A Comprehensive Analysis of Generative Adversarial Networks for Single Image Super Resolution

Jessica Pinto
California State Polytechnic University, Pomona
jmpinto@cpp.edu

ABSTRACT
Within the agricultural field, high resolution (HR) remote sensing data in the form of landscape images has been a prevalent tool for crop monitoring and analysis. However, in recent years, its value has been bolstered by popularity in applications such as environmental monitoring, urban planning, and disaster management. In these domains, where precise details about the image content including crop data, existing structures, and natural disasters are paramount, high-quality images are critical to support decision making. With increasing use of HR data, associated monetary costs for capturing and storing such images are also expected to increase. Alternatively, low resolution (LR) images are less expensive and more frequently updated. One solution is the upscaling of ubiquitous LR images to HR images using Single Image Super Resolution (SISR). Widespread methods of SISR using deep neural networks (DNNs) such as the Laplacian Pyramid Super Resolution Network (LapSRN) have been successful with remote sensing data, however, with newer DNN architectures such as the Real-Enhanced Super Resolution Generative Adversarial Network (Real-ESRGAN), model performance is noticeably better for remote sensing applications. When trained on paired LR and HR sentinel images from summer months of 2020 to 2024, Real-ESRGAN's performance surpasses that of LapSRN in terms of perceptual quality, but not in quantitative metrics, as expected. This research provides a performance analysis of GAN based techniques for image super resolution compared to LapSRN based techniques. The results of each model are evaluated using perceptual comparison, mean squared error, structural similarity index measure, and peak signal-to-noise ratio to delineate SISR performance.

Key words: SISR, LapSRN, Real-ESRGAN, GAN

1 INTRODUCTION

In the last thirty years, there has been a global rise in the number of crop calories consumed, driven by high population growth and larger diets. The United States Department of Agriculture expects this number to increase significantly over the next few decades, and unsurprisingly so considering recent advancements in medical technology, accessibility, and modern infrastructure. With this increase in population comes a greater need for food sources, the demand of which cannot be met without adequate agricultural planning, supported largely by remote sensing data.

A more prominent application of such data comes during times of emergency, when satellite imagery helps determine evacuation routes, analyze damage, and monitor reconstruction efforts. The versatility of remote sensing data makes it an important resource used by professionals in a number of fields, and yet its availability is much more restricted. Defined by the United States Geological Survey, "remote sensing is the process of detecting and monitoring the specific characteristics of an area by measuring its reflected and emitted radiation at a distance", and one of the most popular forms of remote sensing data comes in the form of satellite imagery (Sands, 2025).

Despite its ubiquity, these images are not always accessible to those that need it, such as first responders and agricultural planners, largely due to purchasing and storage costs. Gathering data from multiple regions and storing weeks, months or even years worth of high resolution (HR) data is not feasible for most organizations or individuals with limited resources. Moreover, low resolution (LR) images may be available at more affordable costs, but can lack the details necessary for accurate analysis, as seen in Figure 1. To solve this problem and make HR images more accessible, a process known as Image Super Resolution provides a unique solution.

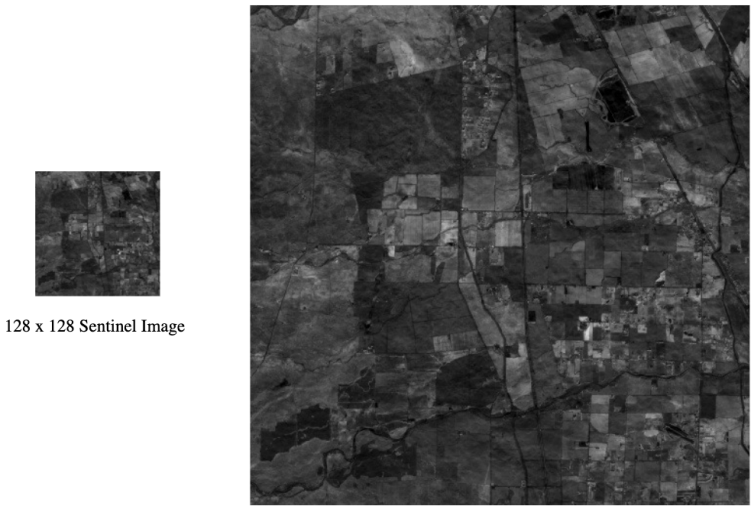

Figure 1: Visual comparison of a low resolution and high resolution sentinel image from December 2022.

