Enhancing Satellite Images of FACSAT-1 through Generative Adversarial Networks for Super-Resolution

Paola Zarate (paola.zarate@fac.mil.co)
Research Center in Aerospace Technologies
Christian Arroyo
Research Center in Aerospace Technologies
Jesús López
Universidad Autónoma de Occidente
Jorge Jiménez
Colombian Aerospace Force

Research Article

Keywords: Artificial Neural Network, Deep Learning, Generative Models, satellite images, Super-Resolution

Posted Date: January 11th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3847860/v1

Additional Declarations: No competing interests reported.

Enhancing Satellite Images of FACSAT-1 through Generative Adversarial Networks for Super-Resolution

Paola Zarate Author[1*], Christian Arroyo Author[1], Jesus Lopez Author[2], Jorge Jimenez Author[3]

[1*]Technological Development, Research Center in Aerospace Technologies, Carrera 8 No. 58-67et, Cali, 760001, Valle del Cauca, Colombia.

[2]Faculty of Engineering, Universidad Autonoma de Occidente, Calle 25 No. 115-85 Km. 2, Cali, 760030, Valle del Cauca, Colombia.

[3]Department of Science and Technology, Colombian Aerospace Force, , Bogota D.C., 61631, Cundinamarca, Colombia.

*Corresponding author(s). E-mail(s): paola.zarate@fac.mil.co;
Contributing authors: christian.arroyo@epfac.edu.co; jalopez@uao.edu.co; jorge.jimenez@fac.mil.co;

Abstract

Satellite images have diverse applications across scientific, commercial, and other domains. As a result, institutions are increasingly deploying Earth observation satellites to cater to their specific needs. This is the case of the Colombian Aerospace Force, which launched its nanosatellite, the FACSAT-1, to contribute to developing the space sector in Colombia. However, in some cases, captured images may need more quality and resolution for their intended purposes. Numerous image processing tools have been developed to enhance and optimize satellite imagery to address this challenge. Deep learning techniques, particularly Generative Adversarial Networks, have recently shown significant advancements in image processing. This paper investigates the widespread application of Generative Networks for satellite and aerial imagery, specifically focusing on super-resolution tasks. Super-resolution involves increasing the resolution of satellite images by up to four times their original size. The study presents the implementation and training of four different generative models and evaluates their performance using qualitative and quantitative measures. Two metrics, namely the maximum signal-to-noise ratio and the structural similarity index, are employed for comparative

analysis. By assessing the output of each generative model, this research aims to determine their efficacy in enhancing satellite imagery resolution.

Keywords: Artificial Neural Network, Deep Learning, Generative Models, satellite images, Super-Resolution

1 Introduction

In Colombia, the space initiative arose from the Colombian Space Commission (Comision Colombiana del Espacio -CCE- in Spanish). Created in July 2006 and headed by the Instituto Geografico Agustin Codazzi, which served as Executive Secretariat of this commission with the main objective of launching a communications satellite called SATCOL-1 in 2011; this way, autonomy in the use of space services would be obtained, avoiding the need to resort to foreign operators.

Subsequently, in 2013 the secretariat of the CCE was assumed by the Colombian Aerospace Force (Fuerza Aeroespacial Colombiana -FAC- in Spanish) previously named Colombian Air Force, where it was announced the development of a project for the creation of an observation nanosatellite, with a useful life of 3 to 5 years, called FACSAT-1 (Martínez, 2018), this nanosatellite integrated in Denmark, consisting of three cubes, with an RGB camera (visible spectrum) with a resolution of 30 m per pixel and a useful life of three to five years; It is currently at an altitude of about 490 km and circles the planet Earth approximately every 90 minutes.

Artificial satellites play a pivotal role as unique tools with diverse and highly practical applications, notably encompassing Earth observation from outer space, capturing valuable snapshots of the planet. These snapshots are instrumental in addressing critical inquiries about climate, atmosphere, oceans and various terrestrial soil conditions. Satellites offer unparalleled insights into forest evolution, deforestation, mining, melting glaciers, sea-level rise, and marine pollution. Many of these applications necessitate high-resolution images of specific areas, as they provide a superior level of detail, granting access to invaluable information for analytical purposes.

In certain circumstances, satellite images may need more requisite resolution due to various factors, including meteorological conditions, posing limitations for research endeavors. This situation may necessitate acquiring higher-resolution images or exploring alternative sources, leading to a significant increase in investment costs for such projects.

The present research employs cutting-edge Deep Learning techniques, specifically Generative Adversarial Networks (GANs), to execute a super-resolution process using freely available satellite images captured by cameras resembling those aboard the FACSAT-1 nanosatellite. Additionally, a dataset derived from the FACSAT-1 nanosatellite enriches the process. The primary objective is obtaining magnified high-resolution images of zones of interest, facilitating their more effective utilization in future research projects.

The paper contains a review of the state of the art in super-resolution with GANs, going through different algorithms called SRGAN, as well as a methodology to apply

the different existing architectures, by training models with a specific dataset and finally presenting a comparison between the results obtained with the different models, defining their strengths and weaknesses.

2 Problem statement

The first Colombian satellite called FACSAT-1, intended for Earth observation, was equipped with a camera with a resolution of 30 meters per pixel, the nanosatellite's mission is to detect and take Earth observation images. With the images and data obtained, it is expected to develop new capacities in science and technology, as well as to initiate the path towards technological autonomy and less dependence on the exterior in this aspect, as well as to support programs for urban development, national security, crop substitution, natural disaster response and scientific studies.

One of the problems encountered with the FACSAT-1 satellite was during the acquisition of the captured images. These images must initially be acquired in low resolution. This choice is due to the fact that downloading information requires considerable time. It is essential to verify the capture conditions, such as atmospheric conditions and location, before determining if the image is really of interest. This prior process avoids wasting time by downloading at a higher resolution.

The download of high-resolution images can take between 5 and 6 days. This is due to the fact that the download can only be carried out when the satellite orbits over Colombian territory for a period of 10 minutes. This temporal limitation significantly prolongs the time required, making it difficult to carry out the various applications for which the satellite images were conceived. This inconvenience delays the initially planned research and technological development objectives. It is essential to seek and implement methods to improve the resolution of images downloaded in low resolution.

This would allow a more detailed analysis of specific captures, facilitating the identification of important features. It would also help to determine possible applications for each image, thus optimizing the use of the valuable time that would be spent to download again the really useful images directly from the satellite, but this time with a higher resolution.

In addition to contributing to the resolution of the FACSAT-1 problem, the implementation of a super-resolution tool based on artificial intelligence for satellite images could offer substantial benefits in other research carried out by the Colombian Aerospace Force. An outstanding example is its application in the detection and segmentation of patterns associated with illegal open-pit mining through the use of convolutional neural networks. The application of super-resolution to the images used in such detection would enable a significant improvement in the quality of target details. This advance would facilitate the precise identification of relevant elements by the neural network, thus improving the efficiency and accuracy of the process of detection and segmentation of illegal mining patterns.

