# Sentinel-2 Single Image Super-Resolution with the SEN2VEN$\mu$S Dataset: architecture, training strategy, performances assessment and application to Water Bodies Detection

Aurélien Lac, Julien Michel, Vincent Poulain, Nafissa Sfaksi

# Sentinel-2 Single Image Super-Resolution with the SEN2VENμS Dataset: architecture, training strategy, performances assessment and application to Water Bodies Detection

Aurélien Lac, Julien Michel, Vincent Poulain, Nafissa Sfaksi

**Abstract**

This paper presents training and validation strategies for the application of a Single Image Super-Resolution (SISR) architecture to Sentinel-2 imagery using the SEN2VENμS open dataset, in order to generate 5m resolution images from the initial 10m and 20m bands. Our approach focuses on preserving radiometry and geometry of the input images and avoiding the introduction of artifacts, and leverages a smaller ESRGAN generator network, as well as high and low spatial frequencies separation in loss calculation. Bands B11 and B12, which are missing in the SEN2VENμS dataset, are also process by means of the Wald protocol. We demonstrate that those adaptations allow to enhance the resolution of all Sentinel-2 bands to 5m without substantial loss in reliability. Benefits of the 5 meter super-resolved images are demonstrated through a water bodies detection use-case, allowing to significantly improve detection performances on smaller water bodies.

**Index Terms**

Sentinel-2, Super-Resolution, Water Bodies

## I. Introduction

Sentinel-2 is probably one of the most famous source of optical satellite imagery along with Landsat-8. This is due to its unique combination of spectral bands, global coverage, systematic 5-days revisit and open-data policy. While the Landsat series also provide the same features, Sentinel-2 has the highest spatial resolution, with 4 spectral bands in visible and Near-Infra-Red provided at 10 meter resolution, and 6 bands in Red-Edge, Near and Short-Wavenlength Infra-Red provided at 20 meter resolution.

Sentinel-2 data are therefore used in a wide range of applications where both revisit and resolution are important factors, such as vegetation monitoring [1], Land Cover and Land Use Mapping [2], European Common Agricultural

MM. Lac And Poulain are with Thales
M. Michel is with CESBIO
Ms. Sfaski is with MEOSS

Policy control [3] and Water Bodies Monitoring [4]. However, as shown during the Sentinel-HR phase-0 study at CNES [5], the 10-meter spatial resolution - at best - of Sentinel-2 spectral bands can be a limiting factor for smaller objects, including small agricultural parcels, narrow landscape features such as hedges, and small water bodies. In order to overcome this limitation while still benefiting from the numerous advantages of the Sentinel-2 mission, Single Image Super-Resolution (SISR), which is the process of predicting a higher resolution version of an image using only this image at inference time, seems a natural choice.

Indeed, driven by the deep-learning trend, several works in the literature address the Super-Resolution of Sentinel-2 image with a Single-Image approach [6]–[9]. In [10], the authors suggest that Sentinel-2 is especially well conditioned for SISR, because of its high Modulation Transfer Function (MTF) value at Nyquist rate combined with the small spatial shifts between detectors in the focal plane. Nevertheless, data-sets used in existing works are often gathered from a limited number scenes and geographical locations, they do not cover the full range of Sentinel-2 spectral bands, and most importantly they are not released to the public, which impairs any attempt to reproduce or consolidate their results. To overcome those limitations, the SEN2VENµS dataset [11] has been released under an open-data licence in 2021. Based on same-day occurrences of Venµs and Sentinel-2 data, SEN2VENµS offers more than 130k patches spread across 29 different locations, and covering all 10-meter and 20-meter spectral bands of Sentinel-2 (excluding SWIR bands) with a 5-meter target image acquired by the Venµs satellite at the exact same date.

Water surfaces represent the largest part of the water resource available to societies and their various uses (food, agriculture, industry, etc.). Continuous monitoring of water surfaces is essential for water resource management, to ensure a fair distribution between the various uses. It is also an indicator of global and environmental change, and studying it helps to identify levers for action and adaptation to cope with the effects of climate change. Detailed knowledge and monitoring of water bodies over time is therefore paramount for water managers [12]. Small water bodies monitoring is a challenge, as a trade off must be found between noise and false detection on one hand and accuracy and integrity of water extraction [13] on the other. SISR super-resolution of Sentinel-2 images therefore represents an interesting solution for going beyond the current limits (0.5 ha) and offering complete monitoring of inland water resources.

This paper presents the adaptation and performances assessment of the ESRGAN network trained using the SEN2VENµS dataset in the course of ESA ITT project "Super-Resolution Enhanced Sentinel-2 Data for EO Applications and Services"[1]. The main contributions of this work are as follows. First, a network architecture based on ESRGAN that can jointly super-resolve 10-meter and 20-meter Sentinel-2 bands is proposed. Super-resolution of SWIR band is also achieved by means of the Wald protocol. Second, limitations of the SEN2VENµS dataset in terms of radiometric and spatial consistency are identified and mitigation strategies are proposed. Third, detailed performances assessment is presented. And Fourth, the benefit of the super-resolved data for Water Bodies Detection (WBD) is demonstrated.

The remaining of the paper is organised as follows. Section II briefly summarizes the state-of-the art. Section III

[1]https://eo4society.esa.int/projects/super-resolution-enhanced-data-for-eo-applications-and-services-topic-a-sentinel-2-super-resolution/

presents materials and methods used in this work. Section IV-B presents the performances assessment, while section IV-D presents the application to WBD.

## II. State of the art

Deep learning has revolutionized the field of super-resolution by introducing various architectures that can effectively learn complex mappings between low-resolution (LR) and high-resolution (HR) images. Many training dataset are composed of HR images only, and the proper simulation of the LR degradation model is therefore paramount for model performances and good generalization to real data. For instance, in [14], the Pléiades NEO LR images are carefully simulated from the airborne HR reference images. In the current work however, the dataset consists of LR and HR pairs of real-world satellite data, which allows to focus on optimization paradigms and cost-effective architectures capable of producing sharp and reliable results in the context of SISR. Some of the most notable deep learning architectures for SISR are reviewed in the reamaining of this section.

### *A. Convolutional Neural Networks (CNN)*

This family of models includes all architectures that learn a correspondence between LR and HR images, optimized from end to end and using a single classic CNN model. SRCNN [15] is recognized as the first successful use of a CNN applied to super-resolution. It uses a simple but pioneering 3-layers convolutional neural network with an early up-sampling scheme designed to learn the super resolution mapping. ESPCN [16], also uses a simple CNN architecture, but with a late up-sampling design. The authors also introduce up-sampling with a sub-pixel convolution layer, which has proved to be very efficient and is therefore used by many subsequent methods. Successive advances in neural networks were then applied to this type of architecture. For instance, EDSR [17] and CARN [18] are based on modified and improved versions of residual connections allowing increased depth of super-resolution networks. In the same way, skip connections are massively used in the DenseNet [19] architecture and subsequently applied to super-resolution in SRDenseNet [20]. As implemented in RDN [21], it is also possible and now common to combine residual connections and skip connections. In addition, architectures such as RCAN [22], [23] have successfully used channel attention mechanisms to adaptively rescale channel-level characteristics.

