

Application of Artificial Intelligence Techniques to Improve Sentinel-3 Spatial Resolution

María Peña Fernández¹, Daniel García Díaz¹, and Fernando Aguilar Gómez¹

¹Instituto de Física de Cantabria

Abstract—The Sentinel-2 mission satellites provide multispectral images with 13 spectral bands at three different spatial resolutions (10, 20 and 60m Ground Sample Distance - GSD). In contrast, the Sentinel-3 mission products have 21 spectral bands at a minimum spatial resolution of 300m. Therefore, the article's objective is to combine the data of two satellite missions to improve the spatial resolution of the latter. We use a convolutional neural network (CNN), which has already been proven to improve the resolution of Sentinel-2 bands from 20 and 60m GSD to 10m, and a generative adversarial network (GAN), both of which are trained with data from different latitudes and terrains at lower resolution, i.e., from 9km to 300m, to predict the step from 300m to 10m. The results of both neural networks are compared with those of the traditional pansharpening and bicubic interpolation super-resolution algorithms. Thus, it shows that the newly proposed methods improve the previous ones both through quantitative analysis and visual comparison. In particular, the outstanding performance of the GAN used is remarkable, which manages to improve the global numerical results of traditional algorithms by around 30%.

Index Terms—Sentinel-2, Sentinel-3, super-resolution, multispectral images, neural networks, deep learning

I. INTRODUCTION

Earth observation from space has seen significant advancements in recent decades, thanks to the proliferation of multispectral satellites like MODIS, VIIRS, Worldview-3, Sentinel-2, and Landsat 8. These satellites are widely used for a broad range of applications, including climate monitoring, agriculture, natural resource management, and general Earth observation. Multispectral satellites are typically equipped with a panchromatic sensor (PAN) and a multispectral sensor (MS), which capture information in diverse electromagnetic spectrum products. The main differences between them are their spatial and spectral resolutions. For example, Sentinel-2 ([1]) offers an excellent spatial resolution (up to 10m GSD), but each image is composed of 13 spectral bands, so its spectral resolution is limited. On the contrary, Sentinel-3 ([2]) images have 21 bands providing more information at the spectral level, but their spatial resolution is much lower, at 300m GSD. Using advanced image fusion methods, as well as deep learning techniques, the goal of this work is to improve the spatial resolution of images captured by Sentinel-3, leveraging the multispectral imagery of Sentinel-2, so that the 21 bands would achieve a resolution of 10m GSD which corresponds to the best resolution that the Sentinel-2 mission can reach in some of its bands. Moreover, in order for these images to be used for scientific purposes, we must also ensure that the radiometric

information of the super-resolved (SR) spectral bands agrees with that of the originals.

In particular, we will use two different neural networks: a convolutional neural network (CNN) and a generative adversarial network (GAN) of super-resolution, which will be trained with lower resolution data, i.e., from 9km to 300m, to predict with 300m GSD images. The data will be sampled over a wide range of geographic locations to obtain models that generalize over different climatic zones and terrains. The predictions made by these models will be compared visually and by quantitative evaluation with the results of bicubic interpolation and pansharpening, traditionally used in up-sampling problems. The approach to be used around neural networks is inspired by the 2018 paper of C. Lanaras et al. (2018), which aimed to improve the spatial resolution of the 20 and 60m GSD bands of Sentinel-2 down to 10m. To our knowledge, none of these algorithms has been used to improve the spatial resolution of satellite images by a factor of 30, combining Sentinel-2 and Sentinel-3 products. This is one of the biggest challenges of this study since, as we will see, the products of the two missions are very different.

II. INPUT DATA

The *Sentinel-2 mission* provides data-feeding services for applications in the Copernicus priority areas of land surveillance, emergency management, and security. It has the MSI (MultiSpectral Instrument), which takes high spatial resolution data (10, 20, and 60m GSD¹) to monitor the Earth's surface. Each image it collects consists of 13 spectral bands with different spatial resolutions with a revisit time of five days.

The *Sentinel-3 mission* is multi-instrumental, consisting of a radar (SAR Radar ALtimeter -SRAL-), which facilitates the study of ocean topography, a radiometer (Sea and Land Surface Temperature Radiometer -SLSTR-) that measures sea and land surface temperatures, and an optical sensor (Ocean and Land Color Instrument -OLCI-). OLCI measures the radiation reflected by the Earth's surface to screen the ocean and land surface, which is used to harvest information related to biology (e.g., phenology of marine and terrestrial biomass). For this work, only products from the latter instrument are collected, which have 21 different bands.

Therefore, the amount of spectral information this instrument provides is greater than that of the Sentinel-2 mission,

¹Ground Sample Distance (GSD) is the distance between image pixel centers measured on the ground.

but it has a lower spatial resolution of 300m GSD. Due to the bands spectral distribution, Sentinel-2 bands B2, B3, and B4 correspond to Oa04, Oa06, and Oa08, respectively, from Sentinel-3. The combination of these three bands gives rise to true color images.

A. Requirements and collecting process

Aiming at combining data from both Sentinel-2 and Sentinel-3 to train Machine Learning models, robust criteria need to be adopted to ensure the initial data coherence and quality. For this reason, the following requirements are adopted to download the datasets:

- *Only Level-1C Sentinel-2 products and Level-1B in Full Resolution Sentinel-3 are collected.* To ensure that the products of both missions have the same corrections (making them comparable), we collect Level-1C Sentinel-2 products, which store top-of-atmosphere reflectances in cartographic geometry (after applying the necessary geometric corrections). As for the Sentinel-3 ones, we download Level-1B in Full Resolution (300m spatial resolution) products, which provide information about the radiance of the imaged area and additional data associated with the pixels. The first ones are available in SAFE format (the Standard Archive Format for Europe), while the second ones have SEN3 format.
- *Images should not have clouds.* This condition can only be imposed on Sentinel-2 products.
- *The collection time interval of Sentinel-2 and Sentinel-3 must be short,* for example, one day. It is a way to ensure that no noticeable meteorological changes between each pair of images appear that would cause them to mismatch.
- *Both image products should capture the same spatial region.* Both images must coincide spatially to train the networks and improve the resolution.
- *The land surface of the images to be collected does not always have the same characteristics.* I.e., there must be forests, deserts, and cities, among others, in the downloaded products. Therefore, networks can be trained with higher terrain variability, making them suitable for any subsequent region prediction.

Thus, the first task that is performed is a list of 5 different regions (see Figure 2²): coastal, desert, city, snow, and forest areas, and for is established (the number of images of each type depends on their size since the total number of pixels should be similar in all regions). These are provided to the function that downloaded them, and a period between January 2021 and July 2023 is set to find Sentinel-2 images within those areas on non-cloudy days. For each of the products from that mission, a Sentinel-3 image is searched for in the same region and with a one-day window with the reference one.

An effort is also made to collect images from different latitudes since the farther the region is from the equator, the

more distorted the projection of the products obtained is. In this way, distinct projections' conditions can be considered when training the network, which can then be encountered in reality when trying to improve the resolution of a product. To simplify the implementation and the training and testing processes of the network, we select images that do not have too many empty pixels, using a condition on the weight of the file. The below figure shows the geographical areas we include in the list mentioned above:

B. Cleansing and curation processes

Due to the differences in Sentinel-2 and Sentinel-3 data formats, downloaded data is harmonized in GeoTIFF format. As the study's main objective is to improve the resolution of the Sentinel-3 images from 300m to 10m, tif files with just the Sentinel-2 bands that have the latter resolution (bands 2, 3, 4, and 8) were generated. However, for Sentinel-3, saved tif files contain information of the original 21 bands.

In the case of the first mission mentioned, the products collect information from a much smaller geographical area than those of Sentinel-3 (to be precise, they are restricted to fields of 110x110 km²), so the coordinate system used is more specific according to the time zone of the images, known as WGS84/UTM, measured in meters. Regarding the Sentinel-3 mission, its products are projected with the EPSG (European Petroleum Survey Group) coordinate system with the identifier 4326, which corresponds to the WGS84 latitude and longitude coordinate system, already explained measured in degrees. To conveniently check the correspondence between the layers of both missions, we proceed with a coordinate transformation for the Sentinel-2 images from the WGS84/UTM system to EPSG 4326.

Once this is done, we see that some of the edges of the image show null values that were not there before. In order for the networks to be trained correctly, it is necessary to avoid as many of these values as possible, so the images were cropped. If we understand the images as multidimensional matrices, the criterion consider for cropping them on the left is that we cut by the first column (from left to center) that fulfills that the sum of null values of the middle row up or down is less than that of the next half column. The reason for proceeding in this way is that the empty values do not generally appear parallel to the edges of the image but in oblique positions. We repeat the same procedure with the right side and the columns from the right to the center. For the rows, once the images are restricted to the columns of interest, it is only necessary to search for those that do not have null values among their elements. In addition to this criterion, it is sought that the number of pixels of the height and width of the Sentinel-2 images is multiple of 30 so that the cropping for both images does not cause problems with the Sentinel-3 image. Once the coordinates of the images of the first mentioned mission that satisfy both conditions are known, we continue by cropping Sentinel-3 products as the intersection of their coordinates with those described. The resulting clippings are saved in tif format. An example can be seen in Figure 3.

The next step is to perform a *resize* of the Sentinel-3 bands, that is, to change the number of pixels that make up the images

²The images that appear in the dissertation as a color composite have an enhancement on the colors displayed, which is the default QGIS enhancement: each displayed band is first normalized based on the range between the 2nd percentile and the 98th percentile.

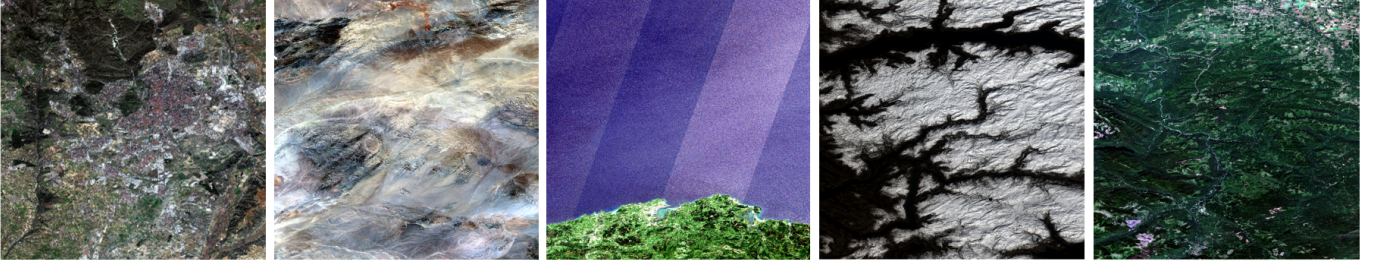


Fig. 1. Sentinel-2 true color image selection of each terrain type. From left to right: urban, desert, coastal, snowy and forested region.

