Sentinel-3 Super-Resolution Based on Dense Multireceptive Channel Attention

Rafael Fernandez, Ruben Fernandez-Beltran ⓘ *, Senior Member, IEEE*, Jian Kang ⓘ *, Member, IEEE*, and Filiberto Pla ⓘ

Abstract—The unprecedented availability of remote sensing data from different complementary Sentinel missions provides increasing opportunities to alleviate the spatial limitations of Sentinel-3 (S3) from an intersensor perspective. Nonetheless, effectively exploiting such intersensor synergies still raises important challenges for super-resolution (SR) algorithms in terms of operational data availability, sensor alignment and substantial resolution changes, among others. In this scenario, this article sets a new SR framework for spatially enhancing S3 ocean and land color instrument (OLCI) products by taking advantage of the higher spatial resolution of the Sentinel-2 (S2) multispectral instrument (MSI). To achieve this goal, we initially study some of the most important deep learning-based approaches. Then, we define a novel Level-4 SR framework which integrates a new convolutional neural network specially designed for super-resolving OLCI data. In contrast to other networks, the proposed SR architecture (termed as SRS3) employs a dense multireceptive field together with a residual channel attention mechanism to relieve the particularly low spatial resolution of OLCI while extracting more discriminating features for the large spatial resolution differences with respect to MSI. The experimental part of the work, conducted using ten coupled OLCI and MSI operational data, reveals the suitability of the presented Level-4 SR framework within the Copernicus programme context as well as the advantages of the proposed architecture with respect different state-of-the-art models when spatially enhancing OLCI products. The related codes will be publicly available at https://github.com/rufernan/SRS3.

Index Terms—Convolutional nural network (CNN), level-4 data processing, ocean and land color instrument (OLCI), sentinel-3 (s3), super-resolution (SR).

I. INTRODUCTION

OVER the last decades, the technological evolution of air-borne and space-borne image acquisition instruments has allowed to considerably improve the spatial resolution of multispectral (MS) sensors in order to face new challenges and societal needs by means of remote sensing (RS) images [1]. From the limited spatial resolution of the moderate resolution

imaging spectroradiometer (MODIS) [2] and the medium resolution imaging spectrometer [3], to the expanded spatial resolution of Landsat [4] and Sentinel-2 [5] instruments, multiple types of RS images with different spatio-spectral features have been successfully used in many valuable applications, such as crop and urban mapping [6]–[8], natural hazards discovery [9], [10], environmental management [11], [12], and detailed land-cover analyses [13]–[15]. With these and other important tasks in mind, the spatial resolution of remotely sensed imagery is often a key factor to achieve better performances since the higher the level of spatial details the more visual information is logically available for consideration [16].

In general, improving the spatial resolution of RS data has always been one of the most important concerns within the RS community. Nonetheless, the physical limitations on the incoming radiation when decreasing the pixel size beyond an specific limit [17] together with the high cost and complexity of such technology often make that earth observation (EO) programmes opt to include different specialized satellites to cover different spatial and spectral needs. The Copernicus programme is not an exception in this regard. Being one of the most important EO programmes, Copernicus [18] is a joint initiative of the European Commission, the European Space Agency (ESA), the European Member States and Agencies in order to provide global monitoring information from space useful for environmental and security applications. In order to ensure its operational provision, seven complementary Sentinel missions have been planned as well as several additional contributing missions [19].

Among all the programme resources, Sentinel-2 (S2) and Sentinel-3 (S3) missions share special synergies since both families of dedicated satellites are focused on the global monitoring of the earth surface using for this purpose mid-resolution and high-resolution MS imagery. On the one hand, S2 [5] includes two identical satellites (S2A and S2B) that incorporate the multispectral instrument (MSI), which provides a total of 13 spectral bands (B01-B12) covering the wavelength region from 443 to 2190 nm of the electromagnetic spectrum, with an spatial resolution ranging from 10 to 60 m. On the other hand, S3 [20] also comprises a pair of satellites (S3A and S3B) that carry the ocean and land color instrument (OLCI), which captures the earth surface using 21 bands (Oa01-Oa21) in the spectral range between 390 and 1040 nm, and an spatial resolution of 300 m. From a general perspective, S2 and S3 missions are able to provide operational products of vegetation, soil, and water cover. However, the existing spatial-spectral differences between both

MS instruments make their corresponding data products more effective for specific RS applications [21]. Whereas the OLCI sensor is optimized to measure spectral features focused on oceans, inland waterways, and coastal areas [22], the substantially higher spatial resolution of MSI makes S2-derived products more suitable for land cover characterization tasks [23].

Despite the fundamental spatial differences between S2 and S3 missions, the unprecedented availability of Sentinel data [24], [25] together with the constant development of the digital image processing technology provide widespread opportunities to overcome the spatial limitations of the corresponding operational products from an image processing-based perspective. Specifically, it is possible to distinguish between two general paradigms when enhancing remotely sensed imagery [26]: pansharpening and super-resolution (SR). Whereas pansharpening algorithms [27] work for merging a given high-resolution (HR) panchromatic (PAN) image and a low-resolution (LR) MS image in order to generate a HR version of the input MS data, SR techniques [28] provide a more general framework since the spatial relationships between the HR and LR domains can be explicitly specified into an imaging model or even learned from the own data.

The increasing popularity of pan-sharpening and SR techniques within the Copernicus programme context exemplify the success of these two image enhancing paradigms for Sentinel data. For instance, Vaiopoulos and Karantzalos evaluated in [29] different pan-sharpening algorithms to spatially enhance the red-edge, narrow-near-infrared and short-wave-infrared bands of S2, using the HR information provided by the 10 m bands. Park et al. [30] also proposed in some modifications over this straightforward pan-sharpening scheme to optimize the synthesis of a simulated PAN image from the 10 m bands of S2. Notwithstanding the good performance of pan-sharpening techniques, they have the important constraint of requiring a real or simulated PAN image as input, which may be difficult to obtain from the operational perspective of Sentinel ground-segment units [31]. Hence, other authors suggest using image enhancing techniques based on the SR paradigm instead. For example, it is the case of Lanaras et al. [32] who presented in a signal reconstruction SR approach for S2 data that recovers HR spatial details using a two-fold regularization term based on the correlation between spectral channels and the image gradients learned from the 10 m bands. Based on this idea, Paris et al. [33] also proposed in an extension that makes use of a patch-based regularization to promote the self-similarities of S2 images.

From the most traditional signal reconstruction methods to the most recent learning-based approaches, different SR mechanisms have been successfully applied to spatially enhance RS optical data [28]. However, deep learning-based SR methods certainly represent the most successful trend to super-resolve RS imagery [34] and, in particular, Sentinel products [35]. This is primarily due to the great potential of convolutional neural networks (CNNs) to uncover high-level features from complex visual data. As a result, several SR methods based on CNNs have been presented and tested in the context of the Copernicus programme. A clear example can be found in, where Lanaras et al. [36] defined a CNN model to effectively up-sample S2

products. Specifically, the authors focus their efforts on developing, for the first time in the literature, a CNN-based SR algorithm optimized for S2 in order to super-resolve all the bands to 10 m. Gargiulo et al. [37] also presented in a CNN architecture specially designed for super-resolving S2 data by introducing additional convolutional layers and a revisited loss function.

In these and other relevant works, CNN-based SR models are mainly used to spatially enhance S2 products from a Level-2 (L2) data processing perspective because the MSI sensor acquires some spectral bands at a more reduced spatial resolution than others. However, one can find that there is a serious lack of research on other related instruments, such as the OLCI sensor, which could be super-resolved using CNNs at higher data processing levels. While regular L2 SR data products are constrained by the physical properties of the acquisition instrument, Level-3 (L3) and Level-4 (L4) products are able to provide more powerful results since they combine multitemporal and multisensor information to produce enhanced remotely sensed imagery [38]. In other words, the spatial limitations of an specific instrument could be potentially relieved, without additional costs, by using other complementary satellites together with image enhancement and fusion tools. Some authors in the literature have shown the viability of this idea. It is the case of Pouliot et al. [39] who were able to super-resolve Landsat data using a CNN model and S2 data for training. In the particular context of Sentinel, one of the most important constraints when considering S3 data is related to the limited spatial resolution of the OLCI sensor, which eventually confines the usage of the available L2 S3 products to coarse RS image analyses [21]. Precisely, this is the reason why the research community and industry, with the EU support, are interested in developing enhanced operational Sentinel data to deal with current and future societal challenges and needs [24].

The growing development of L3 and L4 Sentinel products exemplify this emerging trend. For instance, it is the case of Wang and Atkinson who presented in [40] an spatio-temporal fusion approach for S2 MSI and S3 OLCI data, which aims at improving the availability of S2 imagery by taking advantage of the substantially shorter S3 revisit time. In addition, Wu et al. proposed in [41] combining multitemporal biophysical information extracted from MODIS and the MSI sensor to detect cotton plant diseases. Fernandez-Beltran et al. [42] developed a multimodal fusion framework, which merge synthetic aperture radar (from Sentinel-1) and S2 data to effectively conduct an unsupervised land-cover categorization. Additionally, Palsson et al. designed in [43] an intrasensor fusion approach for S2 data, which pursues to improve the spatial resolution of the lowest resolution MSI bands. Other recent works, such as [14], [44], also show the advantages of considering an intersensor scheme to refine the available data products offered by Sentinel's operational chain.

With all these considerations in mind, CNN-based SR models can play a fundamental role in relieving OLCI's spatial limitation while enabling the generation of L4 S3 data products using the MSI imagery as spatial reference. However, this important question has not yet been addressed in the literature, which is

precisely the gap that motivates this work. Inspired by [36] and [39], this research work focuses on the study and development of CNN-based SR techniques to generate spatially enhanced L4 OLCI products exploiting the synergies between the S2 and S3 missions from an operational perspective. In this regard, it is important to highlight that we use the term operational to indicate that the intersensor information is only accessible for training purposes, unlike regular pan-sharpening approaches. That is, the spatial reference of the MSI sensor is only available when training the models and then the SR process can be conducted under demand within the S3 operational chain using exclusively OLCI data.