3 Background

3.1 Super-resolution

Super-resolution (SR) (Yue et al, 2016) is defined as a process mainly aimed at generating a high-resolution (HR) image from a low-resolution (LR) one. HR images provide greater detail in their contents when compared to LR images, as the former has a higher density of pixels. The need for a HR image is very common in applications of computational vision for reconnaissance and analysis for many areas of interest, such as medicine, surveillance, and forensic science, besides their use in satellite images. When considering the high cost and limitations of improving image resolution through hardware, especially for large-scale imaging devices, SR methods, essentially algorithms for signal processing, have become a potential way to obtain HR images. SR makes it possible to improve the observation of LR images without improving hardware devices.

The general premise in the super-resolution process is that an LR image is the result of the degradation of an HR image. Therefore it is possible to reconstruct it based on the source image and a model to do the opposite of the degradation function. The model's accuracy is crucial for super-resolution, and an incorrect model may further degrade the image.

3.2 Generative models

Generative models (Ruthotto and Haber, 2021) are used as methods with Artificial Neural Networks aimed at learning the true distribution of any kind of data in its respective training set, they can thus generate other samples with slight variations. Some common architectures used as the foundation to develop generative models is the Convolutional Neural Network (CNN) and Generative Adversarial Networks (GAN). In addition, there are other notable approaches. Auto-regressive Models, such as PixelRNN and PixelCNN (Van Den Oord et al, 2016), generate data sequentially, conditioning each step in the generation on the previous ones. Variational Autoencoder Models (VAE) (Kingma and Welling, 2013) combine autoencoder techniques with variational inference methods to learn latent probability distributions in observed data.

Generative models have been used in several areas, such as realistic image synthesis or coherent text generation. However, many challenges remain, such as training stability and interpretation of latent representations. Continued research into these techniques has sought to address these limitations to further extend the applicability of generative models.

Generative models can be considered as a powerful tool for optimization in data generation as well as data processing. With the continuous development and refinement of algorithms, these models promise to play an increasingly crucial role in solving complex problems in various domains.

3.3 Convolutional Neural Network

Convolutional Neural Networks (CNN) use filters or kernels to operate convolution on an image to detect different features. The operation of convolution is carried out by

moving the kernel over regions of the image which are the same size as the kernel, and it calculates the sum of products, element by element. As a result, a numerical value is obtained which indicates the likelihood of the feature represented by such a filter to be present in the region of the image that is being analyzed.

Such kind of CNN architecture was employed in the development of the first deep learning applications intended for the super-resolution process. This gave way to the super-resolution convolutional neural network (SRCNN) (Rath, 2020). The SRCNN process consists of, first, preprocessing the input with bicubic interpolation to magnify an LR image up to its objective size. Then, the blurred input image is enhanced by the algorithm, by processing the convolution network to obtain the SR image, thus making it look as close as possible to what the original HR image should look like. Such a process is illustrated in Figure 1.

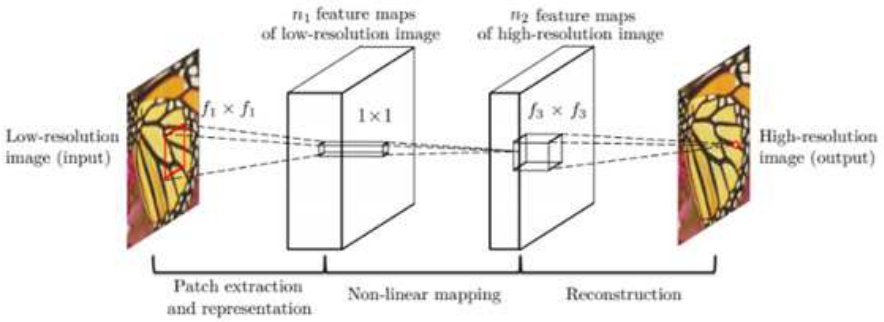

Fig. 1 Example of SRCNN. (Rath, 2020)

3.4 Generative Adversarial Networks

GANs (Karim, 2018) are a clever way to use automatic learning to generate realistic imaging. As shown in Figure 2, GANs use two deep neural networks. Such two networks are adversarial, that is, they play a zero-sum game: one network wins, whereas the other one loses. The discriminator's role is to classify whether an image is authentic or fake. The role of the generator, then, is to create images that look authentic and to overcome the discriminator. GANs have proven highly versatile, with applications ranging from creative content generation to data synthesis for enhanced training datasets.

In the GAN framework, the aforementioned generator (G) aims to learn the distribution of the input data (x). For this purpose, it maps the input noise variables sampled from a distribution (z) to the data space by means of a differentiable function represented by a multilayer perceptron. At the same time, a discriminator (D), another multilayer perceptron, evaluates the probability that a given instance comes from the actual data distribution and not from the generated one. The adversarial training process tries to minimize the logarithmic probability of misclassification for both the real and generated samples.

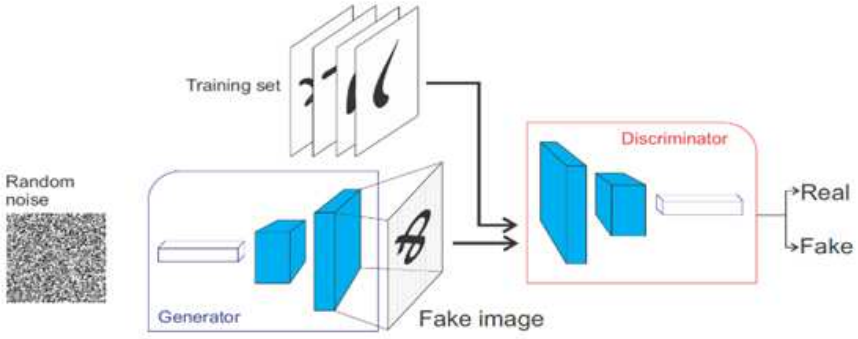

Fig. 2 Example of GAN. (Agrawal, 2023)

The optimization procedure consists of alternating between the discriminator refinement step and a denerator improvement step. This iterative approach avoids over-tuning the discriminator with respect to the current generator, guaranteeing the stability of the training process. The theoretical analysis of GANs demonstrates that, given sufficient capacity and training time, the framework converges to an optimal estimator of the data generating distribution.

In practical implementations, the generator implicitly defines a probability distribution, which represents the generated samples. In particular, early in the learning process, training the generator to minimize log(1 - D(G(z))) may face problems when the discriminator confidently rejects samples that are markedly different from the training data. In such cases, an alternative objective function, maximizing log D(G(z)), is employed to provide stronger gradients during the initial stages of learning.

Although GANs present new advantages, such as dispensing with Markov chains and relying solely on backpropagation for gradient computation, they also pose problems. In particular, there is no explicit representation of the data, and proper synchronization between G and D during training is crucial.

GANs represent an innovative paradigm in generative modeling. As computational capabilities advance, research to refine GANs and explore their full potential remains a vibrant avenue within the field of machine learning.

3.4.1 Super-resolution with Generative Adversarial Networks

The use of GANs for super-resolution originated the SRGAN architecture (Ledig et al, 2016), which became popular as it showed improved results, compared to SRCNN. Despite having shown very acceptable results, SRCNN methods keep having the same hiccup, the images generated seem false: they are clearly of higher resolution, but they seem blurry and lack fine high-frequency details when compared to the real HR images (Jiang and Li, 2020). Instead, GANs are models that are capable of creating more perceptually satisfactory and less blurry images, as perceptive differences between the original HR image and the SR generated may be captured.

For such a model, generator architecture (3) consists of a residual network instead of deep convolutional networks. Residual networks are easier to train and may go

substantially deeper to generate better results, as residual networks use a kind of connection called skip connections.