### *B. Generative Adversial Networks (GAN)*

GANs were introduced in 2014 [24], paving the way for a whole new branch of generative model research, and have rapidly become the benchmark for image synthesis, including super-resolution. In general, a GAN consists of a generator that creates synthetic data and a discriminator (or a critic, in the WGAN [25] formulation) that distinguishes between real and generated data. They are trained together in a competitive manner to generate realistic data. SRGAN [26] was a pioneer in the use of GANs applied to super-resolution. The generator is a ResNet-type CNN, while the discriminator is a rather classical CNN ending with dense layers and using LeakyReLU activation functions. In addition, SRGAN is optimized using a combination of GAN loss and content loss calculated on feature maps of a VGG [27] network. ESRGAN [28] builds upon SRGAN, but removes the generator batch normalization layers, replaces the basic ResNet with residual in residual dense blocks (RRDB), and brings other improvements

such as residuals scaling. More recently, Real-ESRGAN [29] defined a new state of the art in blind super resolution, using the same generator architecture as ESRGAN but employing larger-scale training, a U-Net discriminator with spectral normalization [30], [31], and highly diverse image degradation model for training.

### C. Transformers

The Transformer architecture [32] originated from Natural Language Processing (NLP). Its novelty lies in the intensive use of attention mechanism without the need of recurrence or convolution. This type of architecture recently appeared in computer vision with Vision Transfomers (ViT) [33], further improved with Shifted Windwos (Swin) Transformers [34], [35]. This latest architecture was quickly applied to super-resolution through SwinIR [36], [37], which surpasses Real-ESRGAN on real images when using the GAN loss.

### D. Diffusion models

Diffusion models [38]–[40] are a family of probabilistic generative models that progressively destroy data by injecting noise, then learn to reverse this process for sample generation. These models very quickly demonstrated record-breaking results in many fields, including super-resolution. Several supervised approaches for SISR have been proposed, including SR3 [41], SRDiff [42], and LDM [43]. Although these methods seem to offer the best results, they are also considerably more expensive in terms of computing resources.

## III. Materials and methods

### A. Data preparation

The SEN2VENµS dataset comprises 256x256 Venµs patches along with corresponding 128x128 (for 10 meter bands) and 64x64 (for 20 meter bands) Sentinel-2 patches. Venµs patches are pair-wised spatially registered to their corresponding Sentinel-2 patch, and their radiometry is linearly transformed so as to minimize discrepancies between both sensors [11]. It is noteworthy that despite this processing, spatial and spectral discrepancies are still present especially for Venµs sites with higher viewing angles. In the present work, the dataset has been further split by retaining a random selection of 119,915 patches for the training and 13,040 patches for validation. Sentinel-2 tile T30TYP was then used to assess the benefits of super-resolved data for the WBD application, as presented in section IV-D. Patches of Sentinel-2 B11 and B12 bands are missing in the original SEN2VENµS dataset, and have been retrieved from the SentinelHub.

### B. Network architecture

In this work, the generator and discriminator architectures are based on those presented in Real-ESRGAN [29]. However, the proposed adaptations could be applied to other state-of-the-art architectures presented in section II as well. The generator is based on the ESRGAN generator architecture also adopted in Real-ESRGAN, with a suite of Residual in Residual Dense Blocks (RRDB), as described in figure 1. The number of BB initially proposed in ESRGAN is 16 or 23. However, we observed that such very deep network can learn the residual spatial distortion related to the parallax effects on rugged terrain for Venµs site with higher viewing angles. Since the super-resolution

factor is small in this work, the number of BB has been reduced to 6, which allows to limit the network ability to capture spatial distortion. The generator input has been adapted to integrate images of different resolution, namely Sentinel-2 10 and 20 meter bands (figure 1). Similar to what is done in [44], the 20-meter bands are concatenated with the 10-meter bands after going through a bicubic upsampling. The upsampling block used at the end of the architecture is a sub-pixel convolution, using Pytorch PixelShuffle function. In addition, we use the mirror padding mode in the convolution layers, so as to limit edge effects, which are particularly noticeable at the start of training. One important parameter in this architecture is the number of RRDB, which conditions the depth of the network. The discriminator used is a U-Net with spectral normalization used in [30], [31], which takes as input the predicted bands at 5-meter resolution. The relativistic discriminator formulation described in ESRGAN [28] is employed.

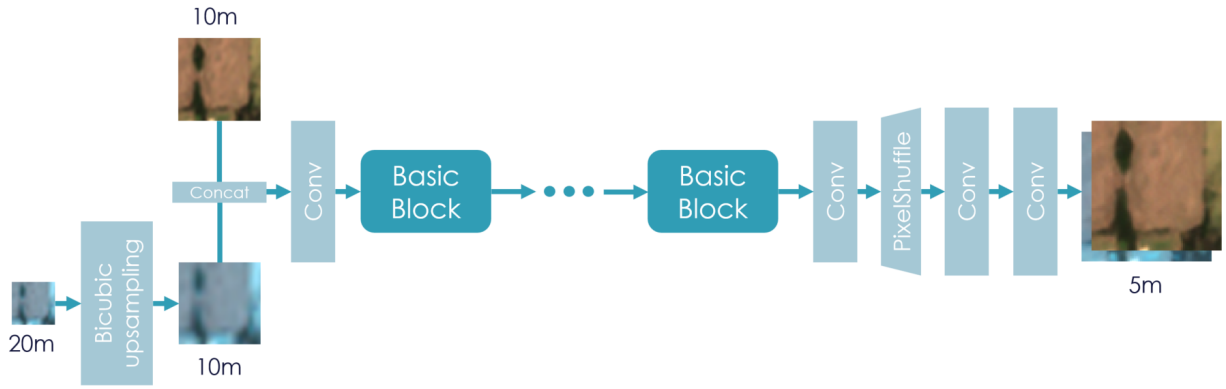


Fig. 1. Overview of the proposed generator architecture. Basic blocks are Residual in Residual Dense Blocks (RRDB).

### C. Proposed LR and HR loss functions

ESRGAN uses the $L_1 = \|I_{SR} - I_{HR}\|_1$ loss as the generator loss, with $I_{SR}$ the reconstructed image and $I_{HR}$ the HR target image), along with the GAN loss $L_{GAN}$ for the discriminator, as defined in equation 1:

$$L_{simple} = L_1 + \lambda L_{GAN} \tag{1}$$

However, during initial tests with $L_{simple}$, $\lambda = 5 \times 10^{-3}$, and with a 18 BB network, radiometric distortions have been observed in the super-resolved images, which were closer to Venμs radiometry than to Sentinel-2 input radiometry. In order to overcome this issue, we designed a new loss term that separately ties the low spatial frequencies of the super-resolved image to the input Sentinel-2 patch and the high spatial frequencies to the Target Venμs patch, as defined in equations 2 and 2:

$$L_1^{lowfreq} = \|I_{SR}^{lowfreq} - I_{LR}\|_1 \tag{2}$$

$$L_1^{highfreq} = \|I_{SR}^{highfreq} - I_{HR}^{highfreq})\|_1 \tag{3}$$

where $I_{SR}^{lowfreq}$ and $I_{SR}^{highfreq}$ are obtained by convolution with an isotropic gaussian kernel whose standard deviation is related to the Modulation Transfer Function (MTF) values at Nyquist rate for each Sentinel-2 spectral band.

In addition, the GAN loss for discriminator is limited to the high frequencies of both the predicted and reference HR images, which gives the final loss function as defined in equation 4:

$$L_{final} = L_1^{lowfreq} + \mu L_1^{highfreq} + \lambda L_{GAN}^{highfreq} \tag{4}$$

This formulation allows GAN loss to be applied only to the high spatial frequencies of the images, and limits the hallucination potential in the low spatial frequencies domain. In addition, it enforces faithfulness to the input Sentinel-2 patch $I_{LR}$.

