









Hyperspectral Image Super-Resolution: Task-Based Evaluation

Michał Kawulok , Senior Member, IEEE, Paweł Kowaleczko , Maciej Ziaja, Jakub Nalepa , Senior Member, IEEE, Daniel Kostrzewa , Member, IEEE, Daniele Latini , Davide De Santis , Member, IEEE, Giorgia Salvucci, Ilaria Petracca , Student Member, IEEE, Valeria La Pegna, Zoltan Bartalis, and Fabio Del Frate , Senior Member, IEEE

Abstract—The need for enhancing image spatial resolution has motivated the researchers to propose numerous super-resolution (SR) techniques, including those developed specifically for hyperspectral data. Despite significant advancements in this field attributed to deep learning, little attention has been given to evaluating the practical value of super-resolved images in specific applications. Most methods are validated in application-independent scenarios, often using simulated low-resolution images, resulting in overly optimistic conclusions. In this article, we propose task-based evaluation strategies for hyperspectral image SR and we present results obtained with various approaches that include pansharpening, multispectral–hyperspectral data fusion, and single-image SR. We demonstrate that the proposed framework allows us to highlight both benefits and limitations of each method and can, therefore, guide the development of SR techniques suitable for real-world applications.

Index Terms—Hyperspectral images (HSIs), spatial resolution, spectral consistency, super-resolution (SR), task-based evaluation.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain both spatial and spectral information, which allows for understanding the structure, as well as the chemical–physical composition of

Received 29 April 2024; revised 5 September 2024; accepted 23 September 2024. Date of publication 7 October 2024; date of current version 23 October 2024. This work was supported in part by European Space Agency (the PIGEON project realized by KP Labs and GEO-K), under Grant 4000137227. The work of Michał Kawulok was supported by the National Science Centre, Poland, under Research Grant 2022/47/B/ST6/03009. (Corresponding author: Michał Kawulok.)

Michał Kawulok, Maciej Ziaja, Jakub Nalepa, and Daniel Kostrzewa are with the Silesian University of Technology, 44-100 Gliwice, Poland, and also with the KP Labs, 44-100 Gliwice, Poland (e-mail: michal.kawulok@polsl.pl; maciej.ziaja@polsl.pl; jakub.nalepa@polsl.pl; daniel.kostrzewa@polsl.pl).

Paweł Kowaleczko is with the KP Labs, 44-100 Gliwice, Poland, and also with the Warsaw University of Technology, 00-661 Warsaw, Poland (e-mail: pawel.kowaleczko@pw.edu.pl).

Daniele Latini is with the GEO-K s.r.l., 00133 Rome, Italy (e-mail: daniele.latini@geo-k.co).

Davide De Santis, Giorgia Salvucci, Ilaria Petracca, Valeria La Pegna, and Fabio Del Frate are with the GEO-K s.r.l., 00133 Rome, Italy, and also with the Tor Vergata, University of Rome, 00133 Rome, Italy (e-mail: davide.de.santis@uniroma2.it; giorgia.salvucci@uniroma2.it; ilaria.petracca@uniroma2.it; valeria.la.pegna@uniroma2.it; fabio.del.frate@uniroma2.it).

Zoltan Bartalis is with the European Space Agency, 00044 Frascati, Italy (e-mail: zoltan.bartalis@esa.int).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSTARS.2024.3475644>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2024.3475644

the captured surfaces. This makes HSIs a valuable source of information that can be exploited in numerous remote sensing applications [37]. However, especially for acquisitions from space, their high spectral resolution is commonly achieved at a cost of reduced spatial resolution. This poses a serious limitation in those cases that require high accuracy not only in the spectral dimension but also in the spatial domain [9]. If the problem of insufficient spatial resolution cannot be solved by employing a sensor of higher resolution or by approaching the investigated object, which is often the case in remote sensing, then the captured images can be postprocessed relying on computational imaging techniques. In particular, the goal of super-resolution (SR) is to reconstruct a high-resolution (HR) image from a low-resolution (LR) observation [55].

In the case of HSI SR, the existing techniques can be roughly categorized into the following:

- 1) those that exploit another source of HR information—either a panchromatic (PAN) [1] or multispectral image (MSI) [59];
- 2) those that process a single hyperspectral cube without any auxiliary data [15].

As capturing an image of high spectral resolution is often subject to a tradeoff with the spatial resolution, many hyperspectral sensors are coupled with a PAN camera that captures images at a higher spatial resolution. This allows for performing spectral sharpening that consists in enhancing the resolution of spectral bands relying on HR information. Spectral sharpening that exploits a PAN image is commonly termed as *pansharpening*.

While the field of SR has received considerable research attention [55], most of the techniques are evaluated following an artificial procedure: a certain image is considered as an HR reference, which is downsampled to obtain a simulated LR input image. This is also the case for HSI SR [59], where LR images are obtained with Wald’s protocol [49] that is supposed to preserve the spectral properties while decreasing the spatial resolution of an HSI. Such evaluation often leads to overestimating the actual capabilities of SR techniques—even though they obtain promising results for the simulated LR images, their performance is significantly worse when they are fed with real-world (i.e., not degraded) imagery [5]. Therefore, in an increasing number of cases, SR outcome is validated using real-world datasets that contain original LR and HR images showing the same area of interest [8], [25], [33]. However, such datasets

would be very difficult to elaborate for HSIs, especially for satellite imagery. Another option for determining the quality of super-resolved images is to exploit them for a specific computer vision task, whose performance can be quantitatively evaluated. Such task-based evaluation attempts were already reported for natural [31] and satellite images [40], but they were not proposed so far for HSI SR.

A. Contribution

In this article, we present a new task-based approach toward validating SR algorithms, applied to PRISMA and Tropospheric monitoring instrument (TROPOMI) HSIs. Our contribution can be summarized in the following points.

- 1) We propose to evaluate the super-resolved HSIs relying on three distinct test cases that exploit hyperspectral imagery for precision farming, estimating the quality of inland and coastal waters, and assessing air pollution. To our best knowledge, this is the first attempt to assess HSI SR in a task-based manner.
- 2) We report the results of both application-independent and task-based evaluation for three different approaches to enhancing HSIs: single-HSI SR for PRISMA data that operates without any auxiliary HR information, pansharpening of PRISMA HSIs, and fusion between HSIs of different spatial resolution (TROPOMI with higher spectral and lower spatial resolution and PRISMA with lower spectral and higher spatial resolution). This overall assesses spatial and spectral aspects of the super-resolved HSIs.
- 3) For HSIs acquired within the PRISMA mission, we provide a comparison between single-HSI SR that operates without any auxiliary HR information, enhancing the spatial resolution from 30-m ground sampling distance (GSD) to 15-m GSD and pansharpening that exploits a PAN image of 5-m GSD. In the latter case, we consider two variants, in which we enhance the resolution to 5-m GSD and to 15-m GSD.

In order to ensure the reproducibility of our study, we published the source code of the investigated enhancement techniques¹ and of the considered test cases along with the trained models.²

B. Article Structure

The rest of this article is organized as follows. Section II outlines the state of the art in HSI SR, including techniques that exploit additional source of HR information and those that are limited to processing a single HSI. The proposed test cases alongside the data exploited in our study are presented in Section III and the SR techniques selected to address these cases are outlined in Section IV. The results of application-independent and task-based evaluation are reported in Section V and they are discussed in Section VI. Finally, Section VII concludes this article.

II. RELATED WORK

In this section, we present the state of the art in fusion-based techniques that exploit an auxiliary PAN image or an MSI of higher spatial resolution (see Section II-A) as well as single-HSI SR techniques (see Section II-B). Also, in Section II-C, we shed more light on the commonly adopted approaches to evaluate HSI SR techniques.

A. Enhancement Based on Auxiliary HR Images

The existing approaches to pansharpening can be roughly categorized into four groups [7], namely: component substitution (CS) techniques, multiresolution analysis (MRA), model-based approaches, and deep-learning-based methods. The CS approaches employ a suitable technique, e.g., principal component analysis (PCA), to decompose an HSI into separated spatial and spectral components. Subsequently, the projection process is inverted after substituting the spatial components with the HR PAN image to generate the super-resolved HSI [41]. Alternatively, in MRA, high-frequency information can be extracted from the PAN image and injected into the original HSI [34]. The model-based methods are underpinned with an optimization problem whose solution retrieves the pansharpened image [44].

The recent advancements in the field of pansharpening are mostly attributed to deep learning. Wu et al. [57] proposed a triple-branch CNN—in the first branch, a PAN image is fused with an HSI upsampled by a factor of $4\times$, while the remaining branches process an original HSI and the HSI upsampled by a factor of $2\times$. A similar approach was proposed in [22], where a hyperspectral and PAN images are processed at different scales using a Laplacian pyramid. Bandara et al. [2] utilized the deep image prior (DIP) to estimate the backbone HR HSI and afterwards they proposed an attention mechanism in a transformer-based HyperTransformer architecture for HSI pansharpening [1]. In [63], TNet and PNet networks are introduced. TNet learns to reconstruct the gradient map from a PAN image (in both horizontal and vertical directions). PNet, trained afterwards to perform the pansharpening, is guided during training by the pretrained TNet, exploited as a component of the loss function.

Some recent techniques based on deep learning also exploit the conventional approaches. Deng et al. [11] proposed a technique that combines CS and MRA approaches—MRA-Net and CS-Net learn the residuals based on the PAN image, which is afterwards added to the upsampled LR HSI. A PCA-based network was introduced by Guarino et al. [18] and a model-based optimization regularized with a deep prior was proposed in [60].

All of the aforementioned methods were developed and validated using simulated data in a fully supervised manner. Alternatively, pansharpening networks can be trained in an unsupervised way without using HR references. A method by Zhou et al. [68] is based on a generative adversarial network (GAN), which at first generates the super-resolved HSI using an HSI and a PAN image, and then, restores the original (LR) HSI and the PAN image from the super-resolved HSI. Unsupervised learning can also be performed by exploiting the attention mechanism [39]. Inspired by DIP, Uezato et al. [47] proposed a guided deep

¹[Online]. Available: <https://github.com/kplabs-pl/hsi-sr-evaluation>

²[Online]. Available: <https://github.com/geo-k-devs/PIGEON>

decoder. The features of the PAN image are extracted by the U-Net architecture at various levels and injected into the DIP decoder, without needing an HR reference.

Alternatively to a PAN image, an MSI of higher spatial resolution can be exploited to super-resolve an HSI. Palsson et al. [36] used 3-D convolutions for this purpose. Ma et al. [32] employed deep unfolded neural networks to learn both spatial and spectral priors that are exploited for the fusion. Zhang et al. [67] proposed to first upsample the HSI to the resolution of an MSI and combine them to obtain a preliminarily fused image. It is subsequently processed in a sequential manner by two networks that reconstruct spatial and spectral information. Several techniques have been inspired by conventional approaches implemented using deep networks [12], [13]. In [53], the fusion process is composed of learning the observation and fusion models that perform opposite mappings between the input HSI and MSI, and the super-resolved HSI. Xie et al. [58], [59] formulated the fusion in a form of an optimization problem. The problem can be solved using an iterative algorithm whose subsequent steps were unfolded into network layers of the proposed MSI/HSI Fusion Net (termed MHF-net) [59]. The network has been later enhanced to improve its interpretability [58]. Commonly, the existing techniques require the MSI and HSI to be accurately coregistered before proceeding with the fusion. Recently, Qu et al. [38] proposed to combine the registration and fusion steps within a single architecture that super-resolves the HSI in an end-to-end manner without the need for registering MSI and HSI beforehand.

Similarly as for pansharpening, unsupervised learning was also explored for fusing HSIs with MSIs. Gao et al. [16] proposed a self-supervised approach based on adaptive Gram-Schmidt transformation, which renders a backbone HR HSI that is used for computing spectral and spatial loss functions.