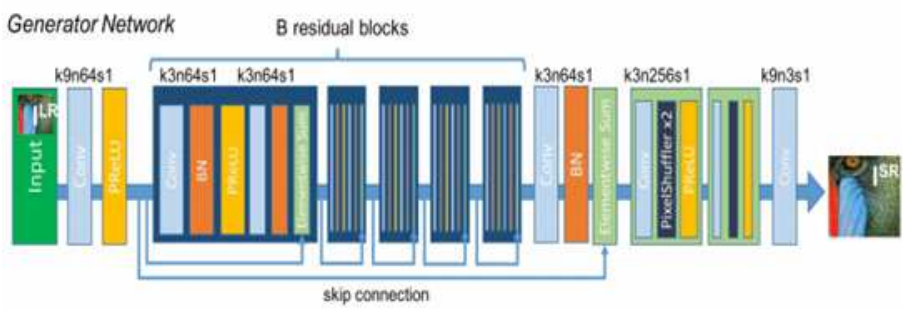

Fig. 3 Generator architecture. (Ledig et al, 2016)

Discriminator architecture (Figure 4) usually consists of a network containing eight convolutional layers with a 3 x 3 filter, which increases by a factor of 2, from 64 up to 512 filters. The discriminator architecture used in this network is like deep CNN architecture (Radford et al, 2015), with LeakyReLU as an activation function.

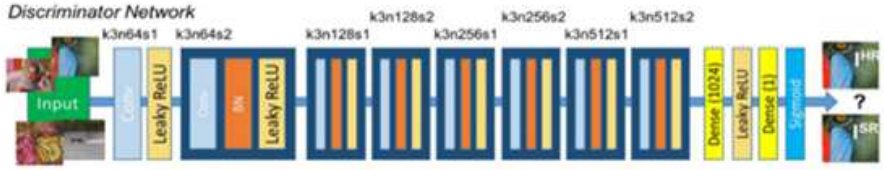

Fig. 4 Discriminator architecture. (Ledig et al, 2016)

The loss function for the generator consists of content loss and adversarial loss (Equation 1), where the latter forces the generator to produce SR images that are more similar to HR ones using the discriminator, which is trained to distinguish between those images.

Loss function for generator:

$$L^{SR} = L_x^{SR} + 10^{-3} L_{Gen}^{SR} \tag{1}$$

Where L_X^{SR} is the content loss and L_{Gen}^{SR} is the adversarial loss.

The loss function for the discriminator is also the adversarial loss as expressed in Equation 2.

$$L_{Gen}^{SR} = \sum_{n=1}^{N} -logD_\theta(G_\theta(I^{SR})) \tag{2}$$

Where D_θ is the discriminator, G_θ is the Generator and I^{LR} is the low-resolution image.

4 Related work

Some research projects have already applied GANs for satellite image processing, for example in the correction of radiometric errors such as banding (Zárate et al, 2023).

Similarly, super-resolution techniques with deep learning (Zhang, 2019; Gupta et al, 2020; Abdelmaksoud et al, 2020), have shown results that are similar to those expected from the present project.

Super-resolution has also been approached previously to process satellite images. In 2018, European Space Agency (ESA), together with Advanced Concept Team (ACT), issued a challenge aimed at the scientific community, to apply the latest AI and image processing techniques to near-infrared (NIR) images acquired by PROBA-V minisatellite, which is dedicated to the study of Earth's vegetation, aimed at contributing in the struggle against deforestation. In 2019 Zhang (2019), had a meaningful progress, but Gupta et al (2020) in 2020 they managed to win the challenge by obtaining the results shown in Figure 5. Such results won the challenge proposed by ESA and were obtained by a network called DeepSUM, an end-to-end trainable CNN which processes a series of low-resolution input images to provide an SR image at its output.

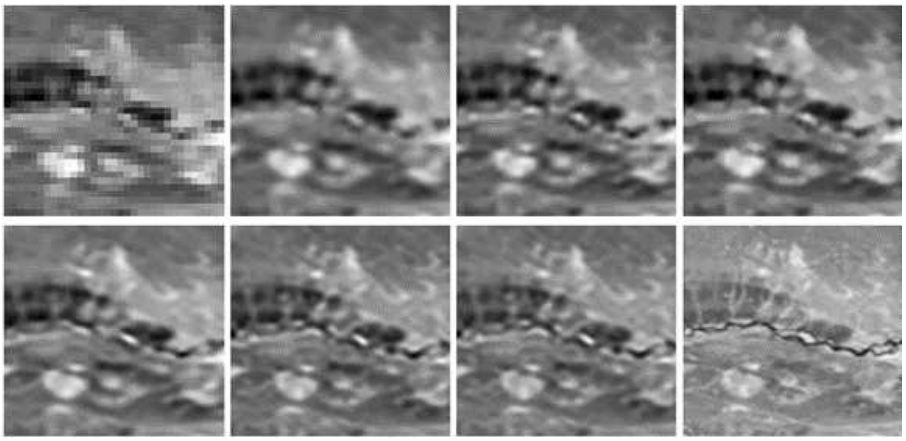

Fig. 5 From top to bottom, and from left to right: one of the LR images, bicubic + mean, iterative back projection (IBP), bilateral total variation (BTV), dynamic upsampling filters (DUF), proposed method without RegNet, DeepSUM, and original HR image. (Molini et al, 2020)

5 Methodology

Advances in satellite imaging technology have enabled and facilitated increased acquisition of earth data that provides valuable insights for a variety of applications. To improve the resolution of satellite imagery, this project employs state-of-the-art techniques focused on generative adversarial networks (GANs) and super-resolution generative adversarial networks (SRGANs). The methodology is developed in three key phases, which address the creation of datasets, the implementation and training of models using SRGAN algorithms and a meticulous evaluation of the results.

5.1 Creation of a dataset for the training of models of generative adversarial networks

To build the dataset that would be available for network training, high-resolution satellite imagery was collected. In a first instance, after searching for already available repositories, a collection of satellite imagery was found on Kaggle, called WiDS Datathon 2019. This dataset, organized in collaboration with Planet and Figure Eight, was built for a challenge focused on addressing the global impact of deforestation and oil palm plantations using high-resolution satellite imagery. Consisting of 3-meter spatial resolution imagery, it contains a total of 1,316 HR satellite images, all with dimensions of 2448 x 2448 pixels, and several of these were of rural areas.

Then, more images were collected, but specifically of the Colombian territory, with the purpose of including images with similar characteristics to those taken over areas with similar geography to those to be processed from FACSAT-1. For this purpose, the Russian program SASPlanet was used, zooming the Bing and ESRI satellite map by 18X, so that the images loaded corresponded to a resolution of about 30 meters/pixel, and the images were downloaded with their original resolution. Finally, the WiDS images and those collected with SASPlanet were put together, being a total of 2005 HR satellite images and by using a Python algorithm the images were resized making them four times smaller, in order to obtain each of these images in their LR version. The resizing algorithm uses the Pillow package and the process involves loading the image, specifying the new dimensions (reducing to a quarter of the original size), and then saving the resized image.

After the input (HR images) and reference (LR images) data were available, they were organized into training and validation sets, divided into 80% and 20% respectively, leaving 1604 images for training and 401 for validation. It is important to note that all images had only 3 spectral bands, being Red (R), Green (G) and Blue (B), whose composition results in the satellite image in the common human visible color spectrum.

