Proba-V Multi-Temporal Super-Resolution Guided by Sentinel-2

(Article begins on next page)

PROBA-V MULTI-TEMPORAL SUPER-RESOLUTION GUIDED BY SENTINEL-2

Gabriele Inzerillo[1,2], *Diego Valsesia*[1], *Enrico Magli*[1], *Fabrizio Niro*[2], *Erminia De Grandis*[2]

[1]Politecnico di Torino - Torino, Italy
[2]Serco c/o ESA/ESRIN - Frascati, Italy

ABSTRACT

Multi-image super-resolution (MISR) is a technique used to increase the spatial resolution of images acquired by remote sensing platforms by combining the images acquired through multiple revisits. Supervised training of MISR models requires collecting high-resolution images to be used as ground truth. Except for a few special cases, this involves acquiring images from a different satellite, resulting in a shift in the optical and radiometric characteristics with respect to the sensor to be super-resolved. In this paper, we explore the use of Sentinel-2 images to train a MISR model for Proba-V images and highlight the challenges of this pursuit.

Index Terms— Super-resolution, multitemporal, Proba-V.

1. INTRODUCTION

Multi-image super-resolution (MISR) is a technique used to increase the spatial resolution of images acquired by remote sensing platforms by combining the images acquired through multiple revisits. This can be useful for applications such as mapping, land use analysis, and monitoring the Earth's surface, as it can enhance their accuracy by going beyond the native resolution of the satellite instrument, which is often constrained by the payload design as well as the downlink channel. Compared to single-image super-resolution (SISR), MISR is capable of greater super-resolving power in recovering details hidden in aliased images. Recently, the Proba-V super-resolution challenge [1] issued by ESA has spurred several works on the topic, thanks to a curated dataset with real high resolution (HR) and low resolution (LR) images acquired by the same platform, at resolutions of 100 meters per pixel and 300 meters per pixel, respectively. Thanks to the availability of a HR ground truth, supervised learning algorithms can be used to train effective models [2] for the task. However, the Proba-V dataset is obtained from images of the same platform, thus presenting a marginal domain gap between the LR and HR acquisitions. This is generally not the case, as one needs to resort to use images from a different satellite as HR ground truth for training. Only a few works [3] have currently treated this topic, but issues related to the different radiometric properties as well as spectral responses of the two sensors emerge. This paper presents a case study in which we use Sentinel-2 imagery to super-resolve Proba-V data with a state-of-the-art neural network MISR model. Critically, this work has required the creation of a new dataset of paired Sentinel-2 and Proba-V images. Thanks to the availability of both LR and HR for Proba-V as well as Sentinel-2 HR images, we are able to observe effects on the radiometry of SR products as influenced by the training process of the MISR network.

2. BACKGROUND

Image SR has been a topic of interest for several years, and recently, there have been significant advancements in this field due to deep learning methods. Previous studies [2] have mainly focused on single-image SR (SISR) for both conventional photographs and remote sensing images. These approaches often use supervised training, which requires high-resolution (HR) images at the target resolution, either in a paired or unpaired manner. In the unpaired setting HR images are available but not from the same scenes as LR images, which reduces the data requirement but does not solve the fundamental problem of needing images at the target high resolution. Except for a few specific instances, such as the Proba-V satellite, where same platform can acquire images at various resolutions, the supervised training process needs to rely on data from multiple satellites or aerial images. For example, Cornebise et al. [4] provide a dataset with paired LR images from Sentinel 2 at m/pixel resolution and HR images from SPOT 6 at 6 m/pixel. In the context of multi-image super-resolution, most of the works used the Proba-V dataset [5, 6, 7, 8]. Thanks to the unique Proba-V setting, those works did not focus on the aspects relating to domain gaps between different sensors. The first work in this direction by Razzak et al. [3] who studied multispectral MISR methods using Sentinel 2 LR images and PlanetScope HR images.