### *D. Processing of Sentinel-2 SWIR bands*

Since there is no Venμs equivalent for SWIR bands, a dedicated network has been trained by downsampling Sentinel-2 bands by a factor of 2 (Wald protocol). The model is therefore trained to reconstruct 20 meter Sentinel-2 SWIR bands from 40 meters simulated ones, assuming scale invariance for inference. The same generator and U-Net discriminator were used. Similar to [44], low-resolution 40 meter images for training were generated by first applying a Gaussian blur followed by a downsampling. However, for more robust results, random parameters are used. We first apply a random Gaussian blur with $\sigma \sim \mathcal{U}(1.7, 2.6)$, then a downsampling algorithm randomly chosen between Pytorch interpolate mode "area", "bicubic" and "bilinear", and finally we apply Gaussian noise with $\sigma \sim \mathcal{U}(0.1, 0.01)$. During inference, the generator input is the 10-meter bands and 20-meter bands (including B11 and B12) upscaled to 10 meters using bicubic interpolation.

### *E. Performances assessment*

Performances assessment is a challenging task when training SISR network with dataset that rely on two different sensors for the HR and LR patches. Indeed, traditional metrics such as RMSE, PSNR or SSIM with respect to the HR testing patches will favor algorithms that compensate for spatial and radiometric discrepancies between the input LR Sentinel-2 patches and target HR Venμs patches. From the user point of view however, and especially in the remote sensing field, SISR should not incur any distortion of the input image that may impair downstream applications.

In order to provide insight on the ability of the trained network to preserve the input Sentinel-2 image radiometry and geometry, while effectively injecting higher resolution details we perform a separate evaluation:

- Standard image quality metrics are applied to the low pass filtered prediction, with respect to the input Sentinel-2 image, in order to assess radiometric faithfulness to the input LR image,
- A block-matching algorithm is used to evaluate spatial distortion between low-pass filtered prediction and input LR images,
- In the mean time, simulated input patches are generated from reference HR patches by mean of bicubic downsampling, and fed to the network. Those patches are perfectly consistent with the reference HR patches and standard image quality metrics are applied to measure the network super-resolution performances.

### F. Water bodies Detection

Numerous techniques and methods exist for studying and analysing inland water surfaces using different data sources: optical [12], [13], [45], radar [46]–[49] or both [50]–[53]. In this work, the assessment of the relevance of super-resolution data from a thematic point of view is carried out using a WBD algorithm initially designed for processing native resolution Sentinel-2 data, based on water-specific spectral indices [54]–[57] calculated from Green (B3), NIR (B8) and SWIR (B11 and B12) bands. Resampling to 10 meters is necessary to exploit the Sentinel-2 SWIR bands. A binary threshold is then applied to each of the water indices (table I) followed by a fusion by majority voting to obtain the final classification. The technique allows images to be segmented into two distinct regions, the areas in water (the class of interest) and the rest of the image. Simple thresholding has been chosen for its simplicity, speed, and limited computing power requirements [46], [50], [58]. The threshold applied in the present case has been chosen through the analysis of the original Sentinel-2 histogram over year 2020. In this work, the water detection algorithm is applied to the output of the SISR network without any modification.

| Acronym | Index | Authors | formula |
|---|---|---|---|
| NDMI | Normalized Difference Moisture Index | Gao 1996 | $\frac{B8-B11}{B8+B11}$ |
| MNDWI | Modified Normalized Difference Water Index | Xu 2006 | $\frac{B3-B11}{B3+B11}$ |
| NDPI | Normalized Difference Pond Index | Lacaux et al. 2006 | $\frac{B11-B3}{B11+B3}$ |
| NDWI-2 | Normalized Difference Moisture Index-2 | Gao 1996 | $\frac{B8-B12}{B8+B12}$ |

TABLE I

SPECTRAL WATER INDICES USED BY THE WBD ALGORITHM.

This algorithm has been providing continuous temporal monitoring of water bodies larger than 0.5 ha over the french department of Gers since 2017, which is used as the testing area in this work (see section IV-D). The minimum surface area detected and monitored is 0.5 ha. Smaller objects below this limit are excluded from the monitoring, however their analysis is essential for understanding hydrological processes and monitoring water bodies capacities. There are therefore strong expectations from the land and water resource managers community regarding the improvement of monitoring performances.

## IV. EXPERIMENTS

### A. Experimental setup

For all the experiments, the mini-batch size is set to 16. The training process is divided in two step [28]. The generator optimisation starts only with the $L_1$ loss during 50k steps. Next, based on the observation that the abrupt introduction of the discriminator during training could introduce instability, we use (4) with $\mu = 1$ and $\lambda = tanh(\frac{i}{I})\lambda_{max}$ with $i$ the number of steps since the introduction of the discriminator, $I$ a scaling factor (set to 10,000) and $\lambda_{max} = 5 \times 10^{-3}$. The number of steps for this stage is 250k. The learning rate is set to $1 \times 10^{-4}$ and halfed every 50k steps. We use Adam [59] with $\beta_1 = 0.9$, $\beta_2 = 0.999$. We alternate optimization between the generator and discriminator.

### B. Image quality assessment

*1) Qualitative assessment:* Figure 2 presents the result of the ERSGAN network with 6 BB on our validation dataset, both on 10-meter bands (B2, B3 and B4) and 20-meter bands (B5, B6 and B7). It can be observed that the predicted patches are very similar to the reference Venμs patches. The gain in spatial resolution is especially noticeable for the 20-meter bands. Though the benefits with respect to bicubic up-sampling is less noticeable for 10-meter bands, the network seems to also perform denoising and de-aliasing, which can be observed on figure 3. Though denoising and de-aliasing is stronger when using a higher number of BB, it is already noticeable when using only 3 BB.

Regarding B11 and B12 bands obtained using the Wald protocol, the spatial resolution seems to have improved visually 4. As no high resolution reference is available to assess the performances, the rest of the validation work on these bands relies on the application WBD presented in section IV-D. Furthermore, given that the dataset is simulated, there was no difficulties in preserving LR radiometry.

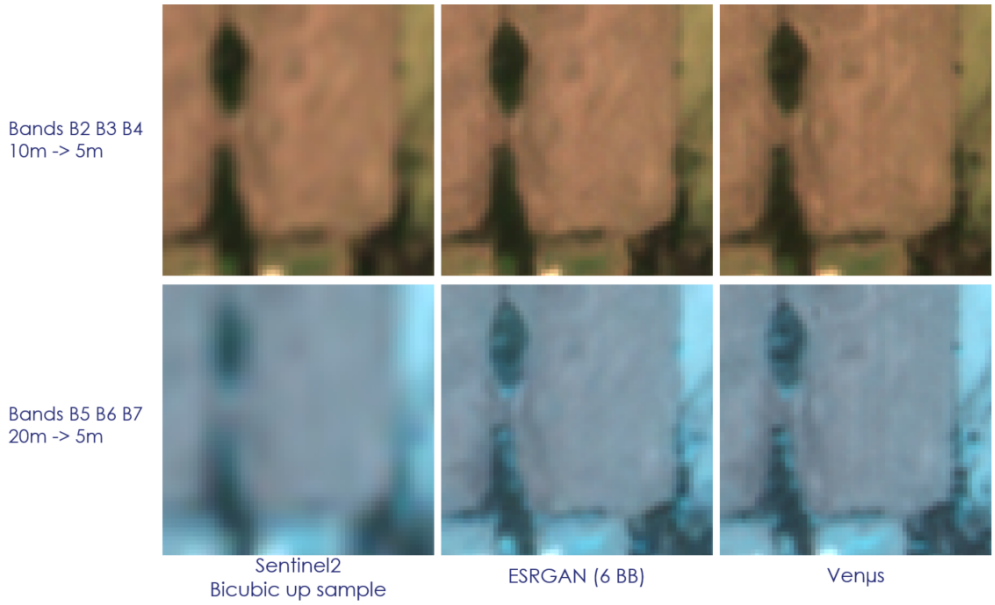


Fig. 2. Example of results with the proposed method. The first line presents a color composition with B2, B3, B4 bands. The second represents a color composition with B5, B6, B7 bands. The left column is Sentinel2 images upsampled with a bicubic algorithm. The middle column is the output of our network. and the right column is the Venμs reference.

