AGILE: GIScience Series, 3, 57, 2022. https://doi.org/10.5194/agile-giss-3-57-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.


# Evaluation of SRGAN Algorithm for Superresolution of Satellite Imagery on Different Sensors

Jaskaran Singh Puri [1] and Andre Kotze[2]

[1,2]NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

Correspondence: Jaskaran Singh Puri (m20211149@novaims.unl.pt), Andre Kotze (m20211199@novaims.unl.pt)

**Abstract.** In recent years, deep learning has quickly evolved to be the go-to solution for any kind of analysis of non-linear data. One such use has been that of Generative Adversarial Networks (GAN) in the field of Computer Vision. GAN models have a variety of applications for image processing, specifically, super-resolution of images. A lot of work has been done to enhance or upscale generic RGB imagery such as the ones taken from a mobile or digital camera. However, in the field of remote sensing, it presents challenges like preserving the spatial resolution of the sensor, which is affected by a wider pixel value range and relation of a pixel to ground sampling distance (GSD). From data preparation to enhancing a complete set of tiles at scale, the upsampling/downsampling requires the ratio of number pixels to the actual area in geography to be preserved. SRGAN model has been proven to be effective for interpolating the pixels based on context. However, it was observed that the same algorithm with or without parameter tuning behaves differently based on the sensor source and target resolution. We evaluate the performance of the model from 10m to 2.5m and 2.4m to 0.6m resolution. The comparison will enable better decision making when using the enhanced images for LULC classification, segmentation, and object detection.

**Keywords.** SRGAN, super-resolution, remote sensing, model evaluation

## 1 Introduction

Several applications have been developed in the remote-sensing domain using the Sentinel-2 as a data source, primarily because of its open-use policy, specifically in the area of vegetation analysis. Due to the high presence of mixed-pixels in Sentinel-2 products especially for the urban regions, the 10m (RGB) resolution does not provide enough spatial variability/information as urban use-cases are dependent on the size of the object (Rabbi et al., 2020), (Zhang et al., 2021) such as that of buildings, roads, railway tracks, and any similar man-made structures.

With the support of major advancements in the deep learning space, several image enhancement algorithms (Wang et al., 2020) have been developed both for traditional RGB and satellite images. The algorithms are mostly supported by CNN (Convolutional Neural Networks) algorithms. Unlike traditional image upsampling techniques like bicubic, bilinear, or nearest neighbors, these algorithms perform calculations that do not consider the context of the objects in the image and neither have a ground truth reference to improve its efficiency. Deep learning algorithms overcome these common pitfalls by making an algorithm learn the different kinds of objects in the image like their texture, contrast, placement in the image, and their pixel values or Digital Number (DN) in the context of satellite imagery. Inspired by CNNs, multiple architectures of these algorithms (Shermeyer and Etten, 2019) have been proposed to enhance the resolution of satellite imagery where a similar observation was made stating that resolution enhancement might not be as effective in coarser resolutions. Some researchers have taken it one step further and used the concept of GANs (Generative Adversarial Network) (Ma et al., 2019) which tries to generate a similar copy of ground truth data, which in this case would be high resolution (HR) images. SRGAN, Edge-Enhanced GAN (Jiang et al., 2019), Ultra-Dense GANs are some of the commonly used GAN architectures used for resolution enhancement.

For achieving 6x and 2x upsampling the idea proposed in (Lanaras et al., 2018) is to first perform downsampling, specifically, the 60m bands by 6 times to 360m and the 20m bands to 40m. However, this approach suffers a major drawback because this algorithm enhances only the non-RGB bands but in fact, it uses the spatial information from the RGB bands since they are at 10m as compared to the remaining bands of 20m and 60m. Hence, it is highly dependent on providing a high-resolution reference image that might not exist for other sensors.

In our work, we not only explore how an SRGAN (Ledig et al., 2017) algorithm behaves if we change the source and target resolution but also propose a modified version of SRGAN where the loss functions can handle pixel values beyond 0-255. Two experiments, each for RGB bands across different sensors with multiple iterations are performed for a target resolution of 2.4m and 0.6m, from 10m and 2m respectively. The resulting enhanced images can then be used as input for further classification or segmentation and a comparison could be drawn to which enhanced imagery may be ideal for what kind of further analysis.