2 OBJECTIVE

The main purpose of utilizing deep neural network (DNN) super resolution models is to make HR images more accessible to those who need it, especially when it comes to costs. Using trained models that produce images with minimal artifacts, significant qualitative improvement, and the ability to be applied to land analysis problems is the standard for good performance.

Through a detailed analysis of both the Laplacian Pyramid Network and Generative Adversarial Network models, the goal is to determine which model is better suited for these applications. Issues such as mode collapse and hallucinations can be detrimental in these scenarios, leaving little to no room for error in image reconstruction. Based on the unique architecture and importance of perceptual similarity that GANs prioritize, they may have an advantage when applied to the task of greyscale satellite image super resolution. However, the widespread use and focus on residual skip connections could place LapSRN in a similarly favorable position.

3 BACKGROUND

Image Super Resolution refers to the techniques used to enhance the quality of digital images or videos from their original resolution, specifically modeling HR images from LR inputs (Tomar et al., 2023). Traditional SISR methods include interpolation and are labeled traditional because they do not employ deep learning approaches but instead stretch images by approximating pixels or aligning multiple frames of the same image (Li et al., 2020). With the introduction of Convolutional Neural Networks (CNNs) and its ability to process graphical input, SISR has improved, enhanced by deep learning (DL) approaches. Some of the most unique architectures applied to this task are the Laplacian Pyramid Super Resolution Network (LapSRN) by Lai et al. (2017) and the Super Resolution Generative

Adversarial Network (SRGAN) by Ledig et al. (2017), with the most recent iteration of GANs being Real-Enhanced SRGAN (Real-ESRGAN) by Wang et al. (2021).

3.1 Laplacian Pyramid Super Resolution Network

Image Super Resolution using LapSRN–as its name suggests–utilizes a Laplacian Pyramid, defined as a pyramid of image residuals. Each residual is calculated as the difference between the upsampled LR image and the ground truth HR image at a specific resolution. A visual of this residual for Sentinel images is depicted in Figure 2. During training, LapSRN studies and learns how to predict the Laplacian residuals and during the testing process, it adds its own predicted residuals to the blurred and upsampled LR image through multiple iterations. Blurring is a byproduct of image upsampling as stretching an image leads to an expected loss in quality, therefore, adding residual details refines the larger image while preserving its perceptual quality. As can be seen in Figure 2, the output of each upscaled image is used not only to reconstruct the image but also extract specific features that allow for super resolution.

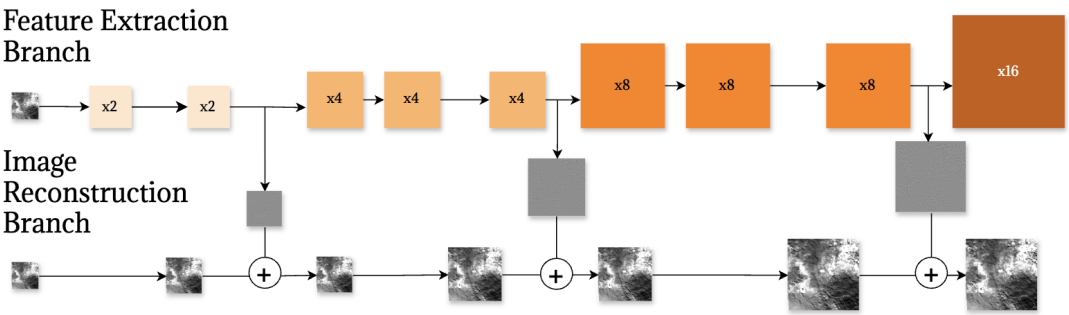

Figure 2: LapSRN Model Training Architecture (Lai et al., 2017).

The loss function used by LapSRN is the Charbonnier penalty function, chosen for its robust nature when it comes to handling outliers. At each layer of the pyramid, as the model continuously predicts the residual of the next level, the Charbonnier loss function is used to stabilize training and accelerate convergence as it represents the sum of losses across all levels of the pyramid.

$$\mathcal{L}(x, y) = \sqrt{(x - y)^2 + \epsilon^2}$$

x : Predicted value (model output)
y : Ground-truth target value
ϵ : Small constant for numerical stability
$\mathcal{L}(x, y)$: Charbonnier loss between x and y

Figure 3: Charbonnier loss function (Charbonnier et al., 1997).