*2) HR performances:* Figure 5 presents the PSNR for each spectral bands, when applying the trained SISR network to LR patches simulated by downsampling the reference Venμs patches. While this only gives a proxy of the actual performances, it is the only mean to obtain HR metrics that are not biased by the dataset discrepancies. It can be observed that the proposed network performs similarly to bicubic up-sampling for 10-meters band, and provide a PSNR gain of between 2 and 3 dBs on the 20-meters bands. This is consistent with the qualitative assessment of section IV-B1, which shows that the benefits of the trained SR algorithm is more noticeable for 20-meter bands. It must be stressed that the de-aliasing and denoising effects at 10-meter that can be observed in

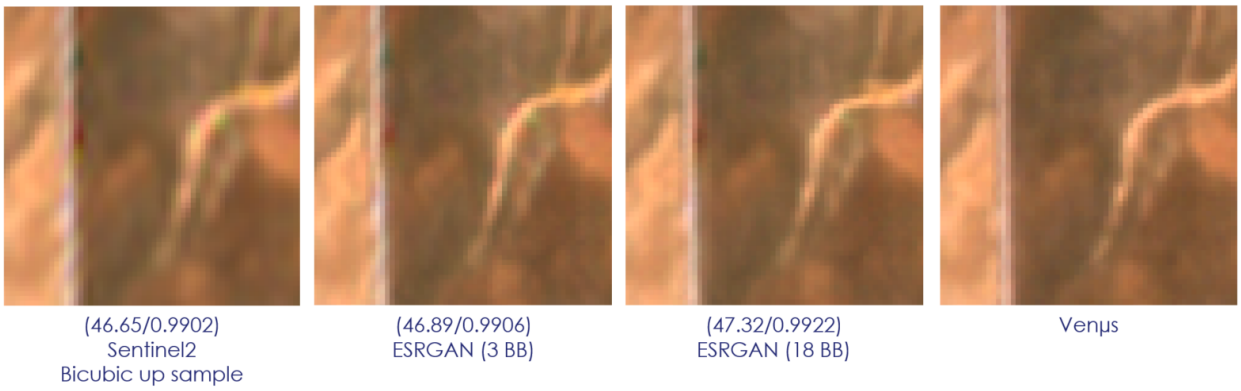


Fig. 3. (PSNR/SSIM) and B2, B3, B4 color composition comparison between Venμs (right) and : Sentinel2 with a bicubic upsample (left), ESRGAN (3 BB) output (2nd), ESRGAN (18 BB) output (3rd)

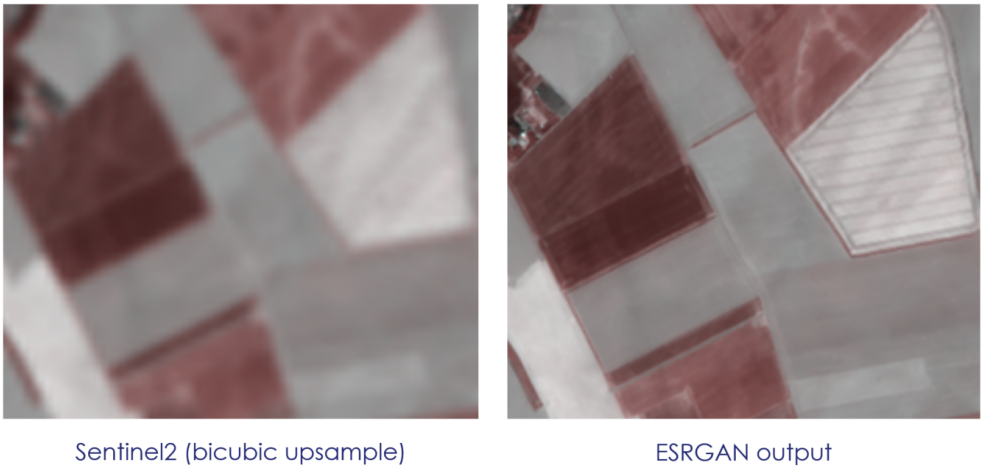


Fig. 4. Example of results with the proposed method using Wald protocol. The color composition used is with B11, B12, B12 bands. The left image is Sentinel2 upsampled with a bicubic algorithm. The right image is the output of our network.

figure 3 are not well accounted for in the measured PSNR: not only will de-aliasing only marginally impact metrics based on global statistics such as PSNR, but the LR patches simulation process used in this experiment is very simple and will not generate realistic Sentinel-2 noise and aliasing. Therefore, PSNR values in figure 5 should be interpreted as a lower bound of the expected performances on real Sentinel-2 images.

*3) LR performances:* Table II gives median and quantiles at 25% and 75% of differences between the initial Sentinel-2 image and the predicted HR image, downscaled back to the Sentinel-2 Resolution. It can be observed that all errors are very low, with a median value within 1e-3 surface reflectance order of magnitude for visible bands and 1e-2 surface reflectance order of magnitude for Near Infra-Red (NIR) bands. It should be noted that the latter have a higher dynamic range which may explain higher errors. In any case, given that the error budget of atmospheric corrections for L2A products is around 0.01 surface reflectance [60], it is safe to say that the SISR model does not incur radiometric distortion to the input Sentinel-2 image.

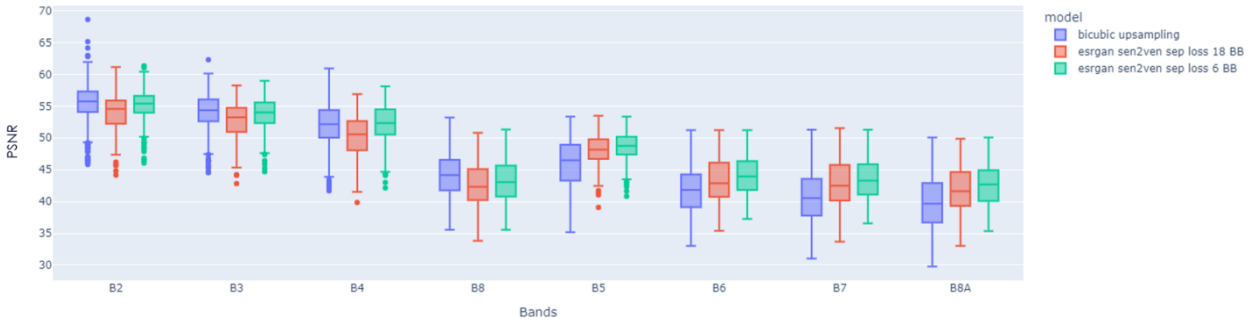


Fig. 5. Box plot of PSNR for each spectral band between Venμs and ESRGAN outputs (for different methods) with simuated Sentinel-2 inputs.

| Band | B2 | B3 | B4 | B8 | B5 | B6 | B7 | B8A |
|---|---|---|---|---|---|---|---|---|
| Median | 106.3 | 147.8 | 139.5 | 725.0 | 435.8 | 466.7 | 725.9 | 903.3 |
| Q1 | 46.7 | 64.6 | 63.6 | 266.1 | 155.0 | 167.6 | 247.4 | 288.9 |
| Q2 | 216.7 | 321.1 | 262.3 | 2373.9 | 2067.1 | 1998.0 | 2784.0 | 3529.2 |