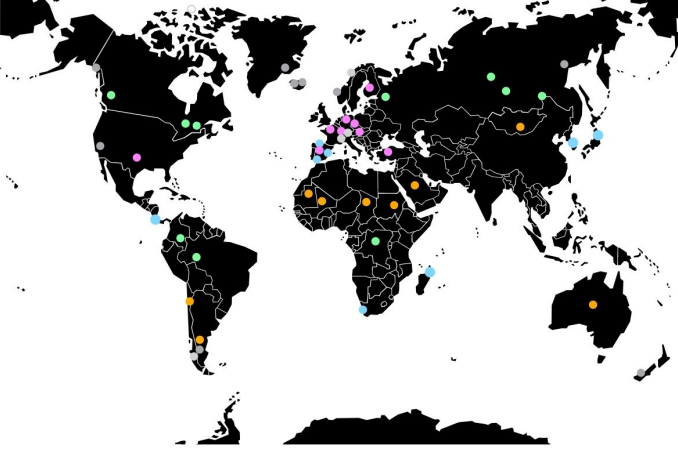


Fig. 2. Locations of the geographic zones used for training and validation datasets. Forest regions are shown in green, snowy regions in gray, coastal regions in blue, desert regions in orange, and urban regions in pink.

without affecting the spatial region they show. The reason to perform the process with those of that mission is just that it is computationally cheaper to do it with those images than with those of the other mission (despite having 21 bands versus 13, the width and height of these is a number close to 30 times smaller). Initially, it was thought that this could be done from the full uncropped images, but because these have an oblique position and the edges are not straight (see Figure 3), the projections after the *resize* were distorted. Therefore, the process is calculated from the already cropped images to avoid this. In this way, the following t factor is computed:

$$t = \frac{30 \times \text{S2PixelSize}}{\text{S3PixelSize}}, \quad (1)$$

such that the new Sentinel-3 image, which was again saved in tif format, had the next dimensions:

$$\text{widthResize} \times \text{heightResize} \times 21 = (\text{width}/t) \times (\text{height}/t) \times 21, \quad (2)$$

where width and height correspond to the spatial dimensions of the original image.

This process is necessary to avoid problems with the direct correspondence between Sentinel-2 and Sentinel-3 images, which we expect it to be 1:30, but it is not always exact because of the coordinate transformation applied to Sentinel-2 data. An example can be appreciated in Figure 4, where the images of Figure 5 appear overlapped and zoomed with some transparency. Instead of 30, slightly more than 27 pixels

per side are observed for each larger pixel. The ratio between the two sizes is not an integer, which also causes some small pixels to fall between two Sentinel-3 pixels. Also, it changes depending on how far away the region is from the Equator latitude, so it is a variable parameter.

Finally, we do a joint cropping of the transformed Sentinel-3 (already cropped) and Sentinel-2 images in EPSG 4326 coordinates by intersecting the coordinates of both missions. Figure 5 shows the result of a pair of images once the cleansing and curation processes have been completed, where the one corresponding to the Sentinel-2 mission is on the left, and the Sentinel-3 one is on the right. Thanks to this, it can be seen that both correspond to the same geographical location. However, while the left image shows the differences in structures and details with good definition, such as the maroon lines in the upper right corner, in the Sentinel-3's, they are less well resolved. This remarkable difference is due to the size ratio between the pixels of both missions.

The final resulting images are the ones we use in the SR algorithms. However, it is still necessary to take a few more steps to introduce them into neural networks. They are described in the continuing subsection.

C. Datasets creation process

Neural networks like the ones we use in this dissertation are among the supervised learning algorithms, which implies that to train the models, it is necessary to have both the inputs and the correct values they should have (known as *targets*). Ideally, the latter would be the 21 Sentinel-3 bands with a resolution of 10m. However, this information is not available, so before training the networks, it is necessary to generate a dataset that satisfies this need.

To work with the networks, it is assumed that the transfer of details from HR bands to LR bands is scale-invariant and only depends on the scale factor between resolutions, which in this case is 30. That is, it is considered that the mappings $300\text{m} \rightarrow 10\text{m}$ and $9\text{km} \rightarrow 300\text{m}$ are equivalent. Therefore, the second case is trained to apply it to the first case.

Before we start the degradation process, we normalize the original bands. This step is essential to ensure that the information of the details of Sentinel-2 bands is properly transmitted to Sentinel-3 bands and that they all have the same importance when training the network (normalization makes the pixel values of all bands have a similar scale). Thus, the normalization that homogenizes the distributions of the bands

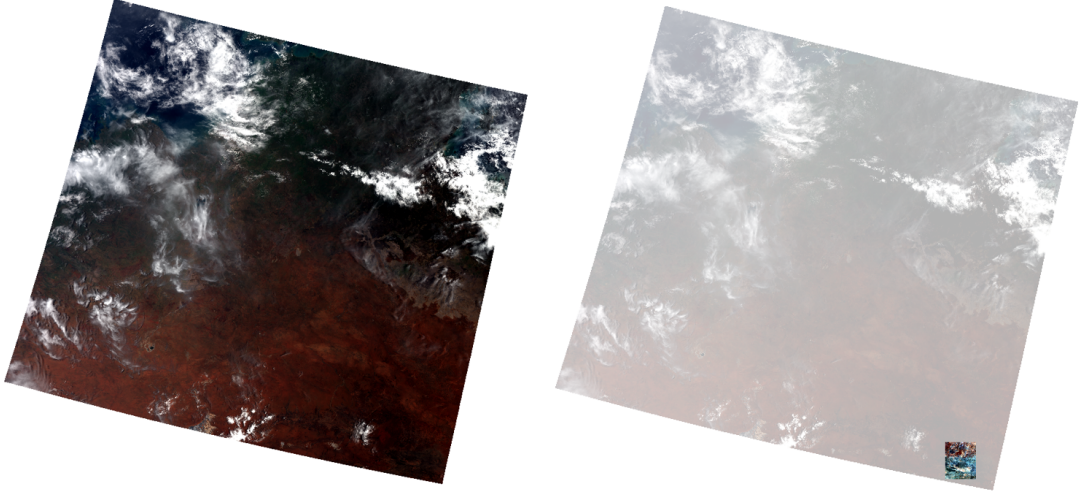


Fig. 3. Cropping example in the Sentinel-3 product corresponding to the Alice Springs Desert region (Australia).

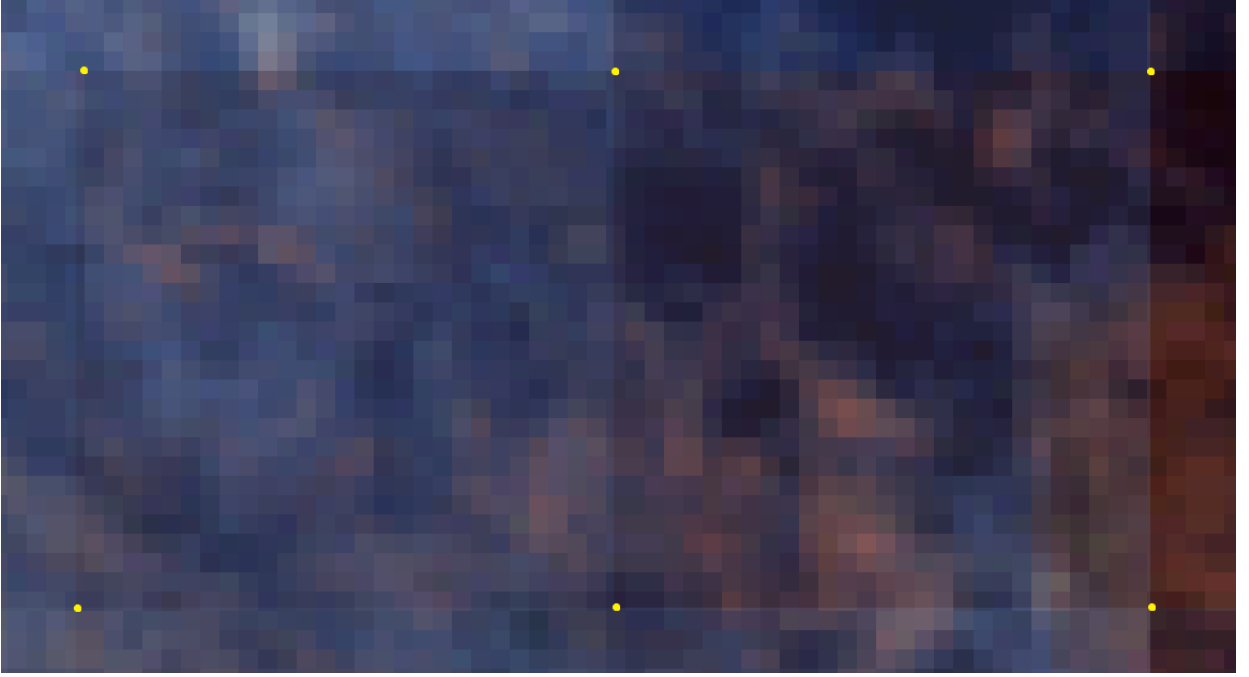


Fig. 4. Example of the difficulty encountered with respect to the ratio between the product pixels of both missions, superimposing both. Yellow dots represent the corners of the Sentinel-3 bands pixels.

to a greater extent is the so-called mean-variance, so these become mean 0 and variance 1.

Once we have this operation done, and in order to generate the dataset with which the network is trained with a scaling factor $s = 30$, the bands of both satellites are first smeared with a Gaussian filter with standard deviation $\sigma = 1/s$. This filter allows emulating the modulation transfer function, which is essential to train with $9\text{km} \rightarrow 300\text{m}$ mappings and use it to super-resolve 300m GSD bands to 10m ones. Then, the average of the pixels of each band in $s \times s$ windows is calculated. Finally, downsampling is performed, reducing the size of the bands by the established scaling factor. The result of this operation is a set of HR images of 300m (those from

the Sentinel-2 mission) and complementary LR images of 9km (from Sentinel-3) generated with the downsampling by a factor of 30.

In addition to the above, it is necessary to insert the images into the network in patches to optimize the computational time. Moreover, due to the cropping procedure considered, not all images have the same size, so it would not be possible to introduce them completely into the network. For simplicity, we take these patches to be squared, with 120 pixels on each side in the HR images (4 pixels in the corresponding LR ones), by sweeping the entire image area.

