

A High-Performance Multispectral Adaptation GAN for Harmonizing Dense Time Series of Landsat-8 and Sentinel-2 images

Rocco Sedona, *Student Member, IEEE*, Claudia Paris, *Member, IEEE*, Gabriele Cavallaro, *Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE* and Morris Riedel, *Member, IEEE*

Abstract—The combination of data acquired by Landsat-8 and Sentinel-2 Earth Observation (EO) missions produces dense Time Series (TSs) of multispectral images that are essential for monitoring the dynamics of land-cover and land-use classes across the Earth’s surface with high temporal resolution. However, the optical sensors of the two missions have different spectral and spatial properties, thus they require a harmonization processing step before they can be exploited in Remote Sensing (RS) applications. In this work, we propose a workflow based on a Deep Learning (DL) approach to harmonize these two products developed and deployed on an High-Performance Computing (HPC) environment. In particular, we use a multispectral Generative Adversarial Network (GAN) with a U-Net generator and a PatchGAN discriminator to integrate existing Landsat-8 TSs with data sensed by the Sentinel-2 mission. We show a qualitative and quantitative comparison with an existing physical method (National Aeronautics and Space Administration (NASA) Harmonized Landsat and Sentinel (HLS)) and analyze original and generated data in different experimental setups with the support of spectral distortion metrics. To demonstrate the effectiveness of the proposed approach, a crop type mapping task is addressed using the harmonized dense TS of images, which achieved an Overall Accuracy (OA) of 87.83% compared to 81.66% of the state-of-the-art method.

Index Terms—Landsat-8, Sentinel-2, dense Time Series (TSs), harmonization, virtual constellation, Remote Sensing (RS), Deep Learning (DL), Generative Adversarial Network (GAN), High Performance Computing (HPC).

I. INTRODUCTION

THE availability of multispectral images systematically acquired by Remote Sensing (RS) satellites is pivotal for the observation of land surface change and dynamic processes [1], such as changes resulting from natural calamities [2], expansion of urban areas [3], vegetation anomaly and phenology changes [4], distribution of surface water resources [5], deforestation [6], etc. The time series (TS) of multispectral images acquired by the NASA/United States Geological Survey (USGS)’s Landsat-8 [7] and the European Space Agency (ESA)’s Sentinel-2 [8] missions are the most

widely accessible moderate-to-high spatial resolution RS satellite images.

Landsat-8 was launched in 2013 and carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). It acquires multispectral images at 30 m spatial resolution, which is suitable for a wide variety of tasks. However, Landsat-8 can only revisit the same area every 16 days, which is not sufficient in applications requiring more frequent observations (e.g., near real-time monitoring of continuous processes [9]). The Sentinel-2A and Sentinel-2B are the two polar orbiting satellites of the Sentinel-2 constellation that were launched in 2015 and 2017, respectively. This constellation can reach a revisit time of 5 days at the equator (and even less for areas covered by more than one orbit) and acquire 13 optical bands with 10, 20 and 60 m spatial resolution.

The starting of the Sentinel-2 mission has opened potential opportunities for combining its data with the ones acquired by Landsat-8 to achieve more dense observations. In particular, their integration can densify the acquired TSs and increase the revisit time up to 3–5 days [10] and obtain more frequent cloud-free surface observations. Furthermore, the spatial resolution and spectral configuration (i.e., placement and number of spectral bands) of the Sentinel-2 sensor were designed to be compatible to analogous bands in Satellite Pour l’Observation de la Terre (SPOT) and Landsat sensors [11]. Consequently, many research works have exploited virtual constellations of Sentinel-2 and Landsat-8 for addressing different types of applications, for example to assess winter wheat yields at regional scale [12], estimate number and timing of mowing events of grasslands [13], monitor aquatic systems [14], retrieve the temporal variations in biochemical and structural vegetation properties [15], estimate inland water quality [16], detect irrigated areas [17], analyse land productivity and yield assessment [18], map land surface phenology at continental scale [19], determine the spatial distribution of evergreen forest in cloudy and rainy areas [20], etc.

Despite the similarity between Sentinel-2 and Landsat-8 observations, the two missions have different spatial resolution, field of view spectral bandwidth, and spectral response function. Consequently, before using together Sentinel-2 and Landsat-8 images, it is necessary to apply models for cross-sensor data integration [21] [22] [23]. Linear regression is the most widely used approach to reduce the spectral differences between the two sensors. The authors in [24] used Bidirectional Reflectance Distribution Function (BRDF)

Rocco Sedona and Morris Riedel are with the Jülich Supercomputing Centre, Wilhelm-Johnen Straße 52428 Jülich, Germany, and the University of Iceland, 107 Reykjavik, Iceland (e-mail: r.sedona@fz-juelich.de; m.riedel@fz-juelich.de).

Claudia Paris and Lorenzo Bruzzone are with the Department of Information Engineering and Computer Science, University of Trento, 38122 Trento, Italy (e-mail: claudia.paris@unitn.it; lorenzo.bruzzone@ing.unitn.it).

Gabriele Cavallaro is with the Jülich Supercomputing Centre, Wilhelm-Johnen Straße 52428 Jülich, Germany (e-mail: g.cavallaro@fz-juelich.de).

correction and data re-sampling to attenuate the difference introduced by the different field of view and spatial resolution, respectively. Other studies designed regional fixed per-band transformation coefficients for applying reflectance adjustment in Australia [25], southern Africa [26], and United States [27].

Since 2018, NASA is producing a Harmonized Landsat and Sentinel (HLS) dataset¹ to further improve the temporal resolution of the combined product [28]. NASA proposed a method that creates global fixed per-band transformation coefficients to reduce the reflectance difference between Landsat-8 and Sentinel-2 and generate smooth spectral TSs. In particular, the approach takes into account the differences in spatial resolution, atmospheric correction approaches, view geometry and radiometric characteristics of spectral bands. ESA has considered this approach as a reference work for the definition of the Sen2Like framework [29]. The objective of Sen2Like is to generate Sentinel-2 like harmonised/fused surface reflectances with higher periodicity by integrating additional compatible optical mission sensors. The current implementation (November 2020) can harmonize Landsat-8 and Sentinel-2 data products². The authors in [30] observed that these methods can reduce the reflectance difference to only some degree. It is possible that the regional or global scale fixed per-band transformation coefficients may not be suitable for all land cover types and at all geographical locations. To mitigate this problem, they proposed a time-series-based approach³ to improve the consistency of the HLS datasets, which uses the TSs of matched Landsat-8 and Sentinel-2 observations to build linear regression models for each pixel. They then conducted the reflectance adjustment for each individual pixel separately.

Instead of using a physical method or fitting the transformation coefficients of a linear regression, in our work we developed an approach based on Machine Learning (ML), and more specifically on a Generative Adversarial Network (GAN) architecture to harmonize the Sentinel-2 and Landsat-8 products, transforming the data acquired by the Sentinel-2 Multi-Spectral Instrument (MSI) sensor into Landsat-8 OLI-like data. In the last decade DL has enabled a leap in the quality of a wide variety of applications in RS [31]. In particular, GANs were first presented by [32] in 2014 and are based on the training with the backpropagation algorithm of two sub-models, a generator and a discriminator. An extension of GANs are the Conditional GANs [33], in which the generator is given additional information to better approximate the distribution of the real samples. The competitive game of one model against the other pushes the generator to create new fake examples that are indistinguishable from real ones. While the generator creates new data from an input distribution, the discriminator is devoted to discern the real and generated examples looking at their distribution. For these reasons, GAN have attracted much research efforts to computer-vision-related tasks [34].