TABLE II

MEDIAN AND QUANTILES AT 25% AND 75% OF ABSOLUTE DIFFERENCES BETWEEN THE INITIAL SENTINEL-2 IMAGE AND THE PREDICTED HR IMAGE, DOWNSCALED BACK TO THE SENTINEL-2 RESOLUTION, WITH ESRGAN TRAINED WITH SEPARATED LOSS AND 6 RRDBS (REFLECTANCE$\times 10^6$)

## C. Ablation studies

*1) Impact of separate HR and LR loss functions:* by comparing errors on the low frequencies between input Sentinel-2 and Venμs reference images on one hand, and between Sentinel-2 and HR prediction from the vanilla network (no separated loss, 18 RRDBs) on the other, we noticed that the magnitudes of errors are similar. This means that the network trained with the vanilla loss learned the residual radiometric bias present in the data set. Table III shows the median of absolute error between Sentinel-2 and the downscaled images generated with ESRGAN, with and without using the separate LR and HR loss functions. Median errors with respect to the input Sentinel-2 radiometric values are 2 times (for NIR and Red-Edge bands) to 4 times (for other visible) bands lower with the proposed separate HR and LR losses than with the vanilla loss. Figure 6 illustrates this trend.

| Band | B2 | B3 | B4 | B8 | B5 | B6 | B7 | B8A |
|---|---|---|---|---|---|---|---|---|
| Median (18 BB, separate HR and LR losses) | 383.5 | 463.9 | 570.1 | 1670.8 | 1584.2 | 2725.4 | 3002.2 | 3435.2 |
| Median (18 BB, vanilla loss) | 2216.2 | 2423.5 | 2488.0 | 5268.0 | 3715.3 | 5816.2 | 6598.7 | 7387.0 |

TABLE III

MEDIAN OF ABSOLUTE ERROR BETWEEN SENTINEL-2 AND THE DOWNSCALED IMAGES GENERATED WITH ESRGAN, WITH AND WITHOUT USING THE SEPARATE LR AND HR LOSS FUNCTIONS (REFLECTANCE$\times 10^6$).

*2) Impact of network depth:* In order to find the correct trade-off between network depth and level of geometric distortion, we tested networks with 3, 5, 6, 8, 10, 16 and 18 RDBBs. Geometric deformations become noticeable

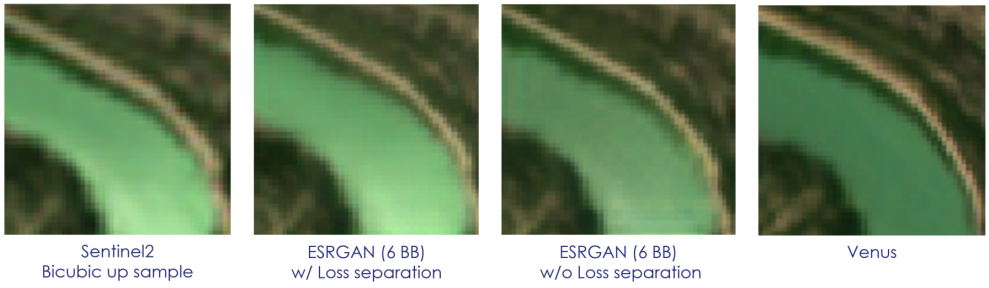


Fig. 6. From left to right: RGB composite of Sentinel-2 patch up-sampled to 5 meter through bicubic zoom, prediction of ESRGAN with vanilla loss, prediction of ESRGAN with separate HR and LR losses, and VEnμs reference.

starting at 8 RDBBs, and there is no difference between 10 and 16 RDBBs. The number of BB was therefore set to 6 RDBBs.

To measure the geometric deformation between Sentinel-2, Venμs, and the super resolved images, we used the block matching algorithm implemented in Orfeo-ToolBox (OTB) [61]. Block matching is performed between the input Sentinel-2 patch and the predicted patch downsampled to 10 meters. Figure 7, illustrate the pixel-wise estimated Y-axis offset for one of the worst areas between Sentinel-2 and the Venμs reference, the predicted image with a 18 RDBBs network, and the predicted image with 6 RDBBs network. It can be observed that the 18 RDBBs network learns the existing deformation between the Sentinel-2 and reference Venμs patch, while the 6 RDBBs shows almost no deformation. Figure 8 shows the distribution of geometric offsets measured on the whole validation dataset, and shows that the 6 BB has far less geometric distortion than the 18 BB, with $90^{th}$ percentile of absolute deformation evaluated at 0.115 pixels.

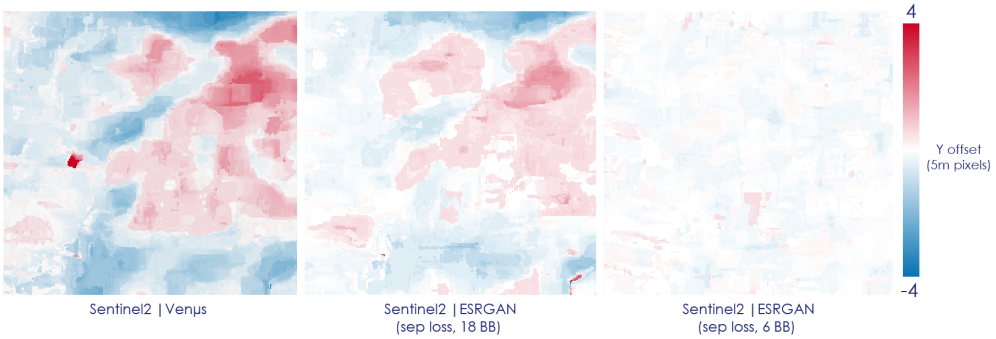


Fig. 7. Y-axis deformation, measured by means of block matching, between Sentinel-2 and the downscaled Venμs reference (left), Sentinel-2 and the 18 RDBBs network prediction (center) and Sentinel-2 and the 6 RDBBs network prediction (right), for a sample patch in the validation set.

### *D. Application to Water Bodies Detection*

The Gers district, France, has been chosen to evaluate the benefit of the super-resolved Sentinel-2 images of the standard ones. Gers is an intensive agricultural area in south-west France, with significant water requirements

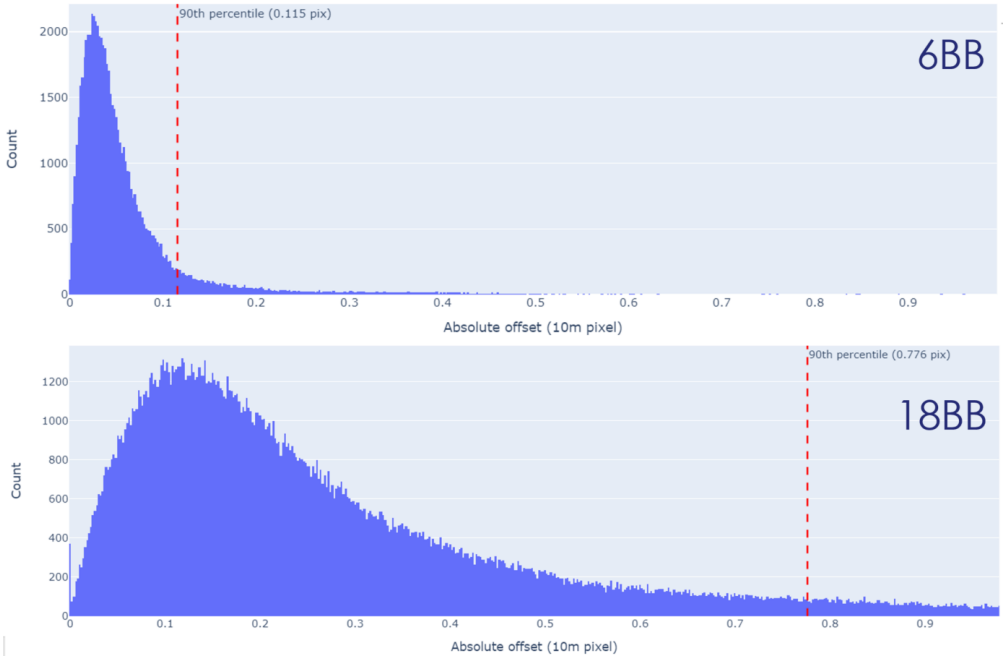


Fig. 8. Distribution of absolute Y-offset between input Sentinel-2 image and prediction by the 6 RDBBs network (bottom) and 18 RDBBs network (bottom), computed on the entire validation set.

and a dense network of water reservoirs of varying sizes, from a few hundred square meters to several hundred hectares. The perimeter covers almost 5,000 km² and includes about 2,000 water bodies. The data used for validation and quantitative assessment comes from a database provided by local users and partners in charge of land management and planning. The database is the result of photo-interpretation work based on Very High Resolution aerial photography (20 cm), corrected and completed by field campaigns, and includes 2,383 water bodies, of which 1,025 are smaller than 0.5 ha (figure 9 ).