Within the images we have to train, there are 44 that constitute the training dataset itself and five composing the

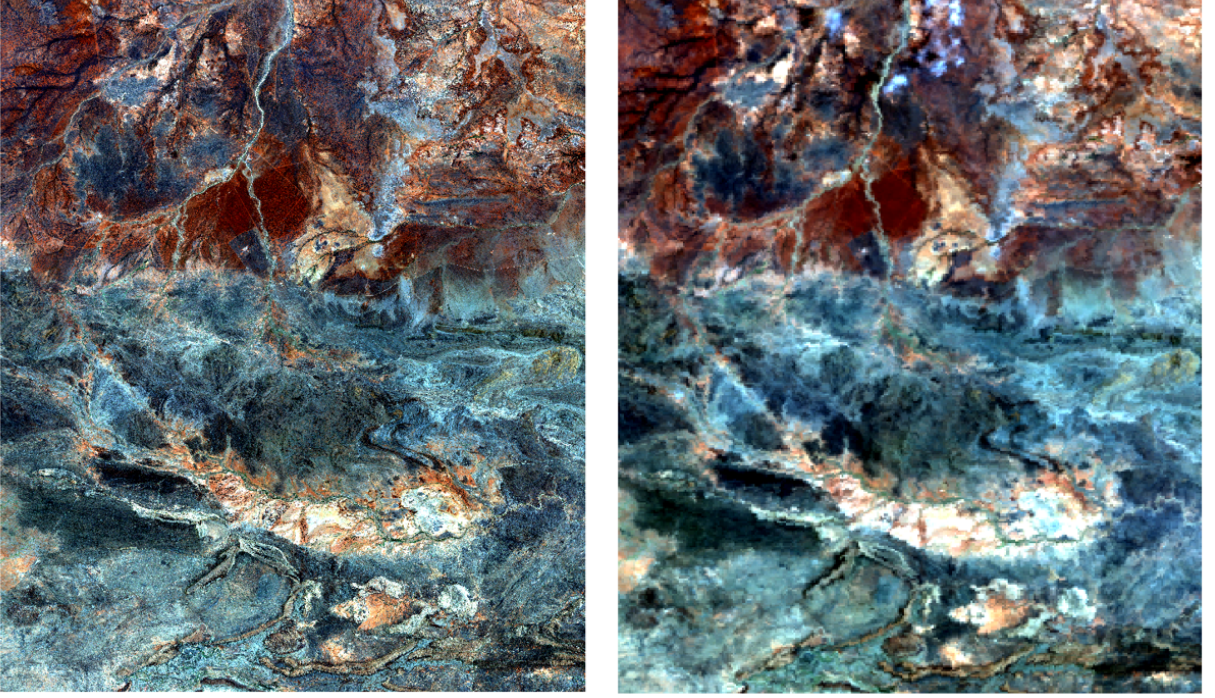


Fig. 5. Comparison between Sentinel-2 (left) and Sentinel-3 (right) images after cropping, representing the RGB bands.

validation dataset. The latter is only used in the CNN to adjust the learning rate value when training this model. In each dataset, there is a similar number of patches of each of the five terrains explained in Subsection II-A (around 440 patches for training and 60 for validation per terrain). In addition, five more image pairs are collected to test the models and compare them with traditional algorithms that do not have a training phase, both at original scale ($300\text{m} \rightarrow 10\text{m}$) and lower scale ($9\text{km} \rightarrow 300\text{m}$).

III. MACHINE LEARNING MODELS

To solve the problem of super-resolution of Sentinel-3 images, we use a deep learning approach. In particular, we utilize two neural networks: a Convolutional Neural Network (CNN) and a Generative Adversarial Network (GAN) that take images of both missions as input (multiresolution input) and enhance the resolution of the bands corresponding to Sentinel-3. They operate as regression machines in such a way that there is a transfer of the details from the high-resolution bands, so that they learn correlations across the spectrum. Thus, the transformation sought can be defined as:

$$\varphi_{30\times}: \mathbb{R}^{W \times H \times 4} \times \mathbb{R}^{W/30 \times H/30 \times 21} \rightarrow \mathbb{R}^{W \times H \times 21}, \quad (3)$$

$$(x_{S2}, x_{S3}) \rightarrow y_{S3}$$

where $x_{S2} \in \mathbb{R}^{W \times H \times 4}$ is the set of HR bands (those corresponding to Sentinel-2), $x_{S3} \in \mathbb{R}^{W/30 \times H/30 \times 21}$ the Sentinel-3 bands and $y_{S3} \in \mathbb{R}^{W \times H \times 21}$ the SR bands. W and H are the width and height, respectively, of the HR bands.

A. Super Resolution Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are powerful tools and constitute the state of the art in Machine Learning. In this case, we use the [3] network, whose authors built based on the Enhanced Deep Super Resolution (EDSR) networks to improve the resolution of the Sentinel-2 bands from 20 and 60m down to 10m. Here, we present some of the most relevant details of the network's architecture.

Before inputting the bands into the network, it is necessary to equalize the size of the bands introduced from both missions, so an upsampling with bilinear interpolation of the Sentinel-3 bands x_{S3} is performed. Thus, $\hat{x}_{S3} \in \mathbb{R}^{W \times H \times 21}$ is obtained, which is the input of the network together with x_{S2} , and on which a skip connection is made at the same time, such that in the final step it is added to the result of the network to give rise to the output. In addition, the network consists of a number of convolutional layers (with 256 3×3 kernels each, as in [3]), nonlinear activation layers ReLU (Rectified Linear Unit) and skip connections, some of them grouped in so-called ResBlocks (we use 32 of them, which is the number of those that had the most successful model of the ones studied in the article). The general architecture of the network takes the form of Figure 6.a.:

The detail of the layer architecture within the ResBlocks is shown in Figure 6.b.. These are defined as a series of layers that operate on some input data to transform and add them to the input data to give rise to an output through a skip connection. They consist of two convolutional layers separated by a ReLU layer. Finally, it has a residual scaling layer, which multiplies the inputs by a constant $\lambda = 0.1$.

Regarding the training details, we use the specifications

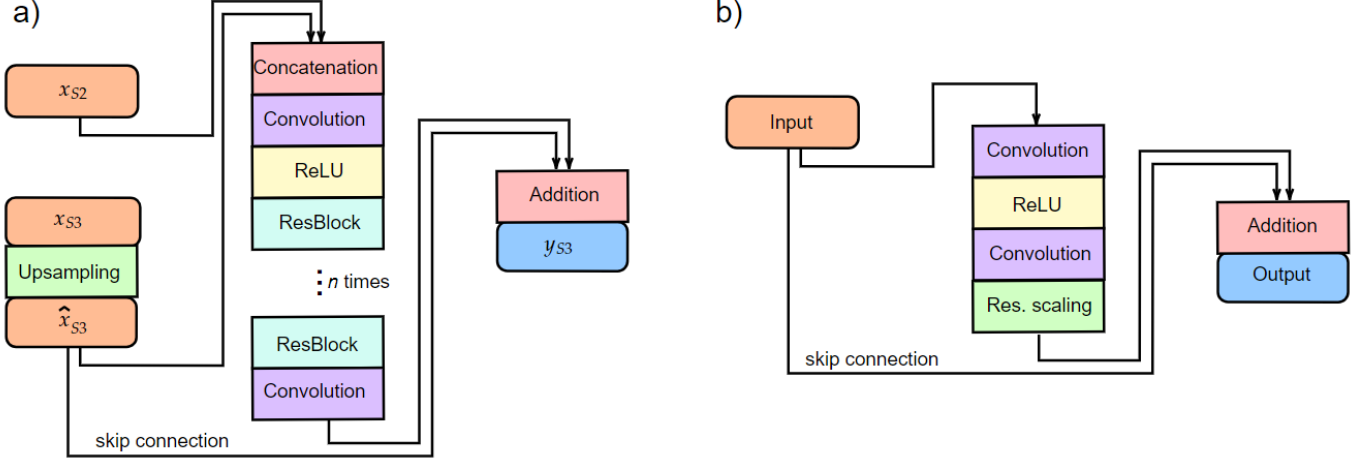


Fig. 6. a) Complete architecture of the CNN. b) Detail of a residual block of the CNN.

provided in the reference article. Initially, the model weights are set using the *HeUniform* method (see [4]). The loss function we use is the Mean Absolute Error (MAE), such that, for each image, given the Sentinel-3 input HR corresponding image \mathbf{x} , with x_i , $i = 1, \dots, N$ being the value of its pixels, and the prediction \mathbf{y} , with elements y_i , returned by the model in a particular iteration, it can be defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - x_i|, \quad (4)$$

where N is the number of pixels in the images. The reason for using this function is that it offers higher robustness against outliers compared to others, such as the Mean Square Error function (MSE), and the satellite images used have some pixels with extreme values of radiance.

Once the loss function is calculated, the weights are updated using mini-batch gradient descent, where the batch size is set to 4. I.e., each gradient step consists of a forward step to calculate the current loss in 4 multispectral images, followed by a backward propagation of the error signal through the network. We utilize the Adam Optimizer with Nesterov momentum (see [5], with $\beta_1 = 0.9, \beta_2 = 0.999$), which aims to converge fast to the minimum by using algorithmic characteristics of other methods (RMSProp [6] and gradient descent with momentum [7]). The learning rate is set initially to 10^{-4} and is reduced by a factor of 0.5 if no improvement in the validation loss is seen for five epochs.

B. Super Resolution Generative Adversarial Networks

Unlike the SRCNN explained in the previous section, Super Resolution Generative Adversarial Networks (SRGANs) apply a neural network in combination with an adversary network to produce higher-resolution images. Each of them is called generator and discriminator, respectively.

On the one hand, the generator's function is to estimate from an HR image, the corresponding HR image, i.e., the network's output is an SR image. The architecture of this part of the GAN is the same as shown in Figure 6, so the

inputs we have to introduce are the same as before. The loss function selected for the generator is also the MAE. The reason for keeping the same architecture is to observe the effect of adding a discriminator on the quality of the super-resolved images. On the other hand, the discriminator is responsible for distinguishing whether the input images are SR or HR images (i.e., whether they have been generated artificially or not, respectively), so it acts as a classifier. For this, it must separately receive as input both the SR image and the Sentinel-3 image that we introduce as input but with high resolution. Thus, during training, the generator becomes progressively better at improving the super-resolution of images, such that they look like real HR images, while the discriminator gets better at differentiating HR from SR ones.

Figure 7 shows the architecture of the GAN we use, detailing the structure of the discriminator, which is based on the one designed in [8] article. As can be seen, it consists of two convolutional layers. They both have 5×5 kernels and a number of them of 64 for the first layer and 128 for the second.

After each of the convolutional layers is an activation layer of the Leaky ReLU (Leaky Rectified Linear Unit) type, which is commonly used when training GANs (see [9] for an example). It is based on the ReLU, with the difference that negative values z are not transformed into 0, but into the original multiplied by a factor $\beta > 0$ (we take $\beta = 0.3$), such that:

$$v = \max(z, \beta z). \quad (5)$$

After each activation layer, we have a *dropout* regularization layer, which randomly sets the input units to 0 with a rate frequency f at each step during the training time, which helps prevent overfitting (i.e., overtraining the network with particular data). For our GAN, we take this frequency as $f = 0.3$, an intermediate probability, not too low to have minimal effect nor too high to result in under-learning by the network.

As we have said, the discriminator can be understood as a classifier, so we use the *flatten* operation to pass from the convolutional part of the network to the classifier part itself. It reshapes the tensor it receives to have a shape equal to

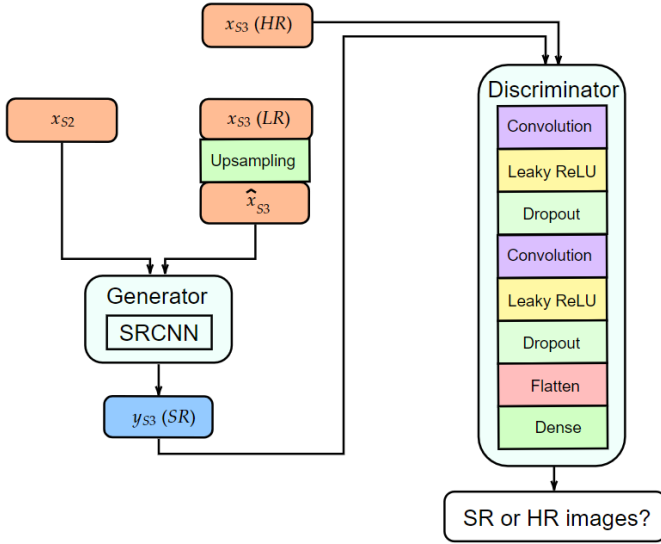


Fig. 7. GAN architecture with the detail of the discriminator.

the number of elements contained in the tensor. The result is finally introduced into a dense layer with just one unit. It is worth mentioning that the output of this last layer is not activated, but it is when calculating the loss function that the sigmoid function, $\sigma(z) := (1 - e^{-z})^{-1}$, is applied. It transforms the values entered z to a scale (0,1), where high values tend asymptotically to 1, and low values tend asymptotically to 0.

