Model-Based Super-Resolution for Sentinel-5P Data

Alessia Carbone, *Graduate Student Member, IEEE*, Rocco Restaino, *Senior Member, IEEE*,
Gemine Vivone, *Senior Member, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

Abstract— Sentinel-5P provides excellent spatial information, but its resolution is insufficient to characterize the complex distribution of air contaminants within limited areas. As physical constraints prevent significant advances beyond its nominal resolution, employing processing techniques like single-image super-resolution (SISR) can notably contribute to both research and air quality monitoring applications. This study presents the very first use of such methodologies on Sentinel-5P data. We demonstrate that superior results may be obtained if the degrading filter used to simulate pairs of low- and high-resolution (HR) images is tailored to the acquisition technology at hand, an issue frequently ignored in the scientific literature on the subject. Because of this, as well as the fact that these data have never been deployed in any previous studies, the primary theoretical contribution of this article is the estimation of the degradation model of TROPOspheric Monitoring Instrument (TROPOMI), the sensor mounted on Sentinel-5P. Leveraging this model—which is essential for applications involving super-resolution—we additionally improve a well-known deconvolution-based strategy and present a brand-new neural network that outperforms both traditional super-resolution techniques and well-established neural networks in the field. The findings of this study, which are supported by experimental tests on real Sentinel-5P radiance images, using both full-scale and reduced-scale protocols, offer a baseline for enhancing algorithms that are driven by the understanding of the imaging model and provide an efficient way of evaluating innovative approaches on all the available images. The code is available at https://github.com/alcarbone/S5P_SISR_Toolbox.

Index Terms— Atmospheric pollution, degradation filter, image enhancement, remote sensing, Sentinel-5P, super-resolution.

I. INTRODUCTION

WORLD Health Organization (WHO) reports that more than 90% of the global population lives in areas where air pollution emissions exceed the recommended levels, which makes air quality monitoring of vital importance for global health [1]. Satellites have recently become commonly employed for this purpose. On its path to their sensors, electromagnetic energy goes through the atmosphere twice, where it is mostly absorbed and dispersed by aerosols. To create concentration maps of these components, it is thus essential to recognize how each of them react to light at distinct wavelengths and locations. Satellites with a high spectral and temporal variation typically measure the whole atmospheric vertical column over a kilometer area of the surface. Since the spatial concentration of pollutants may vary by an order of magnitude over just a few hundred meters [1], providing averaged data in wide areas results in significant degradation of signals and consequently compromises quantitative analyses. Fortunately, the most recent availability of sophisticated technology has enhanced the spatial resolution of satellite imagery.

Sentinel-5 Precursor (Sentinel-5P or S5P), the first European mission intended specifically for observing our atmosphere, is an example of such innovation [2]. S5P is a single satellite mission responsible for monitoring air constituents, specifically ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2), methane (CH_4), formaldehyde (CH_2O), aerosols, and clouds. Inevitably, for some applications, its spatial resolution of about 3.5×5.5 km^2 may not be enough. When monitoring trace gases in urban areas, a higher spatial resolution can help distinguish sources of pollution, track emissions, and assess air quality more precisely. Furthermore, a few studies have demonstrated that applying a high-resolution (HR) model over some areas can help enhance air quality modeling [3] and support mitigation actions [4]; additionally, HR is more significant for studies of local air pollution than for large-scale variations that are relevant to climate studies [5].

Despite the fact that image sensors keep being improved, no device can measure a physical signal with infinite precision [6]. Indeed, the grid of pixels that makes up the image results from the combination of scans in the direction orthogonal to the motion of the platform (cross-track direction) and the one parallel to it (in-track direction) [6]. Thereby, each pixel occurs whenever the sensor digitally samples the continuous stream generated by the scanning, establishing its spatial resolution [6]. Unlike in other applications, spatial resolution in remote sensing is formally defined as the shortest distance that the sensor can resolve on the ground or ground range detected (GRD) [7]. As this range decreases, and therefore spatial resolution improves, less incoming light is detected by the sensor, thus affecting image quality. As HRs are constrained by these physical limitations, certain scales can only be inferred via image processing techniques [8].

Alessia Carbone and Rocco Restaino are with the Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, Italy (e-mail: alcarbone@unisa.it; restaino@unisa.it).

Gemine Vivone is with the CNR-IMAA, Institute of Methodologies for Environmental Analysis, 85050 Tito Scalo, Italy, and also with the National Biodiversity Future Center (NBFC), 90133 Palermo, Italy (e-mail: gemine.vivone@imaa.cnr.it).

Jocelyn Chanussot is with the INRIA, CNRS, Grenoble INP, LJK, University of Grenoble Alpes, 38000 Grenoble, France (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Digital Object Identifier 10.1109/TGRS.2024.3387877

For this reason, spatial resolution enhancement of remote sensing images is currently a very active research area. When no external information is available, other than the original low spatial resolution information, single-image super-resolution (SISR) techniques must be employed to generate an HR image by increasing the number of available pixels while preserving fine details.

A. Related Works

Several approaches have been proposed throughout the years to address SISR [8], [9], [10].

The earliest techniques used to handle the problem were interpolation-based approaches. Theoretically, if the original signal is band-limited and the sampling frequency is above the Nyquist rate, the missing points can be perfectly retrieved with a sinc function. As this would not be possible in real-world scenarios, effective interpolation techniques are usually obtained by approximating the ideal sinc function by piecewise local polynomials passing through $p + 1$ samples [11], as in bicubic interpolators ($p = 3$) [12], [13]. Nevertheless, they are unable to retrieve high-frequency information, leading to solutions with excessively smooth textures.

More powerful approaches are based on image reconstruction, which aims to produce a final super-resolved image that has the same visual characteristics as the low-resolution (LR) image but at a higher scale. Some first examples are the iterative back projection [14] and the gradient profile [15] SR techniques. Other important reconstruction-based approaches used for SISR employ the projection onto convex sets (POCS) [16] and the maximum a posteriori (MAP) [17] methods. Even though the outcomes of reconstruction-based approaches are far better than those of interpolation-based techniques, the a priori employed is quite sophisticated, which results in the need of an extensive level of competence.

Nowadays, super-resolution based on image learning is undoubtedly one of the most important areas of investigation in the field [8]. These techniques learn from an external training set potential relationships between LR and HR domains, then use that information to create the final super-resolved image [18]. Some classic approaches include sparse coding [19] and neighborhood embedding [20]. Other classic approaches are the ANR [21] method and the A+ [22], an advanced version of ANR.

More successfully, learning-based mapping approaches estimate a mapping function from the input LR image to the target HR image. Early works relied on one simple mapping function, like kernel ridge regression [23], while more recent methods [24], [25] employ several mapping functions. The very latest techniques use deep neural networks.

Although there are significant differences between these models, they all basically employ some essential components, including model frameworks, upsampling techniques, network designs, and learning procedures [10]. Early lightweight deep neural networks proposed to address SISR are either convolutional neural networks, like SRCNN [26], FSRCNN [27], VDSR [28], ESPCN [29], and DSRN [30], recursive networks, like DRRN [31], RFDN [32], and DRCN [33], networks that employ the cascading mechanism, like CARN [34] and CBPN [35], or networks that use mul-

tiscale fusion, like MSRN [36] and MCSN [37]. The results of these early networks are satisfactory, but their perceptual quality is relatively bad.

In particular, it has been demonstrated that the interactions between distortion and perceptual features of an image are complimentary yet competitive [38]. According to the perception-distortion trade-off theorem, there is an unreachable region in the perception-distortion plane. So, it is impossible for any super-resolution technique to provide both a low distortion error and a high perceptual quality, regardless of the efforts made to create algorithms that come as close as possible to the boundary. An intriguing finding is that approaches based on generative adversarial networks (GANs), like SRGAN [39], ESRGAN [40], and RankSRGAN [41], [42], converge more closely to the bound. However, when compared to more modern approaches, the batch normalization employed in these models has a negative impact on quality.

Remote sensing data differs from other types of data and over the years, the many different domains in which the images of interest are applied have divided the research in this field independently. In the process, the study of any potential model has quickly progressed, integrating ever-more complex components. For example, some authors have proposed the integration of high-level semantics and low-level details [34], [43], [44], and others local and non-local attention blocks to extract features that capture the long-range relationships between pixels. Different kinds of attention mechanisms are employed for SISR: channel-attention, used in RCAN [45], SeNet [46], SESR [47], and SAN [48], spatial-attention in DRLN [49] and RFANet [50], and hybrid attention in CSFM [51], [52], and HAT [53]. Further attention mechanisms more recently introduced are non-local attention, like in RNAN [54], pixel attention, like in PAN [55], and multiscale attention [56], [57]. Likewise, other important works on this subject employ either diffusion mechanisms [58], transformers [53], [59], residual information [45], [60], [61], or knowledge distillation [62], [63], [64], [65].

Although super-resolution approaches are quite sophisticated, the task surely has multiple challenges and research gaps. Particularly relevant to our work is the fact that, since real-world images undergo a specific degradation, even complex models, if trained on simulated datasets, perform badly in real-world scenarios [10]. A model assessed using such data actually learns to "reverse" a different degradation process, and may therefore be unable to generalize to another distortion scheme. As a result, in real-world settings, state-of-the-art super-resolution approaches often fail to perform as intended [9]. Interestingly, in most studies, the degradation model employed to simulate image pairs is independent of the acquisition sensor, even for rather complex models, and the issue remains to be investigated [9].

B. Contributions

In this article, we thoroughly explore the super-resolution of Sentinel-5P images, a topic that has never been addressed in the entire literature on the subject. As explained at the end of the previous paragraph, the specific characteristics of the sensor have a major influence on the operation of model-based

super-resolution algorithms. So, their design must be adapted to the acquisition technology of Sentinel-5P.

Hence, the majority of our research focuses on modeling the spatial degradation induced by the sensor on the images. Then, leveraging the knowledge of the spatial response of Sentinel-5P's sensors, we develop super-resolution approaches tailored to the satellite at hand and provide a basis for algorithm assessment.

Consequently, the main contributions of this article are as follows.

1) The formulation of a reliable approximation of the overall spatial response of Sentinel-5P's sensors.
2) The improvement of a super-resolution algorithm based on the widely accepted variational optimization.
3) The implementation of a deep-learning approach in which the upsampling step is carried out by a transposed convolution layer, and the upsampled image is then enhanced by a series of convolutional layers.
4) The evidence of the superior performance of these customized approaches in comparison to the methods usually employed in the literature not only at reduced resolution (RR), i.e., when the degradation filter used to simulate the LR image is known, but also at full resolution (FR).

The remainder of this article is organized as follows. In Section II, we formalize the relationship between LR and HR images employing a widespread representation of the imaging model and subsequently formalize the task as an optimization problem. Assuming the issue is non-blind, we then focus our investigation on modeling the sensor degradation that will be used for the proper design and assessment of the super-resolution methods. The proposed SISR solutions, which exploit the derived acquisition model within a classical variational approach and a novel shallow neural network, are described in Section III. After having thoroughly described the images used, the protocols and quality indices, and the training setting (Section IV), as well as justified the design of the deep-learning model proposed (Section V), the experimental results (Section VI) at RR and FR will basically demonstrate how the degradation we have modeled affects super-resolution algorithms, and how this issue, which is commonly overlooked, poses a threat to the proper functioning of such approaches.