The input satellite dataset comprises one clear Sentinel-2 image per month over tile T30TYP in 2020. The results have been evaluated against reference data, both for original Sentinel-2 images and super-resolved predictions at 5 meters resolution, which allows to assess the benefits of proposed SISR for the detection and monitoring of water bodies. The main performances are shown in table IV.

*1) Global statistics:* Overall, the results obtained from super-resolved Sentinel-2 images outperform those obtained from standard Sentinel-2 images at 10m resolution, with a 7-points gain in accuracy for the ”water” class, and a 11-points gain in kappa. More importantly, SISR allows to identify 2132 bodies of water, compared to the 791 bodies detected with standard data, thus allowing for a gain of 160% in number of detected objects. Newly detected bodies include 1003 water bodies smaller than 0.5 ha.

*2) Detected areas and small water bodies:* The 2132 water bodies identified with the super-resolved Sentinel-2 images cover 2818.7 ha, compared with a cumulative 2062.6 ha for water bodies detected at at 10 m resolution, demonstrating an improvement of around 35% (+756 ha) in detected surfaces over the considered area. Smaller water bodies are much better identified, with 16 times more water bodies under 1 ha. In this category, the cumulative

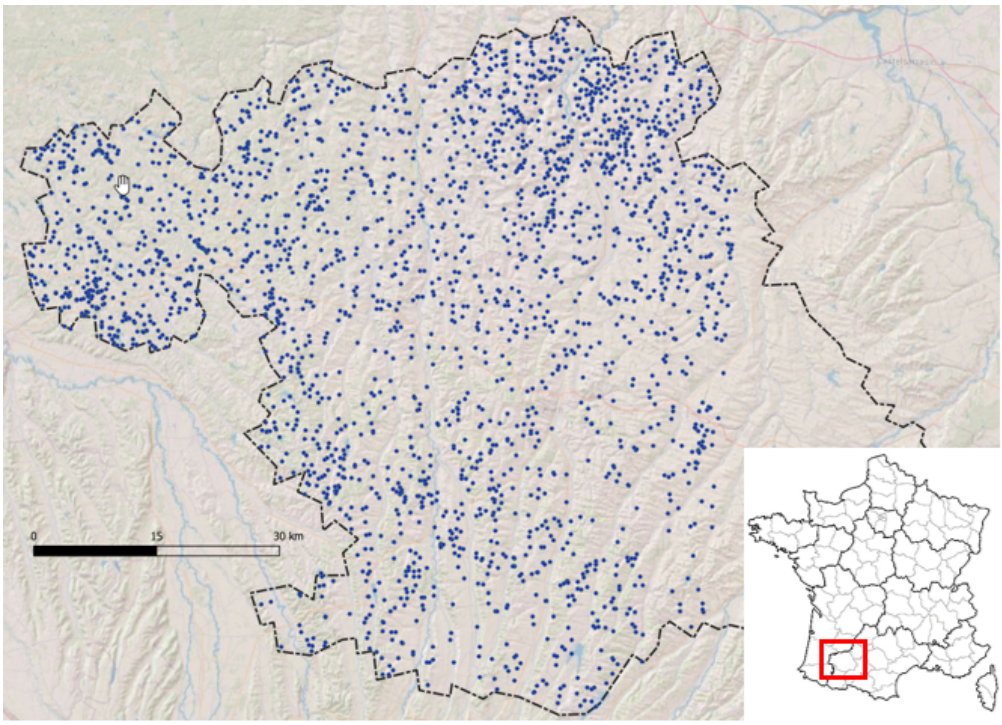


Fig. 9. Local reference database, build from photo-interpretation of Very High Resolution aerial photography (20 cm), corrected and completed by field campaigns, provided by local users of the Gers district, France.

| Metrics | Original Sentinel-2 (10m) | SISR data (5m) |
|---|---|---|
| OA | 98 | 99 |
| UA | 85 | 93 |
| Kappa Index | 0.74 | 0.85 |
| F-Score | 0.73 | 0.85 |
| WBs objects number | 791 | 2132 |
| WBs cumulative area | 2062.6 ha | 2818.7 ha |
| Detection Threshold | 0.5 ha | 0.1 ha |

TABLE IV

MAIN PERFORMANCE METRICS FOR WBD, EVALUATED AGAINST THE REFERENCE DATABASE, FOR STANDARD AND SUPER-RESOLVED SENTINEL-2 IMAGES

surface area represents almost 260 ha in the super-resolved data, compared with 16 ha detected using Sentinel-2 data at 10m. The average improvement ranges from 0.6 ha for small water bodies (less than 1.5 ha) to 1.2 ha for water bodies larger than 10 ha. This represents a significant improvement in performance directly driven by the use of the proposed super-resolved Sentinel-2 images. An improvement in contours is also observed on all water bodies, as shown in figure 11, and is inversely proportional to their size. Bodies of water smaller than 1.5 ha show the greatest increases, with an average improvement in detection of around 40% of their size.

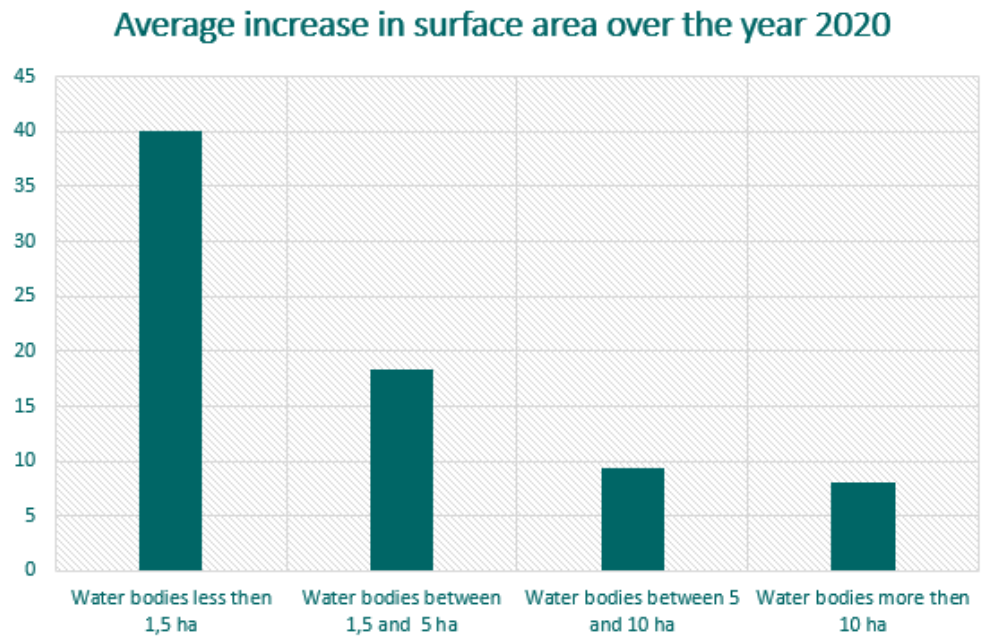


Fig. 10. Average increase in the surface of water bodies between the 5 meter and 10 meter detection, by categories of areas in reference database.