In more details, this article has a two-fold objective. On the one hand, we intend to set an SR framework for spatially enhancing OLCI data in order to provide practical recommendations for actual Sentinel production environments. Considering the spatial reference provided by S2, our initial objective is based on testing to which extent the OLCI sensor can be effectively super-resolved using different CNN-based SR architectures and scaling ratios while maintaining certain performance improvements with respect to a baseline interpolation. In this way, a very valuable contribution consists in analyzing whether it is worth super-resolving OLCI beyond a particular limit, having in mind the important spatial differences between the OLCI and MSI sensors. Note that this kind of study has not yet been conducted in the literature and some important questions, such as what are the most suitable CNN models for OLCI or what are their practical limitations on generating L4 products, remain still unclear in the context of the Copernicus programme. In addition, we aim at designing an end-to-end CNN-based SR architecture (termed as SRS3) optimized for upscaling S3 data products with particular focus on the following aspects:

1) spatial improvements with respect to different standard CNN-based SR networks;
2) preservation of discriminating features considering the low spatial resolution of OLCI;
3) general applicability for operational data by covering multiple scaling scenarios according to spatial reference given by MSI.

To achieve this goal, the newly proposed architecture effectively integrates a dense multireceptive field together with a residual channel attention mechanism to relieve OLCI's spatial limitations as well as MSI's vast spatial differences. In summary, the main innovative contributions of this article can be compiled as follows.

1) We investigate the suitability of generating super-resolved OLCI products using the MSI sensor of S2 as spatial reference.
2) We develop an SR framework to super-resolve operational OLCI products from an L4 perspective in actual production environments.
3) We analyze the performance of different CNN-based models in the task of super-resolving OLCI data when considering up-scaling ratios closer to MSI.
4) We propose a new SR network (termed as SRS3) specially designed to deal with the particularly low spatial resolution of the OLCI sensor.

The remainder of this article is organized as follows. Section II discusses some related works and also describes the considered CNN-based SR networks. Section III details the proposed framework and SR architecture, which has been specially designed to spatially enhance operational S3 OLCI products from an L4 perspective. In Section IV, a comprehensive experimental assessment of the proposed methodology (in comparison with other state-of-the-art CNN models) is conducted over a newly defined multimodal database including S2 MSI and S3 OLCI data. Finally, Section V concludes this article.

II. BACKGROUND

A. SR in Remote Sensing

Within the RS field, the SR process can be understood as an inverse problem, which pursues to recover the loss of spatial information generated by the acquisition instrument when capturing an aerial scene. That is, the continuous signal of the radiation reflected by the earth surface is converted by the imaging sensor to a discrete output image according to the technical features of the optical system. Specifically, the acquisition process can be typically defined as the composition of three operators, i.e., blurring ($\mathcal{B}$), decimation ($\mathcal{D}$), and noise ($\mathcal{N}$), as

$$\mathbf{I}_{\mathrm{LR}} = \mathcal{D}(\,\mathcal{B}(\,\mathbf{I}_{\mathrm{HR}})) + \mathcal{N} \qquad (1)$$

where $\mathbf{I}_{\mathrm{HR}}$ represents the continuous image scene and $\mathbf{I}_{\mathrm{LR}}$ is the resulting discrete image captured by the RS instrument after applying the $\mathcal{B}$, $\mathcal{D}$, and $\mathcal{N}$ operators. In details, $\mathcal{B}$ symbolizes the initial blurring factor introduced by the optical system in order to filter those frequencies beyond the nominal resolution of the sensor. $\mathcal{D}$ is in charge of subsampling the blurred continuous space in order to integrate all the radiation coming from an specific area of the earth surface in a single pixel. Finally, $\mathcal{N}$ represents the unavoidable perturbation that the imaging sensor introduces according to its own optical limitations. These three operators are specified in the technical description of the acquisition instrument where $\mathcal{B}$ and $\mathcal{N}$ are typically defined as Gaussian-like functions.

The main challenge when super-resolving RS imagery is based on the intrinsic ill-posed nature of the problem together with the special complexity of air-borne and space-borne optical data. On the one hand, the degradation process, described by (1), causes an unavoidable signal information loss that makes multiple HR images to generate the same LR image. As a result, SR methods have to assume strong image priors to constrain the inverse nature of the problem. On the other hand, remotely sensed images are typically full-focused shots with plenty of spatial details, which enlarges the amount of high-frequency that are lost in the acquisition process with respect to standard imagery. Considering all these factors, the most critical problem when super-resolving RS data is often generated by the lack of actual ground-truth spatial information. Note that the final objective consists in improving the resolution of the aerial scenes beyond the sensor limits whatsoever and then higher resolution images cannot be provided by the acquisition instrument itself.

From straightforward interpolation kernels, through image reconstruction models, to more elaborated hybrid techniques,

different approaches have been tested within the RS field to enhance aerial scenes from an intrasensor perspective [28]. For instance, one of the most representative reconstruction approaches is the single-image version of the iterative back projection (IBP) [45]. The main idea behind IBP is based on refining an initial guess of the super-resolved image ($\mathbf{I}_{SR}$) throughout an iterative process where the reconstruction error between the $\mathbf{I}_{LR}$ and a low-resolution version of $\mathbf{I}_{SR}$ is minimized. In details, the reconstruction error on the (i)th iteration can be computed as follows:

$$\mathcal{E}^{(i)} = \mathbf{I}_{LR} - \mathcal{D}(\mathcal{B}(\mathbf{I}_{SR}^{(i-1)})) \tag{2}$$

where $\mathbf{I}_{SR}^{(i-1)}$ represents the super-resolved image at the previous iteration. With $\mathbf{I}_{SR}^0$ initialized to an interpolated version of $\mathbf{I}_{LR}$, the reconstruction error, given by (2), is iteratively reprojected onto the current estimation of the super-resolved image until a convergence condition is reached, with a maximum number of iterations or a global threshold in the error value. Another representative approach that has been also used to super-resolve RS imagery is the transformed self-exemplars (TSE) [46]. The idea behind this hybrid method is based on exploiting the self-similarity property across scales to learn a mapping between LR and HR patches without external training data. Specifically, TSE allows geometric variations among scales in order to increase the number of matches and then the final size of the SR dictionary.

In general, these and other unsupervised SR approaches have the advantage of not requiring any HR training data, which may certainly be a valuable feature in RS. However, their resulting performance highly relies on the nominal spatial resolution of the sensor since they generally work at a lower scale where more spatial details are logically lost [28]. Precisely, this aspect can be a very important drawback when super-resolving S3 optical data due to the low spatial resolution of the OLCI sensor [20]. As a result, the possibility of generating L4 super-resolved OLCI products by using supervised CNN-based SR methods and the S2 MSI sensor as spatial reference for training [39] becomes a very attractive but still unexplored option that this article pretends to shed light on. In this regard, the following section reviews some of the most important SR networks.

B. CNN-Based SR

CNNs represent one of the most important supervised paradigms in SR due to their great potential to uncover high-level features from optical images. Consequently, different CNN-based methods have been successfully employed to super-resolve RS data. A pioneering contribution is the SR convolutional neural network (SRCNN) introduced by Dong *et al.* [47]. In details, the SRCNN approach initially up-scales $\mathbf{I}_{LR}$ to the target resolution by a bicubic interpolation as $\tilde{\mathbf{I}}_{LR}$ and extracts 33×33 image patches. Then, a three-layer CNN is used to learn the mapping between the interpolated LR and ground-truth HR domains. From a conceptual perspective, these three layers represent the following operations: 1) patch extraction, 2) nonlinear mapping, and 3) high-resolution generation. The first two layers make use of the rectified linear unit (ReLU) on the filter responses, i.e., $\max(0, W * X + b)$, where W and b are

the filters and biases, respectively, $*$ is the convolution operator and X represent the patches of $\tilde{\mathbf{I}}_{LR}$. The third layer is a regular convolutional layer with filter responses $W * X + b$. Finally, the mean square error (MSE) loss is used to train the network parameters.

Although the traditional SR scheme was oriented to super-resolve grey scale images, Dong *et al.* [47] showed the possibility of extending the standard SRCNN procedure to MS images. Representing B the number of input and output bands, two different options can be considered in this regard. On the one hand, it is possible to train the SRCNN network using a two-dimensional (2-D) input and output kernel over all the bands. That is, each image band can be considered as a 2-D input image with $B = 1$. On the other hand, another possibility is based on using a 3-D input and output kernel to cover the complete spectral range by setting B to the total number of bands. Despite the potential of 3-D kernels, the substantial computational complexity increase of this kind of convolutions together with the curse of dimensionality problem make the use of 2-D filters typically preferred under operational scenarios. A related model extension is also presented in [48] where Dong *et al.* reformulated the network design in order to reduce the computational burden by shrinking the input feature dimension.

In addition to these reference works, alternative CNN-based SR models have been also proposed by other authors. For instance, it is the case of Kim *et al.* [49] who define in the very deep SR (VDSR) method. Specifically, this approach considers a substantially deeper architecture, a data augmenting scheme together with residual and multiscaling learning. In more details, the VDSR first interpolates the input image to the target resolution as $\tilde{\mathbf{I}}_{LR}$ and extracts 41×41 patches. Then, a total of 20 identical convolutional layers with the ReLU activation function are used to map the up-scaled input to the ground-truth HR domain. At this point, it is important to highlight that a final residual connection is also considered in the network topology to learn only those HR details not present in the interpolated image by means of the MSE loss.

Despite the remarkable performance achieved by these methods, the high complexity of the air-borne and space-borne image domain could still limit the resulting performance in those RS applications, where the spatial precision is important [50]. As a result, more advanced and accurate CNN-based SR models have been also developed in the literature. Among all the conducted research, one of the most positive results are provided by the SR residual network (SRRN) presented by Ledig *et al.* [51]. Specifically, this model takes advantage of multiple residual building blocks to achieve the current state-of-the-art performance in supervised SR. In details, the SRRN architecture starts from the LR image ($\mathbf{I}_{LR}$) and extracts 24×24 image patches. Then, an initial convolutional layer together with 16 residual blocks are used to uncover the corresponding residual feature maps. Each residual block is made of the concatenation of six different layers (Conv2D, BatchNorm, ReLU, Conv2D, BatchNorm, and Add) as well as its corresponding skip connection. After all these residual blocks, an additional convolutional layer with a final skip connection is also considered. This network design allows the model to extract more relevant features from the LR

domain since both residual units and skip connections work for uncovering spatial details on both local and global image areas, which logically have a particular impact in RS. Finally, an efficient pixel shuffle convolutional block is used to upscale the feature maps according to the considered scaling ratio (R) and a regular convolutional layer maps those feature maps onto the final HR domain. Regarding the model optimization, the MSE loss is also used in this case.