As for the loss function of the discriminator, since this is a classification problem, we use Binary Cross Entropy (BCE), also called *log loss*. Thus, for each image, taking $p(\mathbf{z})$ as the output of the discriminator to input \mathbf{z} once it has passed through the sigmoid function, and \hat{z} the label (1 for HR images and 0 for SR ones), we have:

$$\text{BCE} = -(\hat{z}\log(p(\mathbf{z})) + (1 - \hat{z})\log(1 - p(\mathbf{z}))). \quad (6)$$

To obtain the discriminator error, it is necessary to calculate the BCE after introducing into the discriminator the SR images created by the generator, BCE_{SR} , and also when introducing the HR Sentinel-3 bands, BCE_{HR} . The error is, therefore, the sum of both quantities.

As for the optimizer, we use Adam with Nesterov moment ($\beta_1 = 0.9$, $\beta_2 = 0.999$) in both the generator and the discriminator. We take a batch size of 2 for computational memory reasons. The learning rate α , which sets the pace at which the algorithms update the weights, is 10^{-4} in both cases, and is constant (to optimize the training time, we did not use a validation dataset to adjust this parameter). Moreover, it is worth mentioning that the discriminator weights are initially set using the *Xavier* method, as in [8].

IV. EXPERIMENTAL RESULTS

Once the data have been preprocessed, they can be introduced into the algorithms explained in previously to super-resolve Sentinel-3 bands. We now proceed to present some of the potential results provided by the different techniques when using the test dataset and the corresponding discussion.

For this purpose, we make a distinction between predictions obtained from low-scale input images (the 9km GSD of Sentinel-3 and the 300m GSD of Sentinel-2), for which we have the targets, and those obtained from original-scale inputs, which are the main objective of the study.

V. EVALUATION AT LOWER SCALE

As we have the actual values that the predictions should take (i.e., the original Sentinel-3 images at 300m), we can do both quantitative and qualitative analysis. The first one is the best way to check if the radiometric information is maintained, while a visual comparison allows us to establish if there is spatial coherence. Before that, the following subsection explains the different metrics we use for the quantitative evaluation.

A. Metrics to evaluate the results

The main evaluation metric of our quantitative comparison is the MAE, estimated independently by each spectral band and globally, which we have already defined in Equation (4), as it is the one we used as a loss function when training the neural networks. In this analysis, as in the reference article, both predictions and targets are introduced without normalization. Thus, the reported MAE values are within the range of radiance values of the original images. As a guideline, it is worth mentioning that within the test dataset, we have an average minimum radiance value of 0.074, maximum of 0.255, mean of 0.148, and standard deviation of $0.028 \cdot 10^{-3} \text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$.

However, although not normalizing does not affect the intra-band comparison between the MAE of different techniques, it does when the results of each metric are evaluated between distinct bands. It is because both mean radiances and standard deviations vary notoriously between bands, as we see in the next section. In particular, a greater standard deviation implies that the radiance values are spread over a wide range, so the predictions may tend to err to a greater extent, leading to a higher MAE. As a way to reduce this effect, we also use the Relative Absolute Error (RAE) metric, which can be expressed for each image as

$$\text{RAE} = \frac{\sum_{i=1}^N |y_i - x_i|}{\sum_{i=1}^N |x_i - \bar{x}|}, \quad (7)$$

where \mathbf{x} is the Sentinel-3 input HR corresponding image and x_i , with $i = 1, \dots, N$ the value of its pixels \mathbf{y} the prediction returned by the model and y_i , with $i = 1, \dots, N$ its elements, \bar{x} the mean of \mathbf{x} and N the number of pixels they have [10]. RAE is expressed in terms of a ratio (it is unitless), whose value can range from zero to one. The closer it is to zero, the better the model is. This metric shows how the mean residual error relates to the mean deviation of the ground truth image from its mean³.

We also compute the Signal-to-Reconstruction Error Ratio (SRE) as an additional error metric. It measures the error

³RAE can also be normalized by dividing by the ground-truth image [11].

	SRGAN	SRCNN	Pansh.	Bicubic
MAE	0.0126	0.0127	0.0178	0.0176
RAE	0.5299	0.5427	0.7842	0.7546
SRE	23.98	23.84	22.11	22.32
SAM	2.30	2.94	3.30	3.04

TABLE I

AGGREGATE METRIC RESULTS FOR ALL TEST IMAGES AND ALL BANDS.
BEST RESULTS ARE REPRESENTED IN BOLD.

relative to the power of the signal, and for each image, it can be computed as follows:

$$\text{SRE} = 10 \log_{10} \frac{\bar{\mathbf{x}}^2}{\|\mathbf{y} - \mathbf{x}\|^2 / N}. \quad (8)$$

It is measured in decibels (dB), and the best model (according to this metric) is the one that provides the highest SRE. It is used to evaluate errors relative to the mean image intensity, making SRE values more comparable between images and bands of varying brightness. Finally, we also calculate the Spectral Angle Mapper (SAM), which measures the angular deviation between real and estimated spectral signatures as:

$$\text{SAM} = \cos^{-1} \left(\frac{\sum_{j=1}^{21} \mathbf{x}'_j \mathbf{y}'_j}{\sqrt{\sum_{j=1}^{21} \mathbf{x}'_j{}^2} \sqrt{\sum_{j=1}^{21} \mathbf{y}'_j{}^2}} \right), \quad (9)$$

where \mathbf{x}'_j , and \mathbf{y}'_j , with $j = 1, \dots, 21$ are the 21 bands that compose the ground truth Sentinel-3 image, and the prediction, respectively [12]. The values of SAM are given in degrees, and the closer to zero, the better the model is. It complements the previous metrics as it measures the fidelity with which the relative spectral distribution of a pixel is reconstructed, ignoring the absolute luminosity.

Note that SAM can only be computed with the whole image, while the other three metrics can be calculated both globally and band-by-band.

B. Quantitative evaluation

First, the following table shows the average results over all five test images and all bands:

Given the results shown in Table I, we first observe that all MAE scores are considerably lower than the test dataset standard deviation, reaching a 55% difference with the GAN's score. Comparing between techniques, the traditional pansharpening and bicubic interpolation ones perform similarly, with the latter slightly better in all error metrics. Furthermore, SRCNN reduces MAE and RAE by approximately 28% of the value of traditional techniques. The SRE metric also confirms an improvement in performance for this technique of about 1.7 dB and 1.5 dB over pansharpening and bicubic interpolation, respectively. Moreover, SAM is slightly lower than that of both traditional methods, which seems to indicate that the overall relative spectral fidelity of the images is better in SRCNN. Even better results are obtained if SRGAN is considered. If we compare them with SRCNN, we have a 0.8% MAE, and 2.4% RAE reduction. The rest of the metrics also reinforce the idea that this neural network is the best for super-resolving Sentinel-3 images of those we considered in the study. Specifically, the SRE improves by 0.14 dB, and the

SAM is almost 22% lower than the other network. Therefore, the overall results in Table I seem to confirm this gulf in performance of the considered neural networks compared to traditional techniques (with the exception of SAM in SRCNN and bicubic interpolation), and in particular, GAN outperforms CNN with all metrics.

For a more accurate analysis, the above metrics have also been calculated for each band separately, again averaging the results of all the images in the test dataset. The obtained values are shown in Tables II-IV. They are also shown in Figure 8 to facilitate the discussion of these values.

The first thing we can see from the values obtained in the tables above is the advantage obtained across all bands with the networks compared to traditional techniques. We can also see that in the first bands, the performance is similar in the case of SRGAN and SRCNN, with the latter standing out slightly. However, the SRGAN stands out remarkably against the convolutional network from band 12 onwards.

Concerning the MAE metric, we can see an initial error growth as the number of bands increases, maintaining the same trend in all techniques. We can also see a greater distance between errors of the different methods when reaching bands with associated wavelengths of around 750-850 nm. For example, in band 16, the MAE for the pansharpening technique is 49% higher than in SRGAN. The most striking feature of the results is the short values obtained for bands 13 and 20 compared to adjacent bands in all techniques.

With the RAE metric, we see that the value between bands is much more stable compared to that of the MAE. We observe that the relative error remains at about 0.52 for the GAN and is slightly higher, around 0.55, for the CNN. As for the traditional methods, it is observed that these values are higher in all bands, the difference with the Artificial Intelligence algorithms being about 0.2. While bicubic interpolation maintains stability in the RAE for all bands, this is not the case for pansharpening. As can be appreciated in Figure 8, there are some very noticeable peaks in both bands 13 and 20, which exceed the RAE of the adjacent bands by more than 0.1. According to what is stated in the referenced article, the pansharpening methods are influenced by the panchromatic band (the high-resolution Sentinel-2 band) that is taken as a reference to infer spatial details. Thus, the performance in all bands may not be the same if we fix the same panchromatic band to reconstruct the whole Sentinel-3 image, as also mentioned in the data analysis of the article. We have always taken B4 band as the reference, so we can appreciate how the bands associated with the visible range, closer to the wavelength range of the panchromatic band, RAE takes the lowest values. In fact, in the bands of that range, the pansharpening technique performs better in terms of RAE compared to bicubic interpolation. However, as we move away from the visible spectrum, the RAE increases, surpassing that of the other traditional technique studied. It also seems that using it for bands 13 and 20 is particularly contraindicated.

In terms of SRE, the worst scores are again those of the traditional techniques, which up to band 12 report a performance that is barely distinguishable between them. As for the trend of the methods, they all follow a similar behavior