II. Super-Resolution of Sentinel-5P Products

This section describes the problem addressed in this study. Its main objective is to formalize the SISR problem of S5P radiance products, highlighting the criticality of the imaging model for both the optimization and the assessment of algorithms. A specific formulation is motivated by the peculiarity of the TROPOspheric Monitoring Instrument (TROPOMI) acquisition process and the steps leading to the physical derivation of the mathematical model proposed in this study is accurately described.

A. SISR Problem Formulation

A powerful mathematical description of the super-resolution problem is based on the assumption that the available LR representation of the scene observed by the satellite, say image $\mathbf{L}$, can be obtained from an HR representation of the same scene $\mathbf{H}$ through a series of consecutive degradations that simulate how the sensing technology acquires the scene [8]. This process can be formalized by the imaging model

$$\mathbf{L} = (\mathbf{H} * \mathbf{K}) \downarrow_r + n \tag{1}$$

where $*$ is the convolution operation with a blurring kernel $\mathbf{K}$, $\downarrow_r$ is the downsampling by a factor r, and n is an additive noise. In all formulations, matrices are denoted with uppercase boldfaced letters.

The operator $\mathbf{K}$ corresponds to the point spread function (PSF) of the sensor, a function that weighs the measured physical signal before it is integrated to produce the output value [6]. If the performance is not limited solely by diffraction and other noise corruptions arise, it can be modeled as a zero-centered Gaussian function, whose standard deviation σ depends on the specific acquisition sensor [8].

The image degradation model described in (1) is usually recast in terms of a matrix formulation, which is easily obtainable by representing the involved images as column vectors. Indeed, the image degradation process can be analogously expressed as

$$\mathbf{l} = \mathbf{S}\mathbf{h} \tag{2}$$

where $\mathbf{l} = \text{vec}(\mathbf{L})$ and $\mathbf{h} = \text{vec}(\mathbf{H})$ are the vector image representations of $\mathbf{L}$ and $\mathbf{H}$, respectively, obtained by stacking their columns on top of one another. Additionally, $\mathbf{S} = \mathbf{CP}$ implements both the blurring and downsampling operations. In all formulations, vectors are denoted with lowercase boldfaced letters.

The imaging model can be expressed as in (2) if and only if the following hypotheses hold for these matrices [66], [67].

1) The blurring matrix $\mathbf{P}$ can be approximated by a block circulant with circulant blocks (BCCB) matrix, i.e., the matrix representation of the cyclic convolution operator. This hypothesis is met if and only if the blurring kernel $\mathbf{K}$ is shift-invariant and periodic boundary conditions are used for the convolution operator.
2) The matrix $\mathbf{C}$ is the downsampling operator, and its conjugate transpose interpolates the decimated image with zeros over the rows and columns, respectively.

The goal of SISR is inverting the acquisition process, i.e., computing $\mathbf{h}$ given that $\mathbf{l}$ is available and $\mathbf{S}$ is assumed to be known. This last assumption means that the SISR problem is supposed to be non-blind.

The model described in (2) is a system of linear equations, whose solution existence depends on whether it is possible to find the inverse matrix $\mathbf{S}^{-1}$. As the amount of pixels available in $\mathbf{l}$ is lower than the number of pixels in $\mathbf{h}$, the matrix $\mathbf{S}$ is not square and, even if it is still feasible to solve the system, the inverse problem is inherently ill-posed, namely there exists an infinite number of HR images corresponding to a single LR image [67].

To compensate for the instability of solutions, it is necessary to include a regularization factor in the optimization problem. As a result, the SISR goal can be better achieved as the finding

of an estimate $\widehat{\mathbf{h}}$ of $\mathbf{h}$, i.e., the solution of [10]

$$\widehat{\mathbf{h}} = \arg \min_{\widetilde{\mathbf{h}}} \left[\mathcal{L}\left(\mathbf{S}\widetilde{\mathbf{h}}, \mathbf{l}\right) + \tau \ \Phi\left(\widetilde{\mathbf{h}}\right) \right] \qquad (3)$$

in which the regularization term $\Phi(\cdot)$ is weighted by the parameter τ, and the loss function $\mathcal{L}(\cdot)$ is a consistency term between the degraded version of the HR image and the available LR image, both in column vector form.

B. PSF Model Formulation

The knowledge of the blurring kernel $\mathbf{K}$ in the imaging model (1), or, equivalently, of the degradation operator $\mathbf{S}$ in (2), is essential to employ the formalization just described in the algorithms design and assessment. To create model-based algorithms and generate images for performance evaluation with the same features as the real images, the first step is to establish an appropriate and manageable approximation of the spatial response of the acquisition system, i.e., TROPOMI.

The TROPOMI is the sole payload of the S5P satellite. It has four spectrometers (UV, UVIS, NIR, and SWIR), each with its own detector and split in two halves (band 1–8), covering almost 4000 overlapping channels between the ultraviolet and the short-wave infrared, and it has a swath of roughly 2600 km on the Earth's surface [2]. TROPOMI operates in a push-broom configuration [2]. This means that the signal from the covered area is sampled as the satellite moves along-track, and it is dispersed onto 2-D detectors: the signal sampled in the direction orthogonal to the motion, i.e., cross-track direction, is projected onto the rows, and the signal sampled in the spectral direction is projected onto the columns. Light from the entire swath is sampled in a period of approximately 840 ms, resulting in a ground pixel size of about 5.5 km along-track. The amount of signals co-added during this period is programmable between 1 and 256. In the cross-track direction, the ground sample interval is equal to roughly 1.75 km. However, row-binning is usually performed and, depending on the number of rows summed during one frame transfer in the detector, the resolution may change in the cross-track direction too, with a lower limit of 3.5 km, corresponding to a row-binning factor equal to two.

When an image of the target area is generated, a conversion from object space coordinates to detector space coordinates occurs. The pixel response function (PRF) $\Psi_{\mathrm{prf}}[r_{\mathrm{det}}, c_{\mathrm{det}}](\phi, \epsilon, \lambda)$ is the function that describes the relative response of a single detector pixel $[r_{\mathrm{det}}, c_{\mathrm{det}}]$, where r_{det} indicates the rows and c_{det} the columns, as a function of the object space dimensions, namely the azimuth angle (ϕ), which is the observation angle in agreement with the cross-track direction, the elevation angle (ϵ), which is the observation angle in relation to the along-track direction, and the wavelength (λ). The instrument transfer function is a helpful idea for understanding the mapping between unequal dimensions. In optical imaging systems, this function is the Fourier transform of the PSF, and hence the mathematical description of the projection of the object space $(\phi, \epsilon, \lambda)$ onto the image space $[\mathbf{X}, \mathbf{Y}]$, which is not necessarily equal to the detector space $[r_{\mathrm{det}}, c_{\mathrm{det}}]$, mainly because of co-addition and row-binning. The magnitude component of this complex
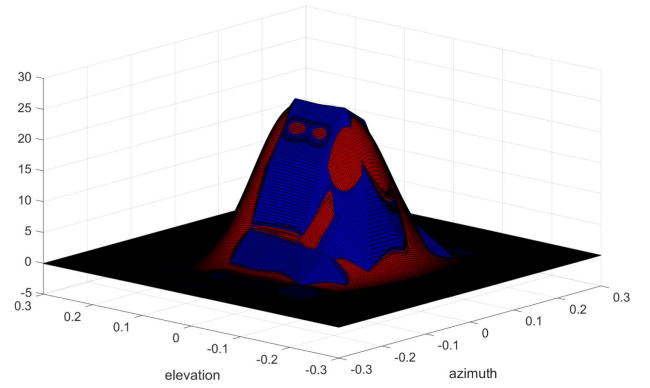

Fig. 1. Illustration of a PRF at a fixed wavelength of the UVIS detector. The response has been interpolated in a predefined grid and centered in both directions. The resulting function (in blue) was fit to a function (in red) obtained by convolving a rectangular window with an asymmetric Gaussian function.

function is the modulation transfer function (MTF) and it is the transformation we require.

As the PRFs are quite similar to asymmetric Gaussian functions, they can be fit to Gaussian-like distributions. The convolution along the elevation direction between an asymmetric Gaussian distribution and a rectangular window yielded the best fitting results. An example of a PRF at a fixed wavelength of UVIS fit to a Gaussian-like distribution is shown in Fig. 1. These responses are different from each other both in spectral and spatial coordinates. With regard to the spectral channel, we will only consider one channel at a time in this article, hence the response taken into account for each image, once a single channel has been picked, will be the response associated with the corresponding channels. Here, we would like to make some general statements about the row coordinate, i.e., the spatial dimension, as the responses are defined for each detector pixel. To this end, we did find that the response at the exact center can be employed as the response of all the detectors' pixels in the central area of the cross-track direction. This is due to the fact that distortions at the swath's center can be ignored; anyways, this is not possible at its furthest borders, where the signal comes from a wider region because of the curvature of the Earth's surface. This phenomenon is shown for the second detector (UVIS) in Fig. 2. An additional experiment demonstrated that averaging multiple responses results in a function that is much more closely matched to the responses in this area, as shown in Fig. 3 for the UVIS detector.

Once dispersed onto the detectors, pixels' signals are mapped onto the image space. As already said at the beginning of this section, the image space is not necessarily equal to the detector space and there is considerable freedom in determining the exact area of the Earth covered by a single image pixel. The following processes all impact on image pixels' responses.

1) Co-addition, which is the programmable sum of measurements along-track in a predefined co-addition period.
2) Row binning, namely the programmable sum of measurements in the cross-track direction.
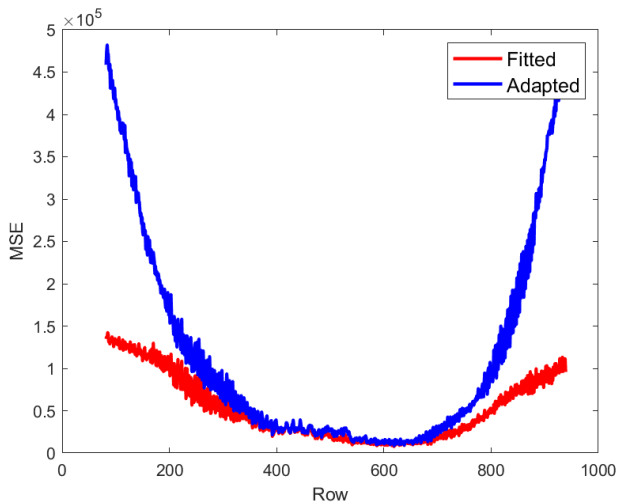
Fig. 2. Plot shows the error computed by comparison of the effective PRF, once it has been interpolated to a predefined grid and centered in both directions, with the fit Gaussian-like function obtained from its discrete points (in red) and with the fit Gaussian-like function obtained from the discrete points of the PRF at the exact central index of the swath (in blue).
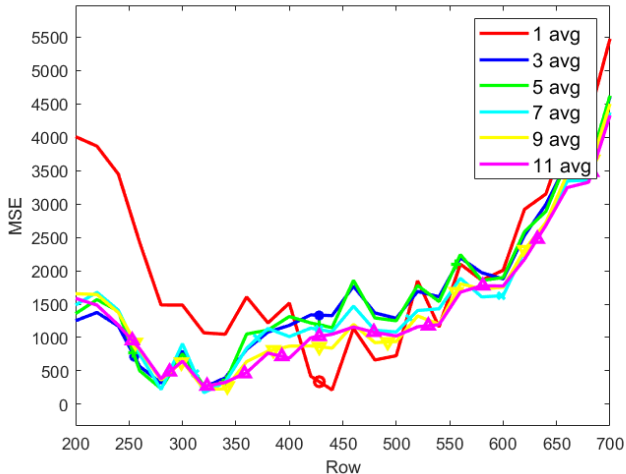

Fig. 3. Plot shows the error computed by comparison of the effective PRF, once it has been centered in both directions, with the distribution obtained by averaging the points of the central response with no other function (in red), with two additional functions (in blue), with four additional functions (in green), with six additional functions (in cyan), with eight additional functions (in yellow), and with ten additional functions (in magenta).