Following a similar design, Lim *et al.* [52] developed the enhanced deep residual network (EDSR), which is an evolution of SRRN that pursues to optimize the residual units by removing some layers while stabilizing the training process using a residual scaling. More specifically, all BatchNorm layers are eliminated from residual blocks to provide the network more flexibility to adapt the functional range of the data. Additionally, a constant scaling layer with a 0.1 factor is introduced after the last convolution of each residual block to make the training of the network more stable from a numerical perspective, especially when considering a large number of convolutional filters. With these modifications, EDRS has shown to provide some performance advantages with respect to SRRN when considering moderate scaling ratios, being successfully applied in different domains [53].

Other successful approach is the deep recursive residual network (DRRN) presented by Tai *et al.* [54]. Specifically, DRRN starts by interpolating the input image to the target resolution as $\tilde{\mathbf{I}}_{LR}$ and extracting 31×31 patches. Then, it applies an initial convolution followed by 25 recursive residual blocks composed of 5 concatenated layers (ReLU, Conv2D, ReLU, Conv2D, and Add). Note that, unlike conventional residual units, all these recursive residual blocks share the same weights and biases in order to drastically reduce the number of model parameters while allowing the use of a substantial network depth. At the end, ReLU, Conv2D, and Add layers are used to generate the final super-resolved output with a global skip connection.

An alternative CNN-based technology is also provided by the residual dense network (RDN). In this case, Zhang *et al.* [55] proposed a SR model that takes advantage of the so-called dense blocks, which pursue to avoid learning redundant information by densely connecting all the layers. Specifically, RDN starts from the from the LR image ($\mathbf{I}_{LR}$) and extracts 32×32 image patches. At the beginning of the network, two convolutional layers are used for feature extraction. Then, RDN employs 3 groups of blocks with the following structure: 18 dense blocks made of 3 layers (Conv2D, ReLU, and Concat) together with a final convolution and skip connection. After these components, four more layers (Concat, Conv2D, Conv2D, and Add) are used to create a final dense and global skip connection. Finally, a pixel shuffle layer serves to implement an efficient subpixel convolution with two more Conv2D layers.

In addition to all these architectures, other authors propose exploiting attention mechanisms into deep neural networks. It is the case of Zhang *et al.* [56], who presented the residual channel attention network (RCAN) which employs 10 different groups of residual blocks and a final up-sampling layer to produce the final super-resolved output. In more details, each residual block is made of three initial layers (Conv2D, ReLU, and Conv2D), a
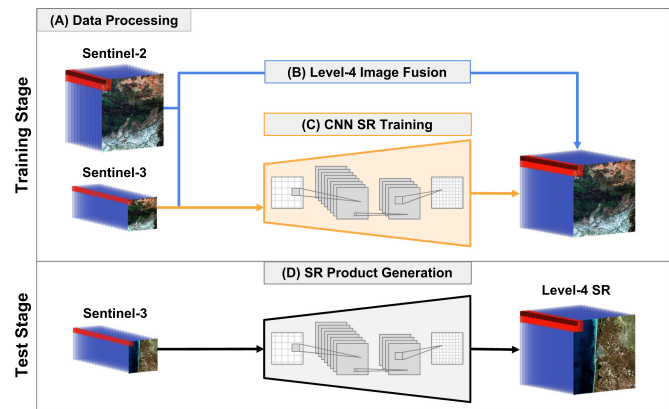

Fig. 1. Proposed Level-4 SR framework for OLCI products.

channel attention block (AvgPool, ReLU, Conv2D, ReLU, and Conv2D) and two final residual connections. With this configuration, RCAN is able to perform two types of skips connections (i.e., short-term and long-term) that allow the model to reach a very large depth. Besides, the considered attention extends the so-called squeeze-and-excitation mechanism [57] to model channelwise relationships with the objective of focusing on the high-frequency information along the SR process.

Notwithstanding the success of these and other relevant models, the particularly low nominal resolution of the OLCI sensor may certainly become an important limitation for super-resolving S3 data from an L4 perspective. Many of the existing CNN-based SR methods are designed assuming small up-scaling ratios and fixed receptive fields where the spatial uncertainty of the SR mapping can be compensated (to some degree) with the spatial details available at the LR image domain. However, unlike other higher resolution acquisition instruments, the lack of high-frequency details in OLCI together with the huge spatial differences with respect to MSI dramatically aggravate the ill-posed nature of the SR problem, which eventually demands an specific network design.

III. METHODOLOGY

This section presents the proposed CNN-based framework specially designed to super-resolve operational S3 OLCI products from an L4 data processing perspective. The rationale behind the proposed approach is based on taking advantage of the higher resolution of the S2 MSI sensor to train a SR network, which can be finally used under demand without the need of having access to S2 MSI data (unlike regular pan-sharpening algorithms). Specifically, the presented scheme consists of four main steps (see Fig. 1): 1) Sentinel data processing, 2) L4 image fusion, 3) CNN-based training, and 4) SR product generation. In the following sections, each one of these parts is depicted in details, however, let us start by defining the notation used for Sentinel data. Let $\mathbf{X} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$ and $\mathbf{Y} = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}\}$ identify two image collections of N S3 OLCI and S2 MSI reflectance products, where each image pair represents the same area on the earth surface. For the sake of simplicity, we assume

that MSI bands have been resampled to an uniform spatial resolution of 20 m. In this scenario, $\mathbf{x}^{(i)}$ contains $(x_1 \times x_2)$ pixels defined over 21 spectral bands whereas the $(y_1 \times y_2)$ pixels of $\mathbf{y}^{(i)}$ include 13 bands. Logically, $x_1 << y_1$ (and $x_2 << y_2$) due to the huge spatial resolution differences between OLCI and MSI sensors, being such difference y_1/x_1 (or equivalently y_2/x_2). Let R be the considered natural scaling factor for super-resolving S3 OLCI products, which can range between 2 and $\lfloor y_1/x_1 \rfloor$. Let $\tilde{\mathbf{Y}} = \{\tilde{\mathbf{y}}^{(1)}, \ldots, \tilde{\mathbf{y}}^{(N)}\}$ be a down-scaled version of $\mathbf{Y}$ (using a bicubic kernel), such that $\tilde{\mathbf{y}}^{(i)} \in \mathbb{R}^{(x_1 R \times x_2 R \times 13)}$. Additionally, let $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}^{(1)}, \ldots, \tilde{\mathbf{x}}^{(N)}\}$ be the up-scaled version of $\mathbf{X}$ with a bicubic kernel, where $\tilde{\mathbf{x}}^{(i)} \in \mathbb{R}^{(x_1 R \times x_2 R \times 21)}$. Finally, let $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(N)}\}$ be the super-resolved version of $\mathbf{X}$ and $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}^{(1)}, \ldots, \hat{\mathbf{y}}^{(N)}\}$ its corresponding spatial ground-truth.

A. Sentinel Data Processing

Although the specific data considered in this article is detailed in Section IV-A, these lines describe the general procedure required to obtain adequate intersensor Sentinel products for training purposes and all the required corrections. Note that the proposed framework tackle the SR problem from an L4 perspective and, hence, it is first necessary to acquire coupled S2 MSI and S3 OLCI training products from the Copernicus Open Access Hub (OAH) at https://scihub.copernicus.eu/. In other words, this initial data processing step aims at obtaining S2 training images that cover the same area of the considered S3 products taking into account the important differences between the OLCI and MSI sensors. Specifically, we propose using the following processing chain for obtaining such intersensor data.

1) *Data collection*: This step consists in selecting and downloading from the OAH platform coupled MSI and OLCI products of interest across the globe. Despite the fact that conducting this task seems trivial, it is important to remark some technical points and recommendations in order to generate a robust intersensor training collection. First, the considered areas of interest should include different types of environments in order to consider a representative data diversity, such as natural parks, coastal areas, mountain ranges, or cities. Second, the products covering the selected regions should not contain clouds since these atmospheric perturbations logically affect the availability of spatial information and may also alter other necessary corrections. In this regard, the OAH platform does not allow filtering the OLCI products by percentage of cloud coverage, hence we advocate doing this filtering process over the MSI sensor. Considering the substantially wider field of view in S3, we recommend considering MSI products without clouds and then selecting the temporally closest available OLCI product that includes that region. Finally, we suggest downloading from the OAH platform L1 data products in order to obtain both MSI and OLCI optical images at the same processing level across the globe.

2) *Product corrections*: Once the initial collection process is completed, it is necessary to apply some additional adjustments to represent both MSI and OLCI products as
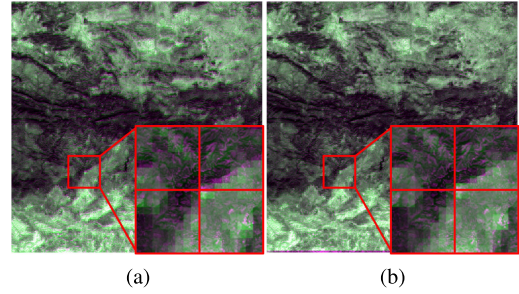

(a)   (b)

Fig. 2. Differences between the S2 MSI (in purple) and the interpolated S3 OLCI (in green) images when considering the original OLCI product (a) and the processed one (b).

corrected reflectance images. To this aim, we recommend using the official sentinel application platform (SNAP) software provided by the ESA. In the case of S2, MSI products can be atmospherically corrected using the Sen2Core processor, which employs the Atmospheric/Topographic Correction for Satellite Imagery algorithm [58]. In the case of S3, such corrections are still not available for OLCI land products, which makes necessary to use the radiance to reflectance processor or even other alternative algorithms [59].