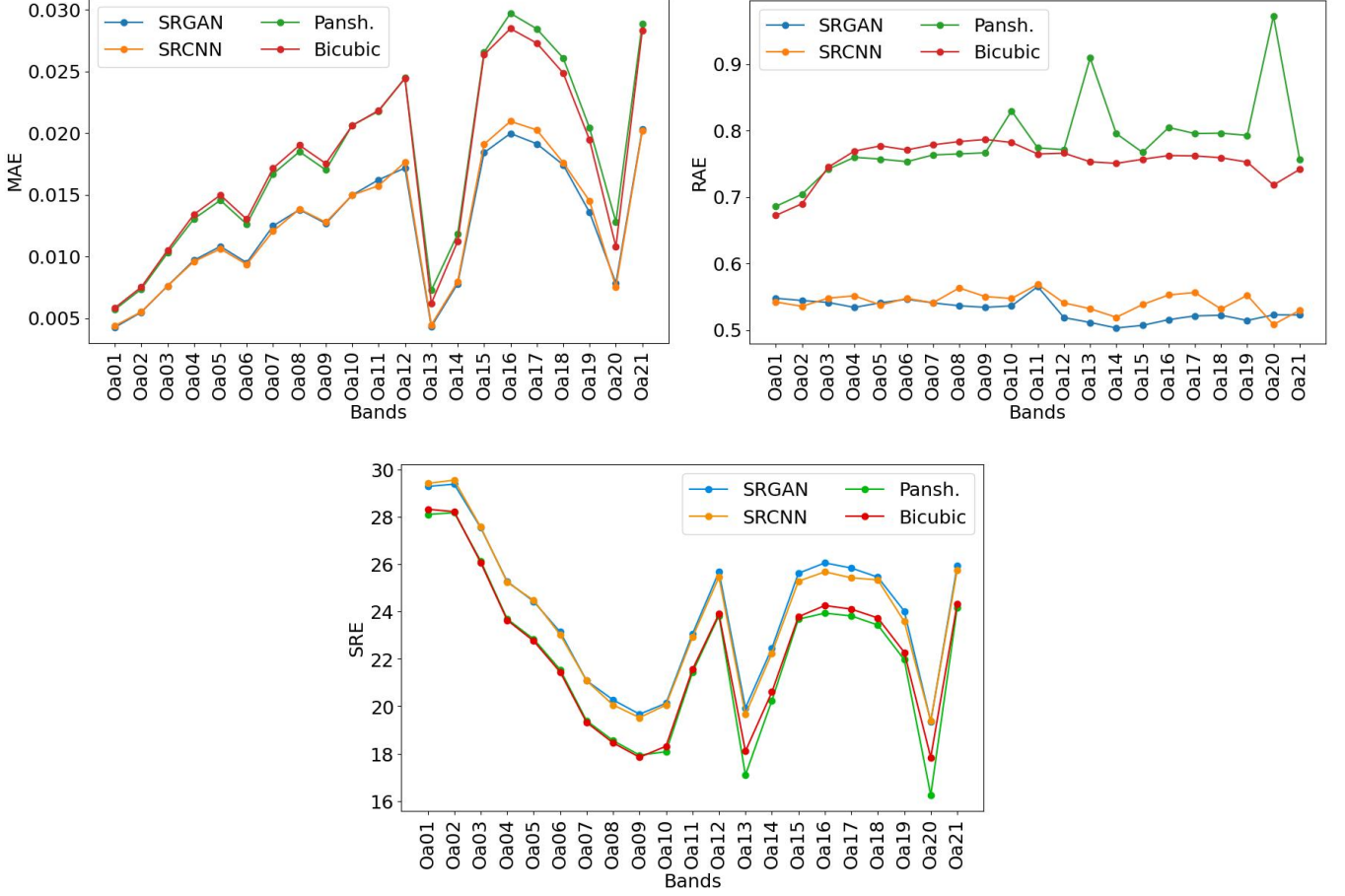


Fig. 8. Per-band error values with MAE, RAE and SRE metrics.

Band	SRGAN	SRCNN	Pansh.	Bicubic	Band	SRGAN	SRCNN	Pansh.	Bicubic
Oa01	0.0043	0.0044	0.0057	0.0058	Oa12	0.0172	0.0176	0.0245	0.0244
Oa02	0.0055	0.0055	0.0074	0.0075	Oa13	0.0043	0.0045	0.0073	0.0062
Oa03	0.0076	0.0076	0.0103	0.0105	Oa14	0.0078	0.0080	0.0118	0.0112
Oa04	0.0097	0.0096	0.0130	0.0134	Oa15	0.0184	0.0191	0.0266	0.0264
Oa05	0.0108	0.0106	0.0146	0.0150	Oa16	0.0200	0.0210	0.0297	0.0285
Oa06	0.0095	0.0094	0.0126	0.0167	Oa17	0.0191	0.0203	0.0285	0.0273
Oa07	0.0125	0.0121	0.0167	0.0172	Oa18	0.0174	0.0176	0.0261	0.0249
Oa08	0.0138	0.0138	0.0185	0.0190	Oa19	0.0136	0.0145	0.0204	0.0195
Oa09	0.0127	0.0128	0.0170	0.0175	Oa20	0.0078	0.0075	0.0128	0.0108
Oa10	0.0150	0.0150	0.0206	0.0206	Oa21	0.0203	0.0202	0.0289	0.0283
Oa11	0.0162	0.0157	0.0218	0.0218					

TABLE II

PER-BAND AVERAGE VALUES OF MAE OVER ALL TEST IMAGES. BEST RESULTS ARE REPRESENTED IN BOLD.

for all bands. As was the case with MAE, the best value of the metric is obtained for the first band, and a worsening of the score is observed until band 9. Thereafter, the value of MAE and SRE grow in roughly analogous fashion. That is, for those bands associated with wavelengths above 700 nm, the closer the values of the radiances are to the original ones, the worse the error relative to the signal power.

In Figure 9, we compare a reconstructed image to ground truth for the test image corresponding to a coastal zone (par-

ticularly attractive because of the change of terrain between the sea and the land), subtracting the pixel values from one image to the ones of the other. That is, the absolute radiance differences between the two are depicted. Yellow color shows high differences, while dark blue color represents zero errors. We display the result of seven bands among the 21 that make up the Sentinel-3 images. As we saw in the MAE, the minor radiance differences are obtained in bands 1 and 13, the darkest squares for all methods. As for bands 4, 7, and 10,

Band	SRGAN	SRCNN	Pansh.	Bicubic	Band	SRGAN	SRCNN	Pansh.	Bicubic
Oa01	0.5478	0.5422	0.6862	0.6726	Oa12	0.5188	0.5409	0.7716	0.7663
Oa02	0.5445	0.5358	0.7044	0.6902	Oa13	0.5114	0.5321	0.9096	0.7534
Oa03	0.5416	0.5481	0.7422	0.7451	Oa14	0.5031	0.5191	0.7957	0.7509
Oa04	0.5341	0.5515	0.7600	0.7694	Oa15	0.5070	0.5384	0.7674	0.7570
Oa05	0.5411	0.5377	0.7573	0.7773	Oa16	0.5157	0.5529	0.8049	0.7627
Oa06	0.5464	0.5477	0.7534	0.7712	Oa17	0.5213	0.5565	0.7958	0.7622
Oa07	0.5409	0.5410	0.7636	0.7788	Oa18	0.5223	0.5317	0.7964	0.7594
Oa08	0.5365	0.5634	0.7652	0.7837	Oa19	0.5145	0.5523	0.7930	0.7528
Oa09	0.5343	0.5504	0.7669	0.7868	Oa20	0.5229	0.5083	0.9726	0.7183
Oa10	0.5364	0.5476	0.8299	0.7824	Oa21	0.5226	0.5296	0.7571	0.7418
Oa11	0.5655	0.5688	0.7742	0.7651					

TABLE III
PER-BAND AVERAGE VALUES OF RAE OVER ALL TEST IMAGES. BEST RESULTS ARE REPRESENTED IN BOLD.

Band	SRGAN	SRCNN	Pansh.	Bicubic	Band	SRGAN	SRCNN	Pansh.	Bicubic
Oa01	29.28	29.41	28.10	28.32	Oa12	25.67	25.45	23.83	23.89
Oa02	29.37	29.54	28.17	28.21	Oa13	19.92	19.65	17.09	18.10
Oa03	27.54	27.56	26.14	26.07	Oa14	22.45	22.24	20.24	20.60
Oa04	25.27	25.23	23.68	23.63	Oa15	25.61	25.28	23.67	23.78
Oa05	24.42	24.47	22.83	22.76	Oa16	26.05	25.67	23.93	24.25
Oa06	23.13	23.02	21.55	21.45	Oa17	25.83	25.42	23.81	24.10
Oa07	21.08	21.08	19.39	19.32	Oa18	25.45	25.33	23.43	23.73
Oa08	20.27	20.05	18.55	18.45	Oa19	24.02	23.58	21.96	22.27
Oa09	19.66	19.52	17.93	17.85	Oa20	19.36	19.37	16.24	17.84
Oa10	20.13	20.05	18.08	18.31	Oa21	25.92	25.76	24.17	24.32
Oa11	23.05	22.91	21.43	21.55					

TABLE IV
PER-BAND AVERAGE VALUES OF SRE OVER ALL TEST IMAGES. BEST RESULTS ARE REPRESENTED IN BOLD.

we see that the relative errors decrease in the case of neural networks compared to traditional methods. The same happens, although with higher errors, in the case of bands 16 and 19. In particular, the largest difference (represented by the most intense yellow) is found in the pansharpening technique for band 16. It coincides with the peak we perceived in Figure 8.

We can also see that both traditional baselines exhibit errors along high-contrast edges. It is particularly noticeable the absolute difference in radiance along the coastline, resembling a high-pass filtering. It implies that the methods have blurred edges or exaggerated the contrast. Presumably (we check it later in the qualitative evaluation), they both act formerly, but pansharpening also includes some noisy details. For their part, the neural networks used only exhibit traces of this behavior in the higher bands. Specifically, SRGAN shows hardly any contrast between the sea and land areas, showing a substantial improvement in that region compared to its competitor SRCNN.

Finally, we can see that marine areas are resolved worse in the case of networks than in traditional techniques. It is because both pansharpening and bicubic interpolation infer the pixel values of the upsampled image by blurring the initial pixels (the first also adds details). However, neural networks predict through the behavior they have learned from the training dataset. Thus, of all the available images, only a fifth of them has a portion of the sea, so its performance in

those areas is reasonable to be worse than on land.

Figure 10 is exposed to introduce more into the differences between the diverse methods considered from the point of view of the radiometric information they provide. It shows histograms of the radiance differences between bands 4, 10, and 16 (one of those associated with RGB, one associated with approximately 680 nm, and one corresponding to near-infrared, respectively, as a representation of the whole set of bands) of the location represented in Figure 9. Thus, the histogram of a perfect model should look like a single bar centered at 0 of height, the number of pixels in the band. As the performance of the technique deteriorates, it should take on a Gaussian shape centered at 0, decreasing in height and increasing in width the worse it gets. We also display in Table V the standard deviations of each technique of the distributions represented.

With this figure, we can see that the Gaussians centered at 0, corresponding to the SRGAN, are the ones with the highest heights. As for its width, both in band 4 and band 10, SRCNN has the lowest standard deviation values but differs very little from those of SRGAN. In the case of band 16, it does obtain the lowest value, differing to a greater extent with SRCNN. All this is consistent with what was mentioned in the error analysis at the beginning of the section. It is also worth noting the difference between the scales of the histograms, which is best reflected in the standard deviation of the corresponding table.

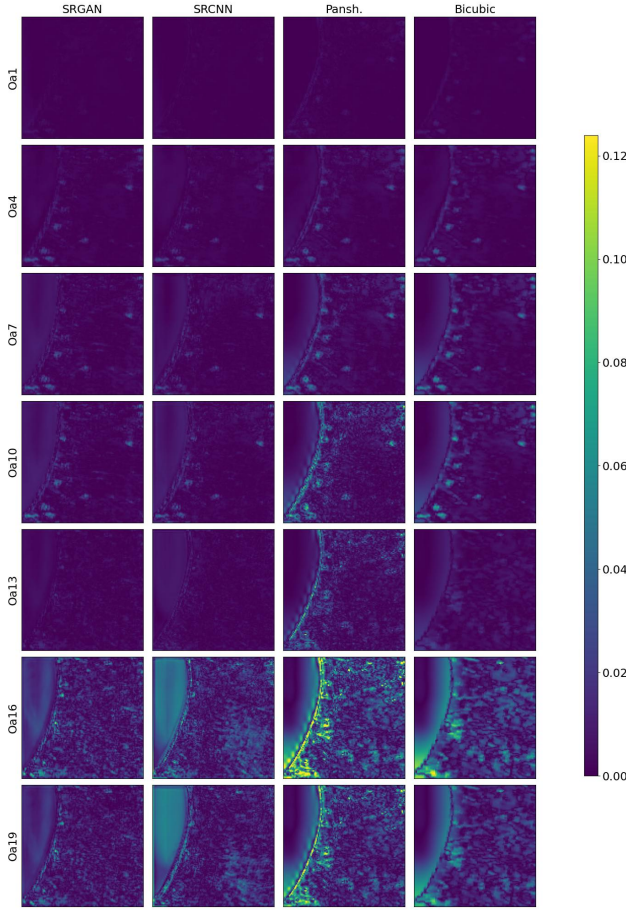


Fig. 9. Absolute radiance differences between ground truth and super-resolved image at 300m GSD measured in $10^{-3}\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$ units. The image depicts the coast city of Setubal, Portugal, near Lisbon.