5.2 Training of generative adversarial networks for super-resolution

The use of SRGAN algorithms, available by Jiang (2021); Molini et al (2020); Ren et al (2021); Wang et al (2021), allows the implementation and use of different super-resolution networks so that the due re-training of models is conducted by using the dataset with satellite images and establishing a comparative reference to select the algorithm showing the best performance with such kind of images.

In reality, the training process of a GAN implies training two neural networks: the generator and the discriminator. The goal with SRGAN is to have the generator process a LR image and to produce at the output an image that is as close to an HR image as possible. While the generator improves with the training, the discriminator's performance worsens as it cannot easily distinguish between real and fake.

The SRGAN algorithms implemented are different in the loss of content that are used to optimize the generator, as well as the network architecture in the case of Enhanced-SRGAN and Real-ESRGAN.

1. The algorithm applied by Jiang (2021), called SRGAN with MSE error, has the architecture of the networks mentioned earlier, in the conceptual framework, and as is shown in Equation 3, the content loss functions are calculated with the mean square error (MSE) between the SR image generated and the HR image of reference.

$$L_{MSE}^{SR} = \frac{1}{N}\sum_{i=1}^{N} \|I^{HR} - I^{SR}\|^2 \tag{3}$$

2. The second algorithm available from Ren et al (2021), called SRGAN with visual geometry group (VGG) error, changes content loss for one based on different VGG layers. Such loss is based on the ReLU activation layers from the pre-trained 19-layer VGG-19 network. Such loss is defined below in Equation 4

$$L_{\frac{VGG}{i,j}}^{SR} = \frac{1}{W_{i,j}h_{i,j}}\sum_{x=1}^{w_{i,j}}\sum_{y=1}^{h_{i,j}}(\phi_{i,j}(I^{HR})_{x,y} - \phi_(i,j)(G_\theta(I^{LR}))_{x,y})^2 \tag{4}$$

   Where ϕ is the VGG network.
3. The third algorithm applied, the one by Huang (2020), corresponds to an Enhanced-SRGAN (ESRGAN) Wang et al (2018). Such GAN changes in the generator and the discriminator, while content loss is the same as VGG. There have been two changes in the generator: First, it eliminates all the batch normalization (BN) layers, as may be seen in Figure 6. Second, it replaces basic residual blocks with residual-in-residual dense blocks (Figure 7), by combining a multi-level residual network and dense connections between layers. On the other hand, the discriminator has been changed to a relativistic one, which estimates the probability that any given input image is real and natural. So, it attempts to predict that a real image x is relatively more realistic than a fake one, as shown in Figure 8.

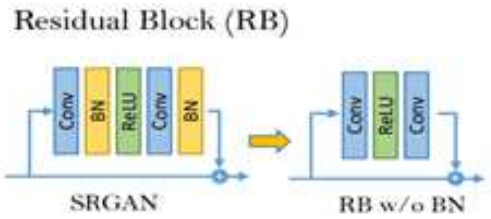

Fig. 6 Elimination of BN layers. (Wang et al, 2018)

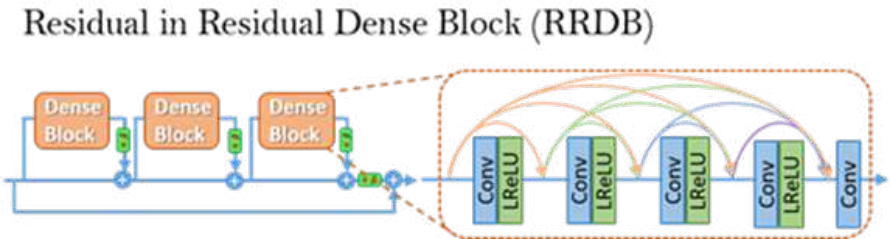

Fig. 7 Change of residual blocks for RRDB. (Wang et al, 2018)

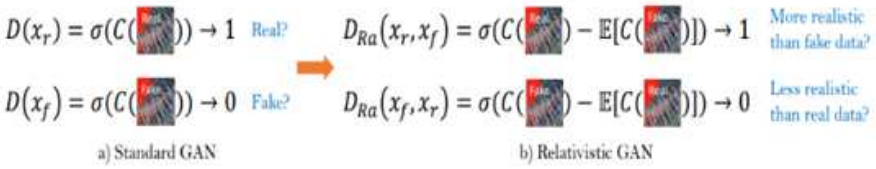

Fig. 8 Difference between standard discriminator (left) and relativistic discriminator (right). (Wang et al, 2018)

4. After having the three usual SRGAN algorithms, one of the most recent ESRGAN updates presented by Wang et al (2021) was incorporated into the experiments, showing an improvement of the ESRGAN architecture. The research of Wang et al (2021) showed considerable improvements in super-resolution with Generative Adversarial Networks. The work was based on extending the powerful ESRGAN to a practical restoration application (Real-ESRGAN) (Figure 9), which is trained on purely synthetic data. In particular, a high-order degradation modeling process is introduced to better simulate complex real-world degradations. Common ringing and overshoot artifacts are also taken into account in the synthesis process. In addition, it employs a U-Net discriminator (Figure 10) with spectral normalization to increase the capacity of the discriminator and stabilize the training dynamics.

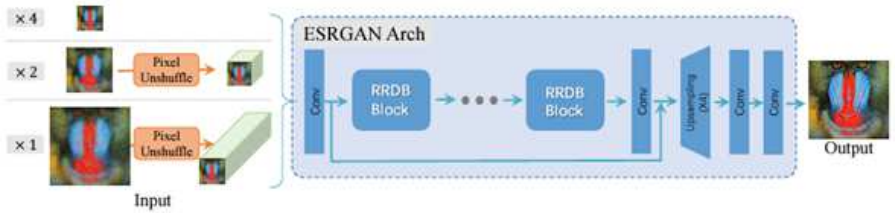

Fig. 9 Real-ESRGAN adopts the same generator network as that in ESRGAN. For the scale factor of ×2 and ×1, it first employs a pixel-unshuffle operation to reduce the spatial size and re-arrange information to the channel dimension. (Wang et al, 2021)

Fig. 10 Architecture of the U-Net discriminator with spectral normalization. (Wang et al, 2021)

Although all the implemented algorithms present differences in the content loss functions used to optimize and update the network weights, as well as in the architectures of the generators and discriminators, the training process remains the same. Regarding that process, Figure 11 shows that during the early stages of training, the generators return images with some errors associated with high content loss, which is evident in the lack of original image features in the generated image. As one progresses to new stages, the results gradually improve, to a point where the generator manages to return images very similar, in terms of color and detail, to the original HR images.