3. PROPOSED METHOD

In this section, we present the main contributions of this paper. They consist in i) a novel dataset of paired Proba-V and Sentinel 2 images; ii) the setting of a comparative investiga-

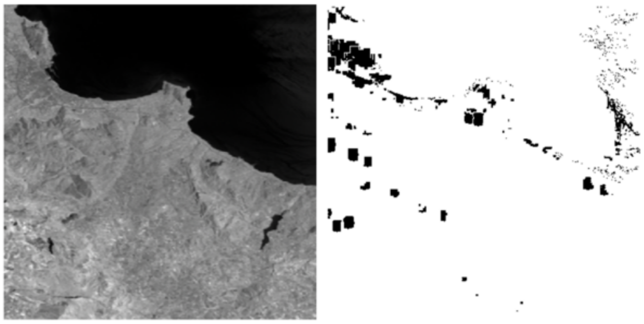


Fig. 1. Example of a HR image from downsampled Sentinel-2 and its corresponding binary quality map (while represents clean pixels). The image was extracted from the RED band at coordinates (38.21130, 13.31250) in May 2020.

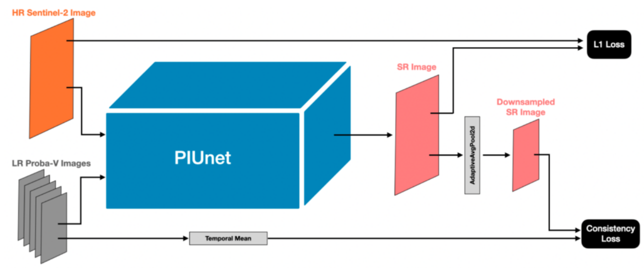


Fig. 2. Mixed training with Proba-V LR images and Sentinel-2 ground truth. The consistency loss tries to match a degraded version of the SR output with the LR inputs to promote similarity with the radiometry of Proba-V rather than Sentinel-2/

tion on the performance of Proba-V MISR with Proba-V HR ground truths and Sentinel 2 ground truths.

3.1. Dataset

In order to understand the impact of ground truth images from a different satellite, namely Sentinel 2, in super-resolving Proba-V images, we first need to create a new paired dataset which allows to run existing MISR techniques and compare the results.

The dataset was constructed using the following procedure. First, we manually selected 636 regions of interest (ROIs) from which to extract images; these images were chosen so as to create as heterogeneous a dataset as possible, with the presence of a wide variety of biomes and spatial features, thus including coastal areas, urban settlements, deserts, vegetation-rich areas, mountains and more. From the coordinates of such ROIs, we extracted a minimum of 16 low-resolution images from Proba-V and one high-resolution image from Sentinel-2 to be used as Ground Truth, spanning a time interval not greater than two months to ensure good temporal consistency. Inspired by the work of Märten et al. [1], for each extracted image, whether high or low resolution, we produced a corresponding quality map, i.e., a binary mask of the same size as the image with the goal of identifying clean reliable pixels (not affected by clouds, artifacts, ...). All the data extracted, both from Proba-V and Sentinel-2, consisted of radiometrically and geometrically corrected Top-Of-Atmosphere (TOA) reflectances. Two processes of reprojection to Plate-Carré and co-registration were carried out in order to avoid as much pixel shifts as possible between the Proba-V and Sentinel-2's images.

In summary, the resulting dataset consists of 1272 image sets, 636 each for the NIR and RED spectral bands; each image set contains the following:

- at least 16 low-resolution images from Proba-V with 128×128 pixels at 300 m/pixel. For each of them, the corresponding quality map was produced;
- one high-resolution from Sentinel-2 with 384×384 pixels image, downsampled to 100 m/pixel from the native resolution. As for the LR images, a quality map was produced;
- one high-resolution from Proba-V with 384×384 pixels image at 100 m/pixel.

Fig. 1 shows an example of a high-resolution image and its corresponding quality map.

3.2. Experimental methodology

Our experimental investigation seeks to analyze the performance of a state-of-the-art MISR method when the HR ground truth for supervision is from the same satellite or a different satellite. Specifically, we used the paired Sentinel-2 and Proba-V dataset described in the previous section with the PIUNet architecture [6] recently proposed for MISR. Supervised training of PIUNet learns a function mapping a set of T LR images into a single HR image of the scene. It is reasonable to expect that, when a different satellite is used as HR ground truth, the learned function will attempt to reproduce the radiometric properties of the ground truth. In essence, if we use Sentinel-2 as HR ground truth, we will produce images that look like images from Sentinel-2 rather than images from Proba-V. This is why in our investigation, we also evaluate the introduction of a "consistency loss" to the training process. Such consistency loss downsamples the SR image produced by the neural network to match the input low resolution. In formulas, the total loss function is

$$L_{\text{tot}} = L_{\text{NLL}} + \lambda L_{\text{con}}$$

$$L_{\text{NLL}} = \sum_i \left(\delta_i + e^{-\delta_i} |x_i^{\text{HR}} - \mu_i|\right)$$

$$L_{\text{con}} = \left\| (\mathbf{k} \circledast \mu)_{\downarrow D} - \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t^{\text{LR}} \right\|_1$$