	SRGAN	SRCNN	Pansh.	Bicubic
St. deviation Oa04	0.0030	0.0028	0.0042	0.0042
St. deviation Oa10	0.0064	0.0058	0.0139	0.0096
St. deviation Oa16	0.0172	0.0210	0.0364	0.0264

TABLE V

STANDARD DEVIATIONS (IN RADIANCE UNITS $10^{-3}\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$) OF THE DISTRIBUTIONS REPRESENTED IN FIGURE 10.

In all models, the lowest value of this magnitude is obtained for band 4, while the highest is in band 16. For example, in SRGAN, the deviation is reduced by 82%, and in SRCNN, it decreases by almost 87%. It is, again, in agreement with the analysis we performed of the MAE in the first part of the quantitative evaluation and with the figure of the absolute radiance differences.

As we have said, we should observe a Gaussian shape in making the difference of the upsamples with the target in each test image for all bands. It is the case of four of the ones we have, but there is one in which this fact does not occur, corresponding to a snow region. We represent the distributions of the differences of bands 4, 10, and 16 with the target (see Figure 11 with the associated standard deviations in Table VI).

By looking at Figure 11, we see that networks still have the largest heights around 0 and the smallest standard deviations. Looking also at Table VI, we see that the difference in standard

	SRGAN	SRCNN	Pansh.	Bicubic
St. deviation Oa04	0.0441	0.0434	0.0597	0.0619
St. deviation Oa10	0.0614	0.0603	0.0805	0.0835
St. deviation Oa16	0.0673	0.0665	0.0894	0.0925

TABLE VI

STANDARD DEVIATIONS (IN RADIANCE UNITS, $10^{-3}\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$) OF THE DISTRIBUTIONS REPRESENTED IN FIGURE 11.

deviations between bands is not as noticeable as before. For example, for SRGAN, band 4 reduces this value by 35% with respect to band 16, while that of 10 and 16 is very similar.

The reason for this different behavior is the low contrast in the input image we have from Sentinel-3 (pixel size is too big to detect most of the contrast differences), which all of them take as a reference to make the super-resolution. The fact that the neural network models have lower standard deviations than pansharpening and bicubic interpolation is due to the details that it can infer from the Sentinel-2 image that is also provided as input.

Therefore, with these latest quantitative analyses, we have also tested the effectiveness of Artificial Intelligence models against traditional techniques to super-resolve satellite images by a factor of 30. In the next section, we visually compare the results of all methods to see if, from a spatial point of view, neural networks also outperform.

C. Qualitative evaluation

Figure 12 shows an area corresponding to a forest zone for different color compositions. Since the visualization is limited to three bands at a time, we select them three by three to view them. The first column shows the true color composition (bands 4, 6, and 8), while the second has bands 7, 13, and 16, and the last column has bands 19, 1, and 10 composition. In the first row, the input of Sentinel-3 of 9km is represented, and in the second, the original Sentinel-3 300m image and the rest of the rows correspond to the different upsamples made.

In all cases, the bands reconstructed by the neural models are clearly sharper and bring out additional details compared both to the respective input bands and the SR images made with pansharpening and bicubic interpolation. If we compare the reconstructions of SRGAN and SRCNN with each other against that of the target, we observe that the details are transmitted correctly from the Sentinel-2 image of 300m GSD in both. However, in the case of SRCNN, the color compositions are not so faithful to reality. As can be seen in the true color tile, garnet regions appear at the bottom that are not in the original. It can also be observed in the yellow upper zone in the composition of bands 7, 13, and 16, although this is also slightly appreciated in the case of SRGAN. It is also remarkable the white dots present at the bottom of the images reconstructed by Artificial Intelligence in front of the target. These are because the original Sentinel-2 image from which the details are inferred had small clouds missing when the corresponding Sentinel-3 product was captured.

With respect to the traditional techniques, it is clear that neither of them is functional for super-resolving images by a factor of 30. In the case of bicubic interpolation, we see that it only acts by blurring the pixels composing the input image.

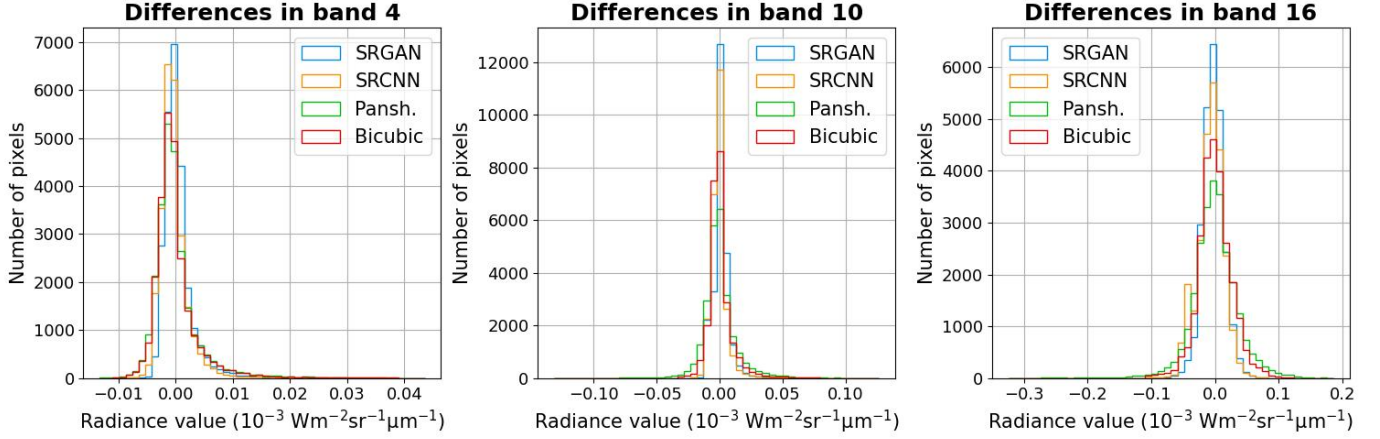


Fig. 10. Histograms of the distributions of the radiance values produced by subtracting the target bands 4, 10 and 16 of Setubal image from the bands produced by the four methods considered.

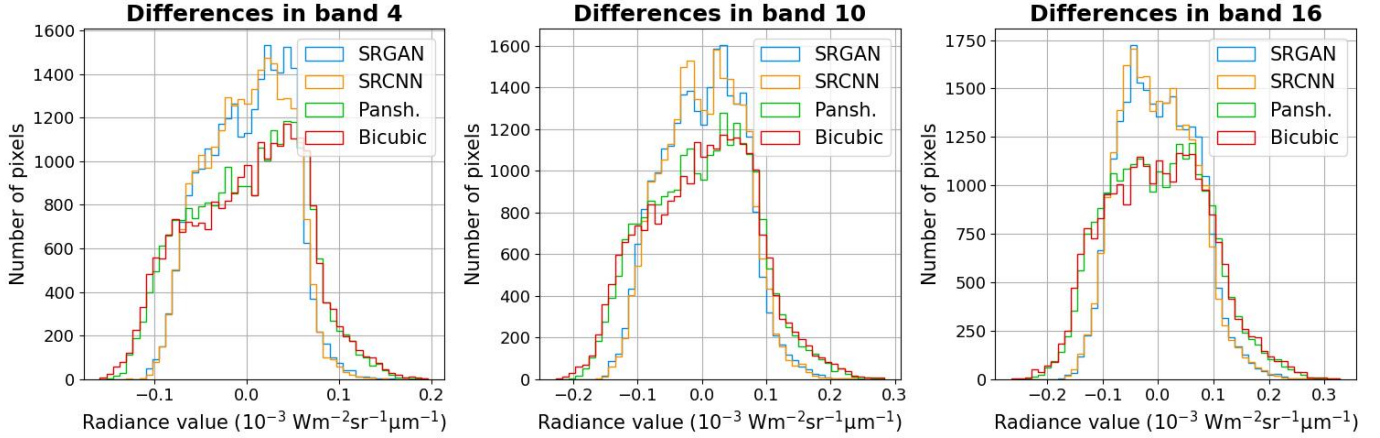


Fig. 11. Histograms of the distribution of the radiance values produced by subtracting the target bands 4, 10 and 16 of Khantaiskoe Ozero image from the bands produced by the four methods considered.

Thus, it is not possible to generate new details that did not appear in the 9km GSD one. A similar effect can be observed in the case of pansharpening, where the base colors of the image are the same as those seen in bicubic. It is because these values are taken with a similar operation from the low-resolution Sentinel-3 image. As for the details picked up from the Sentinel-2 image in the case of pansharpening, these are shown as lighter-colored grainy noise in the image, so its effect is not as expected.

Thus, from a spatial coherence point of view, the effectiveness of Artificial Intelligence techniques in color composition has been demonstrated, with SRGAN standing out from SRCNN. With this result, which also reinforces what arose from the quantitative evaluation, traditional upsampling methods can be discarded to improve the resolution of Sentinel-3 satellite images by a factor of 30.

VI. EVALUATION AT THE ORIGINAL SCALE

Now, to verify that the Machine Learning algorithms can be applied to the original scale of the data from both missions, we super-resolve the same test image as before but

providing the original images to the methods without synthetic downsampling. As previously mentioned, we have no way of obtaining ground truth data for a quantitative comparison that would allow us to perform a spectral coherence analysis, so we move directly to the qualitative evaluation. We plot in Figure 13 the same configuration as in the previous figure, but in the original scale. As the spatial resolutions are 30 times lower than before, the difference between SR images is not as noticeable as before, so we zoom in on the upper left corner of the image, which depicts a narrow part of a river.

First, we see the smearing produced by the interpolation technique and the additional noise caused by the details' transfer from the pansharpening panchromatic band. Concerning the neural networks, we do not have the target image now to be able to make a comparison between the colors it predicts. However, their performance can be inferred from the input image.