GANs have been employed also in different RS applications. Among those, a promising application is super-resolution,

where GANs offer the ability to retrieve high-frequency components that seem not to be captured by existing Convolutional Neural Networks (CNNs) [35], thanks to the contribution of the adversarial loss [36] [37] [38]. Chen *et al.* [39] proposed a GAN-based approach to super-resolve Landsat-8 images and reconstruct them to be Sentinel-2-like using the true color composite of RGB bands. In our approach we propose the opposite direction of the data flow, from Sentinel-2 to Landsat-8 data, as our proposed method focuses on radiometric consistency rather than spatial resolution. Moreover, we also use the near infrared (NIR) and the short wave infrared (SWIR) bands, which are extremely important to perform environmental monitoring (e.g., vegetation biophysical and biochemical variable retrieval, ice detection, etc.). In particular, the NIR and SWIR spectral channels provide key information on vegetation and crops status. GANs have been applied also to other tasks, such as to enhance the detection of small objects in RS data with an adaptation of the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [40], or to change detection with multi-sensor data with the use of a CycleGAN [41]. Conditional GANs were used also for the fusion of acquisitions from Synthetic Aperture Radar (SAR) and optical sensors, e.g., in [42] optical data were reconstructed from SAR and in [43] a GAN was used to fuse SAR and optical multispectral data for cloud removal.

A well known bottleneck of employing DL models is the large amount of computational resources that are needed for the training phase. DL models require to be fed with large amounts of data in order to learn meaningful features, thus implying the need for dedicated pipelines for extraction and handling of such data, which can impact severely the performances of the methods. Despite the great success of CNNs, their deployment on commodity hardware (e.g., desktop computers, laptops) is often challenging, given their computational power and memory constraints. High-Performance Computing (HPC) systems can come at aid in that regard, offering dedicated hardware accelerators to efficiently deploy and scale-up processing workflows and significantly enhancing their computational performance (i.e., reported as Floating Point Operations per Second (FLOPS)). HPC systems are on the verge of entering into the new era of exascale computing in the coming years, as currently the most powerful computers can reach hundreds of PetaFLOPS⁴. A large number of fields of research use HPC systems for addressing data storage challenges and developing scalable data processing workflows: from climatology to astrophysics, medicine and industrial applications [44]. In RS, HPC has been an essential component from the very beginning in the field of EO since its technology and applications include unique data processing, storage or transmission requirements [45] [46]. In the current era of Artificial Intelligence (AI) supercomputers (i.e., HPC systems equipped with specialized hardware accelerators [47]), applications from RS also use them to speed-up the processing of DL models that include a high number of trainable parameters [48].

From this brief analysis of the literature, it turns out that the

¹Online [Available]: <https://hls.gsfc.nasa.gov/>

²Online [Available]: <https://github.com/senbox-org/sen2like>

³Online [Available]: <https://github.com/GERSL/TRA>

⁴Online [Available]: <https://www.top500.org/lists/top500/2020/11/>

integration of the multispectral images acquired by Landsat-8 and Sentinel-2 is extremely interesting from the operational view point due to the complementary properties of the two sensors. While Landsat satellites are approaching 50 years of continuous global data collection with a temporal revisit of 16 days, the recent launch of Sentinel-2 allows for the acquisition of images having a very high revisit time (i.e., 5 days at the equator with 2 satellites which results in 2-3 days at mid-latitudes). In this context, Sentinel-2 images can be used to generate Landsat-8 like images (from the spectral and spatial view point) with the aim of having dense TSs of images compatible with the TSs of real Landsat-8 available in the past. Such long and dense TSs of images allow for long-term environmental analyses, which are extremely important for several applications (i.e., climate change, deforestation analysis, desertification, urban monitoring, etc.). In the literature, the integration of Landsat-8 and Sentinel-2 images has been mainly addressed by the RS community considering physical methods or regression models due to: (i) their capability of properly handling the harmonization problem from the physical view point, and (ii) their low computational burden. The latter is particularly important when working at country or continental scale, where the optical pre-processing has to be applied over a hundred of images. However, such methods can only partially mitigate the reflectance difference and may fail in heterogeneous areas where complex non-linear harmonization problems have to be solved. In this framework, it is necessary to define an automatic system suitable for all land cover types and at all geographical locations, which is able perform the integration of these data in a fast and efficient way.

This paper presents an automatic work-flow which aims to facilitate the integration of the optical satellite images acquired by Landsat-8 and Sentinel-2 spectral sensors at operational level. Differently from the literature, the proposed system architecture takes advantage from the capability of the GAN to accurately learn and model the considered non-linear problem, while preserving the spectral and spatial properties of the two satellite sensors. To mitigate the computational cost of the required DL models, we take advantage of HPC systems to deploy a parallel and scalable processing workflow that encompasses the extraction of the features from the input tiles, the training of the model and the reconstruction of the harmonized Landsat-8 and Sentinel-2 data product. The speed-up of the training of the DL model is obtained thanks to the adoption of a data parallel strategy, which distributes the training of the GAN on multiple GPUs.

The main contributions of this work are: (1) the definition of a multispectral adaptation GAN tailored to the peculiar properties of Sentinel-2 and Landsat-8 in terms of spatial resolution, spectral bandwidth, and spectral response function, (2) the implementation of a fully automatic and unsupervised dedicated pipeline, ready-to-use, being able to ingest Sentinel-2 and Landsat-8 data and to produce a dense TS of optical satellite images, and (3) the efficient implementation of a parallel and scalable processing workflow developed and deployed on an HPC environment on up to 16 GPUs thanks to the adoption of a data distributed strategy, which contributes

to mitigate the computational burden of the training.

II. PROPOSED MULTISPECTRAL ADAPTATION GAN

The aim of this work is to generate harmonized dense TSs of Landsat-8 and Sentinel-2 images. To this end, we propose a Multispectral adaptation GAN (MGAN) model tailored to the specific properties of the considered satellite optical data. Our objective is to model the spatial and spectral properties (Point Spread Function) of the two sensors in order to adapt the Sentinel-2 data to be Landsat-8 like. Indeed, the proposed GAN is tailored to the specific spectral and spatial properties of the considered sensors to facilitate the adaptation of the Sentinel-2 images to the Landsat-8 ones. In particular, the proposed architecture is build upon the established pix2pix conditional GAN [49] that was designed for color and grayscale image-to-image translation. Based on the GAN concept, the adversarial game played by the two models of the original pix2pix architecture [49] can be represented by the formula:

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{X}, \mathbf{Y}} [\log D(\mathbf{X}, \mathbf{Y})] + \mathbb{E}_{\mathbf{X}, z} [\log (1 - D(\mathbf{X}, G(\mathbf{X}, z)))], \quad (1)$$

where \mathbb{E} is the expected value, \mathbf{X} and \mathbf{Y} are the source and target images (having the same resolution), z the input noise of the generator and $V(G, D)$ is the value function. In particular, the generator G and the discriminator D of pix2pix are a U-net encoder-decoder architecture with skip connections and a PatchGAN, respectively. In the U-net encoder-decoder generator [50], the first part contains a number of downsampling convolution layers. The second part is a mirrored version of the first, with a transposed convolution for upsampling the data, which flows from the bottom to the top of the U-net through a bottleneck. The skip connections, which link the inner layers of the encoder and decoder, allow low-level information to pass directly from the first to the last layers of the U-net.