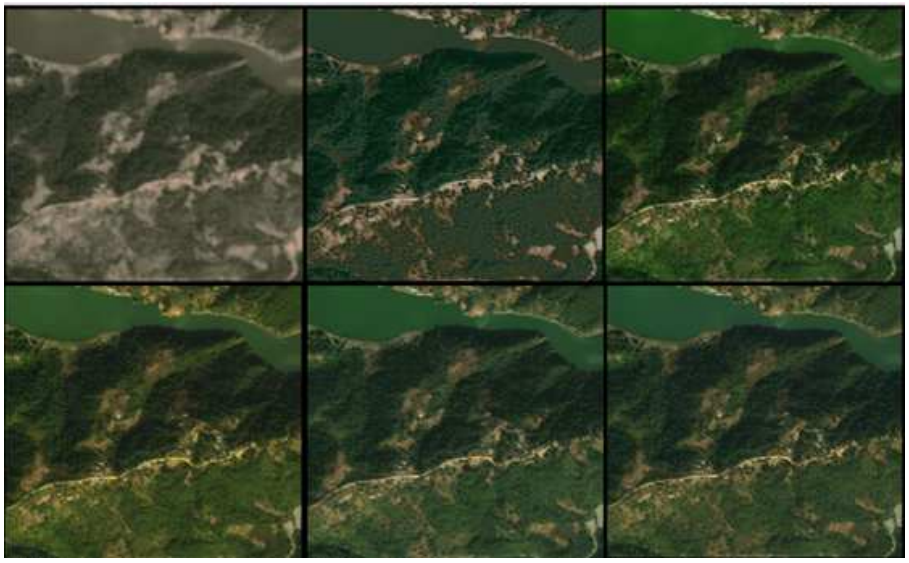

Fig. 11 Images generated as validation of the training process. First row, left in the first epochs of training, center in 10 epochs, and right in 50 epochs. The second row, left in 100 epochs, center in 200 epochs and right is the original HR image.

5.3 Methods for the assessment

Any kind of processing applied to an image may cause important information or quality loss. To measure this, assessment methods may be divided into objective and subjective. Subjective methods are based on human judgment and operate without explicit reference to their criteria. Objective methods are based on comparisons using explicit numerical criteria and include peak signal-to-noise ratio and structural similarity. Such assessment methods allow a comparison to determine which of these three networks trained with the same dataset show better performance and generate better results.

Peak signal-to-noise ratio (PSNR) is an expression used in engineering to define the ratio between the maximum possible power of a signal, and the noise affecting the quality of its representation (Umme et al, 2019). It is usually used as a quantitative measure of reconstruction quality in terms of image compression. Nevertheless, it has been broadly applied to measure algorithm performance and effectiveness, with the purpose of recovering images employing super-resolution. Such metric is measured with a logarithmic scale, taking decibel as its unit.

Based on Horé and Ziou (2010) to define PSNR, an image of reference is used, which, in the case of super-resolution, is an image that was originally HR. And a test image K, which would be the SR image generated by the network from an LR image, both with the same m x n size. PSNR is defined as follows in Equation 5

$$PSNR = 20log_{10}(\frac{MAX_f}{\sqrt{MSE}}) \tag{5}$$

Where MAX_f is the maximum value that a pixel may have in the image, and MSE is the mean square error calculated between reference I and test image K as shown in Equation 6.

$$MSE = \frac{1}{w * h} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} |I_{(i,j)} - K_{(i,j)}|^2 \tag{6}$$

For an image in RGB format, MSE is calculated as the arithmetical average of the MSE of these three colors (R: red; G: green; B: blue) as expressed in the Equation 7

$$MSE_t = MSE_R + MSE_G + MSE_B \tag{7}$$

PSNR value is higher when MSE approaches zero. Therefore, a higher PSNR value provides higher image quality, while a small value implies great numerical differences between the assessed image and its reference.

On the other hand, the structural similarity index measure (SSIM) (Horé and Ziou, 2013), is a metric of perception that quantifies the degradation of image quality caused by some forms of processing, such as data compression, or loss in data transmission. It is a full reference metric that requires two images of the same capture: an image of reference and a processed image. To apply such a metric to a super-resolution assessment, the image of reference is originally HR, and the processed image would be the super-resolution one, generated by the artificial neural network from a low-resolution image.

Instead of using traditional total error approaches, just like PSNR, SSIM is designed to model any image distortion, as a combination of three factors: correlation loss, luminance distortion and contrast distortion. SSIM is defined in Equation 8.

$$SSIM_{(f,g)} = l_{(f,g)} c_{(f,g)} s_{(f,g)} \tag{8}$$

Where the luminance component, l, is in Equation 9, the contrast component, c, is in Equation 10 and the structure component, s, is in Equation 11.

$$l_{(f,g)} = \frac{2\upsilon_f \upsilon_g + C_1}{\upsilon_f^2 + \upsilon_g^2 + C_1} \tag{9}$$

$$c_{(f,g)} = \frac{2\sigma_f \sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2} \tag{10}$$

$$c_{(f,g)} = \frac{\sigma_{fg} + C_3}{\sigma_f \sigma_g + C_3} \tag{11}$$

Being:

- f image of reference and g super-resolution image
- υ_f average of f
- υ_g average of g
- σ_f^2 variance of f
- σ_g^2 variance of g
- σ_{fg} covariance of e f y g
- $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$ two variables to stabilize the
- division with a weak denominator.
- L the dynamic range of the pixel values (usually it is $2^{(bitsporpixel)-1)}$
- $k_1 = 0,01$ y $k_2 = 0,03$ by default.
- $c_3 = \frac{C_2}{2}$

SSIM values are obtained in the [0,1] scale, where a value of 0 means there is no correlation between images, that is, both images are completely different. And 1 means that f = g, that is, both images are exactly the same.

6 Results and Discussion

Once the SRGAN models described earlier were trained to perform the super-resolution process with a scaling factor of x4, that is, they magnify the image four times, some tests were done with satellite images: 30 meters/pixel in real color, 5 meters/pixel in real color, and with infrared. The results obtained are shown in Figures 12, 13 and 14.

In all the different cases, HR images were shot so that they would be taken as a reference, they went through a resizing process to decrease their resolution. An LR image was obtained, which was processed with each of the neural networks. The images generated were compared, to determine which of the networks had a better performance: in terms of the visual aspect, SRGAN with VGG error and ESRGAN generated images with considerably enhanced resolution when compared to the LR one, where the one

generated by ESRGAN showed greater sharpness in detail. Nevertheless, although the ones generated with ESRGAN were sharper, they also showed higher noise levels. And when a quantitative measurement of results achieved with the PSNR and SSIM metrics was done, it showed that in most cases the values of the images generated by the SRGAN with VGG error were better in both metrics. This may be verified in the graphics in Figures 15 (PSNR) and 16 (SSIM), where the orange graphic corresponds to SRGAN with VGG error.

Figure 12 shows the comparison of the results obtained with the networks when a 30 meters/pixel satellite image was processed with the networks generated better results in most tests, with high PSNR and SSIM metrics values and the networks generated better results in most tests, with high PSNR and SSIM metrics values the networks generated better results in most tests, with high PSNR and SSIM metrics values a band combination for RGB to get the real color. Images with such characteristics showed that in most tests the networks generated better results, with high values in the PSNR and SSIM metrics, as may be seen at the bottom of each image. The one generated with SRGAN with VGG error reached the highest values. In the visual aspect, SRGAN with VGG error also featured a better texture and details that were more similar to the ones present in the original HR image. This indicates that such a network is doing a more accurate and faithful reconstruction of the HR image from the LR one.