B. HSI SR Without Auxiliary HR Information

When an auxiliary HR image is not available, HSIs can be super-resolved in a band-wise manner [35] by employing single-image SR techniques [61] to enhance each individual channel. In the reported research, we have exploited a hybrid attention transformer (HAT) [6] for this purpose. However, enhancing each band independently does not ensure preserving spectral consistency between the input image and the super-resolved one. Therefore, a number of techniques have been proposed to specifically deal with HSIs in such cases. Li et al. [27] proposed to train two SR networks (sharing the encoder and having separate decoders) for enhancing HSIs and natural color (RGB) images. In this way, the HSI SR is regularized by the RGB SR network, which makes it possible to exploit RGB data whose accessibility is much larger. In [45], an HSI is split into several groups of adjacent bands, each of which is processed in a recursive manner. Hu et al. [20] proposed to process the HSI channel by channel. Each channel, along with the adjacent channel and residual image of those channels, is processed in a triple-branch architecture. To omit the problem of reference data deficiency, Sidorov et al. [43] employed DIP to reconstruct HSIs from a random noise relying on prior knowledge

embedded into the network structure. An interesting approach combining spectral unmixing with SR aimed at improving image segmentation capabilities was proposed by Wang et al. [51]. This allows for retrieving land-cover maps at subpixel precision, benefiting from spatial-spectral correlation. Such SR mapping can be additionally improved by taking into account the point spread function [50].

In order to exploit spectral correlations among the neighboring bands, 3-D CNNs have also been exploited for single-HSI SR. As the use of 3-D convolutions significantly increases the number of learnable parameters, Li et al. [28] combined 3-D and 2-D convolutional layers to extract the spatial-spectral features while keeping the number of parameters at an acceptable level. To reduce the number of parameters even further, separable 3-D convolutions were used with two orthogonal 2-D kernels instead of full 3-D kernels. A similar approach was reported in [52], where features extracted by the 2-D and separable 3-D convolutions are later fused in the process to obtain the final result. Fu et al. [15] proposed bidirectional 3-D quasi-recurrent neural network (Bi-3DQRNN). In this U-Net-like architecture, the features are processed in both directions along the spectral dimension which helps exploit the correlation between the bands.

The use of attention mechanism [48] has also been reported for single-HSI SR. Li et al. [26] utilize band attention in a generic GAN setup. In [29], group convolutions were employed to reduce the number of parameters—highly correlated bands are grouped and processed by a common kernel. The attention mechanism based on the covariance statistics of features is used to exploit the spectral prior. Such group convolutions, in this case for multiple adjacent channels, were also used in [21]. Each group is processed in a separate branch using 2-D convolutions with the spectral channel attention mechanism. Features in each branch are later upsampled, fused, and processed again in the same manner as in the group branches to obtain the final HR HSI. Wang et al. [54] also divided the consecutive bands of HSI into groups with the same number of bands. These bands are then processed in a recursive manner using a stack of residual attention blocks. Each block receives the reconstructed bands along with the embedding from the previous group. The network is topped with an additional regularization block with separable 3-D convolutions to obtain the final super-resolved HSI. Recently, Zhang et al. [64] employed 3-D convolutions with channel attention to extract spatial-spectral features, whose diversity is additionally enhanced using a dedicated module.

The advancements in general-purpose SR have been successfully deployed into the field of HSI SR. Vision transformers have been exploited for single-HSI SR by Zhang et al. [65], and Dong et al. [14] proposed to exploit the denoising diffusion models to elaborate scale-flexible HSI SR with improved interpretability.

C. Evaluation of SR Outcome

SR models can be evaluated quantitatively or qualitatively [3]. The former is at best performed relying on an HR reference and the similarity of the reconstructed image to the reference indicates the SR quality. The similarity in the spatial domain

is commonly measured with peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [19], [56], as well as with learned perceptual image patch similarity (LPIPS) [66]. To verify the similarity in the spectral domain, the spectral angle mapper (SAM) [62] is employed. If the reference is not available, which is often the case when super-resolving original (rather than simulated) images, then the evaluation can either be carried out relying on no-reference metrics [4], or the resulting image can be downsampled back to the original size to verify the consistency [46]. The qualitative assessment may rely on a mean opinion score, retrieved throughout a survey, whose participants provide subjective feedback on the reconstruction quality. In [25], it was demonstrated that this subjective evaluation aligns well with the quality assessed with LPIPS. However, while qualitative evaluation is helpful in inspecting the spatial features, it can be hardly applied to verify the spectral properties.

An interesting research direction is to use higher level image analysis tasks to evaluate the performance of SR outcome. In [42], the performance of single-image SR [24] applied to satellite imagery was assessed in terms of detecting vehicles of different type, while Razzak et al. [40] proposed to evaluate Sentinel-2 SR relying on building delineation. However, such task-based evaluation has not been reported for HSI SR so far, and the motivation for our study was to address this research gap.

III. PROPOSED EVALUATION TEST CASES

In the research reported here, we investigated a range of HSI enhancement techniques in the context of different specialized scenarios. Our aim was to prepare several SR methods and apply them to existing hyperspectral imagery, then examine how useful the super-resolved products are in a variety of environmental monitoring tasks. We have investigated this idea for a set of experiments that encompass PRISMA and TROPOMI satellite data.

In Section III-A, we briefly describe the characteristics of each of the data sources under consideration. In Sections III-B and III-C, we introduce specific Earth observation use cases that exploit either PRISMA or TROPOMI imagery. The source code for all the use cases is available online.³ These use cases are utilized to validate the models (presented later in Section IV) that enhance the PRISMA and TROPOMI data. This allows us to compare the results of applying Earth observation models before and after subjecting PRISMA and TROPOMI data to image enhancement processes.

A. Data Source

1) *PRISMA Data*: PRISMA satellite imagery comprises two hyperspectral cubes: one for visible and near-infrared (VNIR) wavelengths and another for shortwave infrared (SWIR) wavelengths, alongside a PAN. The VNIR cube features 400–1010-nm range with 66 spectral channels, the SWIR cube encompasses spectral range of 920–2505 nm with 173 channels. Both subcubes are of 30-m GSD and capture 1000×1000 pixels

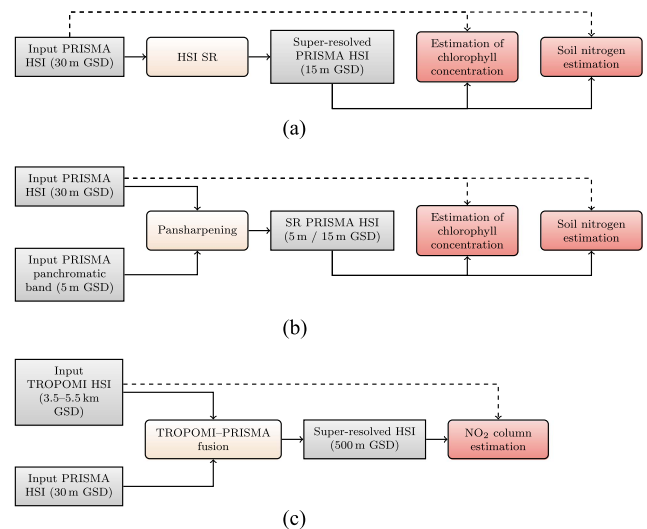


Fig. 1. Flowcharts of the task-based evaluation. The precision farming and water quality scenarios are employed to validate (a) single-HSI SR and (b) pansharpening of PRISMA images, while the (c) air pollution scenario validates the TROPOMI–PRISMA fusion. The dashed lines indicate the processing pipeline performed from the original data, whose outcome is treated as a reference. (a) Single-HSI SR. (b) Pansharpening. (c) HSI–MSI fusion.

covering 900 km^2 . The PAN channel covers 400–700 nm wavelength and features $6 \times$ finer GSD (5 m) than the HSI cubes. We exploit the original Level-1 (air pollution scenario) and Level-2D (precision farming and water quality scenarios) PRISMA products that have been orthorectified and georeferenced.

2) *TROPOMI Data*: The TROPOMI sensor is the payload of the Sentinel-5P mission and it is an imaging spectrometer covering from ultraviolet to SWIR bands of the electromagnetic spectrum. In our study, we focus on ultraviolet-visible (UVIS) band 4 of the TROPOMI data in range 405–500 nm with 497 channels. The instrument operates in a push-broom configuration, with a swath width of 450 pixels spanning around 2600 km. This setup results in a typical TROPOMI hyperspectral pixel size of $5.5 \text{ km} \times 3.5 \text{ km}$ (along \times across track, since 6 August 2019). The push-broom sensing of TROPOMI covers vast parts of the globe often spanning both polar poles.

B. Precision Farming Scenario

The precision farming scenario consists in estimating the soil nitrogen content from original PRISMA Level-2D product at 30-m GSD, from the images enhanced with single-HSI SR [see Fig. 1(a)] to 15-m GSD, as well as from the pansharpened images at 5-m and 15-m GSD [see Fig. 1(b)]. While the possibility to estimate the soil profile from HSIs represents a breakthrough in terms of costs and scale of representation, the 30-m resolution is insufficient for practical precision farming applications. Therefore, on the one hand, exploiting the super-resolved product may enhance the practical significance of such analysis, but on the other hand, it is not clear whether the estimation performed from the SR outcome can be trusted.

To better understand the value of pansharpened images in this context, the estimation performed from the original PRISMA images, as well as from the super-resolved ones, has been

³[Online]. Available: <https://github.com/geo-k-devs/PIGEON>

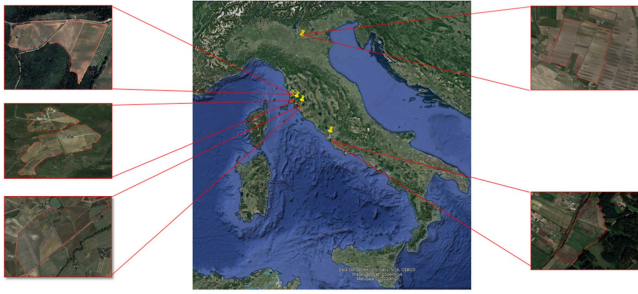


Fig. 2. Sites with measurements available from CREA considered to build the dataset for precision farming scenario.

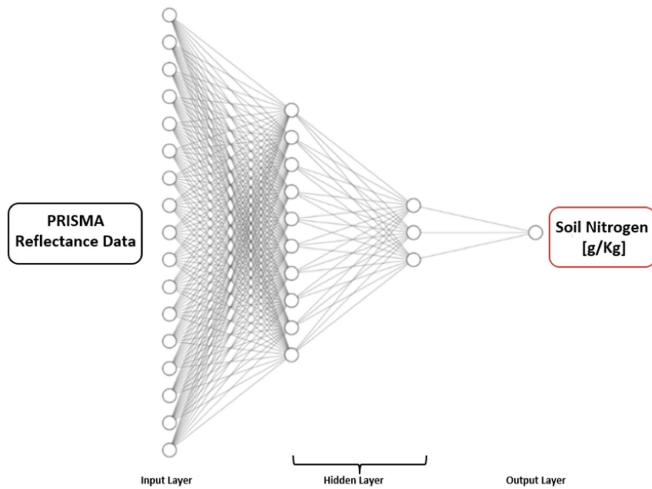


Fig. 3. Architecture of an MLP employed for retrieving soil nitrogen in the precision farming test case.

compared with ground-truth measurements provided by Italian Council for Agricultural Research (CREA). This renders a quantitative assessment of SR performed from original (not downsampled) data. The soil profiles dataset consists of 482 samples acquired between 2019 and 2022, which is a reference operational period of the PRISMA mission. In order to ensure the reliability of the ground-truth data, we assumed that the temporal distance between the ground-truth measurement and HSI acquisition time must be at most two weeks. Considering the availability of the PRISMA images,⁴ this limited the dataset to 55 points overlapping with a set of five PRISMA images (see Fig. 2; the exact locations are provided in the supplementary material). Cloudy points were then excluded from this dataset, reducing the set to 26 samples.

In order to estimate the soil nitrogen content, we have decided to exploit a multilayer perceptron (MLP), as it was successfully tested for this purpose [30]. The architecture of the employed MLP is presented in Fig. 3; it is composed of 17 neurons in the input layer corresponding to PRISMA NIR bands (43–56) and red edge bands (40–42), followed with two hidden layers with 10 and 3 neurons, respectively, and topped with an output layer with a single neuron that represents the soil nitrogen value expressed in grams of nitrogen per kilogram of soil.

⁴We searched the PRISMA catalogue available at <https://prisma.asi.it>



Fig. 4. Test sites considered to build a comprehensive dataset for water quality scenarios. Locations shown on the map include both inland and coastal areas.

To have sufficient data for training the neural network that estimates the amount of nitrogen on the ground, the dataset was expanded through data augmentation techniques generating a set of 87 samples. Specifically, for each range of nitrogen values, synthetic but representative data were created for the variable being inverted, considering variance and standard deviation for each of the spectral bands considered. This dataset augmentation approach allows us to maintain the scientific validity and statistical robustness of the dataset while increasing the training and generalization capability of the neural network algorithm.