Differently from the original implementation of the pix2pix, the input data are no more RGB natural images, but multi-resolution and multiband images with different spectral properties. To handle the peculiarities of the considered RS data, we trained the proposed MGAN from scratch using paired Landsat-8 and Sentinel-2 images. Table I reports the properties of the considered spectral bands in terms of spatial and spectral resolutions for both the considered optical sensors. According to the spectral characteristic of Sentinel-2 and Landsat-8, we focused the attention on the four 10 m bands and the two shortwave infrared spectral channels acquired at 20 m by Sentinel-2 (i.e., the spectral bands consistent with the Landsat-8 ones). Let us focus on the multiresolution Sentinel-2 images. Let $\mathbf{X}_{\text{HR}} \in \mathcal{R}^{d_1 \times d_1 \times L_{\text{HR}}}$ and $\mathbf{X}_{\text{LR}} \in \mathcal{R}^{d_2 \times d_2 \times L_{\text{LR}}}$ be the set of high resolution (10 m) and low resolution (20 m) spectral channels of Sentinel-2, respectively, where \mathbf{X}_{HR} has $d_1 \times d_1$ pixels and L_{HR} bands while \mathbf{X}_{LR} has $d_2 \times d_2$ pixels and L_{LR} bands. Let $\mathbf{Y} \in \mathcal{R}^{d_3 \times d_3 \times L_3}$ be the real Landsat-8 image contemporary to the Sentinel-2 one, having $d_3 \times d_3$ pixels and a number of bands equal to $L_3 = L_{\text{HR}} + L_{\text{LR}}$.

In the considered implementation of the proposed MGAN, the bottom of the generator has been modified to take as input

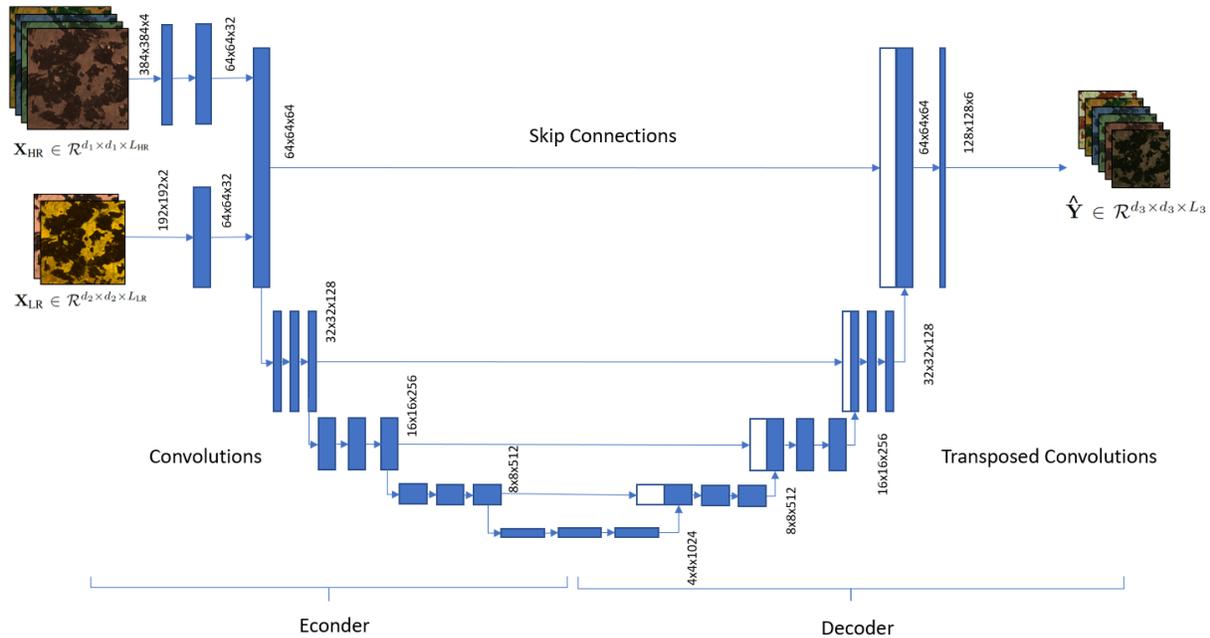


Fig. 1. Flowchart of the modified U-Net tailored to the peculiar spectral and spatial properties of Sentinel-2 and Landsat-8.

the patches of Sentinel-2 at original resolution X_{HR} and X_{LR} (i.e., 10 m and 20 m). To this end, we added one convolutional layer for each initial resolution, concatenating their output before entering into the encoder-decoder structure. Fig. 1 illustrates the modified U-Net tailored to the peculiar spectral and spatial properties of Sentinel-2 and Landsat-8 for facilitating the sensor adaptation performed by the proposed MGAN. The patches of the high-resolution Sentinel-2 spectral channels X_{HR} have size $384 \times 384 \times 4$, while the low-resolution ones X_{LR} have size $192 \times 192 \times 2$. The different convolutions and transposed convolutions lead to the direct production of a Landsat-8 like image having size $128 \times 128 \times 6$, which implicitly includes the 2 channels of X_{LR} and the 4 channels X_{HR} having spatial resolution of 30m. This condition allows us to keep the same number of inner layers of the generator and the discriminator as in the original implementation. Let $\hat{Y} \in \mathcal{R}^{d_3 \times d_3 \times L_3}$ be the downsampled Sentinel-2 image having all the spectral bands at the spatial resolution of the desired target image. Please note that the downsampling convolution layer allows us to directly handle the spatial resolutions of the different spectral bands of Sentinel-2 without the need of performing any pre-processing interpolation step.

The PatchGAN discriminator is designed to capture the patterns at the scale of the input image. Its objective is to classify $N \times N$ patches of $G(\mathbf{X}, z)$ (the input synthetic patch created by the generator) and \mathbf{Y} (the target Landsat-8 patch) as fake or true, encouraging the generator to produce more accurate and realistic outputs. Differently from the standard pix2pix implementation, the considered MGAN does not perform the instance normalization [51], since it is not suited to multispectral images. Indeed, similarly to the case of the standard batch normalization typically used in computer vision, the patches may not be consistent from the

TABLE I
SPECTRAL BANDS OF LANDSAT-8 AND SENTINEL-2 SELECTED ACCORDING TO THE SPECTRAL AGREEMENT OF THE OPTICAL SENSORS.

Landsat-8			Sentinel-2		
Band	Wavelength (μm)	Res. (m)	Band	Wavelength (μm)	Res. (m)
2	0.450-0.515	30	2	0.458-0.523	10
3	0.525-0.600	30	3	0.543-0.578	10
4	0.630-0.680	30	4	0.650-0.680	10
5	0.845-0.885	30	8	0.785-0.900	10
6	1.560-1.660	30	11	1.565-1.655	20
7	2.100-2.300	30	12	2.100-2.280	20

spectral view point. For this reason, in the model we added the spectral normalization right after the instance normalization in the downsampling blocks of the discriminator [52]. The addition of those layers in the discriminator is beneficial for the stability of the training and the spectral content of the obtained synthetic Landsat-8 images. The addition of those layers in the discriminator is beneficial for the stability of the training and the spectral content of the obtained synthetic Landsat-8 images. In greater details, we train the generator and discriminator jointly, employing two losses. The L_1 loss is used in for the training of the generator to learn a low-frequency representation:

$$L_1 = \mathbb{E}_{\mathbf{X}} \left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\|, \quad (2)$$

where $\hat{\mathbf{Y}} = G(\mathbf{X}, z)$ is the generated image obtained considering as input the Sentinel-2 image \mathbf{X} and \mathbf{Y} is the target Landsat-8 image. We adopted a relativistic adversarial loss for the discriminator (shown in equation 3) as a replacement of the original adversarial loss employed in pix2pix. Using the

relativistic loss, instead of the absolute probability that one input image is real or fake, the relative probability that a real image is more realistic than a fake one is computed [53]. The adoption of the relativistic adversarial loss increases for the stability of the training [54]. The discriminator loss is:

$$L_D = -\mathbb{E}_{\mathbf{Y}}[\log(D_{R_a}(\mathbf{Y}, \hat{\mathbf{Y}}))] - \mathbb{E}_{\hat{\mathbf{Y}}}[\log(1 - D_{R_a}(\hat{\mathbf{Y}}, \mathbf{Y}))] \quad (3)$$