The extensive validation and performance of the SRGAN algorithm through the years for computer vision motivated us to use it for our experimental work. The work also accounts for the radiometric resolution of the sensors in use as it is important to preserve the reflectance values of the captured imagery to avoid any information loss. The paper is organized as follows: We initially discuss the data preparation methodology used to train the SRGAN model and briefly introduce the algorithm's working. Finally, we discuss the results achieved and how future improvements can be introduced to make the algorithm fine-tuned toward satellite datasets.

## 2 Data And Methods

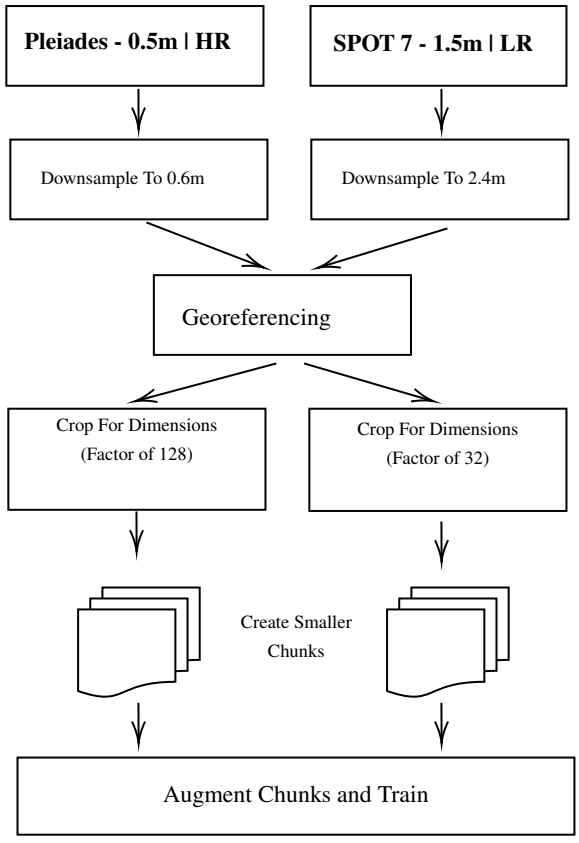


**Figure 1.** Flowchart showing data preparation steps for Dataset 2. A similar approach was taken for Dataset 1

### 2.1 Data Preparation

The methodology to prepare traditional RGB imagery differs from how it is done for satellite images, majorly because when the input size for a deep learning algorithm like CNN is fixed, for eg (512 x 512 x 3), we can simply resize the image to this dimension and because of the small pixel-value ranges of 0-255, the newly averaged values do not change drastically. On the other hand, the remote-sensing images are firstly geo-referenced. Secondly, resizing may result in drastic changes in pixel values because of the wider byte range of satellite images. For eg, the pixel value in Sentinel-2A has 16-bit range as compared to 8-bit of traditional RGB. On doing averaging, the pixel value could signify a major change and hence affect the inference of outputs. In the experiment, we've set up two different test cases:

1. **10m to 2.5m**: Dataset 1

- *Low-Resolution Data:* Sentinel-2 10m RGB imagery for a region near Bangalore, India
- *High-Resolution Data:* Digital Globe imagery for the same region in South-India, available at 2.5m

2. **2.4m to 0.6m**: Dataset 2

- *Low-Resolution Data:* SPOT-7 imagery at 1.5m, downsampled to 2.4m for an area in Saudi Arabia
- *High-Resolution Data:* Pleiades imagery available at 0.5m, for the same area in Saudi Arabia

An important step here to be considered is how do we downsample 0.6m tiles to 2.4m to make it appropriate for 4x upsampling from 10m. We are to extract multiple 128x128 size tiles from the original HR tile, however, while extracting the corresponding LR tiles (32x32), we need to make sure the two tiles are perfectly georeferenced. A visual representation of steps can be seen in Figure 1.