C. Water Quality Scenario

The water quality test case consists in estimating the chlorophyll-a concentration, being one of the main parameters to evaluate water status in aquatic environments. The retrieval accuracy based on the original PRISMA data volume was compared with the one from the SR images, which reach a significant increase in spatial resolution, obtaining for each methodology pixel sizes from 30 to 15 m.

This analysis has been performed for both inland freshwater and sea coastal saltwater. In order to ensure variability, Italian coastal areas near seaports and densely populated areas have been selected, as well as lagoons and lakes of different sizes and located at different latitudes within the country. The distribution of the selected areas over the Italian territory can be observed in Fig. 4 (for exact locations, see the supplementary material). For the inland application, consisting of fresh waters in internal lakes and including lagoons, a number equal to ten PRISMA L2D images for a total of 290 samples were used. As for the coastal case, 12 PRISMA L2D images were analyzed, finally giving 270 samples equally distributed within the displayed scenes. Given the scarcity and high variability of ground-truth

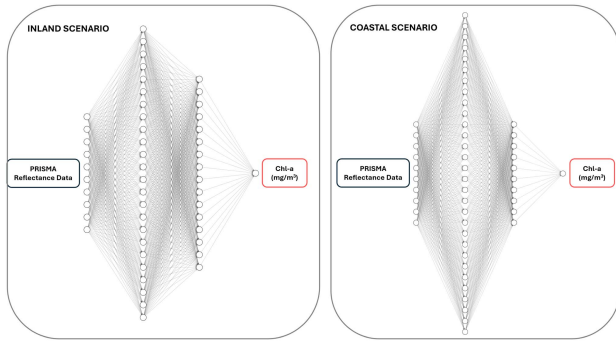


Fig. 5. Architecture of an MLP employed for retrieving chlorophyll-a concentration in inland and coastal areas for the water quality test case.

information regarding the concentration of chlorophyll-a in waters, it was necessary to exploit the available products derived from the Ocean and Land Color Instrument (OLCI) mounted on the Sentinel-3 mission. Chlorophyll-a concentrations, retrieved from Sentinel-3, were collected for each PRISMA image. As temporal resolution of Sentinel-3 is one image per day, it always coincides with the day of PRISMA data acquisition. However, images with cloud cover over the aquatic areas were discarded. Identifying a long-established satellite product as ground truth allows for ensuring coincidence of acquisitions over time and a unique spatial matching for both fresh and sea waters. However, this could introduce a considerable level of uncertainty that derives from the large spatial resolution of these products that present pixels of 300-m GSD.

As in the precision farming test case, here we also developed an MLP, widely used for retrieval prediction analysis of bio/physical parameters in waters [23]. Input data to the model are represented by reflectances of ten bands in the visible channels with wavelengths included in the range 446–764 nm, covering both the absorption and fluorescence windows of chlorophyll-a, having its absorption peaks in the blue (430–450 nm) and in the red (660–680 nm) regions of the spectrum and its fluorescence in the red channels, typically around 680–700 nm [17]. Due to the different aquatic characteristics, as salinity, turbulence, and transparency of the waters to assess, a distinct algorithm for inland and coastal areas has been created, slightly adjusting the selections of the ten bands sensitive to the chlorophyll-a in the visible range. Models architectures are very similar between the two, having the same number of input layers, but using different PRISMA bands, leading to consider the most sensitive wavelengths to chlorophyll-a depending on the aquatic environment. The number of nodes in the hidden layers equals 24 and 16 for the inland case, and 30 and 10 for the coastal one, the final output for both models consists in the value of chlorophyll-a concentration present in the water. The topology of the two implemented networks that retrieve chlorophyll-a concentration from the original data can be observed in Fig. 5.

Standard preprocessing operations were applied to the data. As for the OLCI products from Sentinel-3, these directly provide chlorophyll-a concentration values, expressed in mg/m^3 . Equally distributed points were chosen within the images in

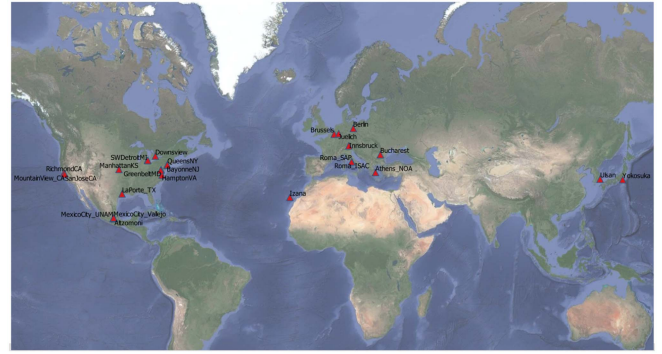


Fig. 6. Pandonia network sites considered to build the dataset for the air pollution scenario.

order to ensure consistent heterogeneity of chlorophyll distribution. For what concerns the PRISMA L2D data, after the coregistration process, the reflectances value for the channels of the VIS cube sensitive to chlorophyll-a were selected in correspondence with the locations chosen as ground truth.

D. Air Pollution Scenario

The air pollution scenario aims at retrieving the NO_2 column from fused TROPOMI Level-1B of low spatial resolution and PRISMA Level-1 images of higher spatial resolution, and the obtained values are compared against the ground-truth measurements at the surface level [see Fig. 1(c)]. Nitrogen dioxide (NO_2) is a key pollutant that receives significant attention in assessing air quality, along with particulate matter. NO_2 is closely linked to human activities, such as traffic and industrial emissions, as well as natural phenomena such as wildfires and lightning. Nitrogen dioxide emissions are carefully monitored and regulated to mitigate their impact on air quality and human health.

The NO_2 products with the highest spatial resolution currently available from space are provided by the Sentinel-5P mission ($3.5 \text{ km} \times 5.5 \text{ km}$). Even though the TROPOMI–PRISMA fusion allows for obtaining the resolution of 30 m, we obtain a TROPOMI SR product of 500-m GSD, as it already represents a very interesting objective to achieve for this first test case of TROPOMI–PRISMA fusion. To validate the estimation performed from the super-resolved images, we have exploited the ground-truth data provided by Network for the Pandonia Global Network (PGN).⁵ We have matched the PGN sites (shown in Fig. 6) with available PRISMA data considering the following conditions:

- 1) image covering a 3-km radius around the ground instrument;
- 2) cloud coverage below 0.3;
- 3) temporal distance between PRISMA image acquisition and the measurement not greater than 2 h;
- 4) temporal distance between TROPOMI data acquisition and the measurement not greater than 2 h.

⁵PGN data are freely available at <http://data.pandonia-global-network.org>

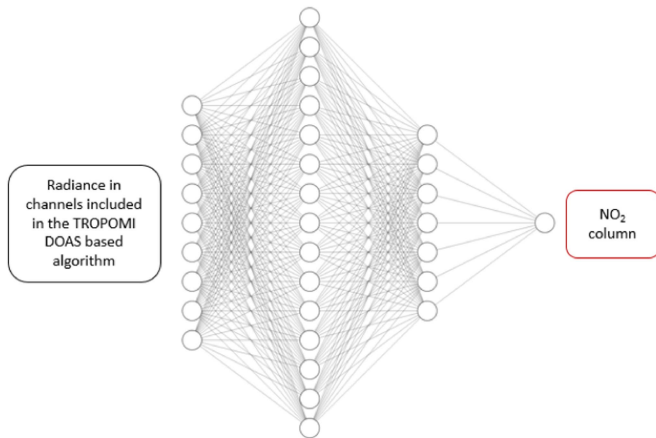


Fig. 7. Architecture of an MLP employed for NO_2 column retrieval.

These criteria were met by 76 PRISMA and TROPOMI images acquired between the year 2020 and 2022.⁶ To augment the dataset, we exploited the wide swath of TROPOMI L1B offline products for extracting additional TROPOMI spectra corresponding to the ground truth available for all the sites included in each of the considered orbits. As a result, we generated 354 samples.

To retrieve the NO_2 level, we employed an MLP (see Fig. 7) with 36 neurons in the input layer. These correspond to 34 radiances of the TROPOMI L1BD4 band, distributed every 2 nm within the 400–466 nm portion of the electromagnetic spectrum. This spectral range aligns with the one used in the algorithm based on differential optical absorption spectroscopy as reported in the TROPOMI documentation.⁷ In addition, latitude and longitude were included in the input vector. Two hidden layers are composed of 1000 and 500 neurons, respectively, considering a dropout factor of 0.5. The output layer contains a single neuron that retrieves the estimated NO_2 column.

IV. SELECTED SR TECHNIQUES

In this section, we outline the methods selected for enhancing PRISMA images without and with PAN image as a source of auxiliary HR information (see Section IV-A), and for enhancing the spectra of PRISMA images based on TROPOMI bands (see Section IV-B). The source code for all the SR techniques is available online.⁸

A. PRISMA Image Enhancement

In our study, we explored both pansharpening and single-HSI SR techniques for enhancing the spatial resolution of PRISMA images that are aimed at the scenarios outlined earlier in Sections III-B and III-C.

⁶The locations for the air pollution test case are reported in the supplementary material.

⁷The TROPOMI documentation on NO_2 retrieval is available at <https://sentinel.esa.int/documents/247904/2476257/sentinel-5p-tropomi-atbd-no2-data-products>

⁸[Online]. Available: <https://github.com/kplabs-pl/hsi-sr-evaluation>

For pansharpening, we selected two methods, namely a simple well-established approach based on PCA [41] and a state-of-the-art HyperTransformer model [1]. The PCA-based technique consists in decomposing an HSI (already bicubically upsampled to the size of the PAN image) into a set of uncorrelated components [41]. Subsequently, the first principal component is substituted with the HR PAN image and PCA projection is inverted to retrieve the super-resolved HSI. This process assumes that the PAN information is gathered in the first component, which is not always the case, especially if the spectral range of an HSI is different from the spectral range of the PAN image.

HyperTransformer utilizes the attention mechanism [48] within a vision transformer architecture, achieving leading results in the field of pansharpening. The HyperTransformer architecture is composed of two separate convolutional branches that extract features from PAN images and HSIs. The features extracted from the original PAN image are used as the *value* for the vision transformer that performs pansharpening, the features extracted from a PAN image subject to downsampling followed by upsampling (which serves as a low-pass filter) are used as the *key*, and the features extracted from an HSI (upsampled to match the PAN image resolution) are used as the *query*. These are then processed through a multihead feature soft-attention module, the output of which is integrated using a spatial–spectral feature fusion module to generate the pansharpened image.

The original HyperTransformer configuration is not directly applicable to the PRISMA data due to the differences in the number of spectral channels and the GSD ratio between the HSIs and PAN images, compared to those used in the original study. To adapt the model to the PRISMA data, we modified the architecture by changing the image upsampling factor from $4\times$ to $6\times$. Originally, the model performs pansharpening of HSI cubes in two steps, each of which doubles the image size, resulting in an overall upsampling factor of $4\times$. Our modified version accommodates the hyperspectral versus PAN image ratio of $6\times$ by employing the factors of $2\times$ and $3\times$ consecutively, achieving a composite upsampling factor of $6\times$.

In order to enhance PRISMA images without exploiting any auxiliary HR information, we selected two different approaches. The first relies on spatial features and enhances an HSI in a band-wise manner without considering spectral correlations between neighboring bands. The second approach analyzes the entire hyperspectral cube and takes advantage of spatial–spectral features. In both cases, we employed recent techniques [6], [15] underpinned with deep learning that we have trained with the data simulated using Wald’s protocol [49].

For band-wise SR, we adopted HAT [6], which is a state-of-the-art SISR technique. It employs a vision transformer to activate more pixels in the receptive field, making it possible to extract more information concerning the image structure. We have exploited the original technique without any modifications, setting the upsampling factor to $2\times$. After training, we used HAT to super-resolve the HSI by enhancing every channel independently.

To exploit the spatial–spectral features, we selected the Bi-3DQRNN network [15], which is a state-of-the-art solution for enhancing HSIs. This network utilizes 3-D convolutions, which

are particularly effective for extracting both spectral and spatial correlations simultaneously. However, the use of 3-D convolutions typically results in increased computational demands and a larger number of parameters. In addition, 3-D convolutions can only process a limited number of adjacent spectral bands at a time, which is insufficient to capture global spectral dependencies, especially given that for the encoder–decoder Bi-3DQRNN architecture, the number of channels increases in the intermediate layers compared to the input and output. To address that problem, Bi-3DQRNN contains a bidirectional quasi-recurrent pooling module, which processes the intermediate features bidirectionally along the spectral dimension, enhancing the network’s ability to manage spectral information effectively. As the number of input spectral channels is configurable in Bi-3DQRNN, we could straightforwardly apply it to enhancing PRISMA images. As in HAT, we set the upsampling factor to $2\times$. Bi-3DQRNN requires the input to be provisionally upsampled to the desired resolution, so we employ bilinear interpolation to each spectral channel.