and the generator loss:

$$L_G = -\mathbb{E}_{\mathbf{Y}}[\log(1 - D_{R_a}(\hat{\mathbf{Y}}, \mathbf{Y}))] - \mathbb{E}_{\hat{\mathbf{Y}}}[\log(D_{R_a}(\mathbf{Y}, \hat{\mathbf{Y}}))], \quad (4)$$

where \mathbf{Y} and $\hat{\mathbf{Y}}$ are the real and the fake generated images, respectively, and D_{R_a} is the output of the discriminator. To properly train the considered MGAN from scratch, we implemented data augmentation. The lack of large amount of data is known to pose several challenges during the training of GAN, since in that setting the discriminator tends to fool the generator easily, which in turn gets stuck and cannot improve anymore. This is particularly true when dealing with RS data. [55] recently introduced a data augmentation technique specifically designed to work with GANs. Differentiable augmentation addresses this issue by applying the same set of transformations on both the generated and real images, regularizing the discriminator and reducing training instability. We adopted the color (contrast, brightness, saturation), translation (the images are translated and zero padded) and cutout (masking a region of the images) policies.

III. DATASET DESCRIPTION AND DESIGN OF EXPERIMENTS

In this section, we present the considered study area and the RS data employed to test the proposed approach. Then, we describe in detail the procedure designed to generate the harmonized TS of Sentinel-2 and Landsat-8 images.

A. Dataset Description

Fig. 2 presents the considered study area, which covers the valley of the Donau in the proximities of Linz, Austria (tile 33UVP of Sentinel-2, tile 191/026 of Landsat-8). Such area is characterized by a heterogeneous landscape typical of the Alpine region, where the topography ranges from high mountain to lowlands areas. The land cover is characterized by the presence of many crop types, which model a complex scenario since crops rapidly change their textural and spectral features. Moreover, the study area is heavily affected by cloud and snow coverage. Due to high temporal resolution of Sentinel-2, several pairs of real Landsat-8 and Sentinel-2 images acquired at the same date (or a one day of distance) are used to train the GAN network from scratch. Table II reports the acquisition dates of the considered images collected in Spring and Autumn. Only images having low cloud coverage (smaller than 30%) were used to train the MGAN.

To assess the capability of the trained MGAN to correctly generate synthetic Landsat-8 data from Sentinel-2 images, the Sentinel-2 data acquired on 03/07/2018 was not involved in

TABLE II
LANDSAT-8 (TILE 191/026) AND SENTINEL 2 (TILE 33UVP) IMAGES USED IN THE EXPERIMENTS. FIVE IMAGES WERE USED TO TRAIN THE MGAN, WHILE THE SENTINEL 2 IMAGE ACQUIRED ON THE 03/07/2018 WAS USED FOR PREDICTION ONLY.

	Landsat-8 images	Sentinel-2 images
Training	04/04/2018	04/04/2018
	20/04/2018	21/04/2018
	06/05/2018	06/05/2018
	27/09/2018	26/09/2018
	13/10/2018	13/10/2018
Prediction	-	03/07/2018

TABLE III
NUMBER OF SAMPLES FOR EACH CROP TYPE.

Crop Type	# Samples
Grassland	2600
Maize	1668
Winter Barley	2400
Winter Caraway	400
Rapeseed	868
Beet	972
Spring Cereal	766
Winter Wheat	600

the training but used for prediction only. Indeed, the Landsat-8 acquisition available in July 2018 are all strongly affected by cloud coverage; thus, they cannot be used to train the model. This real test case demonstrates the importance of the proposed method from the operational view point. The use of Sentinel-2 data to generate synthetic Landsat-8 images having a good temporal sampling of the whole year. These TSs are extremely important to correctly handle multitemporal tasks such as crop type mapping. To this end, a 2018 reference dataset of crop types of the considered study area is used to accomplish this peculiar classification task. Table III reports the set of crop types of the considered classification problems together with the number of samples per class. The training and test sets are statistically independent, since training and test samples have been extracted from spatially disjoint portions of the considered study area. An example of ground reference data used to perform the crop type mapping task is reported in Fig. 2, where in the zoom the different crop types are highlighted in different colors.

B. Design of the Experiments

To train the considered MGAN, both the Landsat-8 and Sentinel-2 images are split into patches. Fig. 3 reports the different stages of our method, from the training of the model to the prediction and reconstruction of the entire tiles. First, the Landsat-8 images are warped to extract the region overlapping with the Sentinel-2 tiles, by applying a nearest neighbor resampling strategy that does not affect the spectral content of the image. Then, possible spectral outliers are removed

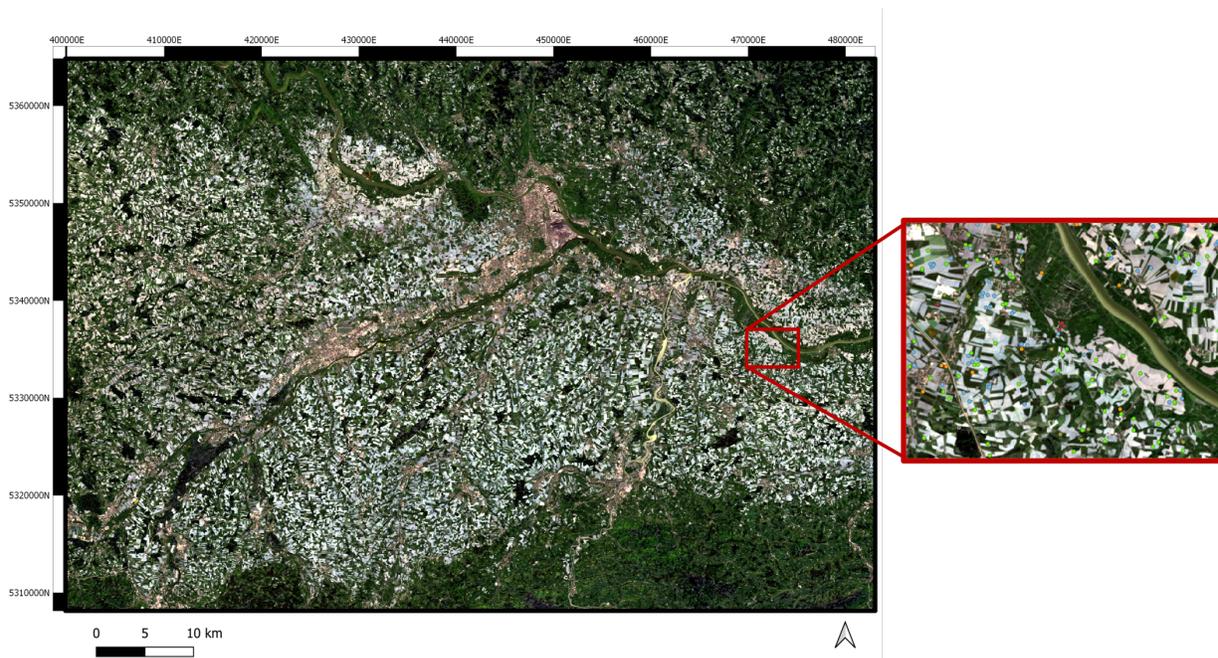


Fig. 2. True color representation of the Sentinel-2 image acquired on the 21/04/2018 over the considered study area (coordinates are reported in the UTM WGS84 33N system). An example of the reference data used to perform the crop type classification task is reported in the zoom area highlighted in red.