For the following result comparison and assessment, introduced in Figure 13, the process was performed on an image with a 5 meter/pixel spatial resolution, and RGB band combination to get the real color. The result was that SRGAN with MSE showed a low performance for images with such characteristics, while the other two (with VGG error and ESRGAN) were capable of reconstructing the LR image and magnifying it up to an x4 scaling factor while keeping the image information. With the 5 meters/pixel images, it was observed that although SRGAN with VGG error and ESRGAN do the super-resolution properly, their performance is not the same as with the 30 meters/pixel images, where the enhancement of the details in the SR is more evident when compared to the input LR image.

The behavior with 5 meters/pixel images is probably due to the fact that the satellite images shot with such spatial resolution have higher quality details and more information, which get lost in a process of degradation when the original image is resized to obtain the LR one. And, given that the super-resolution process takes the LR image to increase its size and resolution, it cannot rebuild all the details that were present in the original HR image. This is shown by the decrease in the values obtained with the metrics, especially in SSIM.

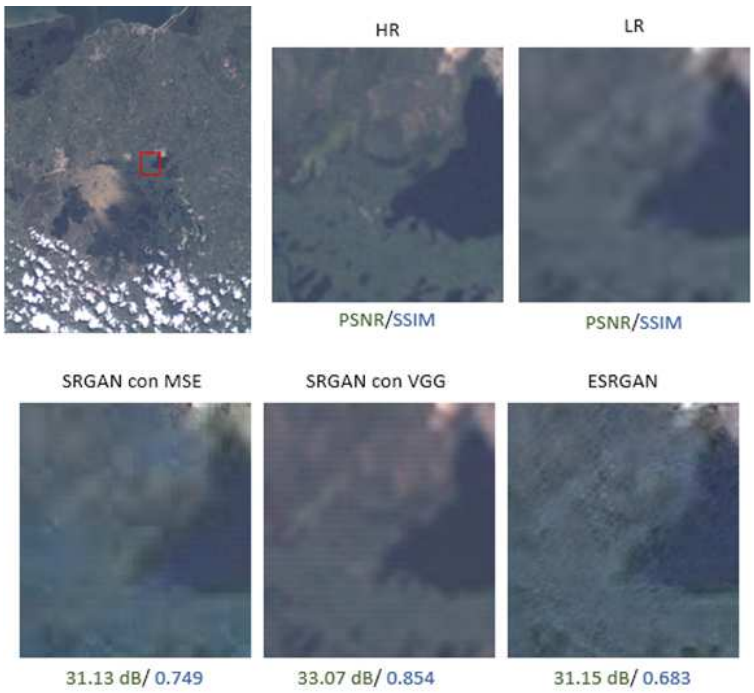

Fig. 12 Comparison of SR results with 30 meter/pixel image in real color. First row: left, the full image of FACSAT-1 obtained by CITAE in 2019; center, approach to HR image; and right, approach to LR image. Second row: left, SRGAN with MSE error result; center, SRGAN with VGG error result; and right, ESRGAN result.

Fig. 13 Comparison of SR results with 5 meters/pixel image in real color. First row: left, the full image of Crater Lake downloaded from Planet Gallery; center, approach to HR image; and right, approach to LR image. Second row: left, SRGAN with MSE error result; center, SRGAN with VGG error result; and right, ESRGAN result.

Additionally, tests were done with images including the near-infrared (NIR) band, so that an analysis could be made of the performance of networks with multispectral satellite images. Comparisons were also made of the results achieved with such tests, as may be seen in Figure 14, which shows that the results obtained with the different networks are similar to those mentioned above. SRGAN with VGG error and ESRGAN generated the best SR images. On the other hand, ESRGAN kept generating sharper images. In these cases, the brighter red zones, which represent vegetation, are better rebuilt with SRGAN with VGG error. Nevertheless, in other aspects of the image, details are very altered when compared to the HR reference. Those other details are better in the images generated by SRGAN with VGG error, and that is shown by the values obtained with PSNR and SSIM metrics. More faithful images, in terms of content, were obtained with SRGAN with VGG error.

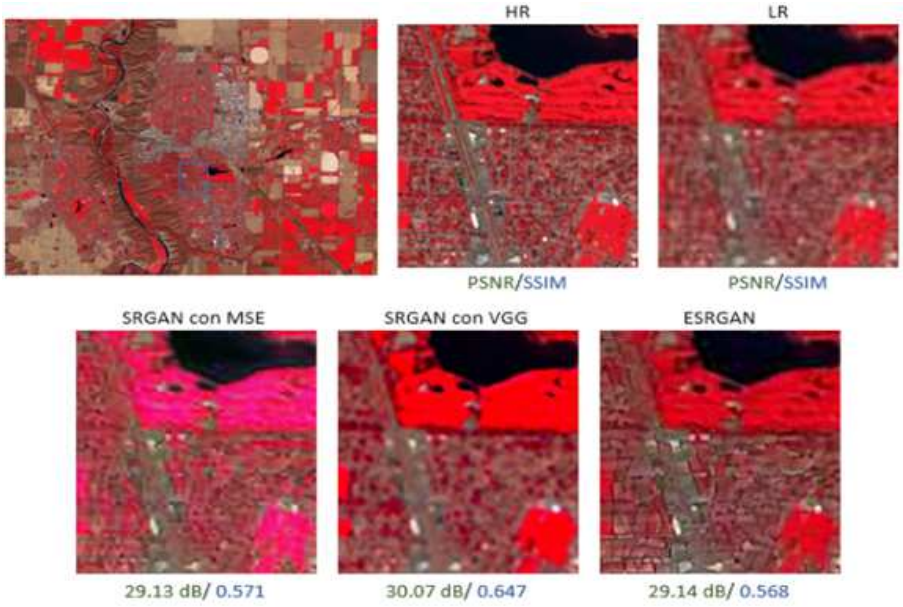

Fig. 14 Comparison of SR results with 5 meters/pixel image with infrared band (NIR). First row: left, the full image of Lethbridge - Alberta downloaded from Planet Gallery; center, approach to HR image; and right, approach to LR image. Second row: left, SRGAN with MSE error result; center, SRGAN with VGG error result; and right, ESRGAN result.

In spite of the fact that, as has been said before, ESRGAN generated sharper images, it did not obtain the best results with the metrics (see Figures 15 and 16), as these assess the noise present in the images (PSNR) and their structural similarity (SSIM), after having an HR original version of each assessed image. When the super-resolution images obtained with ESRGAN were analyzed, it could be observed that they generate details that are different from the ones present in the original image. This involves information alteration, which is harmful to satellite images that require accurate information for land surveys.

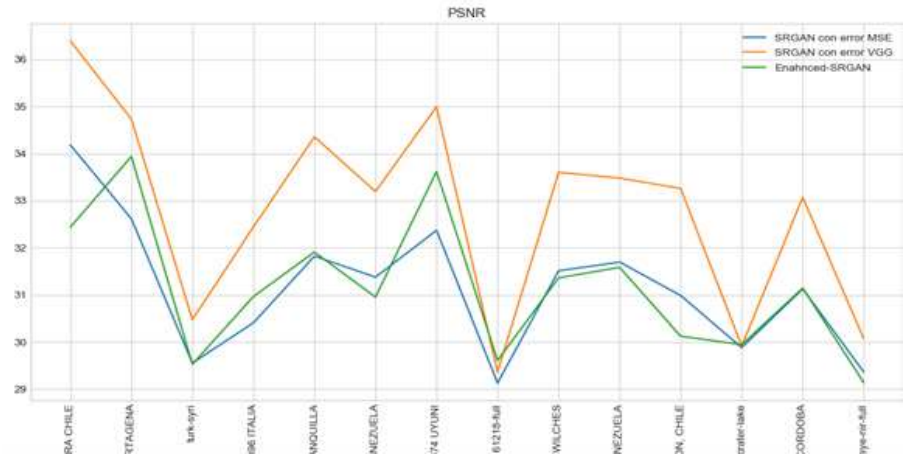

Fig. 15 Graphic of the values obtained by assessing the results with the PSNR metric. Blue: SRGAN with MSE error; orange: SRGAN with VGG error; green: ESRGAN.