3) Satellite motion. It will be assumed that the Earth is an exact sphere and the satellite speed is constant, so that the response can be perfectly elongated along-track. Under these assumptions, satellite motion is represented by a convolution between the response and a rectangular window in the elevation direction along the angle that the satellite covers during a certain co-addition period.

Thus, to accurately determine the actual spatial response function in image space, the PRFs of all individual detector pixels $[r_{det}, c_{det}]$ that contribute to each image pixel $[\mathbf{X}, \mathbf{Y}]$ must be combined. In this way, the spatial response of one image pixel during a single co-addition period can be formally obtained.

Then, depending on the downscaling factor, such a response can be sampled to estimate the function responsible for the
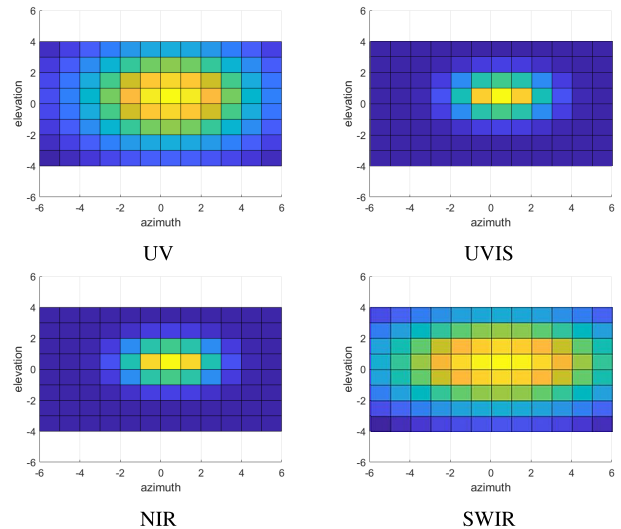

Fig. 4. Illustration of the PSFs for all the detectors for $r = 4$.

blurring, i.e., the PSF. We found that the blurring kernel $\mathbf{K}$ resulting from the described operations can be fit very closely by a Gaussian function with different standard deviations in the along-track and cross-track directions; in particular, the ratio of the sum of squared differences between the exact and the Gaussian-shaped representation of the PSF and the sample variance of the exact values of the PSF is lower than 1%. So, the PSF expression can be easily obtained by computing the gains of the MTF at the Nyquist frequency, i.e., the frequency that is half the sampling rate, in both spatial directions. Accordingly, at the end of all the experiments, the MTF of Sentinel-5P was described as a Gaussian function with gains at the Nyquist frequency depending on the considered detector. A graphical representation of the PSFs for all detectors, with $r = 4$, is provided in Fig. 4. As shown in the figure, the second and third detectors are more similar to each other than the other two. Indeed, the estimated values of the gains in across-track and along-track directions, respectively, are equal to 0.37 and 0.36 for UV, 0.44 and 0.74 for UVIS, 0.45 and 0.74 for NIR, and 0.15 and 0.20 for SWIR.

III. SUPER-RESOLUTION TECHNIQUES

The methods used in the experimental phase to point out the essential role of the degradation model for super-resolution algorithms are described in this section.

A. Non-Blind Deconvolution

As the operation $*$ of the imaging model described in Section II-A is a convolution, the inverse operation, i.e., a deblurring, can be addressed as a deconvolution operation. As already explained in Section II-A, if $\mathbf{S}$ is known, as it is assumed to be in this case, it is possible to degrade an estimate $\widehat{\mathbf{h}}$ of the HR scene $\mathbf{h}$ with $\mathbf{S}$ and determine how far apart its LR equivalent pixels are from $\mathbf{l}$ pixels. The Euclidean distance can be easily determined, although other types of distances can be used. Its minimization, or what is known as the objective function, is intuitively what one desires to achieve.

This kind of regularization is known as least-squares method. From the least-squares method, Tikhonov regularization can be obtained by considering a shrinking parameter μ that is used to reduce the norm of the estimate $\widehat{\mathbf{h}}$. This regularization alone can easily result in Gibbs artifacts and so an additional regularization may be included to the prime differences of the estimate through a smoothing parameter λ. In short, the objective function can be formalized as [67], [68]

$$Q = |\mathbf{l} - \mathbf{S}\widehat{\mathbf{h}}|^2 + \mu|\widehat{\mathbf{h}}|^2 + \lambda|\mathbf{D}\widehat{\mathbf{h}}|^2 \qquad (4)$$

where $|\mathbf{x}|$ denotes the Euclidean norm of a generic vector $\mathbf{x}$, the values μ and λ can change the contribution of each regularization term, $\mathbf{D}$ is the matrix implementing the prime differences, and $|\mathbf{l} - \mathbf{S}\widehat{\mathbf{h}}|$ is the data fidelity term.

The solution $\widehat{\mathbf{h}}$ can be efficiently computed via an iterative algorithm. In this article, we employ the conjugate gradient algorithm (CGA) [68].

B. S5Net

Mapping-based learning approaches handle the SISR problem by observing a training set of pairs of LR and HR images and they attempt to find a mapping function between the LR and the HR representations.

We have defined a neural network, which we will refer to as S5Net, which conceptually matches the phases that make up this transformation.

1) As image upsampling particularly affects the task, a transposed convolution layer is employed. Transposed convolution increases the size of the feature map by inserting $r - 1$ zeros between two consecutive pixels and performing a zero padding. Then, it convolves the image obtained with the kernel weights, which are initialized to the 2-D version of the cubic polynomial kernel [13], as it yielded the best results when compared with other kernels.

2) Image representation is the process that takes each patch in the upsampled image and represents it as a high-dimensional vector in a given basis [69]. A linear representation can be obtained by convolving each channel of the upsampled image n_1 times with filters of dimension $f_1 \times f_1$ and adding a vector of n_1 biases to the result. At the end, the rectified linear unit (ReLU) can be applied to the output, allowing a nonlinear generalization of this step [70].

3) In nonlinear mapping, each n_1-dimensional vector is nonlinearly mapped onto an n_2-dimensional vector. This process can be seen as the convolution of the output of the previous layer with a set of n_2 filters of size $n_1 \times f_2 \times f_2$ and the subsequent addition of a vector of n_2 biases. Then, the ReLU function is applied. Theoretically, additional layers may be added at this point to increase nonlinearity, but this makes the model even more complex.

4) Finally, in the reconstruction phase, the patch-wise HR representation is combined to produce the super-resolved image. Typically, for traditional methodologies, the final image is generated by averaging the HR patches. In other words, the n_2-dimensional representation is convolved with kernels of dimension $n_2 \times f_3 \times f_3$ and then added to a vector of c biases, where c is the number of channels in the image.

We combined the four steps to create the neural network depicted in Fig. 5, where the architecture that gave the best results (and the architecture used in the experiments) is shown, corresponding to the parameter selection $n_1 = 64, n_2 = 32, f_1 = 9, f_2 = 5, f_3 = 5$. All the weights and biases in this model must be tuned to the images employed for the specific task in order to achieve optimal performance. The model's loss function is the mean-square error (MSE) and the learning rate is set to 10^{-2}. In addition, the network is trained considering the same scaling factor r taken into account in the test phase, as we observed that this choice has a significant impact on the weights and biases learned by the neural network.

IV. EXPERIMENTAL TESTBED

In this section, we will go over the test configuration used for assessing the algorithms' performance: the datasets used, the quality assessment protocol, and the experiments' training and testing settings.

A. Sentinel-5P Radiance Products

In this work, we specifically focus on the super-resolution of Sentinel-5P Level-1B radiance products. They are available for each orbit and band as netCDF4 files and contain the Earth's spectral radiance expressed as the rate of photons received per unit area as a function of three dimensions: the index in the cross-track direction, the index in the along-track direction, and the index in the spectral direction. They additionally contain the instrument configurations within the orbit. Each combination of programmable instrument settings, like the number of measurements co-added and the row-binning factor, is uniquely identified by the combination of configuration ID and version.

Each orbit spans eight bands (BD1 to BD8), each of which has approximately 500 channels, and thus totals half of a spectrometer, either UV, UVIS, NIR, or SWIR. The overall orbit, as previously stated, comprises around 4000 channels. Only a single channel at a time will be taken into account in this article, hence each image is monochromatic. In the cross-track direction, only the central area of the swath, i.e., the range where the spatial responses are roughly comparable to each other, as explained in Section II-B, was considered. In this area and for the considered channel the row-binning factor is always equal to two. As a result, no experiment on the spatial response was undertaken as this parameter changed. On the other hand, as already said in Section II-B, the co-addition factor is programmable between 1 and 256. An investigation conducted on the spatial responses when the co-addition parameter varies revealed that gains at the Nyquist frequency changed minimally.