from both images. To this end, we considered the standard procedure of saturating the pixel values below and above the 1 and 99 percentiles of the spectral distribution computed per band. Finally, the paired patches were extracted from the two original TSs of Landsat-8 and Sentinel-2. For the Landsat-8 data, the dimension of the patches is $128 \text{ px} \times 128 \text{ px}$ (30 m resolution), while for Sentinel-2 they are $192 \text{ px} \times 192 \text{ px}$ (for the 20 m resolution bands) and $394 \text{ px} \times 394 \text{ px}$ (for the 10 m resolution bands). In particular, a stride of half the dimension of the patch is considered to generate overlapping patches, thus increasing the number of samples. Patches with a significant cloud or snow coverage are not used during training and are excluded with the usage of the available masks. The information provided by the cloud masks of Landsat-8 (i.e., pixel_qa band) and Sentinel-2 (i.e., SCL band) are used to define the valid patches for training. The pixel values of each patch are normalized per band by subtracting the mean and dividing by the maximum value. Once that the GAN is trained, it can be used to predict synthetic Landsat-8 images by using Sentinel-2 data. During prediction, each original Sentinel-2 patch is fed into the generator and the corresponding synthetic Landsat-8 patch is produced. The final step is the reconstruction of the entire image from the predicted patches. We applied a buffer equal to $1/4$ of the dimension of the patch when fusing them. The tile is then reconstructed using only the central part of the patches, skipping the buffers to limit distortions caused by the convolution operations at the edges of the patches.

IV. IMPLEMENTATION AND EXPERIMENTAL SETUP

In this section, details are given on the implementation and computational setup. Moreover, the quality indexes used to quantitatively evaluate the proposed method are reported in the experimental setup section.

A. Implementation Setup

Of the two main families to distribute the training of a model [56], we used the data distribution approach (i.e., data parallelism). Among the different frameworks that exist to integrate a data distributed strategy into existing code we adopted Horovod [57], a library that offers a flexible API that works on top of most DL libraries, i.e., TensorFlow, Keras, PyTorch and MXNet. Horovod makes use of Message Passing Interface (MPI) and the NVIDIA Collective Communication Library (NCCL) to implement a decentralized and efficient ring-allreduce algorithm [57], which allows the computation of the gradients in a distributed fashion. We used ADAM with base learning rate $lr = 0.0001$ for the optimization of both the generator and the discriminator, which we scaled linearly w.r.t. the number of Graphics Processing Units (GPUs), without warm-up phase and learning rate schedulers. The training was performed for 100 epochs, as after that point the L_1 loss begins to diverge and the quality of the predicted patches deteriorates. The weights of the U-Net and of the PatchGAN were initialized with the default Glorot uniform distribution [58]. The local batch size used for each GPU is 16, therefore the resulting maximum global batch size used in the present work, computed as $\text{global_batch_size} = \text{number_gpus} \times \text{local_batch_size}$, is equal to 256.

B. Experimental Setup

The experiments were carried out on the Extreme Scale Booster (ESB) partition of the of the Dynamic Exascale Entry Platform - Extreme Scale Technologies (DEEP-EST) and on the booster partition of the Jülich Wizard for European Leadership Science (JUWELS) supercomputers at the Jülich Supercomputing Centre (JSC) [59]. The training was scaled on up

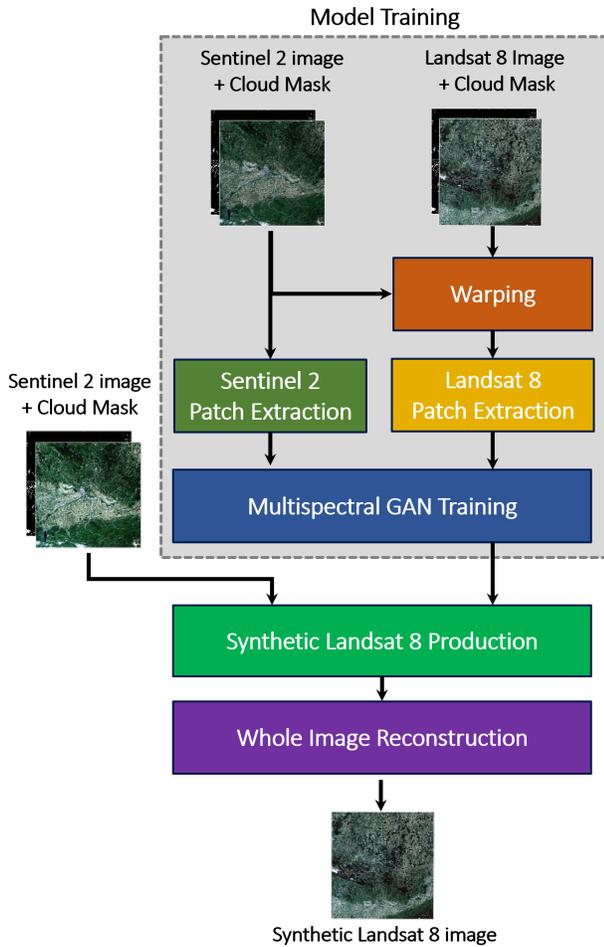


Fig. 3. Flowchart of the different stages of the proposed method. It receives as input TSs of Landsat-8 and Sentinel-2 acquired over the same geographical area. Firstly, the warping step aligns the two TSs. Then, the patch extractor generates paired and overlapping patches (i.e., training samples) that are trained by the proposed MGAN. The final stage reconstructs the whole synthetic Landsat-8 image.

to 16 Nvidia Tesla V100 and A100 Graphics Processing Units (GPUs). We used Horovod data-parallel framework on top of TensorFlow2, with a custom made training loop. The data pre-processing was deployed on the JUWELS system [47]. We used the Geospatial Data Abstraction Library GDAL 2.3.2 through its Python API.

To quantitatively evaluate the results obtained we considered several spectral distortion metrics typically used in the literature. In particular, we considered the Relative Dimensionless Global Error (ERGAS), the Spectral Angle Mapper (SAM), the Root Mean Square Error (RMSE), the Universal Image Quality Index (UIQI) and the Peak Signal-to-Noise Ratio (PSNR) measures on the valid patches (i.e., low cloud coverage). SAM [60] measures the spectral distortion in terms of angle between the vectors of the reference image and generated image:

$$\text{SAM}(\mathbf{Y}, \hat{\mathbf{Y}}) \triangleq \arccos \left(\frac{\langle \mathbf{Y}, \hat{\mathbf{Y}} \rangle}{\|\mathbf{Y}\|_2 \cdot \|\hat{\mathbf{Y}}\|_2} \right), \quad (5)$$

where \mathbf{Y} is the real input and $\hat{\mathbf{Y}}$ the predicted input. The

lower is the value of SAM, the lower the presence of spectral deviations between the two images. ERGAS measures the quality of the generated image compared to the reference image as a normalized mean square error between each band of the two images [61]:

$$\text{ERGAS}(\mathbf{Y}, \hat{\mathbf{Y}}) \triangleq 100 \frac{1}{S} \sqrt{\frac{1}{L_3} \sum_{l=1}^{L_3} \frac{\text{MSE}(\mathbf{Y}_l, \hat{\mathbf{Y}}_l)}{\mu_{\hat{\mathbf{Y}}_l}^2}}, \quad (6)$$

where $\frac{1}{S}$ is the ratio between the pixel sizes (i.e., equal to one in our case), \mathbf{Y}_l and $\hat{\mathbf{Y}}_l$ are the l th bands of the generated image and of the reference image, respectively; the $\text{MSE}(\mathbf{Y}_l, \hat{\mathbf{Y}}_l)$ is the mean squared error between \mathbf{Y}_l and $\hat{\mathbf{Y}}_l$ and $\mu_{\hat{\mathbf{Y}}_l}$ is the mean of $\hat{\mathbf{Y}}_l$. As for SAM, a low value of ERGAS implies a low presence of distortion in the generated image compared to the reference. The RMSE is defined as:

$$\text{RMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) \triangleq \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|}{d_3^2} \quad (7)$$

where \mathbf{Y} is the original input, $\hat{\mathbf{Y}}$ the predicted input.