3.2 Real Enhanced Super Resolution Generative Adversarial Network

On the other hand, Real-ESRGAN follows the classic GAN approach of adversarial training, where the generator's goal is to create the most realistic looking images, and the discriminator's goal is to distinguish which image is real and which one is fake. The generator is first given the LR input image, and then makes its best attempt to generate an HR version of the smae image that looks genuine. The discriminator is given the ground truth (GT) HR image as well as the image produced by the generator and must determine which image is real using a probabilistic score, where a value closer to 1 represents a higher probability of the image being "real". The structure of Real-ESRGAN's generator is depicted in Figure 4.

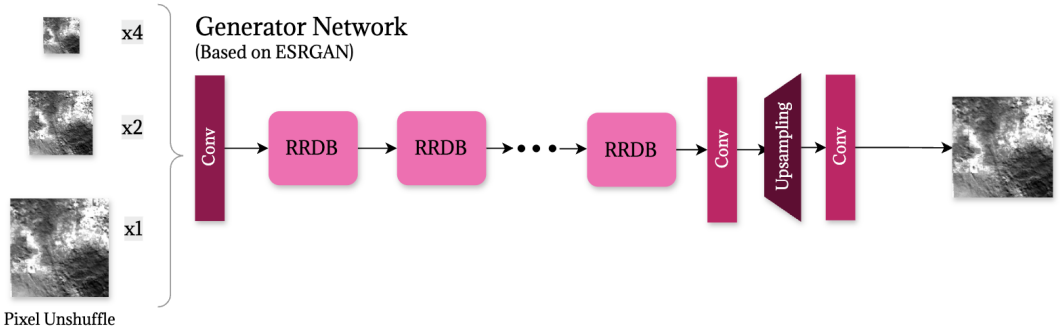

Figure 4: Architecture of Real-ESRGAN's Generator Network (Wang et al., 2021).

The Real-ESRGAN generator network is made up of residual-in-residual dense blocks (RRDBs) that each contain multiple convolution layers, receiving all previous feature maps as input. These RRDBs can be stacked upon each other, forming a residual path and communicating using skip connections, the structure of which is shown in Figure 5. Real-ESRGAN also employs the interpolation method of nearest-neighbor upsampling, which reduces artifacts and helps prevent mode collapse. Mode collapse is a common issue when working with GANs, where gradients become unstable and cause the generator output to collapse into a set of repeated patterns. The improvement in training stability provided by the nearest neighbor interpolation indirectly allows the circumvention of this issue.

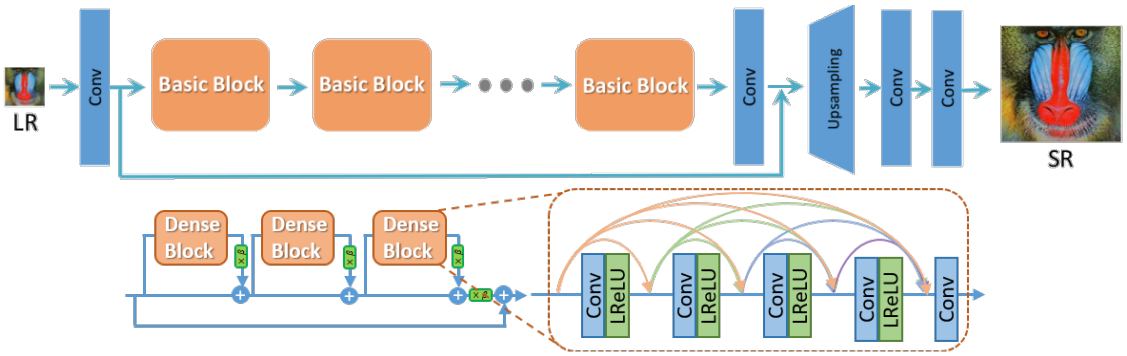

Figure 5: Architecture of Residual-in-Residual Dense Blocks (Wang et al., 2021).

3.3 Dataset

The images used for training and testing are grayscale landscape satellite images provided by the European Space Agency and accessed through Google Earth Engine. They include "all-weather radar images from Sentinel-1A and -1B, high-resolution optical images from Sentinel 2A and 2B, as well as ocean and land data. . . from Sentinel 3" (Gorelick et al., 2017). Images can be grouped and pulled from the database by month, allowing for the creation of custom datasets over the course of several years, but not prior 2019. For each month, more than 100 images can be fetched from Google Earth Engine and it is these images that will be used to train LapSRN and Real-ESRGAN. As a note, Google does not explicitly mention the exact time interval between each image that is captured, but a combination of daytime and nighttime images of the same area are gathered each year, as exemplified by Figure 6.