In the experiments, two examples of such orbits are employed, specifically orbits 28 317 and 29 729 taken on January 4, 2023 and September 7, 2023, respectively. We selected the central frequency of the second band of each
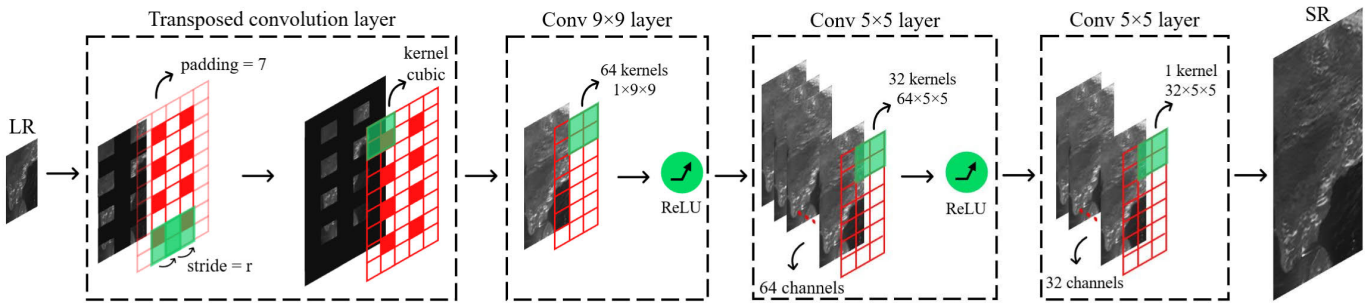
Fig. 5. Illustration of the S5Net. Given a single channel of the LR image LR and the scaling factor r, the network is tasked to map it to a super-resolved image SR. The first layer is a transposed convolution, whose weights are initialized to the cubic kernel. Then, the output image undergoes three convolutional layers: a convolution with 64 kernels of dimension $1 \times 9 \times 9$, a convolution with 32 kernels of dimension $64 \times 5 \times 5$, and a convolution with a kernel of dimension $32 \times 5 \times 5$. The activation function employed is ReLU.

detector (BD2, BD4, BD6, and BD8), obtaining a total of eight monochromatic images. Each image is 512×256 pixels in size and it is obtained by cropping the orbit in two directions: along-track between indices 2561 and 3073, and across-track between indices 66 and 322, a portion that corresponds roughly to the central part of the swath. In across-track, the exception is the SWIR, where the indices are chosen in full because the row-binning procedure does not occur. Anyways, during the training step, all the images were considered in their whole. Once cropped, the images retrieved from the first orbit cover an area over India and Sri Lanka, while the images obtained from the other show California and a small part of Mexico. To keep things simple, the two datasets, each comprising four images (called UV, UVIS, NIR, and SWIR in the experiments), will be referred to as IN and US all through the sections that follow.

B. RR and FR Protocol and Quality Metrics

The success of SISR depends on the quality of the generated super-resolved images. Even though the ideal circumstance to perform an image quality assessment is when there are ground-truth images available, in remote sensing this is not possible considering that the main reason why these techniques are employed is the impossibility to acquire HR images in the first place.

When ground-truth images are not accessible, an RR protocol can be employed [71]. The available LR image can be scaled down using the imaging model described in Section II-A and the SISR algorithm may be tasked to super-resolve it to its original resolution. In this way, a reference can be used to assess the model's accuracy. This protocol is based on the premise that the model performance is scale-independent and must therefore satisfy two important properties: consistency, i.e., the LR version of the super-resolved image must be as comparable to the original input image as possible, and synthesis, namely a coarser LR image can be generated by reducing the original image in accordance with the imaging model.

An alternative approach consists of employing the images at their original resolution, forgoing the use of a reference image. This methodology, known as FR protocol, has the benefit of testing the algorithms' ability to infer missing features at the size used in practice by assessing them at the actual resolution

of the images. This approach should be used in conjunction with the RR protocol to avoid the uncertainties caused by the scale-invariance assumption.

To evaluate the efficacy of image restoration algorithms, a variety of quality metrics have been utilized in the literature. Depending on the availability of a reference HR image, these metrics can be divided into two categories.

1) Distortion measurements can be applied by leveraging a reference image, which is available within the RR assessment protocol. In this article, we will employ the peak signal-to-noise-ratio (PSNR), the Erreur Relative Globale Adimensionnelle de Synthése (ERGAS) [72], the spatial correlation coefficient (sCC) [73], and the universal image quality index (UIQI, or more simply Q) [74]. The optimal value of ERGAS is 0, that of Q and sCC is 1, while PSNR conveys an improving performance as its value increases.

2) Perceptual metrics often exploit only the super-resolved image and, for this reason, they are usually employed in the FR assessment protocol. In particular, we use the blind referenceless image spatial quality evaluator (BRISQUE) [75]. As its values decrease, BRISQUE indicates a better performance.

As they tend to evaluate the "look" of the reconstructed image, perceptual metrics accuracy is lower than that of distortion metrics, especially the ones directly related to the MSE, i.e., PSNR and ERGAS.

C. Setting of Model-Based Algorithms

In the training phase of S5Net, the same procedure utilized for the RR assessment protocol was used for producing a couple of LR and HR images. In particular, to obtain the results at FR, the available image was used as the desired output during the training, while the corresponding input image was generated by a single degradation with a scaling factor $r = 4$. On the other hand, to obtain the results at RR, the available LR image was used as the neural network desired output in the test phase, while the image obtained by a single degradation was used as the input. The latter image was also used as the target image in the training phase, while the network input was generated via a further degradation by a scaling factor $r = 4$.
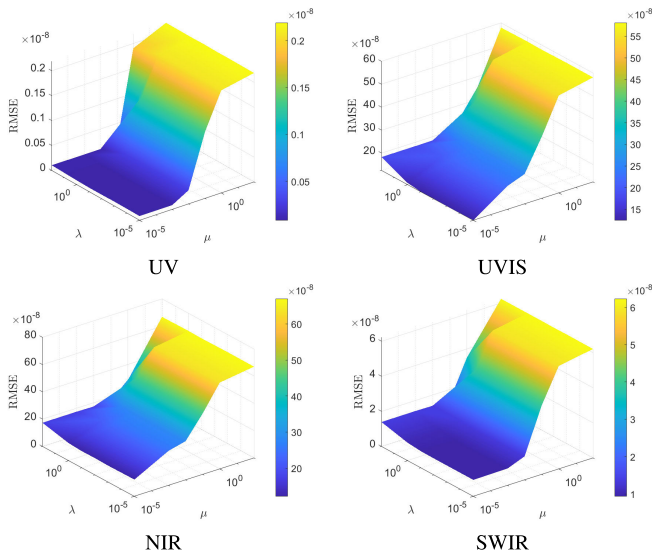
Fig. 6. Logarithmic plot of the averaged RMSE obtained on the two images on each single frequency by the CGA method when λ and μ vary in the interval [0.00001, 100].

TABLE I

Q, ERGAS, sCC, PSNR, AND BRISQUE INDICES FOR THE FOUR IMAGES OF IN DATASET FOR S5NET WITH AND WITHOUT THE TRANSPOSED CONVOLUTION LAYER. THE BEST RESULT IS IN BOLD FONT

	Method	Q	ERGAS	sCC	PSNR	BRISQUE
UV	S5Net_cubic	0.9386	1.0013	0.7713	41.1876	46.6333
	S5Net	**0.9387**	**1.0009**	**0.7738**	**41.1916**	**46.6176**
UVIS	S5Net_cubic	**0.8691**	6.9809	0.8066	31.0803	46.7838
	S5Net	**0.8691**	**6.9792**	**0.8069**	**31.0824**	**38.6241**
NIR	S5Net_cubic	**0.8516**	**5.7396**	0.8034	**31.1665**	45.2257
	S5Net	**0.8516**	**5.7396**	**0.8036**	**31.1665**	**44.3155**
SWIR	S5Net_cubic	0.8499	5.0936	0.7706	29.8836	50.0000
	S5Net	**0.8533**	**5.0494**	**0.7750**	**29.9593**	**48.8522**
Avg	S5Net_cubic	0.8773	4.7038	0.7880	33.3295	47.1607
	S5Net	**0.8782**	**4.6923**	**0.7898**	**33.3499**	**44.6023**

TABLE II

Q, ERGAS, sCC, PSNR, AND BRISQUE INDICES FOR THE FOUR IMAGES OF US DATASET FOR S5NET WITH AND WITHOUT THE TRANSPOSED CONVOLUTION LAYER. THE BEST RESULT IS IN BOLD FONT

	Method	Q	ERGAS	sCC	PSNR	BRISQUE
UV	S5Net_cubic	0.8777	0.9349	0.7613	37.8094	**43.6350**
	S5Net	**0.8779**	**0.9335**	**0.7621**	**37.8219**	43.6381
UVIS	S5Net_cubic	**0.9090**	5.7925	0.8154	31.1584	45.8892
	S5Net	**0.9090**	**5.7904**	**0.8159**	**31.1615**	**44.6014**
NIR	S5Net_cubic	0.8731	4.6685	0.8050	30.6916	53.5800
	S5Net	**0.8732**	**4.6676**	**0.8051**	**30.6932**	**52.4993**
SWIR	S5Net_cubic	0.8349	4.0501	0.7877	28.1004	56.9111
	S5Net	**0.8480**	**3.9188**	**0.7984**	**28.3868**	**54.1321**
Avg	S5Net_cubic	0.8737	3.8615	0.7924	31.9399	50.0038
	S5Net	**0.8770**	**3.8276**	**0.7954**	**32.0158**	**48.7177**

The experimentation section presents the results of the sole fine-tuning phase, which was undertaken to specifically tailor the network to the S5P images by adjusting the initial parameters of standard interpolation and super-resolution networks. Therefore, as Wang et al. [10] explain in great detail, many researchers in this field are moving toward partially online learning: the network that was already trained on an extensive dataset is additionally trained online on the image that will be super-resolved employing the RR protocol. In the non-blind deconvolution, the initial image in the CGA is set to the result obtained from the interpolation with 23 taps, as this choice yielded better results if compared to the method used in [68]. The parameters μ and λ in the objective function were chosen by conducting a sensitivity analysis of the algorithm in respect with their changes for different images. In particular, the analysis was carried out by varying both parameters in the interval [0.00001, 100]. The averaged RMSE obtained on both datasets, reported in Fig. 6, demonstrated that, when μ varies in the interval [0.00001, 0.0001], the error does not change significantly with λ in the interval [0.00001, 1]. For this reason, the values of $\mu = 0.00005$ and $\lambda = 0.1$ were chosen to guarantee a robust behavior for all the images. Moreover, the threshold parameter was set to $\delta = 0.0001$ times the norm of the low resolution image and the maximum number of iterations was chosen to be equal to 200 to avoid a high temporal complexity.

For each protocol and algorithm, two different training approaches were considered. In the first approach, the input image was obtained by degrading the output image via a BCCB matrix $\mathbf{P}$, defined by using the gains at the Nyquist frequency derived from the PSF model described in Section II-B. For the second one, the degradation matrix $\mathbf{P}$ was defined using different values at the Nyquist frequency. We considered a symmetrical filter with both gains at the Nyquist frequency equal to 0.3, a value widely employed in the literature. The scaling-dependent undersampling $\mathbf{C}$ matrix always matches the hypothesis outlined in Section II-A. For simplicity, the two approaches will be referred to as "_match" and "_nomatch" in the following.

The testing phase was conducted for all the techniques described in Section III. The quality of the estimate produced by each method was assessed at the end of the test using the metrics specified in Section IV-B taking into account any flaw caused by how each method performed at the borders of the image. The computing time of each algorithm on all images and for both protocols was additionally taken under consideration. It should be mentioned that PyTorch version 2.0.0 was used for collecting all findings on an Nvidia GTX 1080 Ti GPU.

V. ABLATION STUDY

To evaluate how the performance of S5Net changes in relation to the modules it contains, we carried out an ablation study on all the eight images. Given the importance of the upsampling phase in single-image super-resolution techniques, the ablation study involved the transposed convolution layer, whose effectiveness was assessed by replacing it with the classical bicubic interpolator.