3) *Intersensor projection*: The next step is based on projecting each intersensor product pair onto a common coordinate system to effectively compute the intersection between the MSI and OLCI products. To this extent, the SNAP software is able to carry out this operation in a batch processing mode. In details, the MSI data is initially resampled to a 20 m spatial resolution to enable the use of the projection functionality. Then, the corresponding OLCI products are reprojected onto their corresponding MSI counterparts to obtain the intersection between both products as output. Note that the resulting intersection area is given by the MSI sensor due to its higher spatial resolution. After this process, we obtain an specific OLCI product (300 m) as $\mathbf{x}^{(i)} \in \mathbb{R}^{366 \times 366 \times 21}$ and its corresponding MSI intersection (20 m) as $\mathbf{y}^{(i)} \in \mathbb{R}^{5490 \times 5490 \times 13}$.

4) *Intersensor alignment*: Since there are very important spatial differences between the MSI and OLCI sensors, the last step aims at applying some additional spatial corrections to obtain a more consistent intersensor optical data. In general, the multitemporal registration requirement for operational Sentinel products is 0.3 pixels. However, even a small spatial deviation in OLCI can produce a large spatial displacement in MSI because geometric errors are magnified by the important spatial resolution difference. In this last step, we propose using the intersensor registration technique described in [44] to geometrically refine the OLCI data using the MSI sensor as spatial reference. Fig. 2 shows an example of the differences between the S2 MSI (in green color) and the interpolated S3 OLCI (in purple color) products to highlight the effect of the proposed data processing step. As it is possible to observe, there are important differences between the MSI and the

original operational OLCI product that can be substantially relieved when applying the proposed process.

B. Level-4 Image Fusion

Since the proposed framework aims at generating L4 super-resolved OLCI products within the operational chain of S3, the spatial reference provided by the MSI sensor can be only used for training purposes. Precisely, the target of the previously presented data processing step (see Section III-A) is based on generating such intersensor product collection for training. Nonetheless, it is important to highlight that the final objective of the proposed methodology does not consist in generating MSI data but spatially enhanced OLCI products, which eventually means that the SR mapping cannot be directly learned over MSI spectra. Although a reduced-reference learning protocol could be considered to learn such mapping at a lower resolution scale [28], the very low nominal resolution of the OLCI sensor (with respect to MSI) makes this scheme unfeasible from a practical perspective [14]. As a result, we define this second step to produce a simulated HR version of the OLCI data by means of an L4 image fusion approach. In other words, we make use of a RS image fusion approach to generate ground-truth OLCI data at MSI spatial resolution, i.e., $\hat{\mathbf{Y}}$.

In the RS literature, multiple image fusion techniques have been proposed based on different paradigms, e.g., component substitution, multiresolution analysis, etc. However, Bayesian fusion methods have shown particularly positive results with a sufficient spectral overlap between the fused sensors [60]. In the case of OLCI and MSI, the spectral range of the former sensor (390–1040 nm) is considerably covered by the latter (443–2190 nm), which makes the Bayesian approach based on linear subspace transformations a reasonable choice. Specifically, we select the fast fusion based on Sylvester equation (FUSE) method [61] due to its tradeoff between computational time and qualitative performance according to the experimental results reported in [60]. Nonetheless, it is important to note that any other fusion method could be selected instead since the objective of this second step is just based on generating valid ground-truth imagery for super-resolving OLCI data. Representing $\mathbf{X}$ the collection of OLCI products and $\tilde{\mathbf{Y}}$ their corresponding downscaled MSI counterparts, we define the L4 spatial ground-truth as $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}^{(1)}, \ldots, \hat{\mathbf{y}}^{(N)}\}$ where $\hat{\mathbf{y}}^{(i)} = \text{FUSE}(\mathbf{x}^{(i)}, \tilde{\mathbf{y}}^{(i)})$ [61].

C. CNN-Based SR Training

Once the spatial ground-truth has been generated via L4 image fusion, the third step of the proposed methodology consists in training a CNN architecture to map OLCI images ($\mathbf{X}$) onto their corresponding higher resolution counterparts ($\hat{\mathbf{Y}}$). Although any of the aforementioned CNN-based SR models could be integrated within the proposed framework, the particular nature of the OLCI sensor makes necessary to account for the specific features of this kind of SR application. Specifically, it is possible to identify two issues that play a key role when generating spatially enhanced L4 OLCI products: coarse spatial resolution and large up-scaling ratios. On the one hand, the spatial resolution of the OLCI sensor may certainly become

an important constraint for many of the available CNN-based SR networks. Unlike standard imagery from popular RS data sources [62], OLCI's spatial resolution is rather coarse (300 m), which seriously limits the availability of spatial details in the input data while aggravating the ill-posed nature of the SR problem. In this sense, the probability of facing noninformative patches when super-resolving OLCI products logically increases (with respect other higher-resolution sensors), which eventually demands an special SR model design. On the other hand, the scaling ratios that can be considered for super-resolving OLCI products depend on the spatial resolution of the ground-truth data. Since we use the MSI sensor as spatial reference, OLCI data could be potentially up-scaled to MSI's spatial resolution, which enables the possibility of dealing with huge scaling factors and also studying whether it may be worthwhile. In this regard, it is important to note that many of the existing CNN-based SR methods have been designed assuming small up-scaling ratios (e.g., $2\times$, $4\times$). Therefore, it also becomes necessary to account for the adequate complexity and generality of the developed network under different ratios.

To overcome these limitations, we propose a new SR architecture, which has been specially developed for effectively dealing with OLCI data when considering from small to large up-scaling factors. Due to the huge spatial differences between OLCI and MSI sensors, the first important point to take into consideration is how to implement the up-scaling process within the proposed architecture. Whereas adopting up-scaling layers inside the network (e.g., transposed convolutions, pixel shuffling, etc.) has shown to obtain positive results with standard SR settings [51], [55], the possibility of considering very large scaling ratios to cover the spatial gap between OLCI and MSI sensors strongly constrains the effectiveness of this scheme in our context. Note that the bigger the scaling ratio the more drastic the increase on the number of parameters inside the network and, hence, the more difficult its global convergence. Although the use of more than one up-scaling layer may help, it also reduces the capability of the network to work with certain scaling factors (e.g., prime numbers), which eventually limits the generality of the SR model. To relieve these effects, we opt for starting from the interpolated version of the OLCI image, i.e., the proposed network is focused on learning the mapping between $\tilde{\mathbf{X}}$ and $\hat{\mathbf{Y}}$.

Another important aspect is based on OLCI's nominal resolution. As it was previously mentioned, the coarse spatial resolution of the OLCI sensor may cause important issues in learning the SR projection due to the lack of sufficient discriminating details in the input domain. Besides, this problem is exacerbated by the consideration of larger scaling factors to reach spatial resolutions closer to MSI. In this scenario, it is reasonable to think that increasing the network receptive field can help to produce better HR estimates. However, the direct application of wider convolutions does not guarantee a good SR performance across all scaling ratios since the ideal receptive field may logically vary among images and scales. In response, we develop a new SR network, which pursues to adaptively exploit multiple receptive fields with the objective of increasing the contextual information of the OLCI sensor when required. Fig. 3 shows a graphical visualization of the defined architecture,
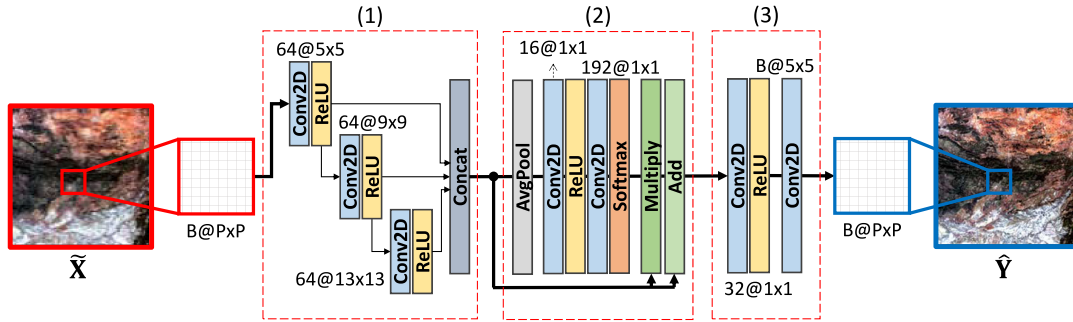
Fig. 3.   Graphical representation of the proposed SR network, including (1) dense multireceptive field, (2) residual channel attention, and (3) HR projection.

which works with $B@P \times P$ input and output patches. In more details, it is made of the following sequential components.

1) *Dense multireceptive field*: The initial part of the proposed network aims at extracting multiple feature representations with different receptive fields and levels of abstraction in order to provide a complete data characterization useful in a wider range of up-scaling settings. Specifically, it is made of the following seven layers: 1) Conv2D, 2) ReLU, 3) Conv2D, 4) ReLU, 5) Conv2D, 6) ReLU, and 7) Concat. As it is possible to see in Fig. 3, all three convolutional layers use 64 filters with an increasing kernel size of 5×5, 9×9 and 13×13, respectively. Besides, they are densely connected via the last concatenation layer. This configuration allows extracting feature maps related to an initially small receptive field, which is hierarchically expanded to provide multiple representations at different abstraction levels. In this way, the obtained characterizations are able to involve broader spatial contexts and features in order to relieve the particularly low spatial resolution of the OLCI sensor.

2) *Residual channel attention*: Once the corresponding feature maps have been extracted, the second part of the proposed network pursues to automatically filter those data characterizations that provide the most informative features for the target super-resolution scale. That is, depending on the input image and the desired up-scaling ratio (R), this module adapts the size of the receptive field and feature abstraction level since the coarse spatial resolution of OLCI can easily make these settings rather volatile. Note that 300 m is an spatial resolution that can often produce image segments with little spatial details due to the hard quantization of highly heterogeneous pixels. In these circumstances, it becomes very adequate to expand the receptive field as well as the abstraction level to uncover more discriminating features that help to better compute the projection onto the HR domain. To achieve this goal, we define a new building block based on the so-called squeeze-and-excitation weighting scheme [63]. As Fig. 3 shows, the defined residual channel attention block contains the following seven layers: 1) AvgPool, 2) Conv2D, 3) ReLU, 4) Conv2D, 5) Softmax, 6) Multiply, and 7) Add. The rationale behind this structure is based on initially averaging each feature channel. Then, an spectral encoder–decoder architecture (with 64 and 192 kernels,

respectively) is adopted to generate the corresponding channel weights after a softmax operation. Finally, these normalized scalar values are used to weight the input channels and a final residual connection works for focusing the attention on discriminating channel dependencies to avoid overfitting.