## V. Conclusion

In this work, the SEN2VENµS dataset [11] is leveraged to train a Single Image Super-Resolution network inspired from Real-SRGAN so as to predict all spectral bands of Sentinel-2 at 5-meter resolution. Proposed adaptations include the bicubic up-sampling of 20 meters bands to 10 meters, so that the network can work with the 8 bands (excluding B11 and B12) together, separate HR and LR loss terms that allows to enforce radiometric consistency with the input Sentinel-2 images, and a reduction of the number of RRDBs in the network, preventing it to learn residual geometric dis torsion from the input dataset. SWIR bands B11 and B12, which are missing in the SEN2VENµS dataset, are processed separately by the same architecture by means of the Wald protocol. Performance assessment shows a very good consistency with input Sentinel-2 surface reflectance, which a median error far below the precision of L2A atmospheric correction, and an improvement in PSNR higher than 2 to 3 dBs for the 20 meters input bands. Ablation studies demonstrate the usefulness of proposed network and losses adaptation.

The benefits of the proposed super-resolved Sentinel-2 images were further assessed for Water Bodies Detection, over a district of France where a reference database is available. Super-resolved images allowed to detect 160% additional water bodies with respect to standard Sentinel-2 images, representing 756 additional ha of detected water surfaces. This benefit is especially high for small water bodies under 1 ha, as 16 times more of them were detected with the super-resolved images, accounting for 260 ha additional ha of detected water surfaces.

This work provides the first SISR performances with the public SEN2VENµS dataset, setting the bar for future competitors and investigating how to fairly evaluate the super-resolved images on a real world dataset combining

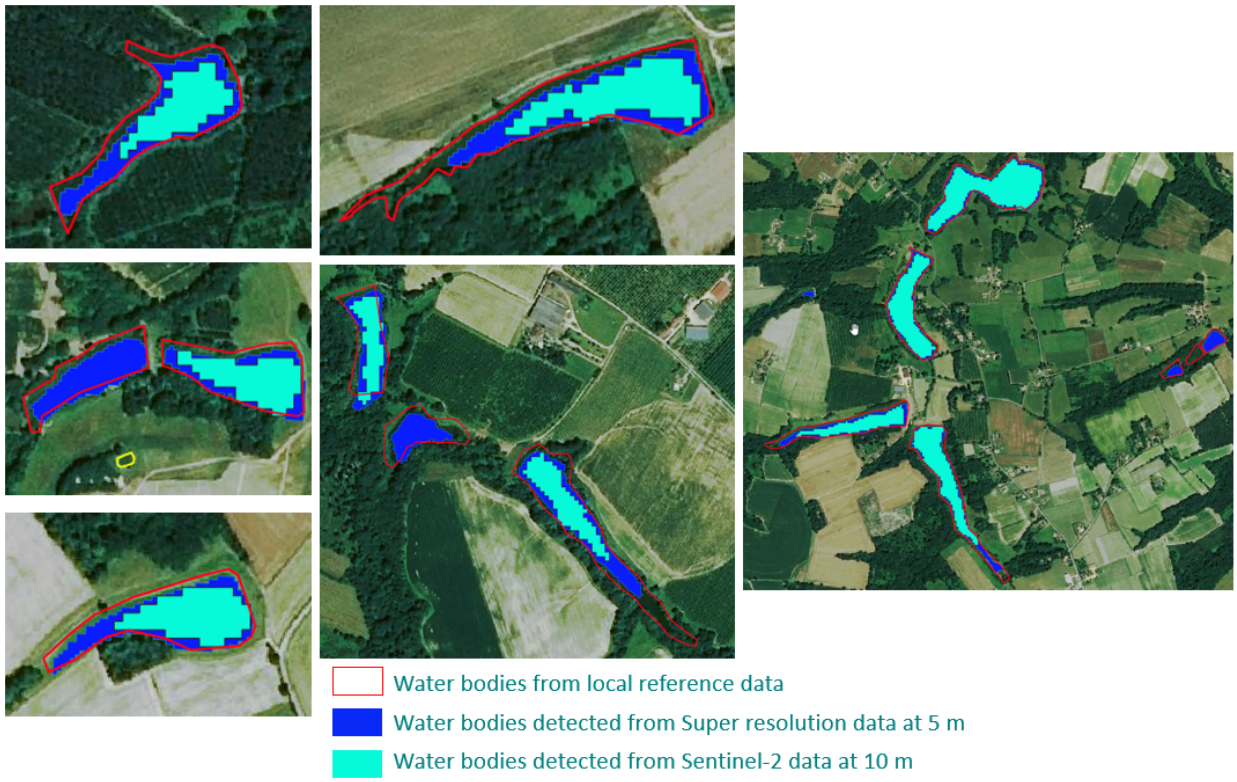


Fig. 11. Improvement of WBD contours detected from the Super-Resolved Sentinel-2 images at 5 meters, with respect to the standard 10 meters Sentinel-2 images, for a selection of water bodies. Red outlines denote the contour from the reference database.

two different sensors. We also hope that those promising results and the demonstrated benefits for an end-user use case will pave the way towards operational Super-Resolution as a standard processing for data providers.

## ACKNOWLEDGMENT

This work was funded by a contract in the framework of the one-year ESA project "Super-Resolution Enhanced Sentinel-2 Data for EO Applications and Services" which is part of the "EO for Civil Security Applications" activity line of the "Development and Exploitation Component" of ESA's FutureEO-1 program (Segment-1). The authors acknowledge the support from the ESA to this project, as well as their advice and technical recommendations.

## REFERENCES

[1] G. Misra, F. Cawkwell, and A. Wingler, "Status of phenological research using sentinel-2 data: A review," *Remote Sensing*, vol. 12, no. 17, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/17/2760

[2] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/14/2291

[3] B. Vajsová, D. Fasbender, C. Wirnhardt, S. Lemajic, and W. Devos, "Assessing spatial limits of sentinel-2 data on arable crops in the context of checks by monitoring," *Remote Sensing*, vol. 12, no. 14, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/14/2195

[4] U. Bhangale, S. More, T. Shaikh, S. Patil, and N. More, "Analysis of surface water resources using sentinel-2 imagery," *Procedia Computer Science*, vol. 171, pp. 2645–2654, 2020, third International Conference on Computing and Network Communications (CoCoNet'19). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920312801

[5] J. Michel, O. Hagolle, J.-M. Delvit, M. Thierry, L. Laurent, D. Laubier, J. P. Malet, L. Jérome, J. Michaud, V.-S. Juan, C. Iliopoulos, S. Le Corre, P. Dusseux, A. M. Kääb, H. Yesou, J. Mouginot, S. Battiston, S. Clandillon, E. W. J. Bergsma, E. Berthier, A. Lucas, J.-F. Dejoux, G. Dedieu, A. Puissant, S. Ferrant, D. Sheeren, A. Lefevre, J.-P. Cantou, J.-F. Crétaux, R. Almar, C. Lelong, P. Maurel, J.-F. Faure, J. Nabucet, T. Corpetti, and P.-A. Herrault, "Sentinel-HR Phase 0 Report," CNES - Centre national d'études spatiales ; CESBIO, Tech. Rep., Apr. 2022. [Online]. Available: https://hal.science/hal-03643411

[6] L. Salgueiro Romero, J. Marcello, and V. Vilaplana, "Super-resolution of sentinel-2 imagery using generative adversarial networks," *Remote Sensing*, vol. 12, no. 15, p. 2424, 2020.