The UIQI [62] has been computed on a sliding window of size 32×32 pixels, and averaged over all window positions per band. Let \mathbf{y}_j and $\hat{\mathbf{y}}_j$ denote the j th windowed segment of a single band of the reference and the simulated images, respectively. The UIQI is given by:

$$\text{Q}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq \frac{1}{W} \sum_{j=1}^W \frac{\sigma_{\mathbf{y}_j \hat{\mathbf{y}}_j}}{\sigma_{\mathbf{y}_j} \sigma_{\hat{\mathbf{y}}_j}} \times \frac{2\mu_{\mathbf{y}_j} \mu_{\hat{\mathbf{y}}_j}}{\mu_{\mathbf{y}_j}^2 + \mu_{\hat{\mathbf{y}}_j}^2} \times \frac{2\sigma_{\mathbf{y}_j} \sigma_{\hat{\mathbf{y}}_j}}{\sigma_{\mathbf{y}_{i,j}}^2 + \sigma_{\hat{\mathbf{y}}_j}^2} \quad (8)$$

where $\sigma_{\mathbf{y}_j \hat{\mathbf{y}}_j}$ is the covariance between \mathbf{y}_j and $\hat{\mathbf{y}}_j$, $\sigma_{\mathbf{y}_j}$ and $\mu_{\mathbf{y}_j}$ are the standard deviation and the mean value of \mathbf{y}_j , while $\sigma_{\hat{\mathbf{y}}_j}$ and $\mu_{\hat{\mathbf{y}}_j}$ are the standard deviation and the mean value of $\hat{\mathbf{y}}_j$, respectively. This index has a range of $[-1, 1]$, being equal to 1 when $\mathbf{y} = \hat{\mathbf{y}}$. To extend the UIQI index to the multiband case, we average the band indexes as follows:

$$\text{Q}(\mathbf{Y}, \hat{\mathbf{Y}}) \triangleq \frac{1}{L_3} \sum_{l=1}^{L_3} \text{Q}(\mathbf{Y}_l, \hat{\mathbf{Y}}_l), \quad (9)$$

The PSNR is defined as:

$$\text{PSNR}(\mathbf{Y}, \hat{\mathbf{Y}}) \triangleq 20 \log_{10} \left(\frac{\lambda^2}{\frac{1}{d_3^2} \|\mathbf{Y} - \hat{\mathbf{Y}}\|} \right) \quad (10)$$

where λ is the number of levels of the images.

V. EXPERIMENTAL RESULTS

In this section, first we present the quantitative results obtained in terms of spectral distortion metrics. The quantitative evaluation is provided together with qualitative examples of the obtained synthetic Landsat-8 images. Finally, the generated TS of synthetic Landsat-8 images is used to perform a crop type mapping task to assess the capability of the network to accurately reproduce the spectral properties of the data. The proposed approach is compared with the physical method HLS [28] developed for reducing the reflectance differences

TABLE IV

SPECTRAL DISTORTION METRICS BETWEEN THE ORIGINAL LANDSAT-8 DATA AND: (1) THE SYNTHETIC LANDSAT-8 IMAGES GENERATED USING THE PROPOSED MGAN, (2) THE HARMONIZED LANDSAT-8 IMAGES GENERATED USING THE BASELINE METHOD HLS, AND (3) THE ORIGINAL CONTEMPORARY SENTINEL-2 IMAGES. THE OBTAINED RESULTS ARE THE AVERAGE VALUES OVER THE 5 IMAGES OF THE CONSIDERED DATASET. RESULTS ARE PROVIDED PER SPECTRAL BAND AND OVERALL. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Data	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Overall
SAM							
Synthetic Landsat-8 (MGAN)	0.22	0.18	0.21	0.16	0.17	0.19	0.19
Harmonized Landsat-8 (HLS)	0.46	0.37	0.48	0.29	0.30	0.39	0.38
Original Sentinel-2	0.32	0.26	0.31	0.23	0.20	0.22	0.26
ERGAS							
Synthetic Landsat-8 (MGAN)	719	674	731	647	669	696	1933
Harmonized Landsat-8 (HLS)	1321	924	1051	1020	882	930	2903
Original Sentinel-2	875	780	884	727	685	740	2180
RMSE							
Synthetic Landsat-8 (MGAN)	185	305	313	1229	806	520	668
Harmonized Landsat-8 (HLS)	275	390	475	1381	946	692	799
Original Sentinel-2	285	371	393	1344	843	607	744
UIQI							
Synthetic Landsat 8 (MGAN)	0.66	0.67	0.66	0.67	0.67	0.66	0.67
Harmonized Landsat-8 (HLS)	0.58	0.63	0.55	0.60	0.64	0.60	0.60
Original Sentinel 2	0.62	0.65	0.64	0.66	0.66	0.65	0.65
PSNR							
Synthetic Landsat-8 (MGAN)	337	333	333	322	325	329	327
Harmonized Landsat-8 (HLS)	331	328	326	317	321	323	321
Original Sentinel 2	331	330	329	319	324	326	324

between Landsat-8 and Sentinel-2, thus generating smooth spectral TSs. Please note that such method is widely used from the operational view point [29].

A. Quantitative and Qualitative Results

Table IV reports the results obtained for different spectral distortion metrics comparing the original Landsat-8 images and: (1) the synthetic Landsat-8 images produced by the proposed MGAN, (2) the harmonized Landsat-8 images generated using the baseline method HLS, and (3) the original contemporary Sentinel-2 images. The best results are highlighted in bold. Please note that the evaluation of the spectral difference between real Landsat-8 data and Sentinel-2 data is reported to evaluate the capability of the methods to reduce the spectral difference of these data.

From the results obtained, one can notice that the metrics computed between Landsat-8 and Sentinel-2 images demonstrate the need of harmonizing these data from the spectral view point. The HLS reduces the spectral distortion for some spectral bands. However, for all the metrics, the best results are achieved by the synthetic Landsat images generated with the proposed MGAN. In particular, the MGAN is able to correctly reproduce the spectral properties of Landsat-8 regardless of the spectral bands. Indeed, similar error metrics are achieved in both the RGB spectral channels (i.e., Band2, Band3 and Band4) as well as the near infrared (Band5) and shortwave infrared bands (Band6 and Band7).

The results obtained from the quantitative view point are confirmed from the qualitative ones. In order to assess the consistency between the generated and the target data, Fig. 4 reports some portions of the: (1) original Landsat-8 image (target), (2) synthetic Landsat-8 data produced by the MGAN, (3) harmonized Landsat-8 data produced by the baseline method (HLS), and (4) contemporary Sentinel-2 image used to generate the Landsat-8 data. The synthetic image produced by the MGAN looks more similar to the original Landsat-8 image than the original Sentinel-2 input data and the harmonized Landsat8 data produced by the HLS method. These results also confirm that the quality of the generated images is good and does not suffer from significant distortions and artifacts. From the results obtained, one can notice that the generated data looks more similar to the original Landsat 8 image than the original Sentinel-2 input data and the harmonized Landsat-8 data produced by the HLS method. For instance, the presence of bright buildings absent in the real Landsat-8 images (see Fig.4a) is visible in the harmonized data produced by the HLS method (Fig.4c) but not present in the synthetic data produced by the MGAN (see Fig.4b).