B. Fusion of TROPOMI and PRISMA Images

In order to enhance spatial resolution of TROPOMI images relying on corresponding PRISMA data, we have selected a state-of-the-art MHF-net [58], [59]. Originally, it was validated for fusing HSIs with color RGB images that convey HR information. MHF-net is underpinned with a comprehensive observation model that is supposed to reflect the low-rankness of the HR HSI, from which the observed HR MSI and LR HSI are supposed to be generated. Reconstruction of HR HSI is achieved by solving the optimization problem using proximal gradient algorithm, unfolded into a deep neural network architecture. Even though MHF-net was tested using MSIs with three channels (RGB), the number of spectral channels is configurable, so we could straightforwardly adapt it to our case. However, the ratio between spatial resolution of HSI and MSI is hardcoded to $32\times$, which is inadequate for direct application to our data due to the significantly higher ratio between TROPOMI and PRISMA resolutions. For this reason, we downsample the PRISMA images to approximately 188 m GSD (160×160 pixels)—around $32\times$ smaller than the TROPOMI’s GSD—and then fuse them with the TROPOMI data. Subsequently, the super-resolved HSI is further downsampled to 500-m GSD, which we deemed sufficient for the air pollution estimation test case.

In order to train MHF-net, we simulate the data from both TROPOMI and PRISMA images to ensure that HR references present either the target spectral features (for TROPOMI) or spatial features (for PRISMA). The data simulation pipeline (see Fig. 8) was based on the official MHF-net code provided by the authors. From input HSIs, we extract 3-D patches of 96×96 pixels and 63 adjacent spectral channels. Spectral resolution of the patches results from the number of PRISMA VNIR channels, which is 66. However, the last three PRISMA bands are corrupted (containing no data), leaving 63 usable channels. To match the number of PRISMA spectral channels, a random subset of 63 adjacent channels out of the 497 TROPOMI channels are selected during each training step. These 63 spectral

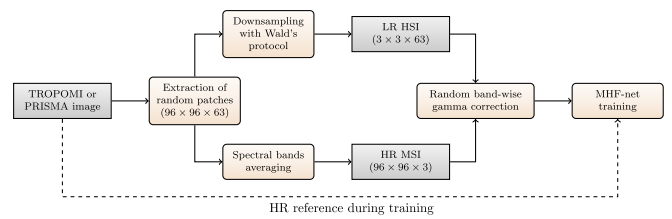


Fig. 8. MHF-net training data simulation pipeline that exploits original PRISMA and TROPOMI images.

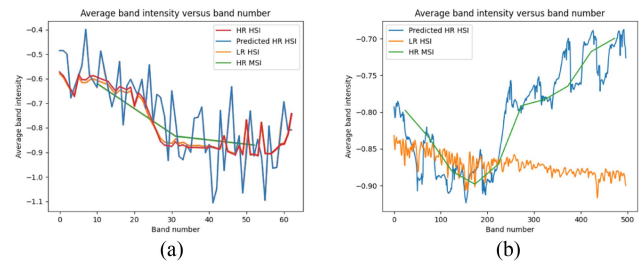


Fig. 9. Spectral curves for a patch in (a) training set and for a (b) real-world PRISMA and TROPOMI sample obtained using the original MHF-net.

bands are then averaged to yield three simulated spectral MSI bands and the spatial resolution of 96×96 pixels remains unchanged. To simulate LR HSI, we employ Wald’s protocol to decrease spatial resolution by a factor of $32\times$ to a size of 3×3 pixels. In addition to standard MHF-net data simulation procedure, we found it important to apply randomized gamma correction to each simulated spectral band of HR MSI, which is justified later in this section. Finally, MHF-net is trained using an augmented dataset with normalized pixel values. Compared with the original MHF-net, we have added a SAM-based consistency component to the loss function (the remaining settings were left unchanged) that penalizes for a spectral angle difference between the super-resolved HSI and the input LR HSI. This component is multiplied by a factor of 0.01 to ensure it is within the same order of magnitude as the original loss.

To justify the use of gamma-based correction and the SAM-based loss, as well as to better understand their impact on the fusion outcome, we demonstrate an illustrative example for a single pair of real-world TROPOMI–PRISMA images. Since there is no ground-truth available in such cases, we downsampled the fusion outcome to the resolution of the input TROPOMI HSI and we verified spatial and spectral consistencies. In Fig. 9, we illustrate the behavior of original MHT-net for a pair of simulated images from the training set (a) and for the aforementioned real-world TROPOMI–PRISMA pair (b). In Fig. 9(a), all images—the target HR HSI (red), the simulated LR HSI obtained with Wald’s protocol (orange), the simulated HR MSI (green), and the fusion outcome (blue)—display a similar global shape of the spectral curve. However, for the real-world case (b), the fused image (blue) clearly follows the HR MSI spectrum (green) rather than the LR HSI spectrum (orange), which would be actually expected. The process for simulating the training data, presented in Fig. 8, results in obtaining LR HSI and HR MSI of similar spectral curves, so during training, the network does not see HSIs and MSIs with different spectral curves, which

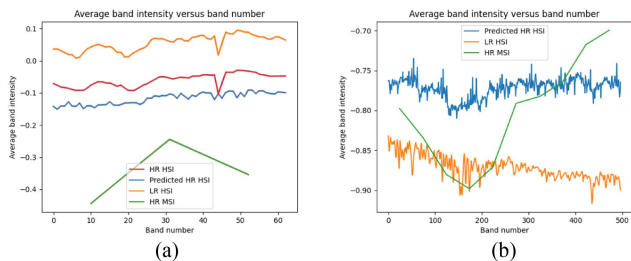


Fig. 10. Spectral curves for a patch in a (a) training set and for a (b) real-world PRISMA and TROPOMI sample obtained using the original MHF-net with random gamma correction applied.

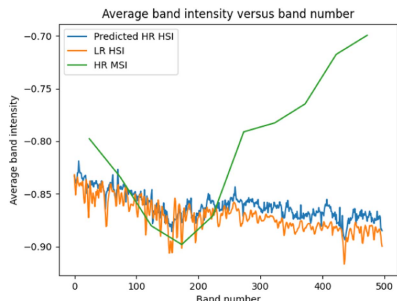


Fig. 11. Spectral curves for a patch in a (a) training set and for a (b) real-world PRISMA and TROPOMI sample obtained using the original MHF-net with random gamma correction coupled with SAM-based loss function.

TABLE I
QUANTITATIVE SCORES OBTAINED FOR THE REAL-WORLD
TROPOMI–PRISMA SAMPLE RELYING ON DIFFERENT TRAINING SETUPS

Training setup	SAM \downarrow	PSNR \uparrow
Baseline MHF-net	8.655	25.57
Gamma correction	3.302	29.49
Gamma correction with SAM-based loss	2.134	31.41

For each metric, we show whether higher (\uparrow) or lower (\downarrow) scores indicate better reconstruction.

may happen in real-world cases (especially if HSI and MSI are captured using different sensors). As a result, the network apparently learned to follow the LR HSI spectral curve. To address this problem, we applied gamma correction independently to each simulated HR MSI spectral band, with a gamma value randomly selected from (0.5; 1.5). The influence of the random gamma correction is shown in Fig. 10. In a patch derived from the training set (a), the simulated HR MSI (green) has, therefore, a different spectral curve than the remaining images that follow a similar pattern. In the real-world sample (b), the super-resolved image (blue) has a spectral signature closer to that of the input LR HSI (orange), but there is a constant shift between them. This issue is mitigated with the SAM-based loss component, which minimizes the spectral angle between the outcome and LR HSI—this is shown in Fig. 11. The quantitative results for the real-world image pair are reported in Table I, and they confirm the observations made in Figs. 9–11.

V. EXPERIMENTAL VALIDATION

In this section, we outline the experimental setup (see Section V-A) and we report the results of application-independent

(see Section V-B) and task-based evaluation (see Section V-C). The results are discussed later in Section VI.

A. Experimental Setup

As the ground-truth data for each of the test cases (outlined earlier in Section III) were collected for different locations, also the input HSIs are specific to each test case. Basically, the precision farming and water quality scenarios operate from PRISMA images, while for the air pollution scenario, we exploit TROPOMI images (being the source for the spectral information) coupled with the PRISMA images (source of the spatial information).

1) *Single-HSI SR of PRISMA Data*: The techniques for single-HSI SR were trained using ten PRISMA L2D VNIR images, covering an area of approximately 9000 km², with LR counterparts obtained with Wald’s protocol. The original images (later treated as HR reference) were downsampled by a factor of 2 \times and blurred using a Gaussian kernel to obtain the simulated LR HSIs. Each LR image was split into overlapping 64 \times 64 patches (with a step size of 32), each of which coupled with an HR patch of 128 \times 128 pixels. We relied on the training pipelines provided by the authors of the investigated techniques and we maintained all settings without modification. The pixel values were normalized before being fed into the network, and standard augmentation techniques, such as random rotation and horizontal and vertical flips, were applied. For model validation, seven additional PRISMA images were used to identify the optimal model configuration. The test set was created based on six PRISMA L2D images (5400 km²) that were exploited in the precision farming scenario, nonoverlapping with the training set.

2) *Pansharpening of PRISMA Data*: In order to train and evaluate the HyperTransformer network [15], we prepared a set of simulated images obtained from 70 PRISMA L2D VNIR scenes. As presented in [69], it is crucial to use level L2D, at which HSIs and PAN images are orthorectified and fully coregistered. The original images were treated as HR references, and LR images were obtained with Wald’s protocol—both HSIs and PAN images were downsampled by a factor of 6 \times . The simulated LR HSIs were split into patches of 50 \times 50 pixels, each coupled with a 6 \times larger PAN patch and a reference HR hyperspectral patch with 300 \times 300 pixels. For training, we used 70% of patches and the remaining ones were used as a validation set. The test set (for HyperTransformer and for PCA-based pansharpening) was created from the same data as for single-HSI SR—from nine PRISMA L2D images exploited also in the precision farming scenario. There was no overlap between the images in the training and test sets.

3) *Data Fusion of TROPOMI–PRISMA Imagery*: For the TROPOMI–PRISMA fusion, MHF-net was trained using a dataset generated from 51 PRISMA L1 images, covering a total area of approximately 45 900 km², and seven TROPOMI L1B UVIS1/BD4 images, each spanning millions of square kilometers. This extensive coverage is due to the large pixel size of approximately 5.5 \times 3.5 km². Each training batch was composed of image patches, 50% of which were retrieved from

PRISMA images and 50% from TROPOMI images. The validation set was obtained from a single TROPOMI image and five PRISMA images. MHF-net was tested on the Pandonia network sites dataset of 69 paired PRISMA L1 and TROPOMI L1B UVIS1/BD4 images collected for the air pollution scenario. For the simulated data setup, the model was evaluated independently on 69 simulated PRISMA and 69 simulated TROPOMI images, and the reported values are the averages of all of the obtained values. For the real-world data setup, the reported values were obtained for TROPOMI–PRISMA pairs.

4) *Task-Based Assessment*: For each of the three scenarios outlined in Section III, we adopted a similar evaluation strategy. Based on the original data coupled with the ground-truth measurements, we trained MLPs to estimate the values based on the spectral signature of a single pixel. The trained model is then applied to process the test data (unseen during training), and the retrieved values are compared with the ground truth to determine the R^2 (for the air pollution case, we report the Pearson coefficient instead), root mean squared error (RMSE), and mean absolute error (MAE). For the test data, we use both the original data, as well as the data from super-resolved images. This provides a quantitative insight into the capabilities of the SR algorithms and helps ensure that the SR process does not affect the spectral features. The latter is also verified by comparing the values retrieved from the input images with those extracted from the super-resolved ones. Furthermore, for each test case, we rendered the estimated values in a form of the parameter maps to inspect the spatial distribution. The MLPs were trained in up to 1000 epochs with an early stopping condition to prevent overfitting and improve the models' generalization capabilities.