The study's findings are shown in Tables I and II, in which "S5Net_cubic" denotes the S5Net without the transposed

TABLE III

Q, ERGAS, sCC, and PSNR Indices for the Four Images of IN Dataset. The Time in Seconds Is Provided Too. The Best Result Is in Bold Font and the Best Result Between the "_Nomatch" and "_Match" Cases Is Underlined

	Method	Q	ERGAS	sCC	PSNR	Time (s)
UV	Cubic	0.9352	1.0410	0.7552	40.8498	0.0150
	SRCNN	0.9386	1.0014	0.7713	41.1871	0.2479
	PAN	0.9371	1.0171	0.7480	41.0517	0.2799
	HAT	0.9347	1.0465	0.7408	40.8046	5.4895
	CGA_nomatch	0.9377	1.0153	0.7592	41.0676	3.2477
	CGA_match	0.9375	1.0165	0.7601	41.0571	4.3347
	S5Net_nomatch	0.9386	1.0019	**0.7740**	41.1828	0.3719
	S5Net_match	**0.9387**	**1.0009**	0.7738	**41.1916**	0.3789
UVIS	Cubic	0.8471	7.3385	0.7895	30.6463	0.0160
	SRCNN	0.8681	7.0734	0.7959	30.9659	0.2469
	PAN	0.8565	7.4682	0.7776	30.4941	0.2879
	HAT	0.8434	7.5724	0.7686	30.3738	2.7338
	CGA_nomatch	0.8496	8.4650	0.7872	29.4059	3.3895
	CGA_match	0.8663	7.0518	0.8036	30.9924	4.8221
	S5Net_nomatch	0.8587	8.2468	0.7949	29.6328	0.2729
	S5Net_match	**0.8691**	**6.9792**	**0.8069**	**31.0824**	0.2949
NIR	Cubic	0.8298	5.9850	0.7853	30.8029	0.0190
	SRCNN	0.8488	5.7899	0.7937	31.0908	0.5428
	PAN	0.8377	6.1204	0.7710	30.6086	0.2959
	HAT	0.8264	6.1573	0.7676	30.5564	2.7047
	CGA_nomatch	0.8341	6.8439	0.7860	29.6381	3.0167
	CGA_match	0.8488	5.7782	0.8024	31.1083	4.7511
	S5Net_nomatch	0.8463	6.4793	0.7953	30.1137	0.2909
	S5Net_match	**0.8516**	**5.7396**	**0.8036**	**31.1665**	0.2699
SWIR	Cubic	0.7985	5.7283	0.7546	28.8636	0.0170
	SRCNN	0.8171	5.4521	0.7703	29.2928	0.2049
	PAN	0.8141	5.5229	0.7764	29.1807	0.2279
	HAT	0.7815	6.0442	0.7318	28.3972	2.6937
	CGA_nomatch	0.8472	5.1185	0.7859	29.8411	2.2019
	CGA_match	**0.8590**	**5.0090**	**0.7880**	**30.0290**	2.4320
	S5Net_nomatch	0.8237	5.3772	0.7689	29.4130	0.2239
	S5Net_match	0.8533	5.0494	0.7750	29.9593	0.2429
Average	Cubic	0.8526	5.0232	0.7711	32.7906	0.0167
	SRCNN	0.8681	4.8292	0.7828	33.1341	0.3106
	PAN	0.8613	5.0322	0.7682	32.8338	0.2729
	HAT	0.8465	5.2051	0.7522	32.5330	3.4054
	CGA_nomatch	0.8671	5.3606	0.7796	32.4882	2.9639
	CGA_match	0.8779	4.7139	0.7885	33.2967	4.0849
	S5Net_nomatch	0.8668	5.2763	0.7833	32.5856	0.2846
	S5Net_match	**0.8782**	**4.6923**	**0.7898**	**33.3499**	0.3019

TABLE IV

Q, ERGAS, sCC, and PSNR Indices for the Four Images of US Dataset. The Time in Seconds Is Provided Too. The Best Result Is in Bold Font and the Best Result Between the "_Nomatch" and "_Match" Cases Is Underlined

	Method	Q	ERGAS	sCC	PSNR	Time (s)
UV	Cubic	0.8737	0.9510	0.7568	37.6607	0.0160
	SRCNN	0.8776	0.9349	0.7613	37.8095	0.5318
	PAN	0.8765	0.9389	0.7374	37.7720	0.7045
	HAT	0.8737	0.9532	0.7522	37.6403	5.3606
	CGA_nomatch	0.8767	0.9392	0.7577	37.7694	3.3226
	CGA_match	0.8765	0.9396	0.7585	37.7652	4.9971
	S5Net_nomatch	0.8776	0.9348	**0.7627**	37.8097	0.5724
	S5Net_match	**0.8779**	**0.9335**	0.7621	**37.8219**	0.5218
UVIS	Cubic	0.8877	6.1331	0.8027	30.6622	0.0160
	SRCNN	0.9070	5.8330	0.8106	31.0979	0.4478
	PAN	0.8921	6.3785	0.7826	30.3214	0.6758
	HAT	0.8814	6.4350	0.7886	30.2447	5.4805
	CGA_nomatch	0.8875	7.0637	0.8103	29.4350	3.0177
	CGA_match	0.9057	**5.7699**	**0.8223**	31.1924	4.4270
	S5Net_nomatch	0.9007	6.7265	0.8127	29.8599	0.3749
	S5Net_match	**0.9090**	5.7904	0.8159	31.1615	0.4089
NIR	Cubic	0.8473	4.8942	0.8000	30.2815	0.0160
	SRCNN	0.8703	4.6682	0.8026	30.6922	0.3988
	PAN	0.8548	4.9811	0.7831	30.1287	0.2959
	HAT	0.8408	5.0924	0.7860	29.9367	2.6997
	CGA_nomatch	0.8540	5.5021	0.8073	29.2646	3.1246
	CGA_match	0.8714	**4.6324**	**0.8183**	30.7591	3.7918
	S5Net_nomatch	**0.8753**	4.7668	0.8177	30.5105	0.2709
	S5Net_match	0.8732	4.6676	0.8051	30.6932	0.2939
SWIR	Cubic	0.7753	4.5510	0.7727	27.0878	0.0130
	SRCNN	0.8008	4.2789	0.7852	27.6232	0.2039
	PAN	0.7938	4.3808	0.7881	27.4188	0.2319
	HAT	0.7523	4.8712	0.7528	26.4971	2.7108
	CGA_nomatch	0.8321	4.0171	0.8028	28.1716	2.5618
	CGA_match	0.8460	**3.9149**	**0.8046**	**28.3954**	2.8859
	S5Net_nomatch	0.8262	4.0911	0.7870	28.0131	0.2249
	S5Net_match	**0.8480**	3.9188	0.7984	28.3868	0.2389
Average	Cubic	0.8460	4.1323	0.7830	31.4230	0.0152
	SRCNN	0.8639	3.9287	0.7899	31.8057	0.3955
	PAN	0.8543	4.1698	0.7728	31.4102	0.4770
	HAT	0.8371	4.3379	0.7699	31.0797	4.0629
	CGA_nomatch	0.8626	4.3805	0.7945	31.1601	3.0066
	CGA_match	0.8749	**3.8141**	**0.8009**	**32.0280**	4.0254
	S5Net_nomatch	0.8699	4.1298	0.7950	31.5483	0.3607
	S5Net_match	**0.8770**	3.8276	0.7954	32.0158	0.3658

convolution layer. In order to simplify the analysis and avoid focusing on the algorithms' distinct performance with the RR and FR protocols, all indices described in Section IV-B are reported in the same table, alongside the averaged value for each index and technique computed by taking into account all channels together. The best results for each index appear in bold in both cases.

With the exception of a single channel of the second dataset, where the difference is still negligible, the network trained with the transposed convolution layer consistently performs better than the network trained with the bicubic interpolator. To be more precise, we attain an appreciable improvement in each quality index for RR alongside a substantial enhancement in FR performance, estimated on average in 1.92 BRISQUE quality classes.

This demonstrates that the transposed convolution layer—which, in contrast to the basic interpolator, learns the weights during the training phase—remains a crucial component of image reconstruction, particularly in cases when the reference image is unavailable.
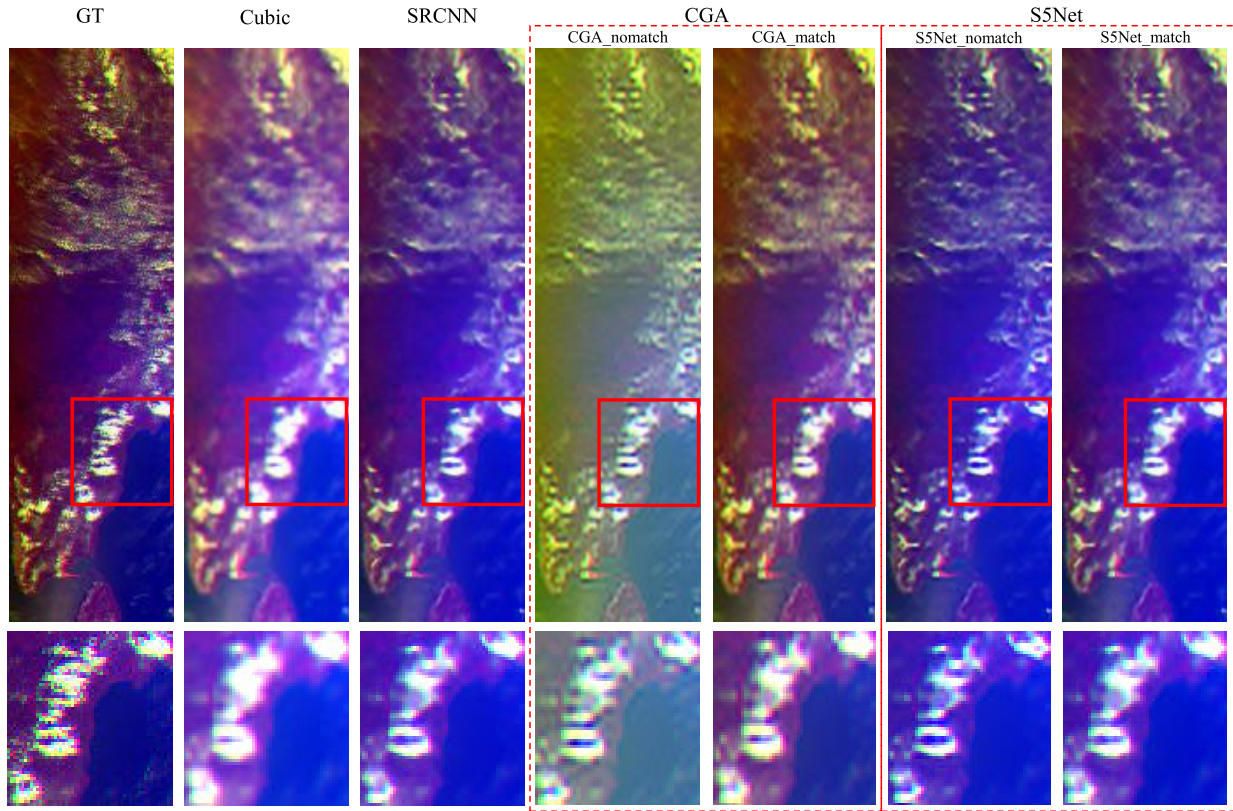
Fig. 7. False-color representation (NIR as red, UVIS as green, and UV as blue) of the ground-truth image and some super-resolved images for IN dataset and the RR protocol.