[7] M. Galar, R. Sesma, C. Ayala, and C. Aranda, "Super-resolution for sentinel-2 images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W16, pp. 95–102, 2019. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLII-2-W16/95/2019/

[8] M. Galar, R. Sesma, C. Ayala, L. Albizua, and C. Aranda, "Super-resolution of sentinel-2 images using convolutional neural networks and real ground truth data," *Remote Sensing*, vol. 12, no. 18, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/18/2941

[9] F. Pineda, V. Ayma, and C. Beltran, "A generative adversarial network approach for super-resolution of sentinel-2 satellite images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2020, pp. 9–14, 2020. [Online]. Available: https://isprs-archives.copernicus.org/articles/XLIII-B1-2020/9/2020/

[10] N. L. Nguyen, J. Anger, L. Raad, B. Galerne, and G. Facciolo, "On the role of alias and band-shift for sentinel-2 super-resolution," *arXiv preprint arXiv:2302.11494*, 2023.

[11] J. Michel, J. Vinasco-Salinas, J. Inglada, and O. Hagolle, "SEN2VENμS, a dataset for the training of Sentinel-2 super-resolution algorithms," May 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6514159

[12] M. C. Cordeiro, J.-M. Martinez, and S. Peña-Luque, "Automatic water detection from multidimensional hierarchical clustering for sentinel-2 images and a comparison with level 2a processors," *Remote Sensing of Environment*, vol. 253, p. 112209, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720305824

[13] Y. Dong, L. Fan, J. Zhao, S. Huang, C. Geiß, L. Wang, and H. Taubenböck, "Mapping of small water bodies with integrated spatial information for time series images of optical remote sensing," *Journal of Hydrology*, vol. 614, p. 128580, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169422011507

[14] F. Chouteau, L. Gabet, R. Fraisse, T. Bonfort, B. Harnoufi, V. Greiner, M. Le Goff, M. Ortner, and V. Paveau, "Joint super-resolution and image restoration for pléiades neo imagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 9–15, 2022.

[15] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," 2015.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," 2016.

[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," 2017.

[18] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," 2018.

[19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018.

[20] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4809–4817.

[21] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," 2018.

[22] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," 2018.

[23] Z. Lin, P. Garg, A. Banerjee, S. A. Magid, D. Sun, Y. Zhang, L. V. Gool, D. Wei, and H. Pfister, "Revisiting rcan: Improved training for image super-resolution," 2022.

[24] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," 2017.

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

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[28] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[29] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 1905–1914.

[30] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018.

[31] E. Schönfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks," 2021.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021.

[35] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin transformer v2: Scaling up capacity and resolution," 2022.

[36] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," 2021.

[37] M. V. Conde, U.-J. Choi, M. Burchi, and R. Timofte, "Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration," 2022.

[38] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," 2015.

[39] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020.

[40] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," 2020.

[41] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," 2021.

[42] H. Li, Y. Yang, M. Chang, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," 2021.

[43] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022.

[44] 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, pp. 305–319, dec 2018.

[45] X. Yang, Q. Qin, P. Grussenmeyer, and M. Koehl, "Urban surface water body detection with suppressed built-up noise based on water indices from sentinel-2 msi imagery," *Remote Sensing of Environment*, vol. 219, pp. 259–270, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425718304279

[46] F. Bioresita, "Exploitation de séries temporelles d'images multi-sources pour la cartographie des surfaces en eau," Theses, Université de Strasbourg, Mar. 2019. [Online]. Available: https://theses.hal.science/tel-02316363

[47] S. Zhou, P. Kan, J. Silbernagel, and J. Jin, "Application of image segmentation in surface water extraction of freshwater lakes using radar data," *ISPRS International Journal of Geo-Information*, vol. 9, no. 7, 2020. [Online]. Available: https://www.mdpi.com/2220-9964/9/7/424

[48] A. Gulácsi and F. Kovács, "Sentinel-1-imagery-based high-resolution water cover detection on wetlands, aided by google earth engine," *Remote Sensing*, vol. 12, no. 10, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/10/1614

[49] F. Bioresita, A. Puissant, A. Stumpf, and J.-P. Malet, "Fusion of sentinel-1 and sentinel-2 image time series for permanent and temporary surface water mapping," *International Journal of Remote Sensing*, vol. 40, no. 23, pp. 9026–9049, 2019. [Online]. Available: https://doi.org/10.1080/01431161.2019.1624869

[50] C. Cazals, "Apport des données Sentinel-1 pour la cartographie des milieux humides," Theses, Université Paris-Est, Nov. 2017. [Online]. Available: https://theses.hal.science/tel-01757083

[51] S. Nativel, E. Ayari, N. Rodriguez-Fernandez, N. Baghdadi, R. Madelon, C. Albergel, and M. Zribi, "Hybrid methodology using sentinel-1/sentinel-2 for soil moisture estimation," *Remote Sensing*, vol. 14, no. 10, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/10/2434

[52] Y. Bai, W. Wu, Z. Yang, J. Yu, B. Zhao, X. Liu, H. Yang, E. Mas, and S. Koshimura, "Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: Demonstration of sen1floods11 benchmark datasets," *Remote Sensing*, vol. 13, no. 11, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/11/2220

[53] S. Peña-Luque, S. Ferrant, M. C. R. Cordeiro, T. Ledauphin, J. Maxant, and J.-M. Martinez, "Sentinel-1amp;2 multitemporal water surface detection accuracies, evaluated at regional and reservoirs level," *Remote Sensing*, vol. 13, no. 16, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/16/3279

[54] F. Yao, C. Wang, D. Dong, J. Luo, Z. Shen, and K. Yang, "High-resolution mapping of urban surface water using zy-3 multi-spectral imagery," *Remote Sensing*, vol. 7, no. 9, pp. 12 336–12 355, 2015. [Online]. Available: https://www.mdpi.com/2072-4292/7/9/12336

[55] Y. Yang, D. Zhao, and B. Peng, "A real time mosaic method for remote sensing video images from uav," *Journal of Signal and Information Processing*, vol. 04, pp. 168–172, 01 2013.

[56] H. Xu, "Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery," *International Journal of Remote Sensing*, vol. 27, no. 14, pp. 3025–3033, 2006. [Online]. Available: https://doi.org/10.1080/01431160600589179

[57] C. A. Pernollet, A. Guelmami, A. J. Green, A. Curcó Masip, B. Dies, G. Bogliani, F. Tesio, A. Brogi, M. Gauthier-Clerc, and M. Guillemain, "A comparison of wintering duck numbers among european rice production areas with contrasting flooding regimes," *Biological Conservation*, vol. 186, pp. 214–224, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006320715001305

[58] J. Li, R. Ma, Z. Cao, K. Xue, J. Xiong, M. Hu, and X. Feng, "Satellite detection of surface water extent: A review of methodology," *Water*, vol. 14, no. 7, 2022. [Online]. Available: https://www.mdpi.com/2073-4441/14/7/1148

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[60] O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, "A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of formosat-2, landsat, vens and sentinel-2 images," *Remote Sensing*, vol. 7, no. 3, pp. 2668–2691, 2015. [Online]. Available: https://www.mdpi.com/2072-4292/7/3/2668

[61] M. Grizonnet, J. Michel, V. Poughon, J. Inglada, M. Savinaud, and R. Cresson, "Orfeo toolbox: Open source processing of remote sensing images," *Open Geospatial Data, Software and Standards*, vol. 2, no. 1, p. 15, 2017.