For precision farming test case, the augmented dataset was divided into training (52%), validation (28%), and test (20%) parts, where the original 26 samples were distributed randomly (the samples derived from the original ones during augmentation always remain in the same part to avoid data leakage). For both inland and coastline water quality assessment, the initial input dataset was divided with 60% of the available dataset dedicated to the training phase, while the remaining 40% was further divided into testing (22% of the total dataset) and validation (remaining 18% of the dataset). In the air pollution test case, the dataset was split into training (71%) and test (29%) parts.

B. Results of Application-Independent Evaluation

Table II presents quantitative results of task-independent evaluation for simulated and real-world data. For simulated data, we employed reference-based evaluation that consists in measuring the similarity between the reconstructed outcome and the reference image relying on PSNR, SSIM, and SAM metrics. For the original real-world images, we verified spatial and spectral consistency—the super-resolved images are downsampled back to the original size to measure their similarity with the input LR images. For simulated data, the results can be compared between techniques at the same upsampling ratio. For real-world data, all techniques at $2\times$ and $6\times$ ratio were tested using the same input data, so their consistency scores can be compared with each other. MHF-net was tested using a different dataset and the

TABLE II
APPLICATION-INDEPENDENT VALIDATION OF INVESTIGATED TECHNIQUES FOR SIMULATED AND REAL-WORLD DATA

Model	Ratio	PSNR \uparrow	SSIM \uparrow	SAM \downarrow
Simulated data				
Bicubic interpolation	$2\times$	45.35	0.993	0.344
HAT	$2\times$	41.71	0.981	1.447
Bi-3DQRNN	$2\times$	48.85	0.995	0.333
Bicubic interpolation	$6\times$	32.23	0.867	3.667
HyperTransformer	$6\times$	33.80	0.923	0.105
PCA-based pansharpening	$6\times$	24.53	0.716	5.723
MHF-net	$32\times$	34.40	0.967	3.048
Real-world data				
Bicubic interpolation	$2\times$	54.72	0.999	0.126
HAT	$2\times$	41.91	0.985	1.385
Bi-3DQRNN	$2\times$	48.52	0.996	0.294
Bicubic interpolation	$6\times$	49.44	0.996	0.573
HyperTransformer	$6\times$	36.44	0.972	0.104
PCA-based pansharpening	$6\times$	24.88	0.746	4.975
MHF-net	$32\times$	32.08	—	8.847

For each metric, we show whether higher (\uparrow) or lower (\downarrow) scores indicate better reconstruction.

SSIM metric cannot be reported for real-world data here, as it is computed in a kernel whose size is larger than the input HSI patch.

Figs. 12–14 show examples of pansharpening and single-HSI SR applied to the simulated (see Figs. 12 and 13) and real-world (see Fig. 14) images. In all these cases, HSIs are presented in a form of RGB composites. Due to different upsampling factors used during simulation, we demonstrate the behavior of single-HSI SR and pansharpening for different examples, but all the methods are tested with the same real-world input images, demonstrated in Fig. 14. For each outcome, we also present its downsampled version that was used for quantitative evaluation reported earlier in Table II. Fig. 15 shows an example of fusing MSI with HSI for simulated and real-world data. In the former case, we show the HR reference used for quantitative assessment, while for the latter we show downsampled MHF-net outcome that is compared with the input LR HSI.

C. Results of Task-Based Evaluation

Quantitative results for the precision farming and water quality scenarios are reported in Table III, while those for the air pollution scenario are shown in Table IV. For precision farming, the model was evaluated relying on the original images (at 30-m GSD), as well as enhanced with single-HSI SR and pansharpening techniques. Single-HSI SR increases the resolution $2\times$, thus reducing GSD down to 15 m and pansharpening renders an enhancement of $6\times$ to 5-m GSD. To compare pansharpening with single-HSI SR, we also downsample the result of the former to 15 m GSD. For water quality, we report the scores only for 15 m GSD, as the degree of uncertainty, resulting from comparison with ground-truth data, with spatial resolution of 300-m, was considered excessive. Moreover, particularly for coastal application, it was considered that an increase in spatial resolution would not add any specific advantage in the chlorophyll-a estimation, as the variability of the parameter is not significant enough in its values to be able to spot such remarkable differences in open aquatic environments. As the

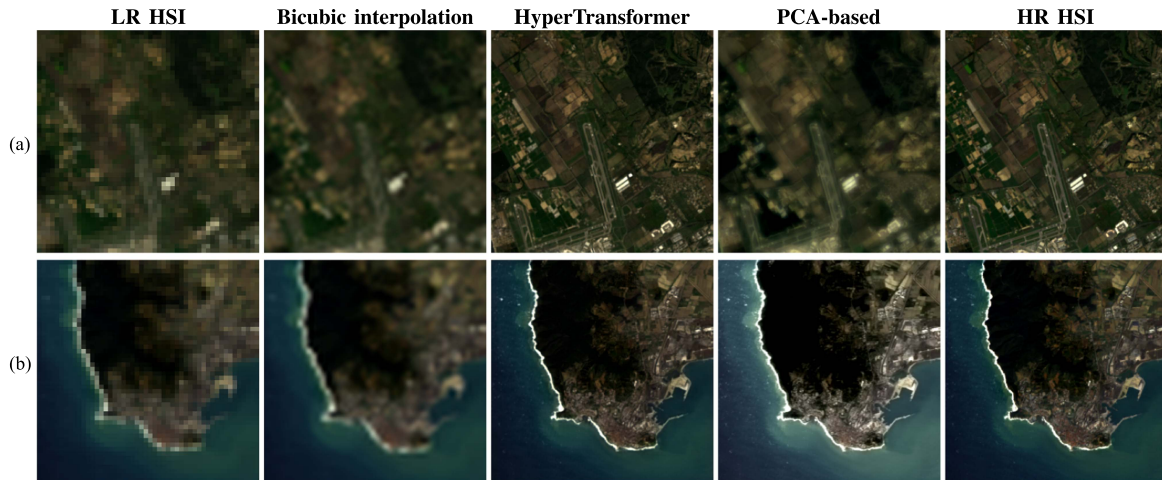


Fig. 12. Examples of pansharpening applied to simulated PRISMA images for (a) suburban and (b) coastal areas in a form of an RGB composite. LR HSIs were simulated from original HSIs (treated as HR reference) with downsampling factor of $6\times$. For comparison, the outcome of bicubic interpolation is also presented.

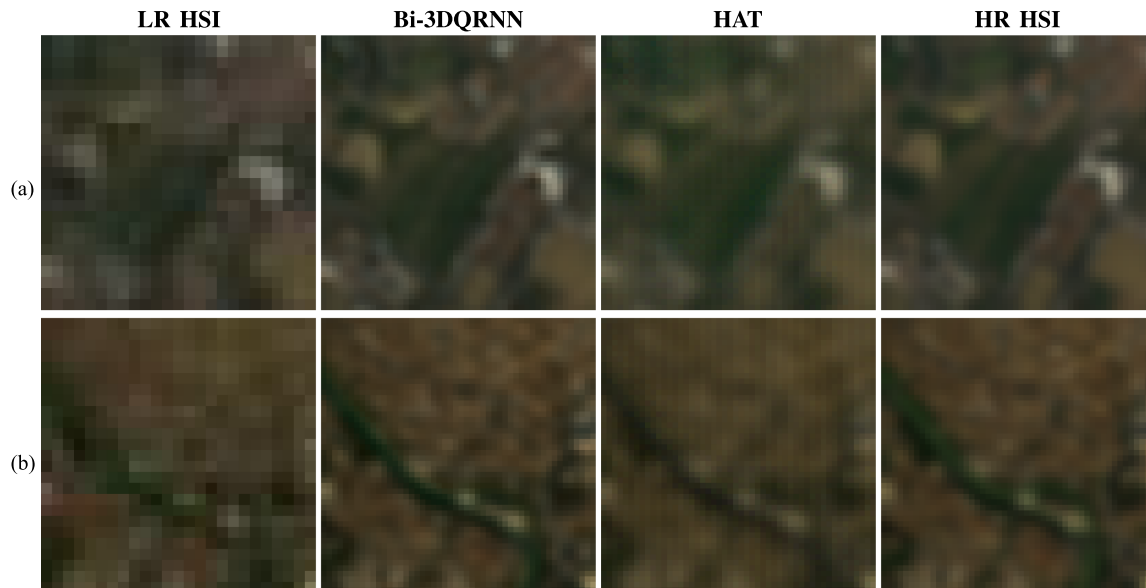


Fig. 13. Examples of single-HSI SR applied to simulated PRISMA data for suburban (a) and urban (b) areas in a form of RGB composite. LR HSIs were simulated from original HSIs (treated as HR reference) with downsampling factor of $2\times$.

results obtained relying on the original data were rather poor (R^2 of around 0.5), in this case, we decided to train the models for each data type. The scatter plots showing correlation between the values retrieved from original and enhanced images are presented in Figs. 16–18, for precision farming, water quality, and air pollution test cases, respectively.

Qualitative results are presented in Figs. 19–21, for precision farming, water quality, and air pollution sites, respectively. In Fig. 19, we present nitrogen concentration maps for the Grosseto site, retrieved from the original and pansharpened images. We show the concentration level in the relevant area, overlapped on the RGB composites. Fig. 20 shows the Trasimeno lake (43.141467N, 12.096280E) with chlorophyll-a concentration maps retrieved from the original and super-resolved images. Finally, in Fig. 21, we present the NO_2 column over the city of Brussels retrieved from the TROPOMI image and from the fusion outcome.

VI. DISCUSSION

In this section, we analyze the results reported earlier in the preceding Section V. At first, we discuss the results of application-independent validation (see Section VI-A), followed by task-based assessment (see Section VI-B), and finally, we discuss the similarities and differences between these two evaluation strategies (see Section VI-C).

A. Discussion on Application-Independent Evaluation

The quantitative results (see Table II) show that for the simulated data, the spatial features are more accurately reconstructed at a smaller upsampling ratio of $2\times$ (indicated by PSNR and SSIM scores), but the spectral features are best preserved relying on pansharpening realized with HyperTransformer (at $6\times$

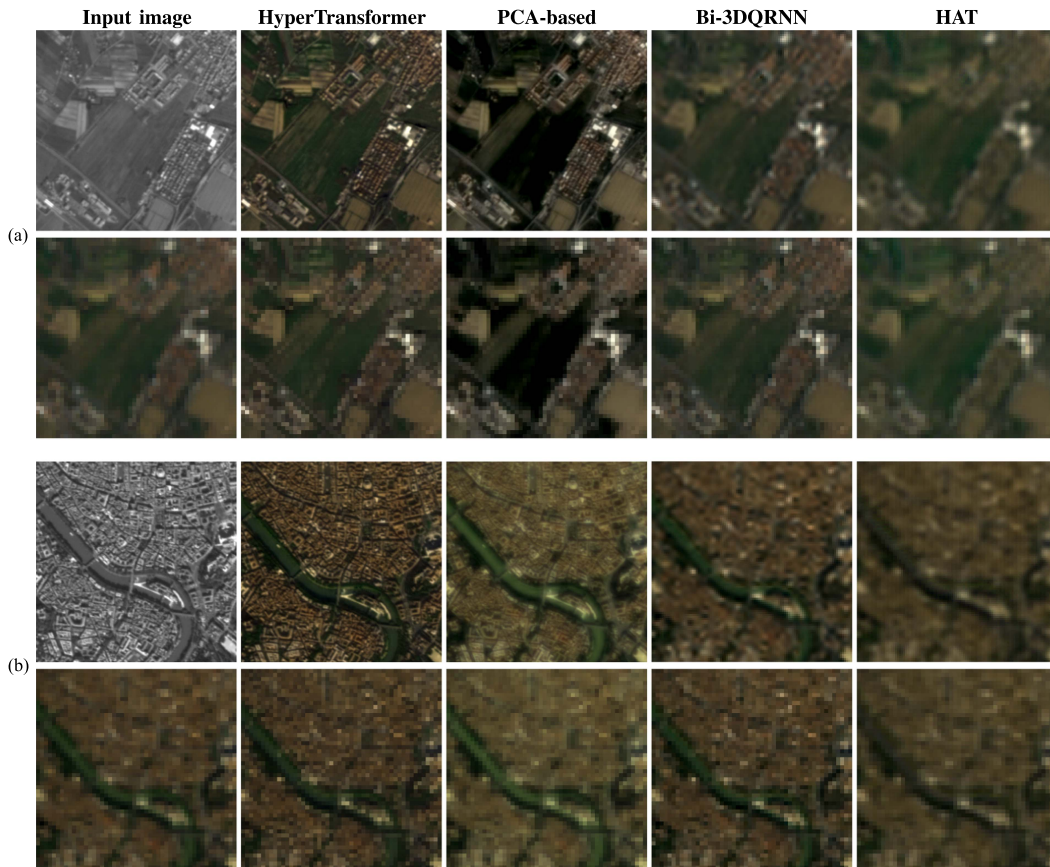


Fig. 14. Examples of enhancing two real-world PRISMA (a) suburban and (b) urban scenes with pansharpening (HyperTransformer and PCA-based methods, 5-m GSD) and single-HSI SR (Bi-3DQRNN and HAT methods, 15-m GSD). For each scene, we show the PAN input image at 5-m GSD (top row) and the input HSI at 30-m GSD (bottom row), followed with the results obtained using different techniques (top row) and the outcomes downsampled to 30-m GSD. HSIs are presented in a form of an RGB composite.