The following Figure 2. shows the relation of pixels between a 2x2 tile and its corresponding 4x upsampled, 8x8 tile. We cannot directly resize images to 32x32 and/or 128x128 as this would affect the spatial properties. The ideal way is to crop a section from the original tile that has dimensions of factors 128 and 32. This way the pixel relation is maintained and the model can learn how to upsample a single pixel 4 times. Finally, we were able to generate thousands of samples using augmentation by flipping and rotating the images. The total amount of training data is shown in Table 1.

The number of files for both the test cases was kept the same to nullify the advantage of having excess data in either of the cases and for better comparison of the metrics.

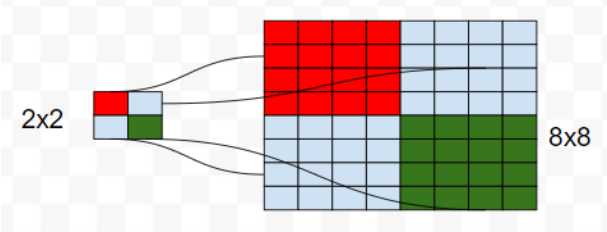


**Figure 2.** Pixel mapping from 2x2 LR Image to 4x Upscaled (HR) 8x8 image

**Table 1.** Training Data Specifications

| | 10m to 2.5m | 2.4m to 0.6m |
|---|---|---|
| Train | 940 | 940 |
| Val | 100 | 100 |
| Test | 100 | 100 |

### 2.2 Time Difference of Acquisition

The change in the image structure and pixel values should be minimal between the two source and target images. In the case where there are significant structural changes in the two images, the training would incorrectly learn the wrong features and eventually generate artificial objects in the outputs. For this reason, it was important to reduce the acquisition time between the two sensors. For our SPOT/Pleiades dataset, the time difference is of 25 days while for the Sentinel-2/Digital Globe dataset it is 30 days. All the acquisitions were made in the year 2021.

### 2.3 Model Training

The SRGAN model proposed involves a two-step training workflow. Firstly, the SRResnet model is trained with the intention of reducing the MSE loss, however, the metrics PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity) are also tracked. This pre-training phase of the two-step stacked model (SRResnet + SRGAN) helps in reducing the number of iterations needed for the SRGAN and proves to be a good starting point with some of the features already learned by optimizing for MSE loss.

On the other hand, MSE loss does not accurately represent the image texture changes that may happen at the pixel level. A pixel can be represented in several different combinations in the high-resolution image, MSE loss will usually output an average of all these combinations which may be unrealistic as compared to the ground truth. To settle this, the original SRGAN model used here is based on a **VGG** loss which is obtained by calculating the euclidean distance between pre-trained feature maps from the VGG-19 model.

Another modification of this SRGAN model was published (https://github.com/xinntao/BasicSR) by removing the Batch Normalization layers from the model architecture because it tended to produce artifacts in the outputs.

### 2.4 Loss Modification

It is important to note that any model architecture is only as good as the metric functions that it should ultimately improve. In the case of satellite images where the pixel values can go beyond 0-255, the current loss functions were required to be modified in our work to compensate for the higher range of values or reflectance from 0 to 1. In the equation of PSNR, where a higher value represents a higher quality of reconstructed image. This is usually computed is *decibals* as it is a measure of noise in signals. The MSE of the reconstructed image and the target image is first computed and provided as input to the following equation, where we use the maximum possible value of reflectance i.e. 1.

$$PSNR = 20 * log10(1./sqrt(mse))$$

## 3 Results And Discussion

### 3.1 SRResNet

Multiple testing scenarios were created to observe what kind of variables trigger the model to give the best and worst metrics, moreover to analyze if the model performs consistently across changing parameters. Table 2 and Table 3 show the 9 test cases that were run using the SRResNet model for the two datasets 10m to 2.5m and 2.4m to 0.6m respectively. Figure 3. and Figure 5. represent the SRResNet performance for learning rate of 0.00003. It can be observed that there is no significant change in the PSNR/SSIM metrics across any of the test cases, meaning the quality of data outweighs the quality of parameter tuning.

Our observations from this dataset are further verified by looking at the metrics of the SRResNet model for the second dataset (2.4m to 0.6m). We see a similar trend of consistent PSNR/SSIM values and the effect of learning rates. The major takeaway here is that providing a higher base resolution (2.4m low-resolution data as compared to 10m) automatically gives the model a better understanding of the data in terms of texture/context which is why we see an increase of around 20 percent in the SSIM values (Two copies of the same image will have an SSIM of 1).