3) *HR projection*: After generating the corresponding feature maps with recalibrated channels, the last part of the proposed approach is targeted at projecting this highly informative data volume onto the final HR space. Note that we started from the interpolated OLCI space (given by $\tilde{\mathbf{X}}$) to relieve the problems of considering large scaling ratios. Hence, there is no need to use any up-scaling layer at this point of the network design. Specifically, we employ the following layers: 1) Conv2D, 2) ReLU, 3) Conv2D, and 4) ReLU, where the first layer employs 32 filters for conducting an spectral convolution and the third layer reconstructs the input data volume using a 5×5 kernel size.

Let us identify the proposed network as $\mathbf{S}(\cdot; \Theta)$, being Θ the network parameters. Finally, $\mathbf{S}$ is trained to optimize the following loss function, based on the MSE figure of merit, which is minimized via the ADAM optimizer using a learning rate of $1e^{-4}$ for a total of 100 epochs as

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\mathbf{y}}^{(i)} - \mathbf{S}(\tilde{\mathbf{x}}^{(i)}; \Theta) \right)^2. \quad (3)$$

D. SR Product Generation

The last step of the proposed methodology is focused on exploiting the trained CNN-based SR model under operational conditions. In this sense, it is important to emphasize that MSI data is not available from the operational processing chain of S3, place where this article puts the focus in order to generate L4 spatially enhanced OLCI products under demand. Being $\mathbf{x}^{(\text{tst})}$ a test OLCI product and $\tilde{\mathbf{x}}^{(\text{tst})}$ its corresponding up-scaled version by an R factor using a bicubic interpolation kernel, this final step computes the super-resolved version of the product as $\hat{\mathbf{x}}^{(\text{tst})} = \mathbf{S}(\tilde{\mathbf{x}}^{(\text{tst})}; \Theta)$. Note that, similarly as training, test images are processed in a patch-based fashion to generate the corresponding results.

TABLE I
DESCRIPTION OF THE CONSIDERED PRODUCTS

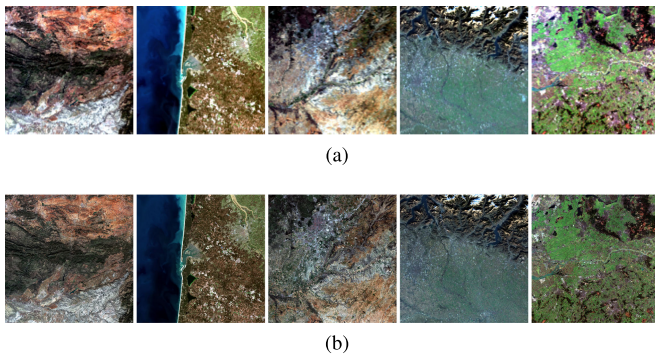| Name		Scene	Location	Sensing dates		Tile (ref. S2)
				S3 OLCI	S2 MSI	
TRA	AN-tra	Natural park	Andujar (Spain)	10/03/2017	10/03/2017	30SVH
	BR-tra	Coastal area	Bourdeaux (France)	19/04/2017	19/04/2017	30TXQ
	MA-tra	Southern Europe	Madrid (Spain)	10/04/2017	09/04/2017	30TVK
	MI-tra	Mountain range	Milan (Italy)	16/02/2017	16/02/2017	32TNR
	UT-tra	Northern Europe	Utrecht (Netherlands)	15/02/2017	27/03/2017	31UFT
TST	AN-tst	Natural park	Andujar (Spain)	30/10/2016	31/10/2016	30SVH
	BR-tst	Coastal area	Bourdeaux (France)	10/03/2017	10/03/2017	30TXQ
	MA-tst	Southern Europe	Madrid (Spain)	23/05/2017	10/03/2017	30TVK
	MI-tst	Mountain range	Milan (Italy)	28/12/2016	07/01/2017	32TNR
	UT-tst	Northern Europe	Utrecht (Netherlands)	27/12/2016	27/12/2016	31UFT

(a)

(b)

Fig. 4. S3 OLCI image scenes (a) and their S2 MSI counterparts (b).

IV. EXPERIMENTS

A. Datasets

In this article, we consider ten overlapped S2 MSI and S3 OLCI operational data products, which comprise five different European areas of interest. Table I describes the selected scenes where five image pairs have been employed for training and the other five pairs for testing. All the images are cloud free operational Level-1 C data products, which have been downloaded from the Copernicus OAH platform (https://scihub.copernicus.eu/) and processed using the SNAP software with its standard settings. On the one hand, S2 MSI products have been atmospherically corrected using the Sen2Cor processor and resampled to 20 m spatial resolution. On the other hand, S3 OLCI products have been corrected using the Radiance to Reflectance processor and they have also been reprojected to their associated S2 tiles in order to obtain the intersection area between both images. After all these steps, the corresponding spatial image sizes are 5490×5490 pixels in S2 and 366×366 in S3. Fig. 4 shows the S3 OLCI and S2 MSI data products used in this article for bench-marking.

B. Experimental Settings

To validate the performance of the proposed network, we make use of the presented methodology to conduct several experiments using a broad range of scaling ratios and state-of-the-art SR models. On the one hand, we test all the possible up-scaling factors between the considered intersensor resolutions, i.e., $R = \{2, 3, 4, \ldots, 15\}$, with the objective of studying

TABLE II
PSNR (dB) ASSESSMENT FOR AN-TST

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	25.15	25.50	25.25	25.50	25.19	25.23	25.48	25.43	**25.69**
3x	23.47	23.74	23.52	23.76	23.56	23.51	23.68	23.68	**23.85**
4x	22.59	22.76	22.61	**22.89**	22.60	22.61	22.77	22.75	22.89
5x	22.03	22.21	22.05	20.97	20.88	22.05	20.93	20.94	**22.28**
6x	21.71	21.87	21.72	19.97	19.85	21.71	19.88	19.90	**21.93**
7x	21.45	21.59	21.47	19.31	19.23	21.46	19.28	19.28	**21.64**
8x	21.21	21.34	21.23	18.88	18.79	21.22	18.78	18.84	**21.39**
9x	21.06	21.19	21.07	18.57	18.44	21.07	18.49	18.51	**21.23**
10x	21.03	21.14	21.04	18.38	18.31	21.04	18.35	18.34	**21.19**
11x	21.02	21.13	21.03	18.22	18.17	21.03	18.21	18.21	**21.18**
12x	21.02	21.13	21.03	18.13	18.05	21.03	18.08	18.11	**21.18**
13x	21.02	21.10	21.03	18.01	17.95	21.02	18.01	18.02	**21.17**
14x	21.02	21.11	21.03	17.97	17.91	21.02	17.92	17.95	**21.16**
15x	20.95	21.04	20.96	17.87	17.82	20.95	17.84	17.85	**21.08**
Avg.	21.76	21.92	21.79	19.89	19.77	21.78	19.84	19.84	**21.99**

to which extent the 300 m spatial resolution of the OLCI sensor can be satisfactorily super-resolved using the 20 m resolution of MSI as operational reference. On the other hand, we assess in the experimental comparison the performance of some of the most relevant SR architectures available in the literature, including SRCNN [47], VDSR [49], SRRN [51], EDSR [52], DRRN [54], RDN [55], and RCAN [56]. In addition, we use the bicubic interpolation kernel (BC) as up-scaling baseline.

Under these settings, we train all the considered SR networks (including the proposed architecture named as SRS3) using the same input data and training configurations. For each scaling ratio R, we extract image patches from the training set using a $14R$-pixel step in order to select over $5e^4$ patches for training regardless the input size defined by each particular SR method. In the proposed approach case, a 33×33 patch size is used (i.e., $P = 33$). Besides, each OLCI band is considered as an input image (i.e., $B = 1$) to reduce the spatial complexity of all the models taking into account the large scaling ratios tested in this article. All the networks have been trained using the ADAM optimizer for 100 epochs with a $1e^{-4}$ learning rate and a 32 batch size. Regarding the evaluation protocol, two different full-reference metrics are used to quantify the quality of the results: the peak signal-to-noise ratio (PSNR) and the spectral angle mapper (SAM). Whereas PSNR is focused on measuring the reconstruction quality from the super-resolved and ground-truth images, SAM calculates the angular deviations between both spectra. Additionally, several visual results are also considered for a complementary qualitative evaluation of the obtained results. The hardware and software environments used in this article are the following: Intel(R) Core(TM) i7-6850 K, NVIDIA GeForce GTX 1080 Ti, 64 Gb of DDR4 RAM, Ubuntu 20.04 $\times 64$, and Pytorch 1.6.0 with CUDA 10.1. The related codes will be publicly available at https://github.com/rufernan/SRS3.