Table 1. Results (cPSNR) with Proba-V training

	Proba-V HR ref.	Sentinel-2 HR ref.
NIR	69.66 dB	45.71 dB
RED	69.43 dB	47.03 dB

Table 2. Results (cPSNR) with Sentinel-2 training (no consistency loss)

	Proba-V HR ref.	Sentinel-2 HR ref.
NIR	47.03 dB	48.18 dB
RED	47.68 dB	50.58 dB

where L_{NLL} is the negative log-likelihood loss from PIUNet [6] being μ the SR image and δ the estimate of aleatoric uncertainty; $\circledast$ denotes convolution with a degradation kernel $\mathbf{k}$ and $\downarrow D$ decimation by a factor D. The purpose of the consistency loss is to match the features of the SR image with those of the LR images, thus serving as a counterbalance to overfitting the features of a different satellite when that is used for $\mathbf{x}^{\text{HR}}$. A schematic depiction of this is shown in Fig. 2.

4. EXPERIMENTAL RESULTS

In the following we present results on the new Proba-V and Sentinel-2 dataset under various testing conditions and analyze them quantitatively in terms of quality of the super-resolved images as well as in terms of their histograms.

4.1. Experimental setting

The results were obtained from the experiments performed on the standard PIUnet architecture, trained with the new dataset including images from both Proba-V (LR) and Sentinel-2 (HR). Experiments were performed on the NIR and RED bands to have matching spectral bands in both Proba-V and Sentinel-2 images. The standard setting with $T = 9$ LR input images was followed. The neural network was trained for about 400 epochs, enough to reach a performance plateau. Training used an Nvidia Quadro P6000 GPU, and required approximately 19GB of GPU memory. Performance evaluation follows established metrics such as cPSNR [1, 5], i.e., PSNR corrected to be invariant to absolute brightness values and small image shifts. When the consistency loss is adopted, we use spatial averaging as degradation operator. Notice that this is a suboptimal choice which could be improved by choosing a degradation kernel more similar to the Proba-V point spread function.

4.2. Results

In our first experiment we establish a baseline by training and on Proba-V data and testing with respect to both Proba-V and Sentinel-2 HR data. Table 1 reports the results of this experiment. Then we train using Sentinel-2 HR images under two different settings: with or without the consistency loss. These results are presented in Table 2 and Table 3. It can be immediately noticed that, compared to the baseline training on Proba-V, cPSNR increases when measured with respect to the Sentinel-2 HR images but decreases with respect to Proba-V, suggesting that the training process is learning to replicate the Sentinel-2 radiometric features more faithfully. Moreover, the introduction of the consistency loss shows improved cPSNR both with respect to Proba-V and with respect to Sentinel-2 ground truth images. This suggest that the consistency loss has a regularizing effect on training by introducing a kind of constraint on fidelity with respect to the original LR observations. However, it remains unclear if it improves faithfulness with respect to the Proba-V statistics rather than Sentinel-2 ones, as we see improvements with respect to both references. However, by observing the image histograms shown in Fig. 3, we see that the pixels values in the Proba-V HR match fairly well the SR pixel values, except for a scaling factor.

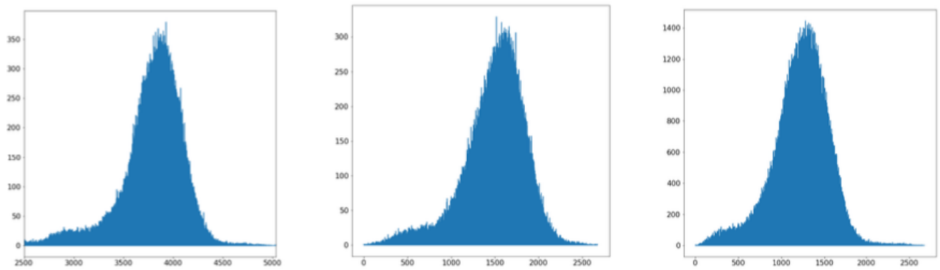
Finally, in Fig. 4 we show a visual result of a super-resolved image compared with one of the LR inputs and the associated uncertainty map produced by the network.