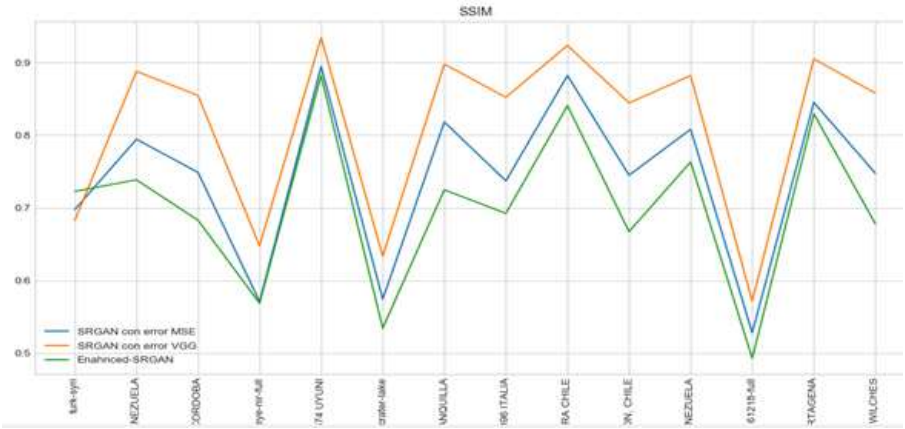

Fig. 16 Graphic of values obtained by assessing the results with SSIM metric. Blue: SRGAN with with MSE error; orange: SRGAN with VGG error; green: ESRGAN.

For the Real-ESRGAN, on this occasion, considering that in the results of the research carried out by the authors of this architecture, the model shows a robust performance, a new model was not trained from scratch, instead it was decided to use the pre-trained model (the one provided by the authors) and perform a fine-tuning with the dataset created with satellite images. Once the fine-tuning was completed, the corresponding tests were performed using FACSAT-1 images and compared with the results obtained with the SRGAN model with VGG presented above. These comparisons can be seen in Figures 17, 18 and 19, where it is visually observed how the Real-ESRGAN generated SR images with better and smoother details, making them more realistic.

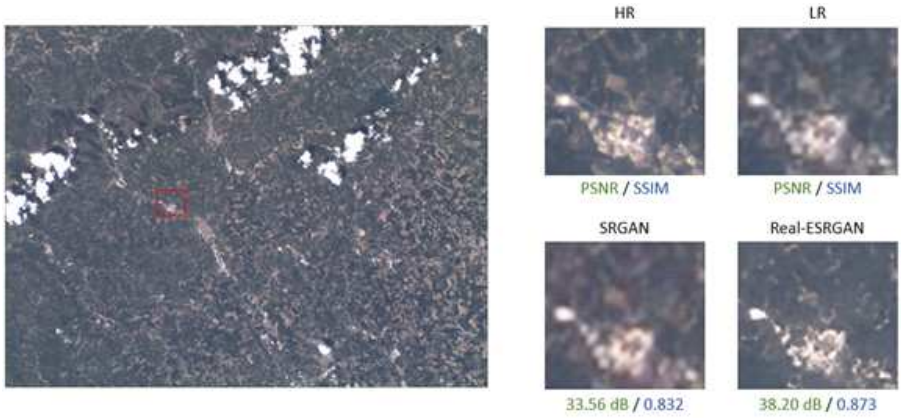

Fig. 17 On the left the complete FACSAT-1 image obtained by JEOES in 2022. On the right at the top a close-up of the HR image, followed by a close-up of the LR image, and on the bottom a close-up of the SRGAN result, followed by a close-up of the Real-ESRGAN result.

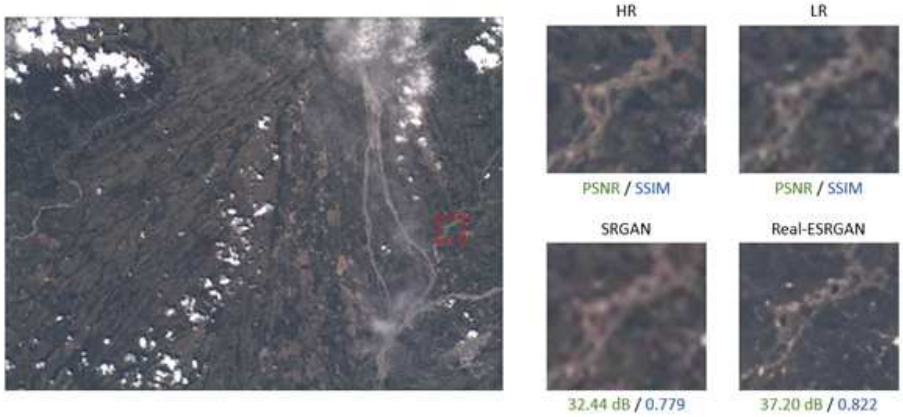

Fig. 18 On the left the complete FACSAT-1 image obtained by JEOES in 2022. On the right at the top a close-up of the HR image, followed by a close-up of the LR image, and on the bottom a close-up of the SRGAN result, followed by a close-up of the Real-ESRGAN result.

As in the previous cases, PSNR and SSIM metrics were used to evaluate the results of SRGAN and Real-ESRGAN, demonstrating that not only qualitatively the improvement of Real-ESRGAN is very significant, but it also quantitatively demonstrates that the generated SR images have lower noise and a more reliable structural similarity to an original HR image. which improves some details, but it is still a bit blurry. While the image generated by Real-ESRGAN is much sharper, allowing it to easily identify an object of interest.

As an alternative test, super-resolution was performed with both networks on an aerial image captured in a mining area, in which it is known that there is the presence of yellow machinery (excavators, backhoes, dredges, etc.). In Figure 20, it is shown that in the original image, the objects are blurred making it a little difficult to identify

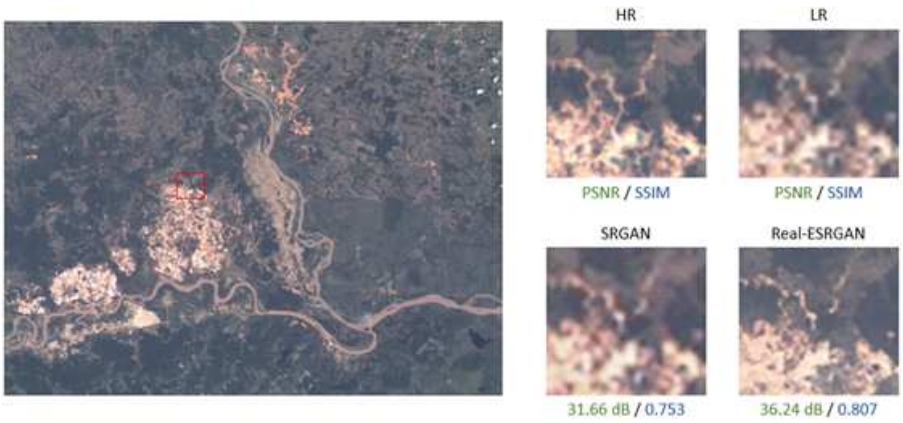

Fig. 19 On the left the complete FACSAT-1 image obtained by JEOES in 2022. On the right at the top a close-up of the HR image, followed by a close-up of the LR image, and on the bottom a close-up of the SRGAN result, followed by a close-up of the Real-ESRGAN result.

them. On the other hand, the image generated with SRGAN has a higher pixel density, which improves some details, but it is still a bit blurry. While the image generated by Real-ESRGAN is much sharper, allowing it to easily identify an object of interest

Fig. 20 On the left the complete aerial image obtained by SUINA. At the right at the top a close-up of the original image and at the bottom a close-up of the SRGAN result, followed by a close-up of the Real-ESRGAN result.