**Table 2.** Metrics: Sentinel 2 - 10m to Digital Globe 2.5m (SRResNet)

| | Feat. | Iterations | Best Iter. | LR | PSNR | SSIM |
|---|---|---|---|---|---|---|
| T1 | 64 | 100K | 20K | 0.0001 | 29.662 | 0.7062 |
| T2 | 64 | 100K | 10K | 0.0002 | 29.642 | 0.7060 |
| T3 | 32 | 100K | 50K | 0.0001 | 29.900 | 0.7091 |
| T4 | 32 | 100K | 20K | 0.0002 | 29.792 | 0.7067 |
| T5 | 16 | 100K | 80K | 0.0001 | 29.935 | 0.7075 |
| T6 | 16 | 100K | 50K | 0.0002 | 29.952 | 0.7080 |

**Table 3.** Metrics: SPOT 7 - 2.4m to Pleadis 0.6m (SRResNet)

| | Feat. | Iterations | Best Iter. | LR | PSNR | SSIM |
|---|---|---|---|---|---|---|
| T1 | 64 | 100K | 5K | 0.0001 | 39.323 | 0.929 |
| T2 | 64 | 100K | 5K | 0.0002 | 39.313 | 0.929 |
| T3 | 32 | 100K | 20K | 0.0001 | 39.389 | 0.929 |
| T4 | 32 | 100K | 10K | 0.0002 | 39.316 | 0.928 |
| T5 | 16 | 100K | 35K | 0.0001 | 39.469 | 0.929 |
| T6 | 16 | 100K | 10K | 0.0002 | 39.372 | 0.929 |

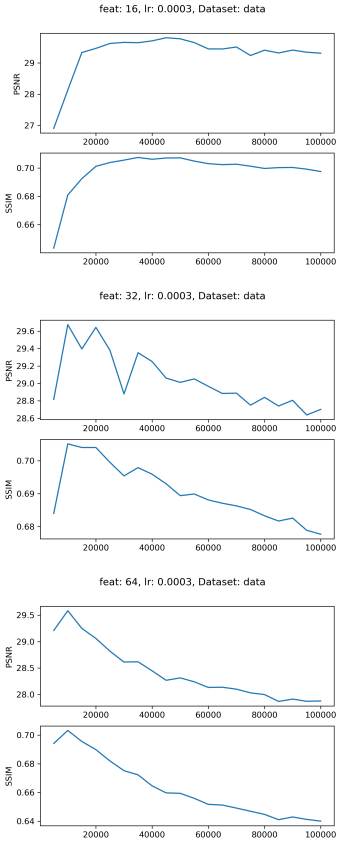


**Figure 3.** Sentinel-2/DG, SRResNet performance for learning rate of 0.0003, top to bottom for feature 16, 32 and 64

## 3.2 SRGAN

For models trained using GAN loss, PSNR should not be used as an evaluation metric as it may not align with human judgment. Similar to that in SRResnet, a similar pattern for the SRGAN performance across the two datasets is observed. Given a better base resolution, the SRGAN performed consistently better across the test cases. SRGAN has been more sensitive to parameter tuning as a result of which we observed the varying metrics.

We also see a drop in SSIM metrics across both datasets which is because of the algorithm's nature to mimic the target imagery, creating unwanted artifacts in the generated images. Table 4 and Table 5 show the 9 test cases that were run using the SRGAN model for the two datasets 10m to 2.5m and 2.4m to 0.6m respectively. Figure 4. and Figure 6. represent the SRGAN performance for learning