We see that for all compositions, the colors of the SRGAN image appear closer to those shown by the LR image pixels than those of SRCNN. In particular, for the latter model, a more greenish and less realistic tone is observed in the true

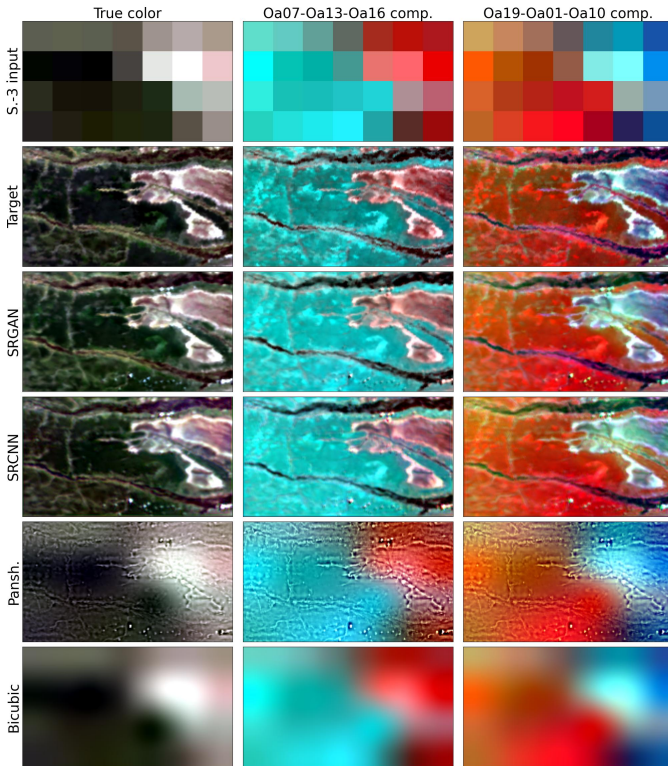


Fig. 12. Results of the different techniques on a low-scale Sentinel-3 image from a forest region of Angola called Bundas. From left to right: True scene RGB (Oa04, Oa06, Oa08), band composition Oa07, Oa13 and Oa16 as RGB and band composition Oa19, Oa01 and Oa10.

color. The same occurs in the composition of bands 7, 13, and 16, where the blue color stands out, with less contrast to the rest of the shades and a violet tone that is not observed in the input one. There is also a lack of contrast between the colors in the last composition considered.

Figures 14 and 15 show the true color complete (and a zoom) reconstruction of the neural networks for two of the images that compose the test dataset, together with the low-resolution Sentinel-3 image and the high-resolution Sentinel-2 one. This way, we can also see if the details' transfer has been carried out correctly. We do not show the ones upsampled by bicubic interpolation and pansharpening because of their poor performance.

As for the image of the Italian city, we see that the one reconstructed by GAN has colors closer to those of the Sentinel-3 input than the CNN one. For example, the green area of the super-resolved by the latter method has a more striking tone compared to the GAN or the Sentinel-3 input. However, we must remember that they collect data with different instruments, so it is possible that there are variations between the radiances they measure. Looking at the zooms, we see something similar to what we explained above with the SRCNN image, since the different details that are rightly inferred from the Sentinel-2 image are colored with little contrast, making everything acquire a greenish tone. Thus, the structural features appear colored and defined with better quality with SRGAN.

It is also noticeable the difference in tone in the image of

the snow region since the one from the convolutional network appears slightly pinker than the two input images and the GAN image. In contrast, the SRGAN SR image is essentially the same as the one we originally got from Sentinel-2, a fact that can also be appreciated in the zoomed images.

VII. DISCUSSION

A. About the results obtained

It is worth mentioning the contrast between the spikes that can be seen in the MAE plot and the small variation in the RAE values when both measure the difference between pixels. The emergence of the peaks in the first case is presumed to be a result of the ranges and standard deviations of the bands. In particular, to give some examples, the average standard deviation of those of the test images is 0.041 for band 12, 0.010 for band 13, 0.019 for band 14, and $0.043 \cdot 10^{-3} \text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$ for band 15. In order to make a better comparison between the different bands, we use the RAE metric, which eliminates the effect of the disparity of dispersions. With its corresponding graph we see that no better results are obtained in the bands associated with the visible range, even though this is the range of Sentinel-2 bands taken as a reference. It could be due to the fact that the normalization in all bands at the training process makes the model all bands have a similar performance.

Thus, although the performance of the models from a band-to-band spectral coherence point of view seems to give different results depending on the band being analyzed, it is clear that the best method for super-resolving Sentinel-3 images is SRGAN, followed by SRCNN. Traditional techniques always have the worst scores, and while the bands associated with the visible range are best upsampled with pansharpening, as we move away from the visible spectrum, its performance is always worse than that of bicubic interpolation for all metrics.

This is also in line with what was obtained in the qualitative evaluation, both in the smaller scale and the original one. It is clear the improvement of the details in the SR images of the networks compared to the input of Sentinel-3, so we have been able to verify the effectiveness of the training of a model in the improvement of 9km GSD at 300m to upsample images from 300m to 10m. It is particularly satisfactory in the case of the SRGAN, being able to appreciate colors more similar to those of the original Sentinel-2 and Sentinel-3 images.

B. SRCNN versus SRGAN

The behavior of networks versus traditional methods resembles that observed in article [3], which has also enabled us to verify that the convolutional network used for improving the resolution in steps of 2 and 5 can also be used for a step of 30. In this case, we have a CNN with a number of layers greater than 100 (considered a very deep network) that performs well for the objective proposed in this article. Despite having a large number of parameters (more than 37 million), it has been verified through the error curves of the training and validation datasets that no overfitting has occurred during the training process.

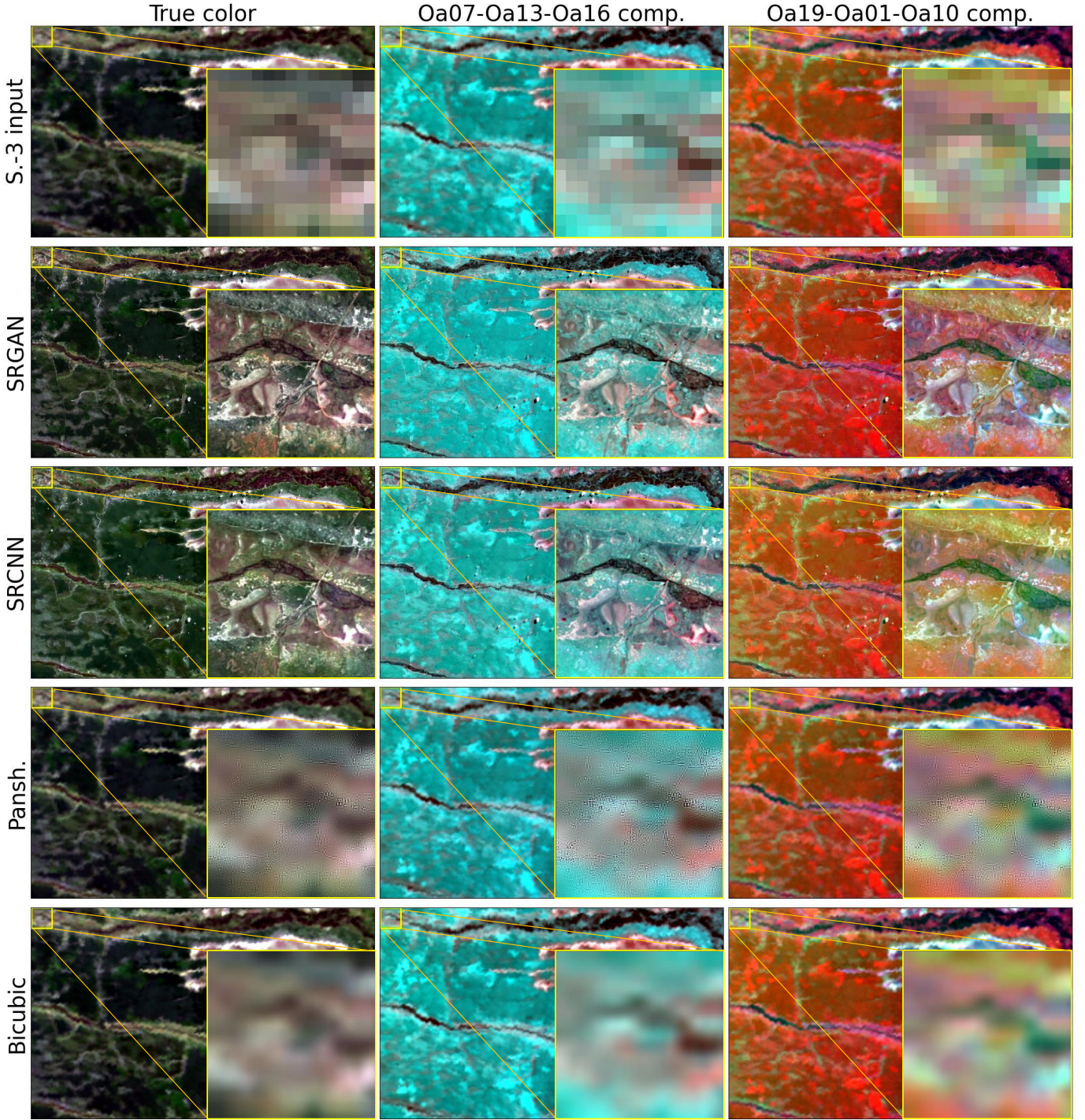


Fig. 13. Results of the different techniques on the original-scale Bundas Sentinel-3 image, zoomed in an area of the image. From left to right: True scene RGB (Oa04, Oa06, Oa08), band composition Oa07, Oa13 and Oa16 as RGB and band composition Oa19, Oa01 and Oa10.

However, unlike what happened in the GAN, in this process the aim was not to make the composition of the super-resolved bands appear to be real, but to keep the difference with the target as small as possible. It is why, although the overall results are slightly better in the GAN with the other metrics, in the case of the SAM, which considers the fidelity of the reconstruction of all the bands at the same time, there is a great improvement with respect to the CNN.

From the results obtained, it can be seen that the modification of the original CNN by adding a discriminator that classifies whether the reconstructions look artificial improves the results, not only quantitatively but also qualitatively. Moreover, this discriminator has slightly more than 350 thousand parameters, so new network's complexity is not so different from that of the convolutional network.