C. Results

Tables II–VI present the quantitative evaluation based on the PSNR metric for the considered test datasets. Note that each table includes the assessment of a different test scene (i.e., AN-tst, BR-tst, MA-tst, MI-tst, and UT-tst) and all the results are expressed in dB. In more details, tables are organized with

TABLE III
PSNR (dB) Assessment for BR-Tst

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	28.93	28.93	28.88	**29.04**	28.89	28.91	28.65	28.89	29.00
3x	27.44	27.53	27.41	**27.54**	27.53	27.44	27.48	27.46	27.49
4x	26.64	26.64	26.61	26.64	26.62	26.64	26.63	26.71	**26.71**
5x	26.11	26.13	26.09	23.98	23.91	26.11	23.93	23.94	**26.19**
6x	25.80	25.82	25.79	22.55	22.49	25.79	22.53	22.53	**25.86**
7x	25.54	25.56	25.52	21.75	21.70	25.53	21.73	21.74	**25.59**
8x	25.29	25.30	25.27	21.19	21.20	25.28	21.19	21.20	**25.33**
9x	25.12	25.14	25.11	20.84	20.78	25.11	20.83	20.82	**25.16**
10x	25.10	25.09	25.08	20.62	20.60	25.09	20.59	20.60	**25.12**
11x	25.09	25.10	25.08	20.42	20.41	25.09	20.41	20.42	**25.11**
12x	25.09	25.10	25.08	20.26	20.24	25.09	20.28	20.26	**25.12**
13x	25.10	25.04	25.08	20.17	20.12	25.09	20.14	20.15	**25.11**
14x	25.10	25.09	25.09	20.08	20.07	25.09	20.05	20.07	**25.10**
15x	**25.01**	24.97	25.00	19.96	19.96	25.00	19.95	19.96	25.00
Avg.	25.81	25.82	25.79	22.50	22.47	25.81	22.46	22.48	**25.85**

TABLE IV
PSNR (dB) Assessment for MA-Tst

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	23.31	23.48	23.27	**23.49**	23.33	23.32	23.49	23.46	23.42
3x	21.95	**22.08**	21.92	21.98	21.99	21.94	22.06	22.06	22.01
4x	21.13	21.20	21.09	**21.22**	21.12	21.12	21.21	21.21	21.18
5x	20.57	**20.65**	20.54	19.50	19.50	20.56	19.54	19.54	20.60
6x	20.23	**20.30**	20.21	18.58	18.54	20.22	18.57	18.58	20.26
7x	19.95	**20.01**	19.93	17.98	17.96	19.94	17.99	17.99	19.98
8x	19.68	**19.72**	19.66	17.54	17.53	19.67	17.53	17.56	19.71
9x	19.51	**19.56**	19.49	17.25	17.20	19.50	17.23	17.24	19.54
10x	19.48	**19.52**	19.46	17.09	17.07	19.47	17.09	17.09	19.51
11x	19.47	**19.52**	19.46	16.95	16.94	19.47	16.97	16.97	19.51
12x	19.47	**19.52**	19.46	16.86	16.84	19.47	16.86	16.87	19.51
13x	19.47	19.51	19.47	16.77	16.75	19.47	16.79	16.79	**19.51**
14x	19.47	**19.52**	19.47	16.71	16.71	19.47	16.71	16.73	19.51
15x	19.39	**19.43**	19.38	16.60	16.61	19.39	16.63	16.63	19.42
Avg.	20.22	**20.29**	20.20	18.47	18.43	20.22	18.48	18.48	20.26

TABLE V
PSNR (dB) Assessment for MI-Tst

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	28.78	28.71	29.05	28.66	28.69	28.89	28.61	28.71	**29.62**
3x	27.01	27.10	27.27	26.68	27.04	27.14	27.03	27.03	**27.54**
4x	26.08	26.20	26.27	25.79	26.08	26.18	26.10	26.10	**26.54**
5x	25.47	25.60	25.63	24.32	24.47	25.54	24.52	24.51	**25.87**
6x	25.10	25.22	25.23	23.38	23.53	25.14	23.58	23.58	**25.43**
7x	24.79	24.91	24.91	22.89	22.94	24.83	22.98	22.99	**25.08**
8x	24.51	24.61	24.60	22.42	22.55	24.53	22.54	22.56	**24.77**
9x	24.32	24.38	24.39	22.23	22.22	24.34	22.22	22.21	**24.55**
10x	24.28	24.35	24.34	22.07	22.08	24.29	22.08	22.10	**24.49**
11x	24.27	24.32	24.32	21.86	21.96	24.28	21.97	21.98	**24.46**
12x	24.26	24.31	24.31	21.81	21.83	24.27	21.87	21.89	**24.43**
13x	24.25	24.24	24.30	21.71	21.79	24.26	21.82	21.82	**24.41**
14x	24.25	24.30	24.29	21.67	21.75	24.26	21.73	21.76	**24.39**
15x	24.16	24.16	24.20	21.54	21.64	24.17	21.64	21.65	**24.29**
Avg.	25.11	25.17	25.22	23.36	23.47	25.15	23.48	23.49	**25.42**

TABLE VI
PSNR (dB) Assessment for UT-Tst

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	25.88	25.98	25.87	**25.99**	25.94	25.89	25.93	25.97	25.96
3x	23.94	**24.05**	23.96	23.46	24.04	23.97	24.04	24.03	24.05
4x	22.88	22.93	22.90	22.74	22.92	22.90	22.95	22.98	**23.02**
5x	22.18	22.24	22.19	21.26	21.33	22.19	21.36	21.37	**22.31**
6x	21.74	21.80	21.76	20.24	20.38	21.74	20.42	20.44	**21.85**
7x	21.38	21.43	21.39	19.74	19.78	21.38	19.83	19.84	**21.47**
8x	21.03	21.07	21.04	19.29	19.36	21.03	19.35	19.37	**21.12**
9x	20.81	20.85	20.81	19.03	19.01	20.80	19.04	19.04	**20.88**
10x	20.76	20.78	20.77	18.87	18.89	20.76	18.90	18.91	**20.84**
11x	20.75	20.79	20.76	18.68	18.78	20.75	18.80	18.80	**20.82**
12x	20.75	20.79	20.76	18.66	18.67	20.75	18.72	18.71	**20.82**
13x	20.75	20.77	20.76	18.52	18.63	20.75	18.65	18.66	**20.81**
14x	20.75	20.77	20.76	18.55	18.60	20.75	18.59	18.61	**20.81**
15x	20.64	20.64	20.64	18.42	18.48	20.63	18.48	18.49	**20.68**
Avg.	21.73	21.78	21.74	20.25	20.34	21.74	20.36	20.37	**21.82**

TABLE VII
Average PSNR (dB) Assessment for Tst Datasets

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	26.41	26.52	26.46	26.54	26.41	26.45	26.43	26.49	**26.74**
3x	24.76	24.90	24.81	24.68	24.83	24.80	24.86	24.85	**24.99**
4x	23.86	23.95	23.90	23.85	23.87	23.89	23.93	23.95	**24.06**
5x	23.27	23.37	23.30	22.01	22.02	23.29	22.06	22.06	**23.45**
6x	22.91	23.00	22.94	20.94	20.96	22.92	21.00	21.01	**23.07**
7x	22.62	22.70	22.64	20.33	20.32	22.63	20.36	20.37	**22.75**
8x	22.34	22.41	22.36	19.86	19.89	22.35	19.88	19.90	**22.46**
9x	22.16	22.22	22.18	19.58	19.53	22.16	19.56	19.57	**22.27**
10x	22.13	22.18	22.14	19.41	19.39	22.13	19.40	19.41	**22.23**
11x	22.12	22.17	22.13	19.22	19.25	22.12	19.27	19.28	**22.22**
12x	22.12	22.17	22.13	19.14	19.12	22.12	19.16	19.17	**22.21**
13x	22.12	22.13	22.13	19.03	19.05	22.12	19.08	19.09	**22.20**
14x	22.12	22.16	22.13	19.00	19.01	22.12	19.00	19.02	**22.19**
15x	22.03	22.05	22.03	18.88	18.90	22.03	18.91	18.92	**22.10**
Avg.	22.93	22.99	22.95	20.89	20.90	22.94	20.92	20.93	**23.07**

TABLE VIII
Average SAM (rad) Assessment for Tst Datasets

R	BC	SRCNN	VDSR	SRRN	EDSR	DRRN	RDN	RCAN	SRS3
2x	0.104	0.102	0.104	0.102	0.103	0.104	0.103	0.103	**0.101**
3x	0.125	0.123	0.125	0.126	0.124	0.125	0.123	0.123	**0.123**
4x	0.138	0.136	0.138	0.138	0.138	0.138	0.137	0.137	**0.136**
5x	0.148	0.146	0.148	0.169	0.147	0.148	0.168	0.168	**0.145**
6x	0.154	0.152	0.154	0.191	0.153	0.154	0.190	0.189	**0.152**
7x	0.159	0.157	0.159	0.205	0.158	0.159	0.204	0.204	**0.157**
8x	0.164	0.163	0.164	0.215	0.164	0.164	0.215	0.215	**0.162**
9x	0.168	0.166	0.167	0.223	0.167	0.168	0.223	0.223	**0.166**
10x	0.168	0.167	0.168	0.227	0.168	0.168	0.227	0.227	**0.166**
11x	0.168	0.167	0.168	0.232	0.168	0.168	0.231	0.231	**0.167**
12x	0.168	0.167	0.168	0.234	0.168	0.168	0.234	0.233	**0.167**
13x	0.168	0.167	0.168	0.237	0.168	0.168	0.236	0.236	**0.167**
14x	0.168	0.167	0.168	0.238	0.168	0.168	0.238	0.237	**0.167**
15x	0.170	0.169	0.170	0.241	0.170	0.170	0.240	0.240	**0.169**
Avg.	0.155	0.154	0.155	0.198	0.155	0.155	0.198	0.198	**0.153**

the considered scaling ratios in rows and with the tested SR methods in columns. Besides, the last row of each table provides the corresponding average PSNR values and the best results are highlighted in bold font. To complete the quantitative analysis, Table VIII also details the average SAM results (in radians) per scaling factor and method. For qualitative purposes, Fig. 5 displays the super-resolved OLCI products generated by the considered methods over the MI-tst scene using a 2× up-scaling ratio.