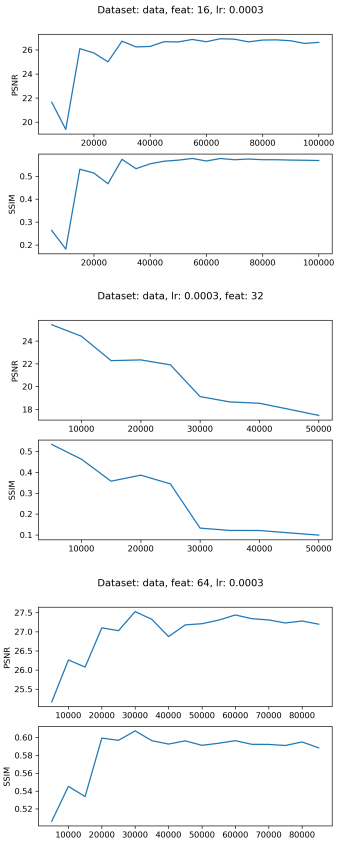


**Figure 4.** Sentinel-2/DG, SRGAN performance for learning rate of 0.0003, top to bottom for feature 16, 32 and 64

rate of 0.00003. The final results from the SRGAN model can be seen in the Figures 7. to 10.

**Table 4.** Sentinel 2 - 10m to Digital Globe 2.5m (SRGAN)

| | Feat. | Iterations | Best Iter. | LR | PSNR | SSIM |
|---|---|---|---|---|---|---|
| T1 | 64 | 80K | 55K | 0.0001 | 27.23 | 0.6022 |
| T2 | 64 | 50K | 5K | 0.0002 | 23.157 | 0.4968 |
| T3 | 32 | 50K | 5K | 0.0001 | 25.428 | 0.534 |
| T4 | 32 | 100K | 10K | 0.0003 | 25.08 | 0.457 |
| T5 | 16 | 100K | 40K | 0.0002 | 26.588 | 0.5575 |
| T6 | 16 | 100K | 30K | 0.0003 | 26.718 | 0.5739 |

**Table 5.** SPOT 7 - 2.4m to Pleadis 0.6m (SRGAN)

| | Feat. | Iterations | Best Iter. | LR | PSNR | SSIM |
|---|---|---|---|---|---|---|
| T1 | 64 | 100K | 40K | 0.0001 | 37.177 | 0.8852 |
| T2 | 64 | 50K | 40K | 0.0002 | 37.45 | 0.8931 |
| T3 | 32 | 100K | 5K | 0.0001 | 36.721 | 0.8798 |
| T4 | 32 | 50K | 5K | 0.0002 | 36.447 | 0.8889 |
| T5 | 16 | 100K | 5K | 0.0001 | 36.981 | 0.8883 |
| T6 | 16 | 100K | 5K | 0.0002 | 36.488 | 0.8638 |