7 Conclusions

With the present project, it is possible to determine the high potential of the application of GAN algorithms for processing satellite images with satisfactory results.

It may be concluded that, from the algorithms employed in super-resolution, the most appropriate and optimum for satellite and aerial images of different spatial resolutions (30 meters/pixel, 5 meters/pixel, etc.) is Real-ESRGAN, as this network obtained the best results in PSNR and SSIM metrics, and the SR images generated do

not show much alteration of details, which, unlike ESRGAN, guarantees greater information conservation when a satellite image resolution is increased, being an important feature of such network. Similarly, SRGAN with VGG error showed good performance with 5 meters/pixel images and with the multispectral ones, however, as demonstrated in the experiments, Real-ESRGAN performs better in generating more detailed SR images and it is even possible to improve its performance as the model is refined by including more images of the required features in the dataset.

GAN models are capable of learning the characteristics and patterns of satellite images, thus establishing a landmark for future research projects aimed at implementing models which are capable of executing different specific kinds of processing on satellite images, such as radiometric, atmospheric and geometric corrections.

Author contributions P.Z. conceived and designed the methodology, data analysis and co-wrote the paper; C.A. contributed to the data collection, software implementation, experiments and co-wrote the paper; J.L. contributed with the original idea and technical design of the requirements for software implementation; J.J. contributed to analysis of results and definition of research impact on the Colombian Aerospace Force.

Funding Open access funding provided by Colombian Aerospace Force within the FAC space project. No funding was received to assist with the preparation of this manuscript.

Data availability The datasets generated during and/or analysed during the current study are either publicly available online (https://www.kaggle.com/c/widsdatathon2019) or available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Abdelmaksoud M, Nabil E, Farag I, et al (2020) A novel neural network method for face recognition with a single sample per person. IEEE Access 8:102212–102221. https://doi.org/10.1109/ACCESS.2020.2999030

Agrawal R (2023) An end-to-end introduction to generative adversarial networks(gans). URL https://www.analyticsvidhya.com/blog/2021/10/an-end-to-end-introduction-to-generative-adversarial-networksgans/

Gupta R, Sharma A, Kumar A (2020) Super-resolution using gans for medical imaging. Procedia Computer Science 173:28–35. https://doi.org/https://doi.org/10.1016/j.procs.2020.06.005, URL https://www.sciencedirect.com/science/article/

pii/S1877050920315076, international Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020

Horé A, Ziou D (2010) Image quality metrics: Psnr vs. ssim. pp 2366–2369, https://doi.org/10.1109/ICPR.2010.579

Horé A, Ziou D (2013) Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? Image Processing, IET 7:12–24. https://doi.org/10.1049/iet-ipr.2012.0489

Huang K (2020) esrgan-tf2. URL https://github.com/peteryuX/esrgan-tf2

Jiang Y, Li J (2020) Generative adversarial network for image super-resolution combining texture loss. Applied Sciences 10(5). https://doi.org/10.3390/app10051729, URL https://www.mdpi.com/2076-3417/10/5/1729

Jiang Z (2021) Srgan-pytorch. URL https://github.com/ziwei-jiang/SRGAN-PyTorch

Karim R (2018) Java deep learning projects, generative adversarial networks. URL https://www.oreilly.com/library/view/java-deep-learning/9781788997454/60579068-af4b-4bbf-83f1-e988fbe3b226.xhtml

Kingma DP, Welling M (2013) Auto-encoding variational bayes. CoRR abs/1312.6114. URL https://api.semanticscholar.org/CorpusID:216078090

Ledig C, Theis L, Huszár F, et al (2016) Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 105–114. URL https://api.semanticscholar.org/CorpusID:211227

Martínez J (2018) El facsat-1 plan piloto para el desarrollo de doctrina espacial en la fuerza aérea colombiana. URL http://www.centrodeestudiosaeronauticos.edu.co/cea/Investigacion/InvestigacionCEA/Anexos%20Memorias%20Encuentro/ANEXO%2018%20EL%20FACSAT-1.pdf

Molini AB, Valsesia D, Fracastoro G, et al (2020) Deepsum: Deep neural network for super-resolution of unregistered multitemporal images. IEEE Transactions on Geoscience and Remote Sensing 58(5):3644–3656. https://doi.org/10.1109/TGRS.2019.2959248

Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR abs/1511.06434. URL https://api.semanticscholar.org/CorpusID:11758569

Rath SR (2020) Image super-resolution using deep learning and pytorch. URL http://www.army.mil/article/70358/Nano_technology_marches_on/

Ren H, Kwon Y, Acharjee J, et al (2021) Srgan. URL https://github.com/leftthomas/SRGAN

Ruthotto L, Haber E (2021) An introduction to deep generative modeling. GAMM-Mitteilungen 44(2):e202100008. URL https://api.semanticscholar.org/CorpusID:232168940

Umme S, Morium A, Shorif UM (2019) Image quality assessment through fsim, ssim, mse and psnr—a comparative study. Journal of Computer and Communications 07:8–18. https://doi.org/10.4236/jcc.2019.73002

Van Den Oord A, Kalchbrenner N, Kavukcuoglu K (2016) Pixel recurrent neural networks. In: International conference on machine learning, PMLR, pp 1747–1756

Wang X, Yu K, Wu S, et al (2018) Esrgan: Enhanced super-resolution generative adversarial networks. In: ECCV Workshops, pp 0–0, URL https://api.semanticscholar.org/CorpusID:52154773

Wang X, Xie L, Dong C, et al (2021) Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp 1905–1914, https://doi.org/10.1109/ICCVW54120.2021.00217

Yue L, Shen H, Li J, et al (2016) Image super-resolution: The techniques, applications, and future. Signal Process 128:389–408. URL https://api.semanticscholar.org/CorpusID:13113451

Zhang T (2019) Research and improvement of single image super-resolution based on generative adversarial network. Journal of Physics: Conference Series 1237(3):032046. https://doi.org/10.1088/1742-6596/1237/3/032046, URL https://dx.doi.org/10.1088/1742-6596/1237/3/032046

Zárate P, López J, Arroyo C, et al (2023) Correction of banding errors in satellite images with generative adversarial networks (gan). IEEE Access 11:51960–51970. https://doi.org/10.1109/ACCESS.2023.3279265, URL https://ieeexplore.ieee.org/document/10131946