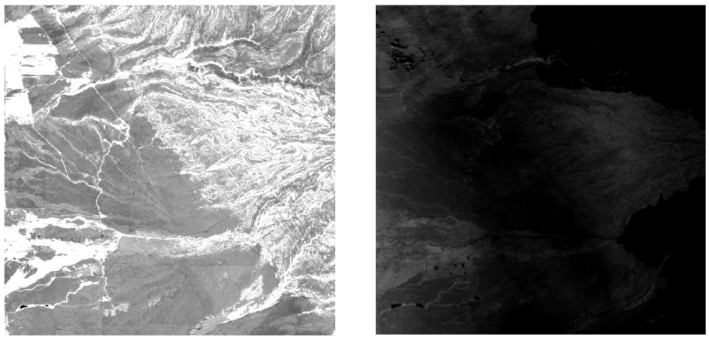

Figure 6: Image captured by Sentinel satellite during the day (left) verses at night (right).

4 METHODOLOGY

The datasets used to train both LapSRN and Real-ESRGAN were taken from Google Earth Engine's API, which provides geospatial datasets and near-real-time (NRT) data. The specific dataset used to train both models included landscape images from the coordinates defined in Figure 7 and labeled as Region 0. Specifically, Region 0's data was collected from 2020 to 2024 for the months of May, June, July, and August in an attempt to focus the training process on seasonal features.

$$\begin{bmatrix} -122.3226 & 40.1907 \\ -122.3226 & 39.8753 \\ -121.9700 & 39.8753 \\ -121.9700 & 40.1907 \end{bmatrix}$$

Figure 7: Coordinates of Region 0, from which all the training data is collected.

The LapSRN implementation provided by Jiu Xu on GitHub takes in the training dataset as a HDF5 file, explicitly differentiating between LR and HR images. During the training process, it identifies patterns between the LR and HR images distinctively, as opposed to considering both HR and LR images equally. This specification is what differentiates paired and unpaired training. Since both LR and HR images exist and are explicitly mapped to each other, LapSRN's training takes place using paired data.

Conversely, Real-ESRGAN's implementation provides both a paired and non paired training version that follows the same idea. Whereas LapSRN's paired input was in the form of a HDF5 file, Real-ESRGAN's setup is simpler, taking in a folder with LR images and one with HR images, then mapping them to each other based on the filename.

For a fair comparison, this experiment focuses on the performance of LapSRN and Real-ESRGAN's paired training performance, with both models provided LR images of size 128x128 and HR images of size 512x512–a total of 640 images. The results of the experiment is shown in Figure 9.

Each model is evaluated based on the test image's perceptual similarity and performance in quantitative metrics of Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR). MSE calculates the average squared difference between the reconstructed and ground truth image using the formula in Figure 8. A lower MSE represents a smaller error and better reconstruction. SSIM measures the image similarity based on luminance, contrast, and structure, working on patches of the image at a time, using the formula in Figure 8. Its values range from -1 to 1, with a number closer to 1 representing a better match. Finally, PSNR is derived from MSE with the same goal of determining the similarities of reconstruction to the ground truth image. It is done using the formula in Figure 8 where a higher PSNR represents a better image quality.

$$\text{MSE} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(I(i,j) - \hat{I}(i,j)\right)^2 \qquad \text{PSNR} = 10\log_{10}\left(\frac{MAX^2}{\text{MSE}}\right) \qquad \text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Figure 8: Formulas for MSE (left), PSNR (center), and SSIM (right) taken from Wang et al. (2004).

In the MSE and PSNR formulas, *I(i,j)* represents the ground-truth image pixel at location (i,j). $\hat{I}(i,j)$ represents the reconstructed or predicted image pixel at location (i,j), (H,W) represents the height and width of the image and MAX represents the maximum possible pixel value. In the SSIM formula, μ_x, μ_y is the local mean of image patches x and y, while σ_x^2, σ_y^2 is the local variance of patches x and y. Finally, C_1, C_2 are the stabilizing constants to avoid division by zero.

5 RESULTS

After training both the LapSRN and Real-ESRGAN models with the same data, over the same number of epochs (50), results of the experiment show that the performance of the GAN model surpasses that of the Laplacian Pyramid model. Both models were trained using the paired training structure and took around 40 minutes to 1.5 hours to complete training. The test image was taken from Sentinel's October 2024 batch and includes a mix of dark and light elements without unbalanced overexposure. The superiority of Real-ESRGAN is specifically based on perceptual similarity, as Real-ESRGAN's quantitative performance is still lacking in comparison to LapSRN. When looking at the results in Figure 9, it is clear that the image generated by Real-ESRGAN resembles much more closely the ground truth Sentinel image, but the Mean Squared Error, Structural Similarity Index Measure, and Peak Signal to Noise Ratio values beg to differ.