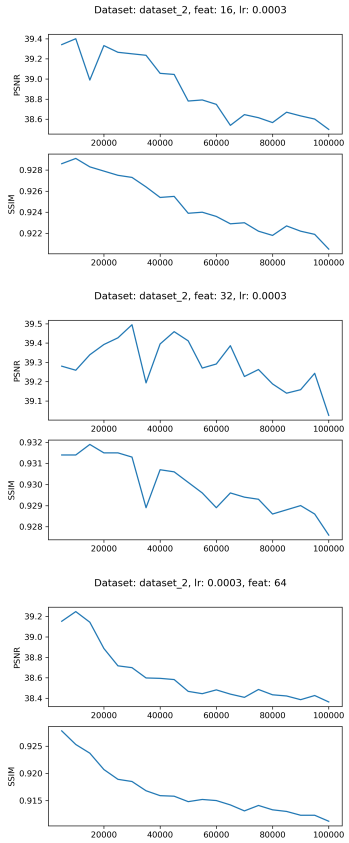


**Figure 5.** SPOT/Pleadis, SRResNet performance for learning rate of 0.0003, top to bottom for feature 16, 32 and 64

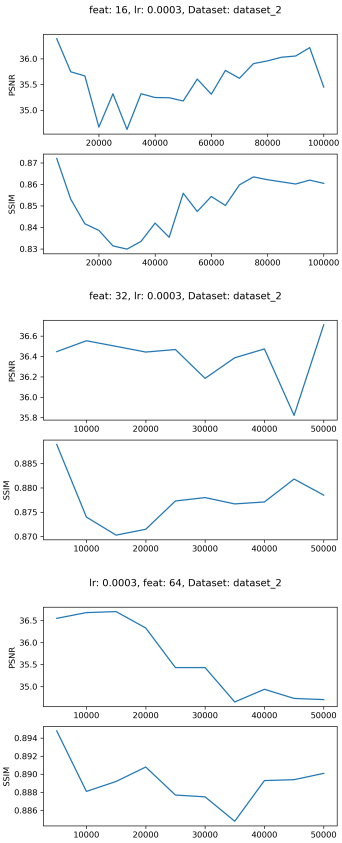


**Figure 6.** SPOT/Pleadis, SRGAN performance for learning rate of 0.0003, top to bottom for feature 16, 32 and 64

## 4 Conclusions and Future Work

In this work, we observed how the GAN algorithms are hugely influenced by the resolution of the data provided. An increase of 54 percent was seen in the SSIM metric by providing a higher base resolution. GAN models however tough to train can learn features quickly with minimum training. Modification of loss functions also helped in retaining reflectance values which can later be converted to DN numbers if needed. Additionally, the colors of the lower resolution images were also preserved with slight changes in the contrast of images which is possible due to the images being taken at different times which makes it harder to have the right dataset for satellite resolution enhancement.

The model however struggles to enhance or understand small buildings or objects, especially when they are present in a cluster, the shadows and roads being mostly black often get confused in the case of Sentinel-2 enhancement. Another major improvement that can be made is in the form of replacing the VGG-19 feature maps which were trained on traditional RGB imagery where the pixel values lie in the range of 0-255. By training a new VGG model for satellite imagery, the GAN models will be able to understand the dynamic range of pixel values and context much more efficiently and can also reduce the chances of overfitting when the dataset is small. Finally, a change-detection preprocess could also be carried out between the source/target imagery to ensure that only similar pixels are enhanced.

**Figure 7.** Location 1, Results for Sentinel-2/DG, left to right (clockwise), low-resolution image at 10m, SRGAN Output at 2.4m, Original high-resolution at 2.4m

**Figure 8.** Location 2, Results for Sentinel-2/DG, left to right (clockwise), low-resolution image at 10m, SRGAN Output at 2.4m, Original high-resolution at 2.4m

**Figure 9.** Location 3, Results for SPOT/Pleadis, left to right (clockwise), low-resolution image at 2.4m, SRGAN Output at 0.6m, Original high-resolution at 0.6m

## Data and Software Availability

The data and code are made accessible through our GitHub repository found here.

## References


Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., and Jiang, J.: Edge-Enhanced GAN for Remote Sensing Image Superresolution, IEEE Transactions on Geoscience and Remote Sensing, 57, 5799–5812, https://doi.org/10.1109/TGRS.2019.2902431, 2019.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., and Schindler, K.: Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network, ISPRS Journal of Photogrammetry and Remote Sensing, 146, 305–319, https://doi.org/10.1016/j.isprsjprs.2018.09.018, 2018.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, pp. 105–114, IEEE, https://doi.org/10.1109/CVPR.2017.19, 2017.

Ma, W., Pan, Z., Yuan, F., and Lei, B.: Super-resolution of remote sensing images via a dense residual generative adversarial network, Remote Sensing, 11, https://doi.org/10.3390/rs11212578, 2019.

Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., and Chao, D.: Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network, Remote Sensing, 12, https://doi.org/10.3390/RS12091432, 2020.

Shermeyer, J. and Etten, A. V.: The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery, pp. 1432–1441, IEEE, https://doi.org/10.1109/CVPRW.2019.00184, 2019.

Wang, Z., Jiang, K., Yi, P., Han, Z., and He, Z.: Ultra-dense GAN for satellite imagery super-resolution, Neurocomputing, 398, 328–337, https://doi.org/10.1016/j.neucom.2019.03.106, 2020.

Zhang, L., Dong, R., Yuan, S., Li, W., Zheng, J., and Fu, H.: Making Low-Resolution Satellite Images Reborn: A Deep Learning Approach for Super-Resolution Building Extraction, Remote Sensing, 13, 2872, https://doi.org/10.3390/rs13152872, 2021.


**Figure 10.** Location 4, Results for SPOT/Pleadis, left to right (clockwise), low-resolution image at 2.4m, SRGAN Output at 0.6m, Original high-resolution at 0.6m