Moreover, we have fixed the number of layers of the

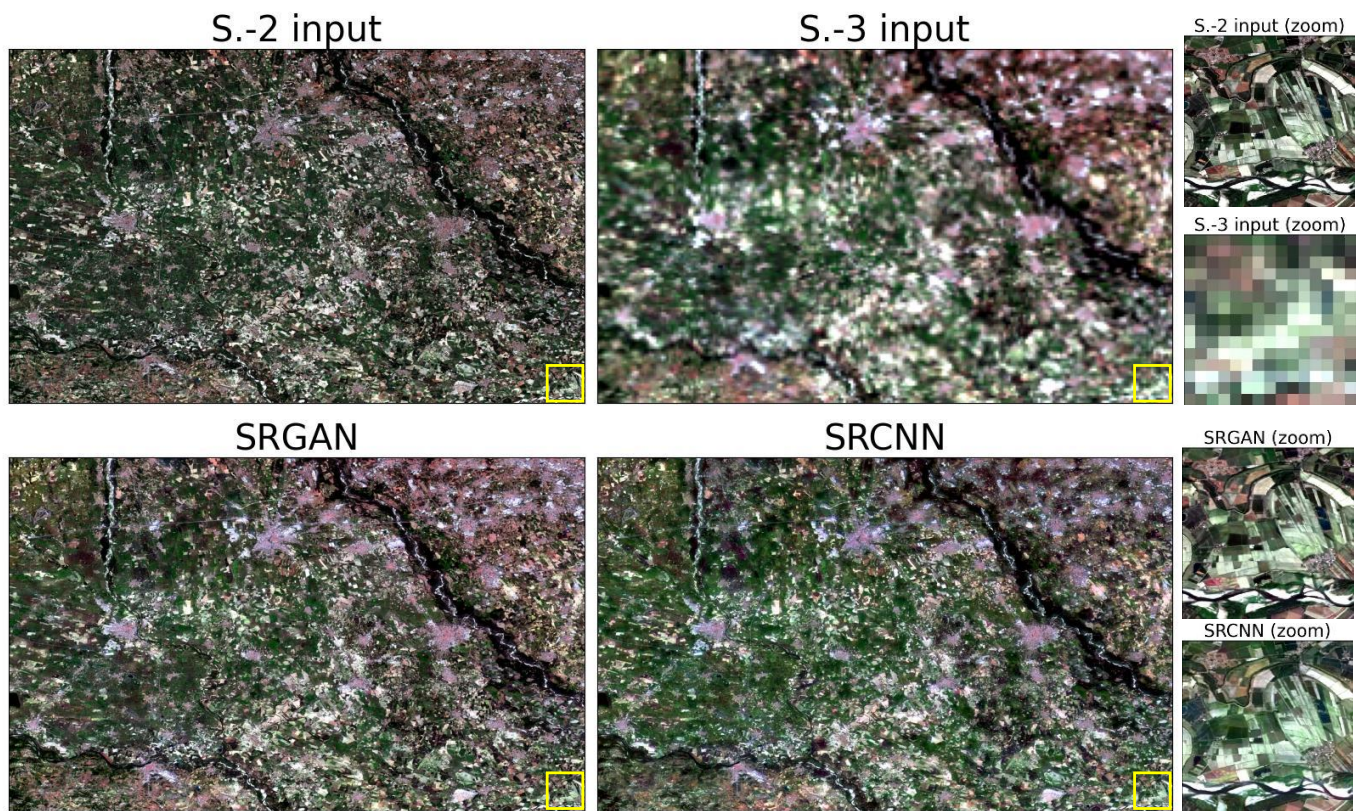


Fig. 14. True color comparison between inputs and neural networks' results on original-scale Sentinel-3 image of the city of Novara (Italy).

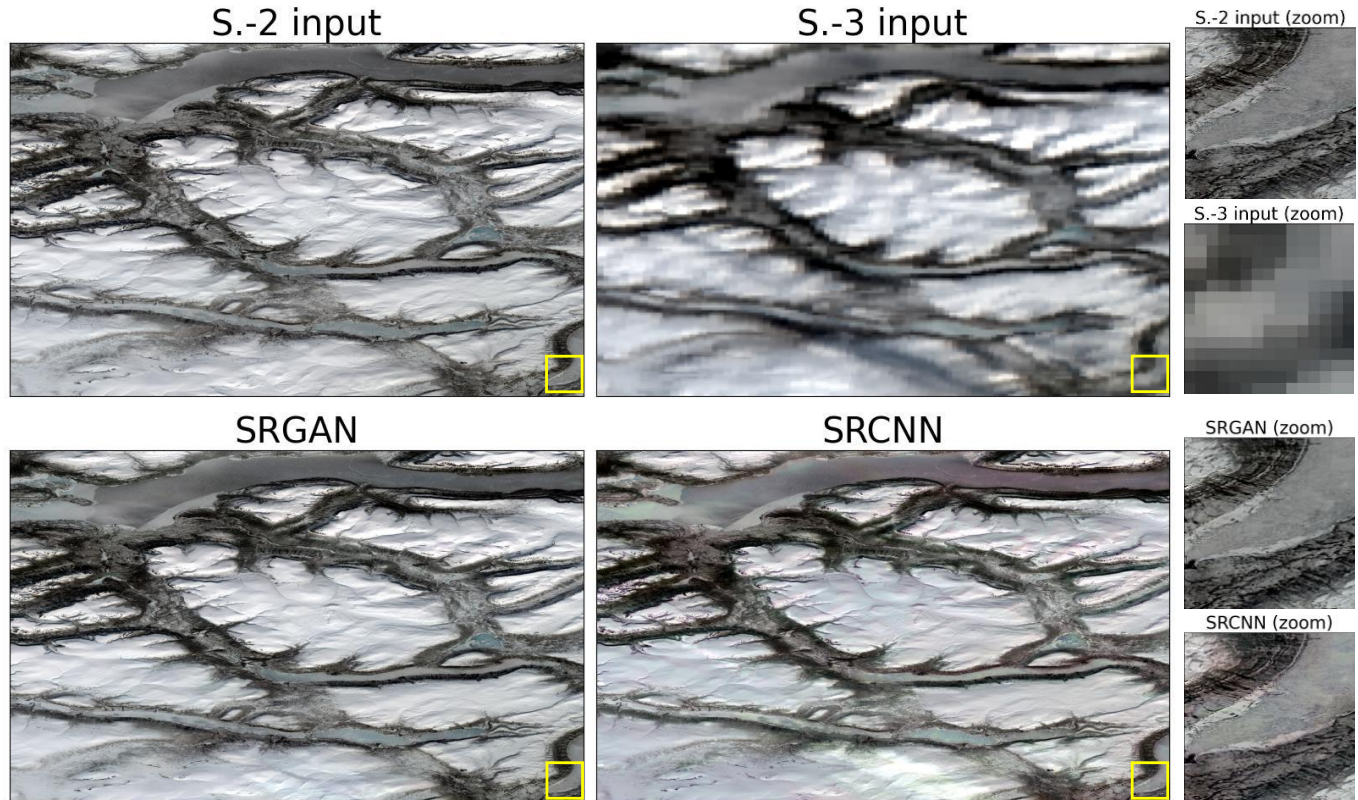


Fig. 15. True color comparison between inputs and neural networks' results on original-scale Sentinel-3 image of the snow region of Khantaiskoe Ozero.

networks according to the reference article. Nevertheless, this is a parameter that can vary depending on the computational and temporal resources we have available. In the case of being limited by any of these factors, the number could be reduced to around 10-20 layers, where Lanaras has already shown that satisfactory results can be obtained.

C. Computational resources and open-source publication

Regarding the training time, it is known that neural networks are expensive, and that it varies depending on both the resources and the structure of the networks themselves. In this particular case, using two Tesla 4 GPUs, it has taken about 45 minutes to train the CNN with 60 epochs.

However, this process is done only once, and then the network can be used already trained for the objective we set, so we do not have to spend so much time each time it is used. Thus, both in GAN and CNN, super-resolving an original scale image takes approximately 15 minutes. It should be noted that due to memory limitations of GPUs, these reconstructions also have to be performed in small patches, so it takes longer.

Finally, it is noteworthy to mention that the developed code is available in the following IFCA Advanced Computing Git-Lab server repository: https://gitlab.ifca.es/penam/sentinel_jae.

VIII. CONCLUSIONS

This study has allowed us to improve the spatial resolution of the Sentinel-3 images so that the 21 bands at 300m GSD reach a resolution of 10m, corresponding to the best one that the Sentinel-2 mission can achieve in some of its bands. Those products are collected with different instruments, which made working with them quite challenging.

First, an automated workflow has been set up to process and homogenize the Sentinel-2 and Sentinel-3 mission products, programmed using Python. It also included collecting images' process from the Copernicus Open Access Hub under certain requirements for downloading. In particular, it has been sought that the dataset to be worked with had a wide variety of land coverages and geographical areas so that the super-resolution tools could generalize to locations in different parts of the world.

Then, two different types of neural networks have been trained: a convolutional network (SRCNN) introduced by C. Lanaras et al. in 2018 [3], and a generative adversarial (SRGAN). For this, we have based on the assumption that the transfer from HR to LR bands is scale-invariant and only depends on the scale factor between resolutions, which in this case is 30. The idea was that the networks learned the mapping $9\text{km} \rightarrow 300\text{m}$ in order to apply it to the $300\text{m} \rightarrow 10\text{m}$ one.

With the different evaluations carried out at lower and original scales, it has been proven that both networks are a great improvement over the traditionally used methods of pansharpening and bicubic interpolation. The first scale allowed us to make not only a qualitative evaluation but also a quantitative one since we had the original 300m Sentinel-3 images as targets. For the numerical evaluation, we have used four metrics that allowed us to compare the performance of the models taking into account different aspects of the images.

They were MAE, RAE, SRE, and SAM. Thus, we have seen that SRGAN is the best technique among those we have tested, followed by SRCNN (with the greatest performance difference in terrains other than land), both by visual comparison and according to the scores obtained (globally and band by band). Both greatly outperform the traditional techniques, reducing the MAE, RAE, and SAM global scores of the latter by almost 30%, and increasing the SRE by up to 1.9 dB. With the visual comparison at the original scale, we could reaffirm this and verify the correctness of the assumption that detail transfer does not depend, at least qualitatively, on the scale at which neural networks learn.

The improved spatial resolution of the multispectral bands of the Sentinel-3 mission may help to extend its use to other applications, such as water quality monitoring of inland water bodies and coasts. It would provide a leap in quality in terms of the accuracy and reliability of these techniques, representing a breakthrough in the scientific field. In this study, the first steps have been taken in an approach that had not been tested with this leap of 30 resolution until now, opening a new range of study possibilities.

As future lines for further research, it would be interesting to obtain a larger number of training images from areas other than land (coasts, and snow regions, for example) to check if the networks' performance there can be improved, especially in the SRCNN case. If networks have more data from these areas during training, they will presumably perform better at super-resolving them. Furthermore, it could be of great interest to modify the structure of the networks to see if increasing the complexity improves the SR images, and also to increase the number of epochs in which the model is trained. These tests have been left out of this study for time and available resources reasons and because we have given priority to comparing the SRCNN with and without the use of a discriminator. Regarding the results obtained so far, it could also be interesting for future work to validate the Sentinel-3 10m SR images through other evaluation methods beyond the qualitative one. For example, through measurements with in-situ instrumentation or by using other satellites to collect the information needed to compare what has been predicted.

REFERENCES

- [1] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spot, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012, the Sentinel Missions - New Opportunities for Science. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425712000636>
- [2] C. Donlon, B. Berruti, A. Buongiorno, M.-H. Ferreira, P. Féménias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra, "The global monitoring for environment and security (gmes) sentinel-3 mission," *Remote Sensing of Environment*, vol. 120, pp. 37–57, 2012, the Sentinel Missions - New Opportunities for Science. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425712000685>
- [3] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, p. 305–319, 2018.

- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.
- [5] T. Dozat, "Incorporating nesterov momentum into adam," *Proceedings of the 4th International Conference on Learning Representations*, 2016.
- [6] Y. Dauphin, H. Vries, J. Chung, and Y. Bengio, "Rmsprop and equilibrated adaptive learning rates for non-convex optimization," *arXiv: Learning*, vol. 35, 2015.
- [7] Y. Liu, Y. Gao, and W. Yin, "An improved analysis of stochastic gradient descent with momentum," *NeurIPS 2020*, 2020, 1533.
- [8] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *International Conference on Learning Representations*, 2016.
- [9] C. Ledig, L. Theis, F. Huszar, and J. e. a. Caballero, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, 2016.
- [10] K. Dineva and T. Atanasova, "Regression analysis on data received from modular iot system," 2019.
- [11] S. Kouadri, A. Elbeltagi, A. Islam, and S. Kateb, "Performance of machine learning methods in predicting water quality index based on irregular data set: Application on illizi region (algerian southeast)," *Applied Water Science*, vol. 11, pp. 1–20, 2021.
- [12] S. Rashmi, S. Addamani, and R. A., "Spectral angle mapper algorithm for remote sensing image classification," *IJISET - International Journal of Innovative Science, Engineering Technology*, vol. 1, no. 4, pp. 201–205, 2014.