Currently, this is a common issue with GANs where quantitative evaluation metrics are unable to accurately capture the performance of the model. Instead, perceptual quality is used for evaluation, where humans are asked to differentiate between the real and fake images. This method has proven to be useful in cases such as this where quantitative evaluation metrics cannot accurately represent the performance of the model.

Model	**MSE**	**SSIM**	**PSNR**	**Notes**
Real-ESRGAN	611.981	0.543	20.263	50 epochs
Real-ESRGAN	624.350	0.554	20.177	24 epochs
LapSRN	442.435	0.568	21.672	50 epochs

Figure 9: Super-resolution performance comparison of Real-ESRGAN and LapSRN.

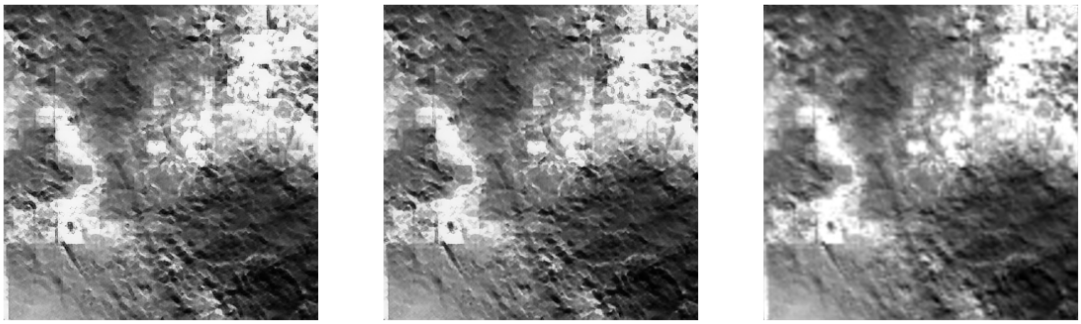

Figure 10: Results from Real-ESRGAN with 50 epochs (left), Real-ESRGAN with 24 epochs (center) and LapSRN with 50 epochs (right).

6 CONCLUSION

Analyzing the performance of Generative Adversarial based and Laplacian Pyramid based Single Image Super Resolution models has revealed the perceptual superiority of GAN performance when applied to remote sensing data in the form of satellite imagery. However, quantitative metrics do not reflect Real-ESRGAN's improvement over LapSRN, revealing a need for specialized quantitative measures optimized for a wider range of deep neural network models, such as GANs. The accessibility of both LapSRN and Real-ESRGAN make them prime solutions to the issue of expensive remote sensing data, and can aid in critical decision making in various real-world applications.

7 ACKNOWLEDGEMENTS

Thank you to Dr. John Korah for his mentorship while working on this paper and Nico Escobedo for guidance on LapSRN models and implementations.

REFERENCES

Charbonnier, P., L. Blanc-Feraud, G. Aubert, and M. Barlaud, 1997, Deterministic edge-preserving regularization in computed imaging: IEEE Transactions on Image Processing, **6**, 298–311.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, 2017, Google earth engine: Planetary-scale geospatial analysis for everyone: Remote Sensing of Environment, **202**, 18–27.

Lai, W.-S., J.-B. Huang, N. Ahuja, and M.-H. Yang, 2017, Deep laplacian pyramid networks for fast and accurate super-resolution.

Ledig, C., L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, 2017, Photo-realistic single image super-resolution using a generative adversarial network.

Li, K., S. Yang, R. Dong, X. Wang, and J. Huang, 2020, Survey of single image super-resolution reconstruction: IET Image Processing, **14**, 2273–2290.

Sands, R., 2025, World crop calories and food calories projected to continue growth trend through 2050: https://ers.usda.gov/data-products/chart-gallery/chart-detail?chartId=109344. (Accessed: 2025-11-01).

Tomar, A. S., K. Arya, S. S. Rajput, and C. R. Rodriguez, 2023, Chapter 9 - comprehensive survey of face super-resolution techniques, *in* Digital Image Enhancement and Reconstruction: Academic Press, Hybrid Computational Intelligence for Pattern Analysis, 213–233.

Wang, X., L. Xie, C. Dong, and Y. Shan, 2021, Real-esrgan: Training real-world blind super-resolution with pure synthetic data.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 2004, Image quality assessment: From error visibility to structural similarity: IEEE Transactions on Image Processing, **13**, 600–612.