According to the reported results, there are some important aspects that deserve to be mentioned. Specifically, the first noticeable point is related to the general performance of the tested methods as the considered scaling ratio increases. Although all the SR architectures logically show a global performance decrease when applying higher scaling factors, it is possible to appreciate significant differences between the models that use an initial interpolation and the rest. Whereas SRCNN, VDSR, and DRRN are able to produce consistent improvements with respect to the bicubic baseline (BC), SRRN, EDSR, RDN, and RCAN certainly suffer an important degradation for up-scaling ratios
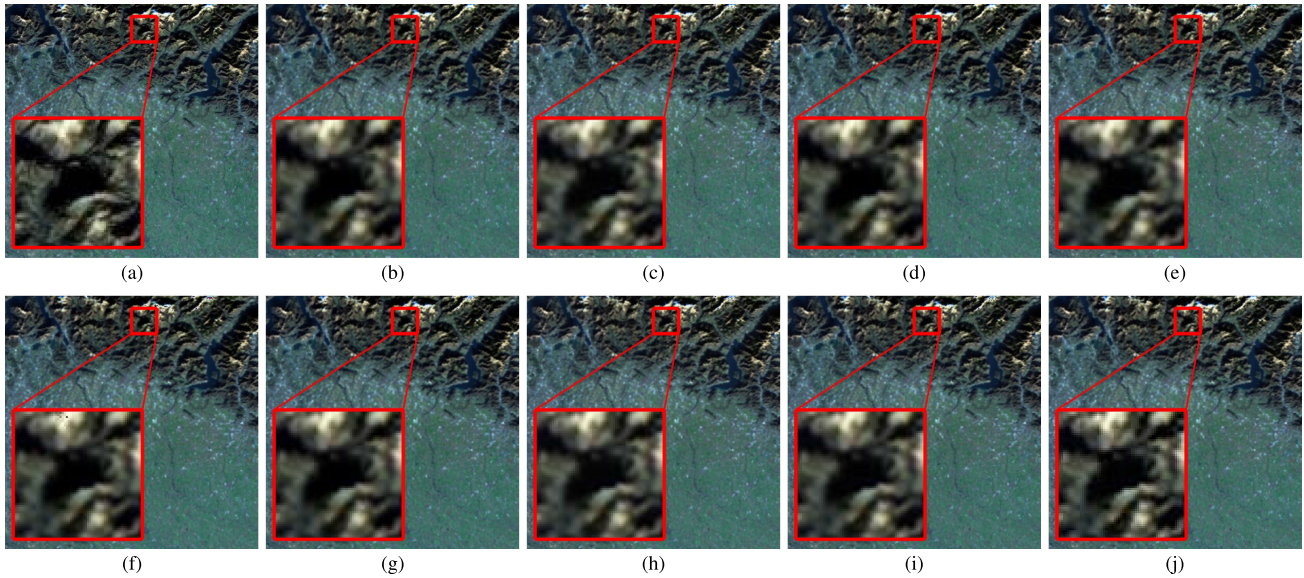
Fig. 5.   Qualitative results for MI-tst with $R = 2$: (a) Ground-truth, (b) BC, (c) SRCNN, (d) VDSR, (e) SRRN, (f) EDSR, (g) DRRN, (h) RDN, (i) RCAN, and (j) SRS3.

higher than $4\times$. As Tables II–VI show, SRRN, EDSR, RDN, and RCAN are able to obtain some PSNR gains (over BC) when using $R = \{2, 3, 4\}$. However, the situation is rather different for larger factors where the other tested methods (SRCNN, VDSR, and DRRN) clearly generate better SR results. As an example, we can see that SRCNN and RDN obtain in Table II the best and third best performances when using $4\times$, however, their results get dramatically worse from $5\times$ to $15\times$. Note that SRRN, EDSR, RDN, and RCAN employ a final up-scaling layer to generate the super-resolved output but this strategy is unable to satisfactorily manage huge ratios due to the dramatic increase on the number of parameters at the end of the network. In contrast, SRCNN, VDSR, and DRRN start from the interpolated original space to keep the feature maps more stable along the network, which eventually becomes more adequate for relieving the huge spatial resolution difference between OLCI and MSI. Precisely, this is the reason why the proposed approach introduces this initial interpolation scheme to generate spatially enhanced L4 OLCI products.

Focusing on the average PSNR results provided in Table VII, we can make some additional observations regarding the performance of the best performing competitors as well as the proposed network. On the one hand, VDSR and DRRN show a very reduced PSNR improvement with respect to the BC when considering high scaling ratios. Despite the fact that both methods are able to consistently outperform the baseline (unlike SRRN, EDSR, RDN, and RCAN), their quantitative gains are rather negligible for factors beyond $7\times$, which reveals the ineffectiveness of these networks in the most challenging scenarios. In more details, both architectures make use of a fix small kernel size (3×3) for substantially increasing the number of convolutional layers and residual connections. Nonetheless, the lack of sufficient high-frequency information with large up-scaling ratios makes these small convolutions unable to extract discriminating features from the interpolated original space.

On the other hand, SRCNN is certainly the best performing competitor for super-resolving OLCI products since it is able to provide more relevant PSNR improvements, which range from $+0.11$ dB ($2\times$) to $+0.05$ dB ($10\times$) on average. The expanded receptive field of SRCNN (9×9) together with its higher topological simplicity make this network to better exploit the limited spatial information provided by the OLCI sensor.

Regarding the performance achieved by the proposed model, Table VII clearly reveals that the newly defined architecture provides the most accurate results, exhibiting average PSNR gains from $+0.33$ dB ($2\times$) to $+0.10$ dB ($10\times$). Although these values could initially seem slight improvements over the bicubic baseline, they twice the average gains obtained by the best performing competitors at each scaling ratio (i.e., SRRN at $2\times$ and SRCNN otherwise), which becomes particularly relevant within the context of this article. Note that the limited resolution of the OLCI sensor together with the huge spatial differences with respect to MSI make this kind of intersensor SR problem particularly challenging while constraining the absolute performance of all the models. However, the proposed approach is the only method that consistently provides competitive advantages from small to large ratios since its topology has been specially designed for enhancing OLCI products from an L4 perspective. More specifically, this has been accomplished by integrating three different components: initial interpolation, dense multireceptive field and residual channel attention. By the initial interpolation, we avoid using any up-scaling layer that may substantially degrade the results when considering large scaling factors due to the drastic increase on the network parameters to reach spatial resolution closer to MSI. With the proposed dense multireceptive field, the presented model works for extracting features at different abstraction levels that hierarchically expand the initial receptive field over the OLCI sensor. In this way, smaller and larger kernel sizes can simultaneously be exploited for smaller and larger up-scaling ratios, respectively. Finally,

the incorporated residual channel attention allows the proposed approach to adequately filter all these dense features in order to focus on the most discriminating features across all the considered scaling factors.

When accounting for the spectral fidelity of the super-resolved results, Table VIII shows a similar general trend. In this case, the best average SAM reduction (with respect to the baseline) is certainly provided by the proposed approach (-0.0020 rad), followed by SRCNN (-0.0014 rad), VDSR (-0.0002 rad), and DRRN (-0.0001 rad). In contrast, SRRN, RDN, and RCAN are not able to reduce the baseline spectral distortion because their results rapidly get degraded with factors higher than $4\times$. In the case of EDSR, it is able to better control the spectral distortion due to the internal use of a constant scaling layer that works for numerically stabilizing the training and the output range. The qualitative results displayed in Fig. 5 also support the conducted quantitative analysis. As it is possible to observe, the ground-truth data for $2\times$ [see Fig. 5(a)] lacks of many spatial details, which indicates the very limited original resolution of the OLCI sensor and the high complexity of this type of SR problem. In general, the visual results show that each particular model tends to highlight specific features on the super-resolved output. On the one hand, SRRN, EDSR, RDN, and RCAN seem to be slightly more robust to aliasing and moire effects due to the absence of the initial up-scaling interpolation. However, it is true that the very coarse spatial resolution of the OLCI sensor makes these differences almost imperceptible from both qualitative and quantitative perspectives. On the other hand, SRCNN, VDSR, and DRRN appear to better enhance those land cover textures that present a certain level of contrast in the scene but, again, visual differences are very small owing to the lack of high-frequency information in the original space. Regarding the proposed approach qualitative performance, it is able to recover more spatial details than others as the magnified image region reveals. The deeper structure of the dense block together with its multireceptive field scheme can effectively attenuate the undesirable effects of the initial interpolation that also affects the best performing competitors across all scales.

In general, the extensive experimental comparison conducted in this work shows the high complexity of spatially enhancing OLCI products via the MSI sensor. It is important to note that SR architectures aim at learning the mapping between LR and HR image domains. Hence, the serious lack of high-frequency information in the input domain may logically aggravate the ill-posed nature of the SR problem since it dramatically increases the uncertainty when learning the corresponding projection. In contrast to standard imagery and other higher resolution instruments, OLCI has rather different spatial characteristics that makes many of the state-of-the-art SR networks fail when trying to achieve resolutions closer to MSI. For instance, although one can see that SRRN, EDSR, RDN, and RCAN provide prominent results in their corresponding papers, these results are based on standard HR imagery using reduced scaling factors. Hence, they cannot directly be extrapolated to our context. When super-resolving OLCI data, the particularly coarse resolution of the sensor together with the huge spatial differences with respect to MSI make other simpler models (e.g., SRCNN)

more effective. However, this type of models have important limitations on the fixed receptive field, which should ideally be adapted within OLCI scenes and across scales. In response, we develop a novel architecture (SRS3) specially designed for super-resolving OLCI products, which tries to densely exploit multiple receptive fields while focusing the attention on the most discriminating features for each input patch and possible scaling ratio. Precisely, these features allow the proposed approach to better characterize LR land cover regions that need additional computations to introduce HR details that cannot be recovered from a global SR perspective where feature maps are equally relevant.

V. Conclusion and Future Work

This article provides a new CNN-based framework for generating spatially enhanced S3 OLCI products from an L4 data processing perspective. Initially, an intersensor framework is presented to super-resolve operational OLCI data by means of the higher spatial resolution of the S2 instrument. Then, a novel SR architecture, based on a dense multireceptive field and residual channel attention, is proposed to alleviate the limited spatial resolution of OLCI while uncovering more discriminating features across large scales. Finally, an extensive experimental comparison is conducted to validate the suitability of the presented intersensor framework and SR architecture with respect to numerous state-of-the-art methods available in the literature.

One of the main conclusions that arises from this article is the high complexity of super-resolving OLCI data via the MSI sensor up to scale factors to obtain MSI resolution. Unlike other SR problems, OLCI has particularly low spatial resolution that makes many state-of-the-art networks fail when trying to reach resolutions closer to MSI. In this sense, adopting a proper multireceptive field design in conjunction with a channel attention mechanism has shown to play a fundamental role to relieve the inherent limitations of standard SR networks, which are unable to adapt the feature extraction process to each input patch and up-scaling factor. Although results are encouraging, further research is required for generating additional improvements. Specifically, our future work will be directed towards the following directions: 1) extending the proposed framework to a sequential up-scaling scheme, 2) improving the loss function by adding multiple cost terms, and 3) expanding this research to other intersensor platforms.

References

[1] A. S. Belward and J. O. Skøien, "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites," *ISPRS J. Photogrammetry Remote Sens.*, vol. 103, pp. 115–128, 2015.

[2] M. A. Friedl *et al.*, "Global land cover mapping from modis: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 287–302, 2002.

[3] O. Arino *et al.*, "GlobCover: ESA service for global land cover from MERIS," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 2412–2415.

[4] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.