B. Crop Type Mapping Results

To assess the capability of the proposed MGAN to accurately model the spectral information of Landsat-8, a crop type mapping task was carried out using the obtained TS of produced synthetic images. This peculiar classification

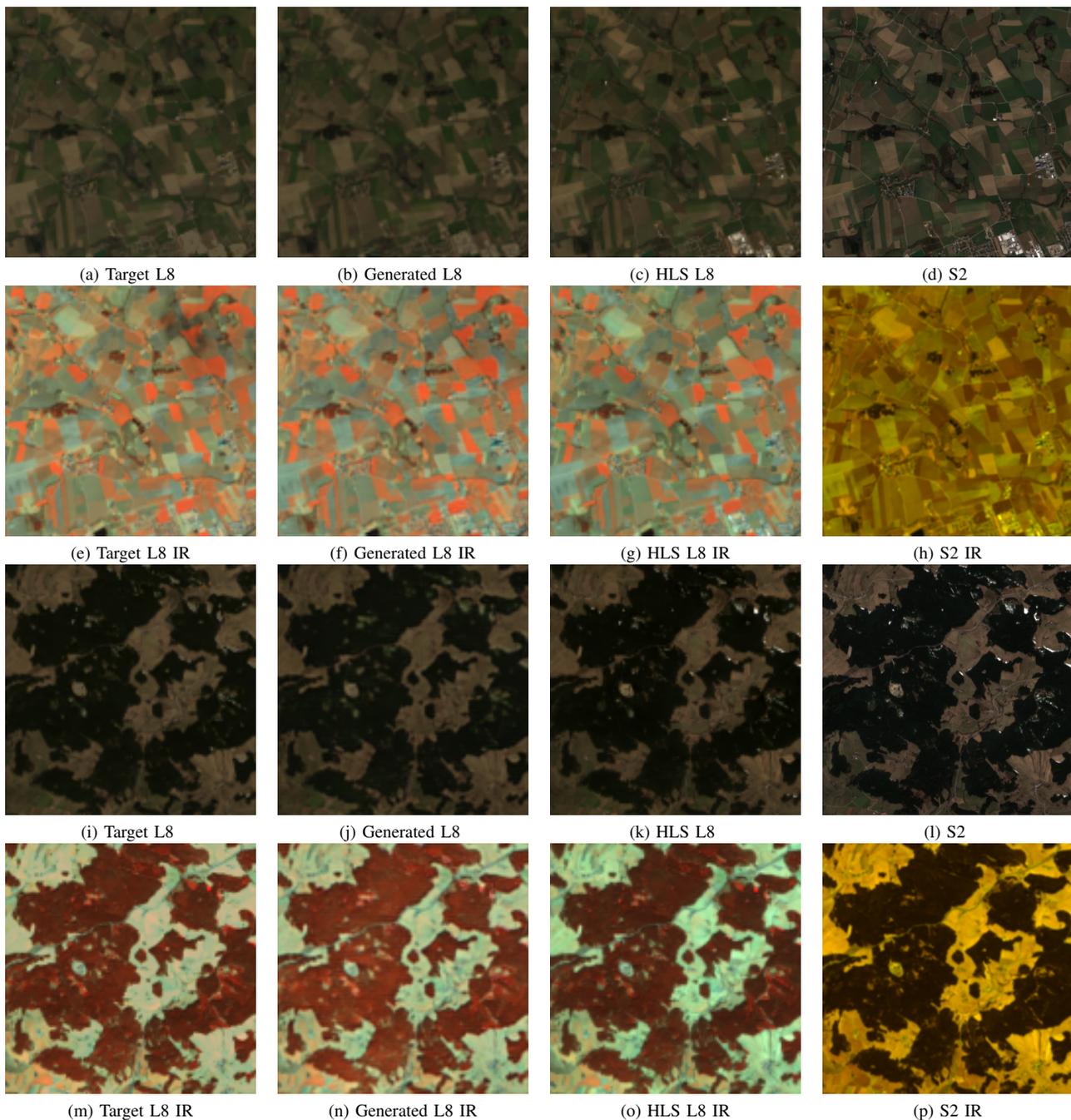


Fig. 4. Qualitative examples of the obtained Landsat-8 images. The true color composite (RGB) is reported for the: (a)(i) target Landsat-8 image, (b)(j) generated Landsat 8 with the proposed MGAN, (c)(k) generated Landsat 8 with the HLS, and (d)(l) original Sentinel-2 image. The false color composite is reported for the NIR and SWIR bands for the: (e)(m) target Landsat-8 image, (f)(n) generated Landsat-8 with the proposed MGAN, (g)(o) generated Landsat-8 with the HLS, and (h)(p) original Sentinel-2 image.

task requires the availability of accurate multitemporal and multispectral information to properly retrieve the crop types present in the scene. Indeed, differently from other land-cover classification tasks that can be performed using mono-temporal data, the temporal information is fundamental to accurately model the phenological trend of the crop types.

Table V reports the classification results obtained by considering TSs of: (1) 5 synthetic Landsat-8 images produced by the proposed MGAN, (2) 5 harmonized Landsat-8 images obtained by using the baseline methods (HLS), and (3) 6

synthetic Landsat-8 images produced by the proposed MGAN. The TSs of 5 images were produced by the proposed and the baseline methods using both the original Sentinel-2 and Landsat-8 images. To generate the TS of 6 images, we considered the Sentinel-2 image acquired on 03/07/2018 for which no corresponding cloudless Landsat-8 data are available. Since no cloud-less images were acquired by the Landsat-8 sensor in July 2018 for the considered tile, no quantitative evaluation can be performed in terms of spectral distortion metrics. However, the PA%, UA%, F1% and OA% confirm

TABLE V

CROP TYPE MAPPING RESULTS OBTAINED BY CONSIDERING TSS OF: (1) 5 SYNTHETIC LANDSAT-8 IMAGES PRODUCED BY THE PROPOSED MGAN, (2) 5 HARMONIZED LANDSAT-8 IMAGES OBTAINED BY USING THE BASELINE METHOD (HLS), AND (3) 6 SYNTHETIC LANDSAT-8 IMAGES PRODUCED BY THE PROPOSED MGAN. THE TS OF 5 IMAGES WERE PRODUCED BY THE PROPOSED AND THE BASELINE METHODS USING BOTH THE ORIGINAL SENTINEL 2 AND LANDSAT 8 IMAGES. TO GENERATE THE TS OF 6 IMAGES, A SENTINEL-2 IMAGE ACQUIRED ON THE 03/07/2018 FOR WHICH NO CORRESPONDING CLOUDLESS LANDSAT-8 DATA ARE AVAILABLE WAS USED. PA IS THE PRODUCER’S ACCURACY OR RECALL, UA IS THE USER’S ACCURACY OR PRECISION, F1 IS THE F1-SCORE AND OA IS THE OVERALL ACCURACY.

Crop Type	TS of 5 images						TS of 6 images		
	Synthetic Landsat-8 (MGAN)			Harmonized Landsat-8 (HLS)			Synthetic Landsat-8 (MGAN)		
	PA%	UA%	F1%	PA%	UA%	F1%	PA%	UA%	F1%
 Grassland	96.38	88.24	92.13	95.46	76.60	85.00	96.85	90.25	93.43
 Maize	87.05	88.00	87.52	79.50	79.31	79.40	92.93	92.70	92.81
 Winter Barley	88.17	86.37	87.26	82.83	82.08	82.45	93.92	90.30	92.07
 Winter Caraway	71.00	96.60	81.84	64.50	94.16	76.56	70.00	95.89	80.92
 Rapeseed	88.25	98.46	93.08	81.57	94.91	87.74	88.71	96.49	92.44
 Beet	93.62	90.46	92.01	85.19	87.71	86.43	95.68	94.90	95.29
 Spring Cereal	77.28	80.65	78.93	65.54	83.39	73.40	82.77	87.81	85.22
 Winter Wheat	64.67	74.33	69.16	49.67	79.68	61.19	78.00	89.31	83.27
OA%	87.83			81.66			91.53		

the quality of the added image. The classification is performed by training a standard Support Vector Machine (SVM) with RBF kernels [63]. The optimal kernel parameters (i.e., the regularization parameter C and the spread of the kernel γ) were selected by a 5-fold cross-validation.