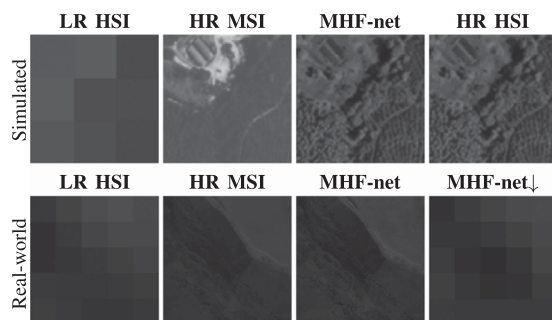


Fig. 15. Example of HSI-MSI fusion performed with MHF-net for simulated and real-world scene. For the simulated pair, HR HSI is used as a reference; for the real-world pair, the outcome is downsampled (MHF-net↓) to the size of input HSI.

upsampling), followed by $2\times$ bicubic interpolation and single-HSI SR with Bi-3DQRNN that exploits spatial-spectral features. It is worth noting that HAT which operates in a band-wise manner achieved worse PSNR and SSIM scores than the bicubic interpolation, probably because the same model was applied to reconstruct each spectral channel. For real-world data, it is not surprising that bicubic interpolation offers best spatial consistency at $2\times$ and $6\times$ enhancement, but again the spectral features are best preserved relying on HyperTransformer.

TABLE III
QUALITY METRICS FOR PRECISION FARMING AND WATER QUALITY SCENARIOS OBTAINED FROM THE ORIGINAL AND SUPER-RESOLVED IMAGES

Data	GSD	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow
Precision farming				
Original	30 m	0.97	0.044	0.046
Bicubic interpolation	15 m	0.83	0.14	0.096
Single-HSI SR (Bi-3DQRNN)	15 m	0.83	0.13	0.11
Single-HSI SR (HAT)	15 m	0.86	0.13	0.099
Pansharpened (HyperTransformer)	15 m	0.88	0.12	0.055
Pansharpened (PCA)	15 m	0.81	0.15	0.11
Bicubic interpolation	5 m	0.92	0.10	0.090
Pansharpened (HyperTransformer)	5 m	0.96	0.062	0.032
Pansharpened (PCA)	5 m	0.90	0.10	0.086
Inland water quality				
Original	30 m	0.483	2.744	1.814
Bicubic interpolation	15 m	0.777	1.804	1.168
Single-HSI SR (Bi-3DQRNN)	15 m	0.419	2.915	2.141
Single-HSI SR (HAT)	15 m	0.694	2.11	1.479
Pansharpened (HyperTransformer)	15 m	0.883	1.306	0.865
Pansharpened (PCA)	15 m	0.987	0.429	0.333
Coastal water quality				
Original	30 m	0.526	1.019	0.563
Bicubic interpolation	15 m	0.503	1.087	0.529
Single-HSI SR (Bi-3DQRNN)	15 m	0.423	1.171	0.62
Single-HSI SR (HAT)	15 m	0.688	0.862	0.467
Pansharpened (HyperTransformer)	15 m	0.897	0.494	0.317
Pansharpened (PCA)	15 m	0.95	0.343	0.245

For each metric, we show whether higher (\uparrow) or lower (\downarrow) scores indicate better reconstruction.

PCA-based pansharpening obtains very poor scores for both simulated and real-world data—this may result from the fact that PRISMA PAN images cover only a small part of HSIs' spectral

TABLE IV
QUALITY METRICS FOR THE AIR POLLUTION SCENARIO OBTAINED FROM THE ORIGINAL AND ENHANCED IMAGES

Data	Size	Pearson coeff. \uparrow	RMSE \downarrow	MAE \downarrow
Original	5.5 km \times 3.5 km	0.87	20.06	17.09
Enhanced	0.5 km \times 0.5 km	0.89	18.54	15.59

For each metric, we show whether higher (\uparrow) or lower (\downarrow) scores indicate better reconstruction. RMSE and MAE are reported in $\mu\text{Mol}/\text{m}^2$.

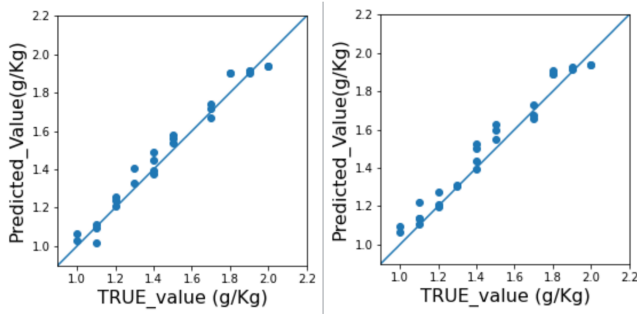


Fig. 16. Scatter plots of predicted versus actual soil nitrogen content retrieved from (left) original and (right) pansharpended PRISMA images.

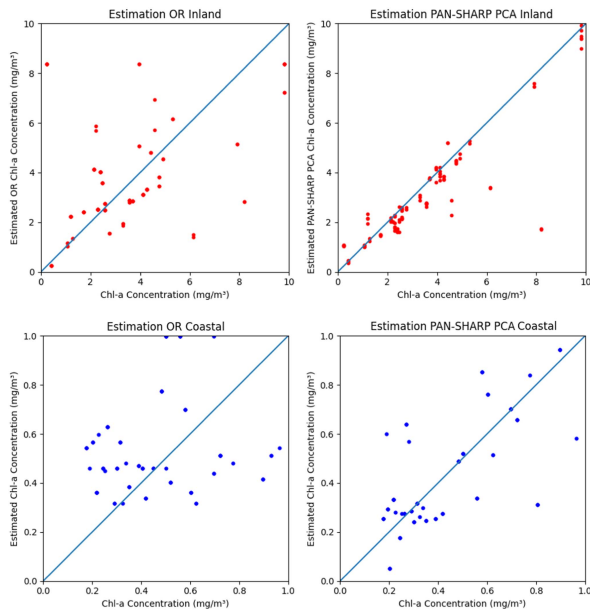


Fig. 17. Scatter plots of predicted versus actual chlorophyll-a concentration retrieved from original (left) and super-resolved (right) PRISMA images for (top) inland and (bottom) coastal waters.

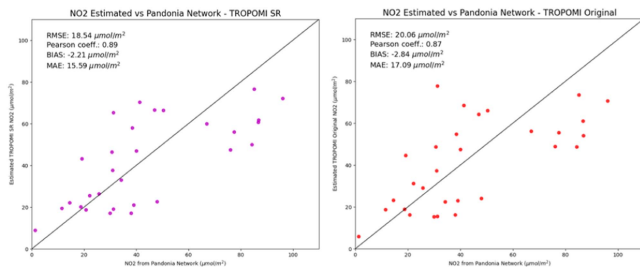


Fig. 18. Scatter plot of predicted NO_2 values retrieved from the (left) super-resolved and (right) original TROPOMI images for the air pollution scenario versus ground truth from the Pandonia Network.



Fig. 19. Example of nitrogen content retrieved from (left) original and (right) pansharpended PRISMA images (Grosseto site) for the precision farming scenario presented over a true color composition of original and super-resolved PRISMA image.

range, so the assumption that the first principal component is correlated with the PAN image may not hold in this case. From the qualitative results (see Fig. 12), it may be seen that PCA-based pansharpending substantially improves spatial features compared with the input HSI, as well as with the bicubically upsampled image. However, the outcome is definitely worse than that of HyperTransformer. Also, the contrast is increased which may account for lower scores than for bicubic interpolation (see Table II). The results for single-HSI SR (see Fig. 13) are coherent with the scores—Bi-3DQRNN renders visually better result than HAT. The latter produces some grid-like artifacts, which is reflected in the scores.

From Fig. 14, it can be seen that both single-HSI SR and pansharpended outcomes downsampled to 30-m GSD are visually quite close to the original input HSI, but naturally the pansharpended images are of much higher visual quality, as they exploit additional source of HR information. Overall, the PSNR and SSIM scores for the test set are high for all the investigated techniques, indicating that they were properly trained before applying them to specific test cases.

From the quantitative scores (see Table II) and qualitative examples (see Fig. 15), it is apparent that MHF-net exploits HR information from an MSI and the super-resolved image is consistent with the input HSI. Clearly, the downsampled SR outcome (MHF-net \downarrow in Fig. 15) used for metric calculations contains significantly fewer pixels in real-world data evaluations than in the simulated datasets, despite representing the same geographic area at different spatial resolutions (as shown in Fig. 15). This difference in pixel density and scale could account for the discrepancy observed in the SAM metric performance (see Table II).

B. Discussion on Task-Based Evaluation

For the precision farming scenario, nitrogen content has been estimated at high accuracy both from the original data and from the pansharpended ones at 5-m GSD (Table III and Fig. 16). The best R^2 and RMSE were obtained for the original data, while the best MAE was retrieved from the HyperTransformer outcome. At the resolution of 15-m GSD, HyperTransformer was the best, followed by single-HSI methods and PCA-based pansharpending. It is worth noting that in this case the PCA-based pansharpending does not offer any gain over bicubic interpolation both at 5- and 15-m GSD. Even though pansharpending retrieves similar quantitative results, the qualitative assessment (see Fig. 19)

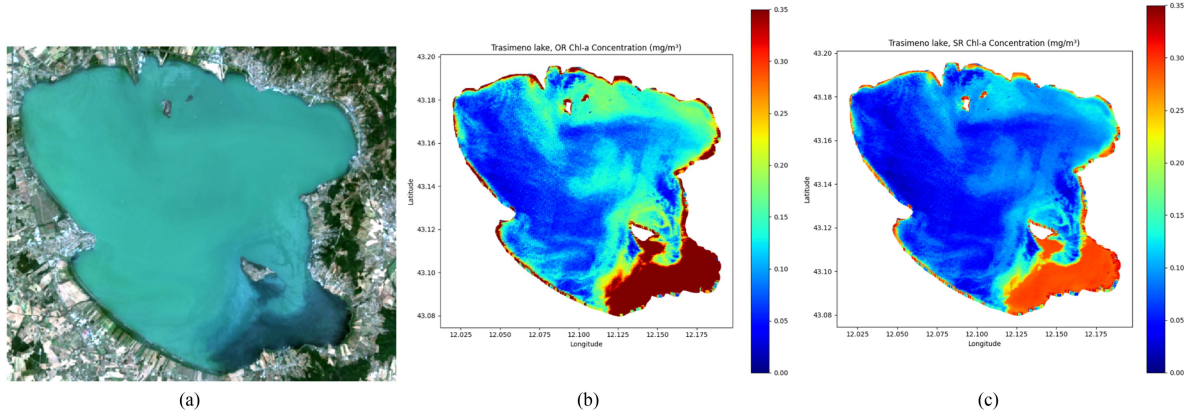


Fig. 20. Example of chlorophyll-a concentration retrieved for the Trasimeno lake for the water quality scenario: (a) RGB PRISMA image, (b) concentration map retrieved from original PRISMA data, and from the (c) outcome of Bi-3DQRNN single-HSI SR.

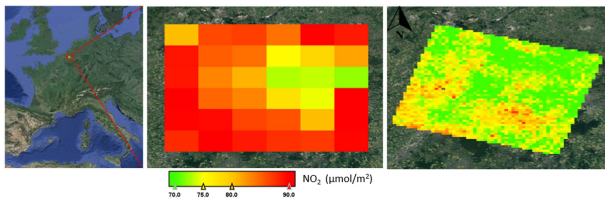


Fig. 21. Example of NO_2 column concentration retrieved from (center) original and (right) super-resolved TROPOMI images for the air pollution scenario over the city of Brussels (left).

indicates that the nitrogen concentration maps present more valuable details compared with those obtained from the original data. This result is of particular importance in the context of the precision agriculture case study, especially considering the need for greater precision in small agricultural fields, such as those typical of the Italian countryside.