VI. EXPERIMENTAL RESULTS

In this section, we will discuss the experimental findings obtained on the eight images introduced in Section IV-A. The results obtained by the algorithm we improved, namely the deconvolution-based approach described in Section III-A, which we will call "CGA," and the network we designed, "S5Net," which is described in Section III-B, will be compared to the outcomes of some well-established approaches: the "Cubic" interpolator [13], the "SRCNN" network, the first neural network proposed for SISR [26], "PAN," a quite recent neural network based on pixel attention, proposed in [55], and "HAT" [53], a very recent neural network based on a hybrid attention transformer. The main goal of the experimental phase is to underline the difference in the results obtained by the experiments in which the simulated images are generated by taking into account the filter discussed in Section II-B, which we will call "_match," and the tests in which standard values for the filter gains at the Nyquist frequency are employed, namely "_nomatch" cases, hence to provide concrete evidence of what has been theoretically examined thus far. These findings are an important step toward the application of such techniques for data that have never been used in the literature for the same intent.

A. RR Assessment

Tables III and IV provide Q, ERGAS, sCC, and PSNR indices of the SR of the four images of each dataset for all the approaches, respectively. The average value obtained for each dataset and index considering all channels together is shown too, alongside the time in seconds taken by each algorithm. Additionally, in both the tables, bold represents the best result for that index among all algorithms, while the best performance between the "_nomatch" and "_match" versions of the same algorithm for a specific index is underlined.

The first important finding of the evidence given is that the proposed network behaves far better than all other approaches examined in most cases, even when compared to well-established networks in the field, especially the most recent ones. As the results reported suggest, while developing a super-resolution technique, an appropriate design of the degradation filter employed to generate the LR simulated image is crucial when the HR ground-truth image is not obtainable. The need for the degradation filter to be appropriately defined arises from the fact that it describes the blurring and downsampling processes that occur during image capture, as it was covered in Section II-A. As a consequence, if image capture is not precisely simulated, the inverse problem given in Section II-A becomes far tougher and could result in misleading solutions. The issue is much more significant when employing satellite images for quantitative analyses, when each pixel value contributes significantly. That is why, in our study, minimizing the error on the pixel, i.e., optimizing the indices directly related to the MSE, is essential. This is particularly relevant for Sentinel-5P radiance data, which are used for creating extremely accurate air pollution maps.

Some super-resolved images are shown in Fig. 7. It contains the false-color representation, employing the NIR channel as
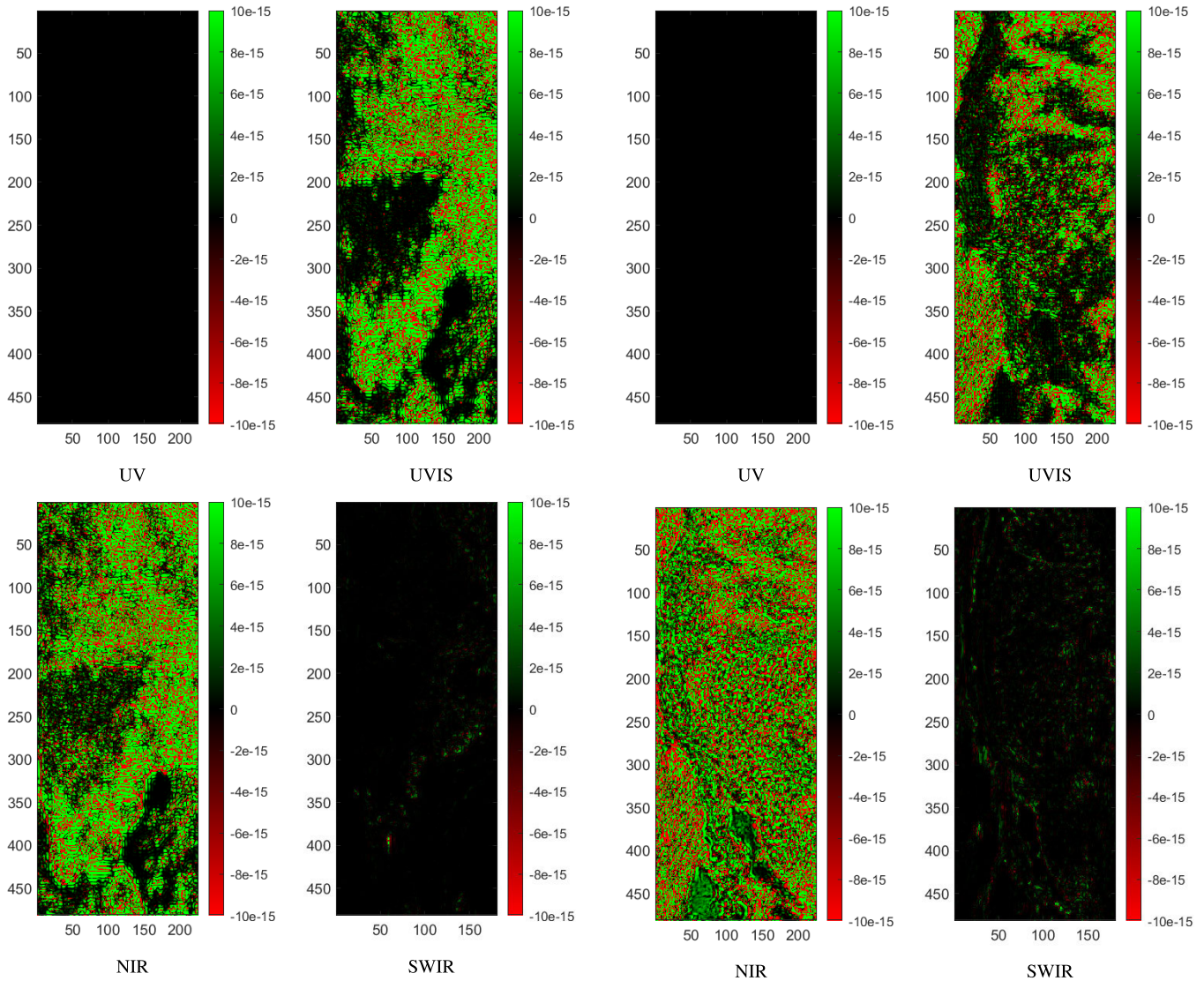
Fig. 8. Difference of MSE values obtained by CGA_nomatch and CGA_match for IN dataset.

Fig. 9. Difference of MSE values obtained by S5Net_nomatch and S5Net_match for US dataset.

red, the UVIS channel as green, and the UV channel as blue, of the ground-truth image and the products of some benchmark methods, namely cubic interpolator and SRCNN, compared to the results of CGA_nomatch, CGA_match, S5Net_nomatch, and S5Net_match for IN dataset. The same portion of the image is zoomed in all the pictures to clearly show the discrepancies in pixel reconstruction of the same algorithm under different conditions. Despite the fact that a visual assessment of the images supports what was previously said about the metrics' results, as already stated, we are more concerned with the preciseness of the pixel values. For this reason, in Figs. 8 and 9, we show the difference of the MSE obtained by CGA and S5Net, with no-matching filters and with matching filters, for the IN and US scenes, respectively. The more green colored the map is (positive values), the bigger the error for the no-matching cases. The maps have been generated at the same scale. As can be observed, when the filters used in the two scenarios have more comparable gains, i.e., the

UV, for which they are practically equal, and the SWIR, the difference has significantly less impact on the performance of both algorithms. In any case, green appears more frequently than red in all the maps, and this confirms what was stated till this point.

We want to underline that the distinction between the two scenarios, which is typically overlooked in the past studies, becomes particularly evident after approximately 400 s of training for the neural network and a maximum of 200 iterations in the deconvolution algorithm. As it has been previously addressed, it is crucial that the simulated data remain consistent with the data provided to the algorithm being evaluated if the results are to be as expected. This is even more important for the Sentinel-5P case, where the gains at the Nyquist frequency tend to strongly diverge from the widely used value (i.e., 0.3), especially for UVIS, NIR, and SWIR detectors.

In summary, a gain of 0.74 dB for S5Net and 0.84 dB for CGA may be obtained by comparing the "_nomatch" and

"_match" scenarios. The proposed method also proves to be computationally efficient, as it is always faster than more sophisticated state-of-the-art networks, in particular PAN and HAT, and often faster than or at least equal to SRCNN.

B. Full Resolution

A good degradation process modeling can significantly improve all super-resolution techniques that depend on it. When there is no ground-truth image available, this is even more vital. To avoid that the validity of the numerical results could be affected by the matching image degradation process in the RR protocol and in the CGA and S5Net algorithms, we used the no-reference metric discussed in Section IV-B to evaluate the algorithms' accuracy at FR too.

Table V provides the BRISQUE index of the SR of the eight images, alongside the time in seconds taken by each algorithm. The same font code, as in Tables III and IV, namely the bold and underlined characters, is used to indicate the best values among all the algorithms and between the "_nomatch" and "_match" versions of the same algorithms, respectively. The averaged value obtained for each dataset considering all channels together is shown too. As pointed out in the preceding section, the proposed network performs better than all traditional methods and neural networks in most cases. We would like to point out that, unlike the prior RR example, for the second dataset, the PAN network exceeds expectations, albeit marginally. Nonetheless, the network in question is extremely complex, and the BRISQUE index, as a no-reference metric, must be supplemented by a visual assessment. Anyways, even in the no-reference evaluation scenario, we would like to draw attention to the difference in the performance of the same algorithm when the degradations in the simulated and test data are different ("_nomatch") in comparison with the case in which they are the same ("_match") as the proper design of the degradation filter employed to generate the LR simulated image is much more vital for the no-reference evaluation scenario.

Some super-resolved images are shown in Figs. 10 and 11. In particular, they show a false-color representation of the products of the same methods shown in the RR scenario, in addition to PAN, to visually analyze its findings, for the IN and US datasets, respectively. In particular, in Fig. 10, the false-color representation is the same as Fig. 7, while, in Fig. 11, the single channel of SWIR is employed as red, that of NIR as green, and that of UVIS as blue, in order to include all detectors. Then, to clearly demonstrate the variations in pixel restoration of the same method under different conditions, the same area of the image is zoomed in all the cases. The analysis of the images supports both the numerical findings and previous assertions made regarding the proposed network and also highlights a superior quality with respect to PAN, although not revealed by quantitative indexes. When the degrading filter for the simulated images is supposed to be a predetermined model that is independent of the sensor acquisition features, the quality of the reconstruction is worse, and this is quite obvious from the images, especially for the S5Net.