Table 3. Results (cPSNR) with Sentinel-2 training (with consistency loss)

	Proba-V HR ref.	Sentinel-2 HR ref.
NIR	48.26 dB	50.53 dB
RED	48.48 dB	52.25 dB

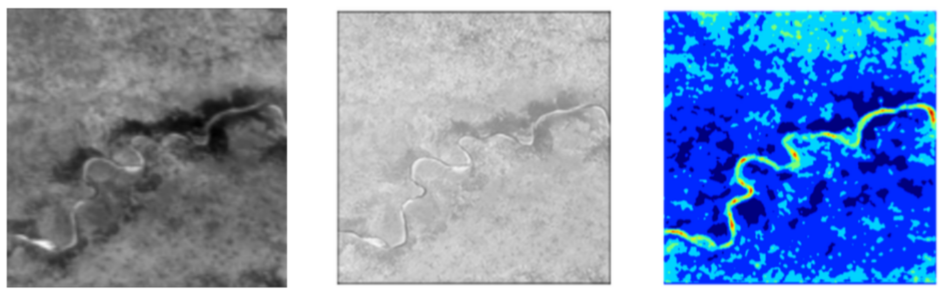
5. CONCLUSIONS

In this paper, we have conducted a preliminary investigation on the effects of training MISR methods with a mismatch between the source of LR images and the source of HR ground truths. This was made possible by a novel dataset comprised of paired Proba-V and Sentinel-2 images. We have indeed verified that training with a mismatch satellite as ground truth affects the generated images by making them more similar to those of the ground truth satellite. We have also shown that a consistency loss can limit this effect and improve overall quality. Further research needs to be performed on this topic by further analyzing the radiometry of generated images, as mismatched training has great practical importance due to the impossibility of having HR and LR images from the same platform. Simultaneously, it is clear that this issue could be entirely avoided by studying unsupervised MISR techniques [9].



(a) SR Image histogram **(b)** Sentinel-2 HR Image histogram **(c)** Proba-V HR Image histogram

Fig. 3. Histograms of pixel values.



(a) LR Image **(b)** SR Image **(c)** Uncertainty Map

Fig. 4. Qualitative results.

6. REFERENCES

[1] Marcus Märtens, Dario Izzo, Andrej Krzic, and Daniël Cox, "Super-resolution of proba-v images using convolutional neural networks," *Astrodynamics*, vol. 3, no. 4, pp. 387–402, 2019.

[2] Brian B. Moser, Federico Raue, Stanislav Frolov, Sebastian Palacio, Jörn Hees, and Andreas Dengel, "Hitchhiker's guide to super-resolution: Introduction and recent advances," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–21, 2023.

[3] Muhammed T Razzak, Gonzalo Mateo-García, Gurvan Lecuyer, Luis Gómez-Chova, Yarin Gal, and Freddie Kalaitzis, "Multi-spectral multi-image super-resolution of sentinel-2 with radiometric consistency losses and its effect on building delineation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 1–13, 2023.

[4] Julien Cornebise, Ivan Orsolic, and Freddie Kalaitzis, "Open high-resolution satellite imagery: The worldstrat dataset – with application to super-resolution," in *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.

[5] Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli, "Deepsum: Deep neural network for super-resolution of unregistered multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3644–3656, 2019.

[6] Diego Valsesia and Enrico Magli, "Permutation invariance and uncertainty in multitemporal image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.

[7] Md Rifat Arefin, Vincent Michalski, Pierre-Luc St-Charles, Alfredo Kalaitzis, Sookyung Kim, Samira Ebrahimi Kahou, and Yoshua Bengio, "Multi-image super-resolution for remote sensing using deep recurrent networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*. 2020, pp. 816–825, IEEE.

[8] Francesco Salvetti, Vittorio Mazzia, Aleem Khaliq, and Marcello Chiaberge, "Multi-image super resolution of remotely sensed images using residual attention deep neural networks," *Remote Sensing*, vol. 12, no. 14, 2020.

[9] Nicola Prette, Diego Valsesia, Tiziano Bianchi, and Enrico Magli, "Towards unsupervised multi-temporal satellite image super-resolution," in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023.