From Table III, it is clear that estimation of chlorophyll-a is improved relying on pansharpening for both inland and coastal waters. The largest gain has been achieved with the PCA-based pansharpening that can also be appreciated from the scatter plots in Fig. 17. However, it is surprising that the results obtained from bicubically upsampled images are quite high for inland waters (better than from images enhanced using single-HSI SR). Also, PCA-based technique appears to be better than HyperTransformer, and Bi-3DQRNN is outperformed in both cases with HAT that was shown to be less effective in preserving spectral features (see Table II). There may be several reasons for that. First of all, the ground-truth reference values may suffer from noisy measurements, as they were retrieved from Sentinel-3 OLCI product with 300 m of spatial resolution. Furthermore, the variability of the chlorophyll-a parameter could be greater in open water environments, due to the increased movement of water masses while maintaining in salt waters concentrations values very low and this could considerably affect the concentration identification moving from a pixel to its neighborhood. From the visualized concentration maps (see Fig. 20), it can be observed that the distribution of chlorophyll-a concentration appears higher along the coastlines and in the southern part of the lake (both for the original and

super-resolved images). The concentration variability for the super-resolved case is smaller, showing concentration values in a more uniform range than the retrieved values for the original case, which instead shows less compact values with a greater differences in terms of concentrations. However, it is clear that the results obtained relying on the super-resolved data are well-correlated with those obtained from the original data, confirming the observations made in the precision farming scenario.

The quantitative results for the NO_2 column estimation (see Table IV) indicate that enhancing spatial resolution helps achieve more accurate estimations compared to ground-based measurements, which have a higher reliability in the vicinity of the ground instrument location. Also, the scatter plots (see Fig. 18) show significant correlation between the estimated values and measurements considered as ground truth, confirming that the MLP represents a suitable approach to estimate trace gases from TROPOMI L1B data, as reported in [10]. A visual example (see Fig. 21) confirms that the values are estimated with much greater precision in the spatial domain.

C. Comparison of the Evaluation Strategies

In general, the increase of spatial resolution is reflected in improved scores for both application-independent and task-based evaluation. For PRISMA images, it is HyperTransformer [1] that offers the best qualitative and quantitative results (especially considering spectral features captured with the SAM metric), and it also renders consistently good results in task-based evaluation aimed at specific remote sensing applications. Surprisingly, PCA-based pansharpening that achieved rather low scores in application-independent assessment, occurred to be the best for inland and coastal water quality scenarios. This may be attributed to the fact that the MLP trained for estimating the chlorophyll-a level exploits a small subset of PRISMA bands that fall within the visible spectrum covered by the PAN image. As PCA-based pansharpening is underpinned with the assumption that the first principal component is highly correlated with the PAN image, the reconstruction may be expected to work better in that part of the spectrum. This also shows that it may be challenging

to identify the most suitable technique relying exclusively on application-independent validation.

Application-independent validation is challenging for real-world (not degraded) HSIs, as usually there is no ground-truth available in such cases. However, comparing the quantitative scores against the performance in specific use cases, it appears that the ability of reconstructing spectral features and preserving their consistency (for real-world data), here captured with the SAM metric, is a more important indicator than the metrics capturing spatial features (PSNR and SSIM). It is important to note that single-HSI SR is not better in the specific test cases investigated here than bicubic interpolation, and this is also reflected in the SAM scores (see Table II). Finally, SAM was relatively low for MHF-net, which was also reflected in the improved scores for the air pollution scenario.

This study also underlines the importance of qualitative assessment. For the precision farming test case, pansharpening renders comparable scores to those retrieved from original images. However, the nitrogen content maps allow for appreciating that they are much more detailed when retrieved based on pansharpened data. Also, qualitative inspection in application-independent evaluation helps understand the difference between the results obtained with single-HSI SR and pansharpening, compared with bicubic interpolation.

VII. CONCLUSION

In this article, we proposed a new approach toward evaluating different types of super-resolved HSI products in specific applications. Two remote sensing test cases concerned with precision farming and water quality were employed to evaluate four pansharpening [1], [41] and single-HSI SR techniques [6], [15] for enhancing PRISMA images. As discussed in this article, the results clearly indicate that the pansharpened images at 5-m GSD lead to quantitatively better results than single-HSI SR performed without any auxiliary HR information. Importantly, this is consistent with the SAM scores obtained within the application-independent evaluation performed for the simulated data, as well as with the subjective qualitative assessment of both the reconstructed images and nitrogen content estimation maps. The air pollution test case has also confirmed that the HSI-MSI data fusion [59], applied to combine selected PRISMA spectral bands with a TROPOMI HSI of low spatial resolution, allows for obtaining highly valuable super-resolved products that improve the analysis outcome (both qualitatively and quantitatively).

It is worth noting that the elaborated test cases primarily exploit the spectral features, so the scores obtained based on the data enhanced using pansharpening and HSI-MSI fusion indicate that SR of HSIs may also increase their value in terms of the spectral information. However, while single-HSI SR improves the images in the spatial domain (as shown qualitatively and quantitatively in Section V-B), they are more likely to distort the spectral features—as a result, the scores for single-HSI SR in the precision farming test case were worse than relying on the original data. This clearly suggests that the future research regarding single-HSI techniques should be focused on

improving their capabilities in terms of preserving the spectral information.

Overall, the reported analysis exposes the importance of applying task-based evaluation, as application-independent assessment may be insufficient to properly assess the value of super-resolved products in specific applications. However, task-based evaluation is challenging due to several reasons. At first, it requires large-scale ground-truth data that may be difficult to collect. Furthermore, it may be time-consuming, especially when the spatially enhanced data are used for training and validating the task-specific models. Finally, the models may be quite selective and focused on specific features, leading to poor generalization. Therefore, in our opinion, more future work should be focused on developing a battery of specific use cases to better understand the correlation between application-independent and task-based assessment. This, in turn, should help in proposing new SR techniques for enhancing the value of hyperspectral data.

ACKNOWLEDGMENT

The authors would like to thank the Italian Space Agency (ASI) for the availability of PRISMA data for this work. The authors would also like to thank the Italian Council for Agricultural Research (CREA) and the Pandonia Global Network for the availability of soil nitrogen and nitrogen dioxide data for the task-based validation of precision farming and air pollution scenarios, respectively. This work is distributed under the LATEX Project Public License (LPPL) (<http://www.latex-project.org/>) version 1.3. A copy of the LPPL, version 1.3, is included in the base LATEX documentation of all distributions of LATEX released 2003/12/01 or later. The opinions expressed here are entirely that of the author. No warranty is expressed or implied. User assumes all risk.

REFERENCES

- [1] W. G. C. Bandara and V. M. Patel, "HyperTransformer: A textural and spectral feature fusion transformer for pansharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1767–1777.
- [2] W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Hyperspectral pansharpening based on improved deep image prior and residual reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5520816.
- [3] P. Benecki, M. Kawulok, D. Kostrzewa, and L. Skonieczny, "Evaluating super-resolution reconstruction of satellite images," *Acta Astronautica*, vol. 153, pp. 15–25, 2018.
- [4] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
- [5] H. Chen et al., "Real-world single image super-resolution: A brief review," *Inf. Fusion*, vol. 79, pp. 124–145, 2022.
- [6] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 22367–22377.
- [7] M. Ciotola et al., "Hyperspectral pansharpening: Critical review, tools and future perspectives," 2024, *arXiv:2407.01355*.
- [8] J. Cornebise, I. Orsolich, and F. Kalaitzis, "Open high-resolution satellite imagery: The WorldStrat dataset with application to super-resolution," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 25979–25991.
- [9] L. De Laurentiis, D. Latini, G. Schiavon, and F. Del Frate, "Integration of IEM_B, ISMN and SAR Sentinel-1 data for accurate soil moisture estimation using neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1272–1275.

- [10] D. De Santis et al., "Volcanic SO₂ near-real time retrieval using TROPOMI data and neural networks: The Dec. 2018 Etna test case," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 8480–8483.
- [11] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6995–7010, Aug. 2021.
- [12] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021.
- [13] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, "Model-guided deep hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 5754–5768, 2021.
- [14] W. Dong, S. Liu, S. Xiao, J. Qu, and Y. Li, "ISPDiff: Interpretable scale-propelled diffusion model for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5519614.
- [15] Y. Fu, Z. Liang, and S. You, "Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2674–2688, 2021.
- [16] J. Gao, J. Li, and M. Jiang, "Hyperspectral and multispectral image fusion by deep neural network in a self-supervised manner," *Remote Sens.*, vol. 13, no. 16, Aug. 2021, Art. no. 3226.
- [17] A. Gitelson, G. Garbuzov, F. Szilagyi, K. Mittenzwey, A. Karnieli, and A. Kaiser, "Quantitative remote sensing methods for real-time monitoring of inland waters quality," *Int. J. Remote Sens.*, vol. 14, no. 7, pp. 1269–1295, 1993.
- [18] G. Guarino, M. Ciotola, G. Vivone, G. Poggi, and G. Scarpa, "PCA-CNN hybrid approach for hyperspectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5511505.
- [19] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.
- [20] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7459–7471, Oct. 2020.
- [21] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1082–1096, 2020.
- [22] C. Jin, L.-J. Deng, T.-Z. Huang, and G. Vivone, "Laplacian pyramid networks: A new approach for multispectral pansharpening," *Inf. Fusion*, vol. 78, pp. 158–170, 2022.
- [23] A. Juna et al., "Water quality prediction using KNN imputer and multi-layer perceptron," *Water*, vol. 14, no. 17, 2022, Art. no. 2592. [Online]. Available: <https://www.mdpi.com/2073-4441/14/17/2592>
- [24] 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.
- [25] P. Kowaleczko et al., "A real-world benchmark for Sentinel-2 multi-image super-resolution," *Sci. Data*, vol. 10, no. 1, 2023, Art. no. 644.
- [26] J. Li et al., "Hyperspectral image super-resolution by band attention through adversarial learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4304–4318, Jun. 2020.
- [27] K. Li, D. Dai, E. Konukoglu, and L. Van Gool, "Hyperspectral image super-resolution with spectral mixup and heterogeneous datasets," 2021, *arXiv:2101.07589*.
- [28] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1660, doi: [10.3390/rs12101660](https://doi.org/10.3390/rs12101660).
- [29] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7711–7725, Sep. 2021.
- [30] Z. Liu, H. Lei, L. Lei, and H. Sheng, "Spatial prediction of total nitrogen in soil surface layer based on machine learning," *Sustainability*, vol. 14, no. 19, 2022, Art. no. 11998.
- [31] J. Ma, S. Guo, and L. Zhang, "Text prior guided scene text image super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 1341–1353, 2023.
- [32] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Learning a 3D-CNN and transformer prior for hyperspectral image super-resolution," *Inf. Fusion*, vol. 100, 2023, Art. no. 101907.
- [33] M. Märtens, D. Izzo, A. Krzic, and D. Cox, "Super-resolution of PROBA-V images using convolutional neural networks," *Astrodynamics*, vol. 3, no. 4, pp. 387–402, 2019.
- [34] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.
- [35] S. Mei, X. Yuan, J. Ji, S. Wan, J. Hou, and Q. Du, "Hyperspectral image super-resolution via convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 4297–4301.
- [36] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.
- [37] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.
- [38] J. Qu et al., "Progressive multi-iteration registration-fusion co-optimization network for unregistered hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5519814.
- [39] Y. Qu, R. K. Baghbaderani, H. Qi, and C. Kwan, "Unsupervised pansharpening based on self-attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3192–3208, Apr. 2021.
- [40] M. T. Razzak, G. Mateo-García, G. Lecuyer, L. Gómez-Chova, Y. Gal, and F. Kalaitzis, "Multi-spectral multi-image super-resolution of Sentinel-2 with radiometric consistency losses and its effect on building delineation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 195, pp. 1–13, 2023.
- [41] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.
- [42] J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1–10.
- [43] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop*, 2019, pp. 3844–3851.
- [44] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [45] Z. Tang, Q. Xu, P. Wu, Z. Shi, and B. Pan, "Feedback refined local-global network for super-resolution of hyperspectral imagery," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1944.
- [46] T. Tarasiewicz et al., "Multitemporal and multispectral data fusion for super-resolution of Sentinel-2 images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5406519.
- [47] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Eur. Conf. Comput. Vis.*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 87–102.
- [48] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, Curran Assoc., Inc., 2017, vol. 30, pp. 6000–6010. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [49] L. Wald, "Assessing the quality of synthesized images," in *Data Fusion. Definitions and Architectures-Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.
- [50] P. Wang, X. Shen, and G. Zhang, "Spatiotemporal super-resolution mapping by considering the point spread function effect," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6001705.
- [51] P. Wang, L. Wang, H. Leung, and G. Zhang, "Super-resolution mapping based on spatial-spectral correlation for spectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2256–2268, Mar. 2021.
- [52] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11276–11285, Nov. 2021.
- [53] W. Wang et al., "Enhanced deep blind hyperspectral image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 3, pp. 1513–1523, Mar. 2023.
- [54] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503113.
- [55] 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.
- [56] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [57] X. Wu, T.-Z. Huang, L.-J. Deng, and T.-J. Zhang, "Dynamic cross feature fusion for remote sensing pansharpening," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 14667–14676.