TABLE V

BRISQUE INDEX FOR THE EIGHT IMAGES OF BOTH DATASETS. THE TIME IN SECONDS IS PROVIDED TOO. THE BEST RESULT IS IN BOLD FONT AND THE BEST RESULT BETWEEN THE "_NOMATCH" AND "_MATCH" CASES IS UNDERLINED

		IN		US	
	Method	BRISQUE	Time (s)	BRISQUE	Time (s)
UV	Cubic	54.9840	0.2660	49.8271	0.2433
	SRCNN	46.6089	0.1939	43.6630	0.3568
	PAN	**43.7375**	1.6625	**43.4582**	1.9890
	HAT	45.7313	5.2233	43.7587	6.7461
	CGA_nomatch	48.1214	62.5790	45.4224	64.4835
	CGA_match	<u>47.8406</u>	64.5019	<u>45.0729</u>	63.8710
	S5Net_nomatch	46.6796	0.6708	43.7081	0.7578
	S5Net_match	<u>46.6176</u>	0.6688	<u>43.6381</u>	0.8527
UVIS	Cubic	52.3721	0.2803	52.4466	0.2515
	SRCNN	46.7850	0.3218	45.8230	0.1899
	PAN	49.2706	1.6585	**44.0912**	1.5751
	HAT	49.4588	5.2395	59.1847	5.2255
	CGA_nomatch	50.0956	14.8691	47.8174	12.9593
	CGA_match	<u>48.3842</u>	15.2548	<u>44.9058</u>	13.0346
	S5Net_nomatch	45.2292	0.6364	45.0965	0.6188
	S5Net_match	**38.6241**	0.6498	<u>44.6014</u>	0.6388
NIR	Cubic	49.5864	0.2807	58.2531	0.2931
	SRCNN	45.1619	0.3819	53.5616	0.4009
	PAN	47.0527	1.8970	**46.0537**	1.6255
	HAT	50.5974	6.1949	43.7587	5.2186
	CGA_nomatch	50.8319	13.1007	57.5507	11.7808
	CGA_match	<u>47.2878</u>	13.3948	<u>56.0637</u>	12.1973
	S5Net_nomatch	44.5277	0.6997	52.7521	0.6214
	S5Net_match	**44.3155**	0.8107	<u>52.4993</u>	0.6748
SWIR	Cubic	57.0316	0.2086	61.1676	0.2283
	SRCNN	49.8339	0.3198	56.7829	0.3418
	PAN	49.9574	1.6835	50.2825	1.3966
	HAT	53.1163	8.3386	54.8469	5.2136
	CGA_nomatch	51.0023	51.5990	51.5988	51.0977
	CGA_match	<u>54.8364</u>	50.6881	**47.0555**	52.5665
	S5Net_nomatch	49.1728	0.5068	55.4208	0.5178
	S5Net_match	**48.8522**	0.5584	<u>54.1321</u>	0.5488
Average	Cubic	53.4935	0.2589	55.4236	0.2540
	SRCNN	47.0974	0.3043	49.9576	0.3223
	PAN	47.7299	1.7253	**45.9714**	1.6465
	HAT	49.9014	6.2490	53.6362	5.6009
	CGA_nomatch	50.0127	35.5369	50.5973	35.0803
	CGA_match	<u>49.5872</u>	35.9599	<u>48.2744</u>	35.4173
	S5Net_nomatch	46.4023	0.6284	49.2444	0.6289
	S5Net_match	**44.6023**	0.6719	<u>48.7177</u>	0.6787

In summary, S5Net and CGA gain 1.2 and 1.4 quality classes on BRISQUE, respectively, when comparing the "no_match" and "_match" scenarios.

The spectral consistency of the channels was not investigated in this work because the four spectra were processed independently. However, a visual inspection reveals that the S5Net output is more similar to systems that process all channels equally, resulting in less spectral distortion. In fact, this premise raises new insights for future research.
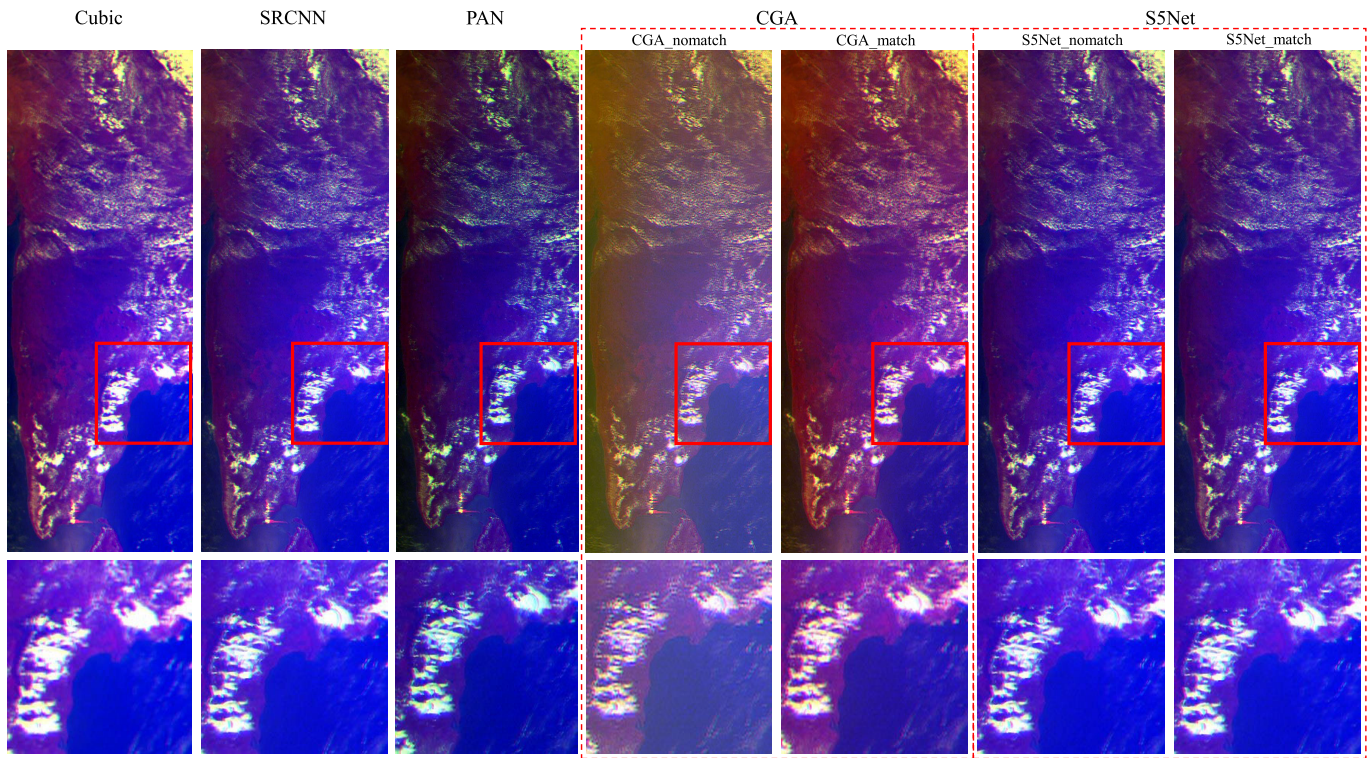
Fig. 10. False-color representation (NIR as red, UVIS as green, and UV as blue) of some super-resolved images for IN dataset and the FR protocol.
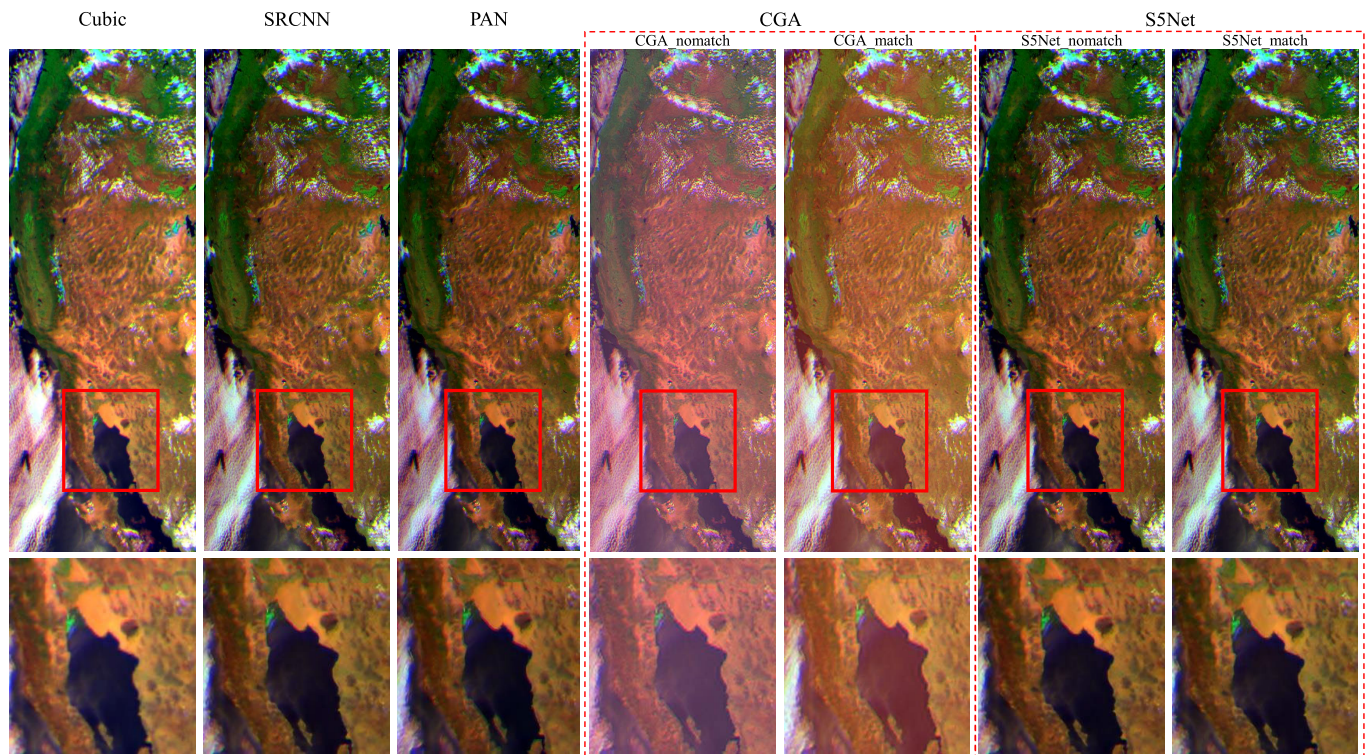

Fig. 11. False-color representation (SWIR as red, NIR as green, and UVIS as blue) of some super-resolved images for US dataset and the FR protocol.