[5] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[6] Y. Chen, X. Song, S. Wang, J. Huang, and L. R. Mansaray, "Impacts of spatial heterogeneity on crop area mapping in Canada using modis data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 451–461, 2016.

[7] D. Marcos, M. Volpi, B. Kellenberger, and D. Tuia, "Land cover mapping at very high resolution with rotation Equivariant CNNs: Towards small yet accurate models," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 96–107, 2018.

[8] R. Fernandez-Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Multimodal probabilistic latent semantic analysis for Sentinel-1 and Sentinel-2 image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1347–1351, Sep. 2018.

[9] L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," *ISPRS J. Photogrammetry Remote Sens.*, vol. 84, pp. 85–99, 2013.

[10] D. Poursanidis and N. Chrysoulakis, "Remote sensing, natural hazards and the contribution of ESA sentinels missions," *Remote Sens. Appl. Soc. Environ.*, vol. 6, pp. 25–38, 2017.

[11] H. M. de Klerk, J. Gilbertson, M. Lück-Vogel, J. Kemp, and Z. Munch, "Using remote sensing in support of environmental management: A framework for selecting products, algorithms and methods," *J. Environ. Manage.*, vol. 182, pp. 564–573, 2016.

[12] T. Sakamoto, "Incorporating environmental variables into a modis-based crop yield estimation method for united states corn and soybeans through the use of a random forest regression algorithm," *ISPRS J. Photogrammetry Remote Sens.*, vol. 160, pp. 208–228, 2020.

[13] R. Fernandez-Beltran, A. Plaza, J. Plaza, and F. Pla, "Hyperspectral Unmixing based on dual-depth sparse probabilistic latent semantic analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6344–6360, Nov. 2018.

[14] R. Fernandez-Beltran, F. Pla, and A. Plaza, "Sentinel-2 and Sentinel-3 intersensor vegetation estimation via constrained topic modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1531–1535, Oct. 2019.

[15] W. Zhao, H. Wu, G. Yin, and S.-B. Duan, "Normalization of the temporal effect on the modis land surface temperature product using random forest regression," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 109–118, 2019.

[16] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proc. IEEE*, vol. 103, no. 9, pp. 1560–1584, Sep. 2015.

[17] N. K. Keppy and M. Allen, "Understanding spectral bandwidth and resolution in the regulated laboratory," *Thermo Fisher Sci., Madison, WI, USA, Tech. Note* 51721, 2008.

[18] J. Aschbacher and M. P. Milagro-Pérez, "The European earth monitoring (GMES) programme: Status and perspectives," *Remote Sens. Environ.*, vol. 120, pp. 3–8, 2012.

[19] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "ESA's sentinel missions in support of earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, 2012.

[20] C. Donlon *et al.*, "The global monitoring for environment and security (GMES) sentinel-3 mission," *Remote Sens. Environ.*, vol. 120, pp. 37–57, 2012.

[21] Z. Malenovskỳ *et al.*, "Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, Cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, 2012.

[22] X. Wang, F. Ling, H. Yao, Y. Liu, and S. Xu, "Unsupervised sub-pixel water body mapping with Sentinel-3 OLCI image," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 327.

[23] P. T. Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, 2018, Art. no. 18.

[24] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, Nov. 2016.

[25] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5901–5904.

[26] N.-B. Chang and K. Bai, *Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2018.

[27] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2014.

[28] R. Fernandez-Beltran, P. Latorre-Carmona, and F. Pla, "Single-frame super-resolution in remote sensing: A practical overview," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 314–354, 2017.

[29] A. Vaiopoulos and K. Karantzalos, "Pansharpening on the narrow VNIR and SWIR spectral bands of Sentinel-2, the international archives of photogrammetry," *Remote Sens. Spatial Inf. Sci.*, vol. 41, pp. 723–730, 2016.

[30] H. Park, J. Choi, N. Park, and S. Choi, "Sharpening the VNIR and SWIR bands of Sentinel-2A imagery through modified selected and synthesized band schemes," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1080.

[31] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, 2016.

[32] C. Lanaras, J. Bioucas-Dias, E. Baltsavias, and K. Schindler, "Super-resolution of multispectral multiresolution images from a single sensor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 20–28.

[33] C. Paris, J. Bioucas-Dias, and L. Bruzzone, "A hierarchical approach to Superresolution of Multispectral images with different spatial resolutions," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 2589–2592.

[34] C. Tuna, G. Unal, and E. Sertel, "Single-frame super resolution of remote-sensing images by convolutional neural networks," *Int. J. Remote Sens.*, vol. 39, no. 8, pp. 2463–2479, 2018.

[35] M. Gargiulo, "Advances on CNN-based super-resolution of Sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3165–3168.

[36] 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 J. Photogrammetry Remote Sens.*, vol. 146, pp. 305–319, 2018.

[37] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2635.

[38] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[39] D. Pouliot, R. Latifovic, J. Pasher, and J. Duffe, "Landsat super-resolution enhancement using convolution neural networks and Sentinel-2 for training," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 394.

[40] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[41] M. Wu *et al.*, "Monitoring cotton root rot by synthetic sentinel-2 NDVI time series using improved spatial and temporal data fusion," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 2016.

[42] R. Fernandez-Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Remote sensing image fusion using hierarchical multimodal probabilistic latent semantic analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4982–4993, Dec. 2018.

[43] F. Palsson, J. Sveinsson, and M. Ulfarsson, "Sentinel-2 image fusion using a deep residual network," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1290.

[44] R. Fernandez-Beltran, F. Pla, and A. Plaza, "Inter-sensor remote sensing image registration using multi-spectral semantic Embeddings," *IEEE Geosci. Remote Sens. Lett*, vol. 16, no. 10, pp. 1545–1549, Oct. 2019.

[45] M. Irani *et al.*, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Vis. Commun. Image Representation*, vol. 4, no. 4, pp. 324–335, 1993.

[46] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.

[47] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[48] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

[49] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[50] J. A. Benediktsson, J. Chanussot, and W. M. Moon, "Very high-resolution remote sensing: Challenges and opportunities[point of view]," *Proc. IEEE*, vol. 100, no. 6, pp. 1907–1910, Jun. 2012.

[51] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[52] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.

[53] B. Ayhan and C. Kwan, "Mastcam image resolution enhancement with application to disparity map generation for stereo images with different resolutions," *Sensors*, vol. 19, no. 16, 2019, Art. no. 3526.

[54] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.

[55] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.

[56] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[57] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[58] R. Richter and D. Schläpfer, "Atmospheric/topographic correction for satellite imagery," DLR, Wessling, Germany, Tech. Rep. DLR-IB 565-01, Feb. 2014. [Online]. Available: http://www.rese.ch/pdf/atcor3_manual.pdf.

[59] L. De Keukelaere *et al.*, "Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using ICOR algorithm: Validation for coastal and inland waters," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 525–542, 2018.

[60] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[61] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[62] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 197–209, 2018.

[63] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

Jian Kang (Member, IEEE) received the B.S. and M.E. degrees in electronic engineering from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively, and the the Dr.-Ing. degree in electronic and information engineering from Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany, in 2019.

In August 2018, he was a Guest Researcher with the Institute of Computer Graphics and Vision, TU Graz, Graz, Austria. From 2019 to 2020, he was with the Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany. He is currently with the School of Electronic and Information Engineering, Soochow University, Suzhou, China. His research interests include signal processing, machine learning techniques, their applications in remote sensing, multidimensional data analysis, geophysical parameter estimation based on InSAR data, SAR denoising, and deep learning based techniques for remote sensing image analysis.

Dr. Kang was the recipient of the first place of the Best Student Paper Award in EUSAR 2018, Aachen, Germany. His joint work was selected as one of the ten Student Paper Competition Finalists in IGARSS 2020.

Rafael Fernandez received the B.Sc. degree in computer science and the M.Sc. degree in intelligent systems from Universitat Jaume I, Castellón de la Plana, Spain, in 2005 and 2019. Since 2019, he has been working toward the Ph.D. degree in computer science from Universitat Jaume I.

His research interests include computer vision and machine learning, with special interest in remote sensing applications.

Filiberto Pla received the B.Sc. and Ph.D. degrees in physics from the Universitat de València, Valencia, Spain, in 1989 and 1993, respectively.

He is currently a Full Professor with the Departament de Llenguatges i Sistemes Informàtics, University Jaume I, Castellón, Castellón, Spain. He was a Visiting Scientist with Silsoe Research Institute, Bedford, U.K., the University of Surrey, Guildford, U.K., the University of Bristol, Bristol, U.K., Centre d'etude du Machinisme Agricole du Ge Rural des Eaux et Forêts France, the University of Genoa, Genoa, Italy, Instituto Superior Técnico of Lisbon, Lisbon, Portugal, the Swiss Federal Institute of Technology ETH-Zurich, Zürich, Switzerland, the IDIAP Research Institute, Martigny, Switzerland, the Technical University of Delft, Delft, The Netherlands, and the Mid Sweden University, Sweden. He is the Director of the Institute of New Imaging Technologies, University Jaume I. His current research interests include color and spectral image analysis, visual motion analysis, 3D image visualization, and pattern recognition techniques applied to image processing.

Dr. Pla is a Member of the Spanish Association for Pattern Recognition and Image Analysis and the International Association for Pattern Recognition.

Ruben Fernandez-Beltran (Senior Member, IEEE) received the B.Sc. degree in computer science, the M.Sc. degree in intelligent systems, and the Ph.D. degree in computer science from Universitat Jaume I, Castellón de la Plana, Spain, in 2007, 2011, and 2016, respectively.

He is currently a Postdoctoral Researcher and a Member with the Computer Vision Group and the Institute of New Imaging Technologies, University Jaume I. He was a Visiting Researcher with the University of Bristol, Bristol, U.K., the University of Cáceres, Cáceres, Spain, and the Technische Universität Berlin, Berlin, Germany. His research interests include multimedia retrieval, spatio-spectral image analysis, and pattern recognition techniques applied to image processing and remote sensing.

Dr. Fernandez-Beltran is a Member of the Spanish Association for Pattern Recognition and Image Analysis, which is part of the International Association for Pattern Recognition, and the Institute of New Imaging Technologies, University Jaume I. He was the recipient of the Outstanding Ph.D. Dissertation Award at Universitat Jaume I in 2017.