- [58] Q. Xie, Z. Minghao, Q. Zhao, Z. Xu, and D. Meng, "MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1457–1473, Mar. 2022.
- [59] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1585–1594.
- [60] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang, "Deep gradient projection networks for pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1366–1375.
- [61] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.
- [62] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. 3rd Annu. JPL Airborne Geosci. Workshop, AVIRIS Workshop*, vol. 1, 1992, pp. 147–149.
- [63] H. Zhang and J. Ma, "GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening," *ISPRS J. Photogrammetry Remote Sens.*, vol. 172, pp. 223–239, 2021.
- [64] J. Zhang et al., "Hyperspectral image super-resolution based on feature diversity extraction," *Remote Sens.*, vol. 16, no. 3, 2024, Art. no. 436. [Online]. Available: <https://www.mdpi.com/2072-4292/16/3/436>
- [65] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, and J. Zhang, "ESSAformer: Efficient transformer for hyperspectral image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 23073–23084.
- [66] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.
- [67] X. Zhang, W. Huang, Q. Wang, and X. Li, "SSR-Net: Spatial-spectral reconstruction network for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5953–5965, Jul. 2021.
- [68] C. Zhou, J. Zhang, J. Liu, C. Zhang, R. Fei, and S. Xu, "PercepPan: Towards unsupervised pan-sharpening based on perceptual loss," *Remote Sens.*, vol. 12, no. 14, Jul. 2020, Art. no. 2318.
- [69] M. Ziája et al., "Hyperspectral image pansharpening: The PRISMA case study," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 1633–1636.



Michał Kawulok (Senior Member, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees in computer science from the Silesian University of Technology, Gliwice, Poland, in 2003, 2007, and 2015, respectively. In 2024, he was promoted to a Full Professor.

He is a Professor with the Silesian University of Technology and he leads the Earth Observation research group at KP Labs, Gliwice. He has been committed to applying algorithmic solutions in a variety of domains, including Earth observation, medical imaging, and machine vision. He has authored and

co-authored more than 150 peer-reviewed papers in journals and conference proceedings. His research interests include image processing, evolutionary computation, pattern recognition, and machine learning, with particular attention given to super-resolution reconstruction, face and gesture recognition, and support vector machines.



Paweł Kowaleczko received the M.Sc. degree in technical physics from the Warsaw University of Technology, Warsaw, Poland, in 2018. He is currently working toward the Ph.D. degree in the field of super-resolution for satellite imagery with the Warsaw University of Technology, Warsaw, Poland.

Since 2021, he has been with KP Labs, Gliwice, Poland, where he has participated in multiple ESA-funded projects. With extensive experience in machine learning, he has previously worked on numerous projects in the fields of computer vision, dig-

ital signal (sound) processing, explainable artificial intelligence, and image generation.



Maciej Ziája is currently working toward the Ph.D. degree in computer science with the Department of Algorithmics and Software, Silesian University of Technology, Gliwice, Poland.

He is currently a Machine Learning Engineer with KP Labs, Gliwice. His research focuses on enhancing image quality using deep learning models, particularly for satellite Earth observation. He has participated in multiple ESA-funded projects, publishing his results at scientific conferences and in journals from IEEE, Elsevier, and Springer Nature. He has

also been responsible for machine learning research, software development, and deployment of deep learning models for the Intuition-1 satellite, which is currently operating in orbit.



Jakub Nalepa (Senior Member, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees in computer science from the Silesian University of Technology, Gliwice, Poland, in 2011, 2016, and 2021, respectively.

He is currently an Associate Professor with the Silesian University of Technology. He is also the Head of artificial intelligence (AI) with KP Labs, Gliwice, where he shapes the scientific and industrial AI objectives of the company. He has been pivotal in designing the onboard deep learning capabilities of Intuition-1 and has contributed to other missions,

including CHIME, Φ -Sat-2, and OPS-SAT. His research interests include (deep) machine learning, hyperspectral data analysis, signal processing, remote sensing, and tackling practical challenges that arise in Earth observation to deploy scalable solutions.

Dr. Nalepa was the General Chair of the HYPERVIEW Challenge at the 2022 IEEE International Conference on Image Processing, focusing on the estimation of soil parameters from hyperspectral images onboard Intuition-1 to maintain farm sustainability by improving agricultural practices.



Daniel Kostrzewa (Member, IEEE) received the Ph.D. degree in computer science from the Silesian University of Technology, Gliwice, Poland, in 2015.

He is an Assistant Professor with the Silesian University of Technology, and has also been working with KP Labs, Gliwice, since 2018, where he is the Head of Bid Management Department and a Researcher. He has authored and coauthored more than 80 papers in peer-reviewed journals and conference proceedings. His research interests mainly include deep learning,

machine learning, and evolutionary algorithms applied to image processing, music information retrieval, and discrete optimization problems.

Dr. Kostrzewa was also a Research Engineer in several ESA-funded and academia projects.



Daniele Latini received the M.Sc. degree in telecommunication engineering from the University of Rome "Tor Vergata," Rome, Italy, in 2010, the Ph.D. degree in geoinformation in 2015, and the post-laurea master degree in "design, application and regulation of unmanned aerial vehicle system (PARS)" from the University of Rome "Tor Vergata," in 2020.

In 2013, he was a Visiting Student Researcher with Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, USA, working on interferometric analysis using NASA UAVSAR data.

In 2016, he obtained the license for UAV piloting in critical scenarios (A2 pilot licence). Since 2010, he has been working in the field of remote sensing with the Earth Observation Laboratory, Tor Vergata University, Rome, and then, with GEO-K srl. As a Project Manager, he has been responsible for several contracts funded by ESA, ASI, European Commission (HORIZON), and Italian Ministries, in matter of EO products and services development in collaboration with various company and research institutes. He was an external Professor with Tor Vergata University for the M.S. degree integrative course in "microwaves" and for postgraduate master program in "geoinformation and geographic information systems." His research interests and publications focus on the analysis of optical and synthetic aperture radar (SAR) remote sensing products, deeply investigating the ensemble of neural networks techniques and architectures for a broad field of application, as well as on integration of different technology solution such as artificial intelligence and on-board data processing.



Davide De Santis (Member, IEEE) received the M.Sc. degree in environmental engineering (cum laude) and the Ph.D. degree in computer science, control and geoinformation from the Tor Vergata University of Rome (TVU), Rome, Italy, in 2017 and 2022, respectively.

From 2020, he is a contract Lecturer with “Image Processing Laboratory” as part of the “Remote Sensing and Cartography” course with TVU. He is currently a Research Fellow with the Civil Engineering and Computer Science Engineering Department, TVU. He is, or has been, involved as a Scientific Advisor in several European Space Agency (ESA) and Italian Space Agency (ASI) funded research projects. His current research activities are focused on air quality monitoring exploiting Earth Observation data, radiative transfer models, and artificial intelligence algorithms. His research interests mainly include optical remote sensing, machine learning, and climate change.

Dr. De Santis is a Reviewer for several scientific journals, including *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *Scientific Reports*.



Giorgia Salvucci received the master’s degree in astronomy and astrophysics from the University of Rome “La Sapienza,” Rome, Italy, in 2011 and the Ph.D. degree in computer science, control, and geoinformation from the University of Rome “Tor Vergata,” Rome, in 2023.

During Ph.D., her research focused on applying AI to satellite data for precision farming. She is an experienced Researcher specializing in remote sensing, satellite data analysis, and Earth observation. Throughout her career, she has worked on several

high-profile projects, including air quality monitoring using the PRISMA satellite and resilience analysis of transport networks. With extensive experience in satellite data processing, neural network algorithms, and Earth observation technologies, she has contributed to both academic and industry-focused initiatives in environmental monitoring, space economy, and climate change. She has worked for organizations such as GEO-K s.r.l., Thales Alenia Space, and the University of Salento, and has a strong background in physics and astronomy.



Ilaria Petracca (Student Member, IEEE) received the Laurea degree in Environmental Engineering in 2020, and the Ph.D. degree in Computer Science, Control and Geoinformation in 2024 from the Tor Vergata University of Rome, Rome, Italy, where she is now a research fellow.

From October 2022 to July 2023, she was a Visiting Student with the Space Telecommunications, Astronomy and Radiation Laboratory (STAR Lab) in the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. Her major research area includes volcanic ash detection and precipitation retrieval by means of neural networks and satellite data, and BRDF (Bidirectional Reflectance Distribution Function) characterization using autonomous aerial systems. She works as EO Data Scientist for GEO-K S.r.l., Rome, Italy.



Valeria La Pegna received the master’s degree in environmental engineering from the Technical University of Catalonia - BarcelonaTech, Barcelona, Spain, in 2019. She is currently working toward the Ph.D. degree in computer science, control and geoinformation with the Tor Vergata University of Rome, Rome, Italy.

She is working as Remote Sensing Data Analyst with GEO-K S.r.l., Rome. Her main research interests include artificial intelligence retrieval algorithms for water quality application from multispectral and hyperspectral satellite data.



Zoltan Bartalis received the M.Sc. degree in space engineering from the Luleå University of Technology, Luleå, Sweden, in 2004 and the Ph.D. degree in satellite remote sensing from the Vienna University of Technology, Vienna, Austria, in 2009.

He works as an Earth Observation (EO) Exploitation Engineer with the EO Directorate, European Space Agency (ESA ESRIN), Italy. His main focus is on growing the prospects of satellite EO data and applications being used in a coherent and sustainable way in the day-to-day activities of various user groups and organizations. He is also actively supporting geospatial information service providers in ESA member states in developing their commercial activities. His science and research background is in microwave remote sensing globally over land, in particular for the hydrological cycle. His current research interests include EO applications for urban development, energy and natural resources, especially in the context of international development.



Fabio Del Frate (Senior Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in computer science from the University of Rome “Tor Vergata,” Rome, Italy, in 1992 and 1997, respectively.

From 1995 to 1996, he was a Visiting Scientist with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. In 1998 and 1999, he was with the ESRIN centre, European Space Agency, Frascati, Italy. Since 1999, he has been with the University of Rome “Tor Vergata,” where he is currently a Full Professor and teaches courses on remote sensing and applied electromagnetism in various master and Ph.D. Programs. In 2006, he co-founded GEO-K srl, the first spin-off company of the University of “Tor Vergata” of which he is now President. In 2021 and 2022, he was a Visiting Professor with the ESA ESRIN Phi-Lab to provide support in the use of artificial intelligence for Earth observation data processing. He has been a Lecturer in different European Universities (Prague, Trier, Cracow, Brasov, Limassol, Naples, L’Aquila) and he is, or has been, a Principal investigator/Project Manager in several European Space Agency (ESA) and Italian Space Agency (ASI) funded research projects. He is author of more than 300 international scientific publications with a special focus on feature extraction algorithms from EO data using neural networks.

Dr. Del Frate has been a Session Organizer and in technical boards of International Conferences and Workshops focused on Geoscience and Remote Sensing. He has been an Associate Editor for the journal “*Geoscience and Remote Sensing Letters*,” and a Guest Editor for the journals “*EURASIP Journal on Advances in Signal Processing*” and “*Remote Sensing*.” He is a Senior Member of the IEEE Geoscience and Remote Sensing Society, where he is also a Member of the Image Analysis and Data Fusion Technical Committee. In 2006 and 2007, he was a member of the group winning the IEEE data fusion contest. In 2015, he was appointed EUMETSAT Associate Scientist for activities regarding the estimation of precipitation rate from satellite data.