR, on the other hand, leverages the power of Swin Transformers to achieve remarkable results in image restoration and super-resolution.

Future research in this domain is expected to focus on optimizing these models to reduce their computational requirements. This optimization will enable deployment for real-time processing, making them viable for time-sensitive projects such as disaster monitoring, urban planning, and environmental tracking. Another key area of improvement lies in the development of user-friendly interfaces. By enhancing the interface, users can gain greater flexibility and control, such as the ability to choose between Real-ESRGAN and SwinIR based on their specific requirements. Allowing users to select the model that best suits their needs could significantly reduce the processing time, making the system more efficient and accessible for diverse use cases.

Additionally, integrating options for customized parameter adjustments could empower users to fine-tune outputs to their liking, ensuring the models cater to a broader spectrum of user preferences and application scenarios. These advancements will broaden the scope of these models and pave the way for their integration into various industries requiring high-quality image enhancement, further solidifying their role as indispensable tools in image processing and analysis.

REFERENCES

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. Advances in Neural Information Processing Systems, 27, 2672–2680.

[2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., & 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), 105–114. https://doi.org/10.1109/CVPR.2017.19

[3] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). *Enhanced Deep Residual Networks for Single Image Super-Resolution*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 1132–1140. https://doi.org/10.1109/CVPRW.2017.151

[4] Wang, X., Yu, K., Dong, C., & Change Loy, C. (2018). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 63–79. https://doi.org/10.1007/978-3-030-11021-5_5

[5] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). *Residual Dense Network for Image Super-Resolution*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2472–2481. https://doi.org/10.1109/CVPR.2018.00262

[6] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). *Learning a Deep Convolutional Network for Image Super-Resolution*. Proceedings of the European Conference on Computer Vision (ECCV), 184–199. https://doi.org/10.1007/978-3-319-10593-2_13

[7] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). *SwinIR: Image Restoration Using Swin Transformer*. arXiv preprint arXiv:2108.10257. https://doi.org/10.48550/arXiv.2108.10257

[8] Zhou, W., 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. https://doi.org/10.1109/TIP.2003.819861

[9] Hore, A., & Ziou, D. (2010). *Image Quality Metrics: PSNR vs. SSIM*. Proceedings of the International Conference on Pattern Recognition (ICPR), 2366–2369. https://doi.org/10.1109/ICPR.2010.579

[10] Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[11] Xie, W., Ma, X., Lu, C., & Liu, Y. (2021). *Advances in Satellite Imagery Super-Resolution: A Survey*. Remote Sensing, 13(4), 724. https://doi.org/10.3390/rs13040724

[12] Bovik, A. C. (2013). *Automatic Prediction of Perceptual Image Quality*. Proceedings of the IEEE, 101(9), 2008–2024. https://doi.org/10.1109/JPROC.2013.2282516

[13] Tao, X., Guo, M., Zhao, J., & Liang, X. (2021). *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*. GitHub Repository: Real-ESRGAN. https://github.com/xinntao/Real-ESRGAN

[14] Zhao, H., Zhang, X., Li, S., & Liu, J. (2017). *Loss Functions for Image Super-Resolution: Which One to Choose?* IEEE Transactions on Image Processing, 26(7), 3228–3243. https://doi.org/10.1109/TIP.2017.2699998

[15] *https://www.kaggle.com/code/kmader/segmenting-buildings-in-satellite-images/input*

[16] Park, S. C., Park, M. K., & Kang, M. G. (2003). Super-Resolution Image Reconstruction: A Technical Overview. *IEEE Signal Processing Magazine*, 20(3), 21–36. https://doi.org/10.1109/MSP.2003.1203207