VII. CONCLUSION

Sentinel-5P's capability to gather a variety of geographical information on atmospheric components has made it an essential instrument in the field of Earth observation. Still, there are constraints when zooming in on smaller geographic areas, and it is tricky to get accurate contextual data. To fill this resolution gap, researchers frequently employ sophisticated techniques like single-image super-resolution. In the scientific literature, assessing super-resolution algorithms often involves pairs of LR and HR images. Though, these HR counterparts

are created by applying degrading filters on the LR data that are typically designed independently of how the sensor acquires the images. The study covered in this article is a first attempt to apply model-based super-resolution techniques to Sentinel-5P data. More specifically, we demonstrate that it is possible to get better results than those commonly reported in the literature by properly generating the degradation filter required to create LR and HR image pairs. The results of the research support the benefit of a newly developed, more realistic degradation model and provide a fundamental accuracy assessment, both of which are critical when comparing more sophisticated super-resolution techniques. Additionally, we improve a well-known variational optimization method and present an innovative mapping-based strategy that makes use of deep learning, S5Net. This strategy outperforms other conventional approaches in the field, including interpolation, the deconvolution-based model and well-known neural networks. This study provides a potential starting point for getting improved air quality monitoring accuracy, and subsequently enabling more extensive and in-depth observations of air contaminants and atmospheric components over different geographic areas.

REFERENCES

[1] World Health Organization. (2021). *Who Global Air Quality Guidelines. Particulate Matter (PM$_{2.5}$ and PM$_{10}$), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. Accessed: Apr. 23, 2023. [Online]. Available: https://apps.who.int/iris/handle/10665/345329

[2] KNMI. *Algorithm Theoretical Basis Document for Tropomi L01b Data Processor*. Accessed: Apr. 11, 2023. [Online]. Available: https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-Level-1B-ATBD

[3] L. S. Neal et al., "A description and evaluation of an air quality model nested within global and regional composition-climate models using MetUM," *Geosci. Model Develop.*, vol. 10, no. 11, pp. 3941–3962, Nov. 2017.

[4] A. L. Goodkind, C. W. Tessum, J. S. Coggins, J. D. Hill, and J. D. Marshall, "Fine-scale damage estimates of particulate matter air pollution reveal opportunities for location-specific mitigation of emissions," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 18, pp. 8775–8780, Apr. 2019.

[5] Ø. Hodnebrog, F. Stordal, and T. K. Berntsen, "Does the resolution of megacity emissions impact large scale ozone?" *Atmos. Environ.*, vol. 45, no. 38, pp. 6852–6862, Dec. 2011.

[6] R. A. Schowengerdt, *Remote Sensing, Models, and Methods for Image Processing*. Cambridge, MA, USA: Academic, 2007.

[7] A. Orych, "Review of methods for determining the spatial resolution of UAV sensors," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 1, pp. 391–395, Apr. 2015.

[8] 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, Jan. 2017.

[9] M. Sdraka et al., "Deep learning for downscaling remote sensing images: Fusion and super-resolution," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 202–255, Sep. 2022.

[10] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[11] B. Aiazzi, S. Baronti, M. Selva, and L. Alparone, "Bi-cubic interpolation for shift-free pan-sharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 65–76, Dec. 2013.

[12] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[13] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[14] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.

[15] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[16] C. Fan, C. Wu, G. Li, and J. Ma, "Projections onto convex sets super-resolution reconstruction based on point spread function estimation of low-resolution remote sensing images," *Sensors*, vol. 17, no. 2, p. 362, Feb. 2017.

[17] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.

[18] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Oct. 1999, pp. 1182–1189.

[19] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[20] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2004, pp. 275–282.

[21] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[22] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Computer Vision—ACCV 2014*, D. Cremers, I. Reid, H. Saito, and M. H. Yang, Eds. Cham, Switzerland: Springer, 2014, pp. 111–126.

[23] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

[24] C.-Y. Yang and M.-H. Yang, "Fast direct super-resolution by simple functions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 561–568.

[25] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. CVPR*, Jun. 2011, pp. 449–456.

[26] 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.

[27] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., 2016, pp. 391–407.

[28] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[29] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[30] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, "Image super-resolution via dual-state recurrent networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1654–1663.

[31] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.

[32] J. Liu, J. Tang, and G. Wu, "Residual feature distillation network for lightweight image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, A. Bartoli and A. A. Fusiello, Jan. 2020, pp. 41–55.

[33] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[34] N. Ahn, B. Kang, and K. A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, in Lecture Notes in Computer Science, 2018, pp. 256–272.

[35] F. Zhu and Q. Zhao, "Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2453–2460.

[36] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Computer Vision—ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 527–542.

[37] W. Li, J. Li, J. Li, Z. Huang, and D. Zhou, "A lightweight multi-scale channel attention network for image super-resolution," *Neurocomputing*, vol. 456, pp. 327–337, Oct. 2021.

[38] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.

[39] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 105–114.

[40] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Comput. Vis.-ECCV Workshops*, vol. 11133, Munich, Germany, L. Leal-Taixé and S. Roth, Eds., 2018, pp. 63–79.

[41] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "RankSRGAN: Generative adversarial networks with ranker for image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3096–3105.

[42] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "RankSRGAN: Super resolution generative adversarial networks with learning to rank," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7149–7166, Oct. 2022.

[43] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4549–4557.

[44] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817.

[45] 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.*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Sep. 2018, pp. 294–310.

[46] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[47] X. Cheng, X. Li, J. Yang, and Y. Tai, "SESR: Single image super resolution with recursive squeeze and excitation networks," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 147–152.

[48] T. Dai, J. Cai, Y.-B. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 11057–11066.

[49] S. Anwar and N. Barnes, "Densely residual Laplacian super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1192–1204, Mar. 2022.

[50] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2356–2365.

[51] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3911–3927, Nov. 2020.

[52] B. Niu et al., "Single image super-resolution via a holistic attention network," in *Proc. ECCV*, A. Vedaldi, H. Bischof, T. Brox, and J. M. Frahm, Eds., 2020, pp. 191–207.

[53] X. Chen, X. Wang, J. Zhou, and C. Dong, "Activating more pixels in image super-resolution transformer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 22367–22377.

[54] Y. Zhang, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–18.

[55] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 56–72.

[56] T. Lu, J. Wang, Y. Zhang, Z. Wang, and J. Jiang, "Satellite image super-resolution via multi-scale residual deep neural network," *Remote Sens.*, vol. 11, no. 13, p. 1588, Jul. 2019.

[57] X. Wang, Y. Wu, Y. Ming, and H. Lv, "Remote sensing imagery super resolution based on adaptive multi-scale feature fusion network," *Sensors*, vol. 20, no. 4, p. 1142, Feb. 2020.

[58] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023.

[59] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5791–5800.

[60] B. Huang, B. He, L. Wu, and Z. Guo, "Deep residual dual-attention network for super-resolution reconstruction of remote sensing images," *Remote Sens.*, vol. 13, no. 14, p. 2784, Jul. 2021.

[61] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[62] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, and Y. Yao, "Deep distillation recursive network for remote sensing imagery super-resolution," *Remote Sens.*, vol. 10, no. 11, p. 1700, Oct. 2018.

[63] Z. He, T. Dai, J. Lu, Y. Jiang, and S.-T. Xia, "FAKD: Feature-affinity based knowledge distillation for efficient image super-resolution," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 518–522.

[64] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 723–731.

[65] Y. Zhang, H. Chen, X. Chen, Y. Deng, C. Xu, and Y. Wang, "Data-free knowledge distillation for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7848–7857.

[66] P. Campisi and K. Egiazarian, *Blind Image Deconvolution: Theory and Applications*. Boca Raton, FL, USA: CRC Press, 2017.

[67] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouame, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for $\ell_2-\ell_2$ problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, May 2016.

[68] P. H. C. Eilers and C. Ruckebusch, "Fast and simple super-resolution with single images," *Sci. Rep.*, vol. 12, no. 1, p. 11241, Jul. 2022.

[69] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[70] L. Wang et al., "Unsupervised degradation representation learning for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10576–10585.

[71] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[72] L. Wald, *Data Fusion. Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.

[73] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[74] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Aug. 2002.

[75] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

Alessia Carbone (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees (summa cum laude) in information engineering from the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Salerno, Italy, in 2020 and 2022, respectively, where she is currently pursuing the Ph.D. degree in information engineering.

Her research was supported by the CNR-IMAA, Tito, Potenza, Italy, the Italian research council's institute of methodologies for environmental analysis, and the University of Salerno. Her main academic interests are remote sensing, image processing, and machine learning.

Rocco Restaino (Senior Member, IEEE) received the Laurea degree in electronic engineering from the University of Naples, Naples, Italy, in 1998, and the Ph.D. degree in information engineering from the University of Salerno, Salerno, Italy, in 2002.

He is currently an Associate Professor at the University of Salerno. His research interests include probability theory, stochastic geometry, and signal processing for remote sensing and networking.

Dr. Restaino received the IGARSS 2015 Symposium Best Paper Award.

Gemine Vivone (Senior Member, IEEE) received the B.Sc. (summa cum laude), M.Sc. (summa cum laude), and Ph.D. (Hons.) degrees in information engineering from the University of Salerno, Salerno, Italy, in 2008, 2011, and 2014, respectively.

He is a Senior Researcher at the National Research Council, Italy. His main research interests focus on image fusion, statistical signal processing, deep learning, and classification and tracking of remotely sensed images.

Dr. Vivone is an Ex-Officio Member of the IEEE Geoscience and Remote Sensing Society (GRSS) Administrative Committee and a member of the IEEE Task Force on "Deep Vision in Space." He received the IEEE GRSS Early Career Award in 2021, the Symposium Best Paper Award by the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2015, and the Best Reviewer Award of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2017. Moreover, he is listed in the World's Top 2% Scientists by Stanford University. He is the Co-Chair of the IEEE GRSS Image Analysis and Data Fusion Technical Committee. He was the Leader of the Image and Signal Processing Working Group of the IEEE Image Analysis and Data Fusion Technical Committee from 2020 to 2021. He is currently the Editor-in-Chief of IEEE GEOSCIENCE AND REMOTE SENSING eNEWSLETTER, an Area Editor of Elsevier *Information Fusion*, and an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL). Moreover, he is an Advisory Board Member of *ISPRS Journal of Photogrammetry and Remote Sensing*, and an Editorial Board Member of Nature *Scientific Reports* and MDPI *Remote Sensing*.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

From 1999 to 2023, he was with Grenoble INP, where he was a Professor of signal and image processing. He was a Visiting Scholar at Stanford University, Stanford, CA, USA; KTH, Stockholm, Sweden; and NUS, Singapore. Since 2013, he has been an Adjunct Professor of the University of Iceland, Reykjavik, Iceland. From 2015 to 2017, he was a Visiting Professor at the University of California, Los Angeles (UCLA), Los Angeles, CA, USA. He holds the AXA Chair in remote sensing and an Adjunct Professor at the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. He is currently a Research Director with INRIA, Grenoble. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008, an ELLIS Fellow, a fellow of the Asia-Pacific Artificial Intelligence Association, a member of the Institut Universitaire de France from 2012 to 2017, and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, since 2018). He is the Founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010 which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS). He was the Chair from 2009 to 2011 and the Co-Chair of the GRS Data Fusion Technical Committee from 2005 to 2008. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society and in charge of meetings and symposia from 2017 to 2019. He is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and PROCEEDINGS OF THE IEEE. He was the Editor-in-Chief of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2014, he served as a Guest Editor of *IEEE Signal Processing Magazine*.