This test case demonstrates the need to densify existing TSs of satellite data. The temporal and spectral information provided by the satellite acquisition of July 2018 sharply increases the classification results by improving the modelling of the phenological trends of the considered crop types. This increases the OA% from 87.83% (TS of 5 synthetic Landsat-8 images) to 91.53 % (TS of 6 synthetic Landsat-8 images). From these results, we can conclude that the proposed MGAN can be used to generate harmonized dense TSs of Landsat-8 and Sentinel-2 images.

C. Scaling efficiency

The adoption of Horovod allowed us to distribute the training on multiple GPUs and significantly reduce the time required to complete the optimization of the model. The maximum number of GPUs used in the present work is 16, a configuration with which we obtained a speed-up of 14x on the JUWELS-BOOSTER and 12x on the DEEP-ESB partitions compared to the use of a single GPU (shown in Fig. 5). The scaling efficiency was close to 90% on the JUWELS-BOOSTER and above 75% on the DEEP-ESB partitions, respectively. In both cases, the scaling efficiency declined more steeply with 8 and more GPUs, possibly due to the increased communication time (time spent to synchronize the gradient among the GPUs) w.r.t. the computation time (time spent to optimize the model on each local GPU, which decreases proportionally to the increase of the number of GPUs, since each GPU is fed a smaller portion of the entire dataset). It can be noted that the efficiency shrinks more prominently on the

DEEP-ESB partition. This behaviour could be explained by the fact that on the DEEP-ESB partition each node is equipped with only one V100 GPU, while each node of the JUWELS-BOOSTER partition has 4 GPUs. This means that when using the DEEP-ESB partition the communication is only inter-node (the nodes are connected through InfiniBand), while on the JUWELS-BOOSTER partition the communication takes place both inter- and intra-node (faster NVLink connections). We performed 3 runs for each experiment, and the reported results are the average and standard deviation. Fig. 6 shows the training time that was reduced from 175 and more than 200 seconds using 1 GPU to 12 and 14 seconds per epoch (16 GPUs) on the DEEP-ESB and JUWELS-BOOSTER partitions, respectively. The JUWELS-BOOSTER, which features the newer A100 GPUs, allowed us to obtain a 20% increase in performances in terms of training time compared to the V100 installed on the DEEP-ESB partition.

VI. CONCLUSION

In this paper we introduced a method to densify and harmonize TSs of images acquired by Landsat-8 and Sentinel-2 satellite. The proposed method, which is based on a multispectral adaptation GAN, was applied to a TS which covers 6 acquisitions in 2018. We designed an experimental setup to validate our approach by comparing it with the well established HLS. The results obtained demonstrate that the proposed GAN is able to accurately reconstruct the spectral properties of Landsat-8 by using the Sentinel-2 images. Moreover, the qualitative comparison with the baseline method confirms the quantitative evaluation of the spectral distortion metrics. Although the physical model employed to harmonize Sentinel-2 and Landsat-8 is a powerful tool to generate long and dense TSs of optical satellite images, the proposed method achieves more accurate results from the spectral view point. Another

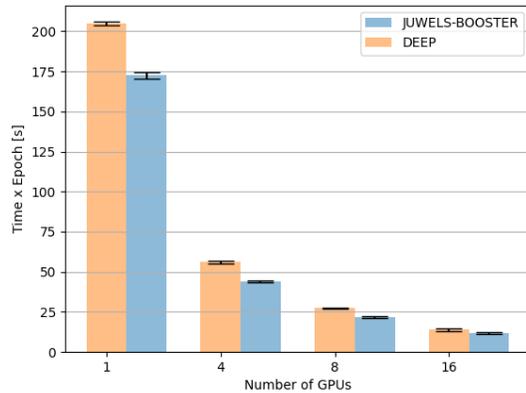


Fig. 5. Training time per epoch w.r.t. the number of GPUs on the JUWELS-BOOSTER and DEEP-ESB partitions.

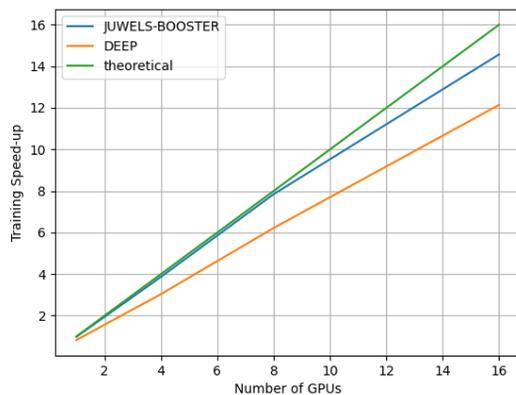


Fig. 6. Time per epoch w.r.t. the number of GPUs on the JUWELS-BOOSTER and DEEP-ESB partitions.

important result is provided by the classification accuracy obtained when considering the TS of 6 images, which allow us to test the capability of the network to accurately predict synthetic Landsat-8 images never used to train the MGAN. The OA% was increased from 87.83% (TS of 5 synthetic Landsat-8 images) to 91.53 % (TS of 6 synthetic Landsat-8 images). Moreover, we deployed the entire workflow in an HPC environment, and with the utilization of Horovod we could make an efficient use of the resources provided by such system, reducing the time required for the training of the model.

Although in this work we demonstrated that our approach can successfully densify TSs of Landsat-8 images, several challenges remain open. We focused our attention on one single region where we could validate our method also in terms of classification; however, our approach should be also extended to include different areas in the future. A strategy to ingest new data from different TSs and scale the training should be drawn up, in order to make the training of the models with larger amount of data feasible in a reasonable amount of time. Further effort should be also put on finding the

optimal hyperparameters of the training, such as the optimizers, learning rate, scheduler. Neural Architecture Search (NAS) could be employed to optimize the structure of the model, i.e., the number and type of layers, the activation functions, etc. Further loss functions should be also added, although this would significantly increase the space of the hyperparameters search, and a trade-off with available computational resources should be found. A repository with the code is available at <https://gitlab.jsc.fz-juelich.de/sedona3/mgan>.

ACKNOWLEDGMENT

Part of this work was performed in the CoE RAISE and ADMIRE projects. The CoE RAISE project received funding from the European Union’s Horizon 2020 Research and Innovation Framework Programme under the grant agreement no. 951733. The ADMIRE project received funding from the European Union’s Horizon 2020 JTI-EuroHPC research and innovation programme under the grant agreement no. 956748. The authors gratefully acknowledge the computing resources from the DEEP-EST project, which received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement no. 754304. The authors are grateful to Manil Maskey and Brian Freitag from NASA Interagency Implementation and Advanced Concepts Team (IMPACT) for the provision of the HLS datasets.

REFERENCES

- [1] Z. Zhu, “Change Detection using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 370–384, 2017.
- [2] B. Feizizadeh, T. Blaschke, D. Tiede, and M. H. R. Moghaddam, “Evaluating Fuzzy Operators of an Object-Based Image Analysis for Detecting Landslides and their Changes,” *Geomorphology*, vol. 293, pp. 240–254, 2017.
- [3] W. Zhang, X. Lu, and X. Li, “A Coarse-to-Fine Semi-Supervised Change Detection for Multispectral Images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3587–3599, 2018.
- [4] M. Meroni, D. Fasbender, F. Rembold, C. Atzberger, and A. Klisch, “Near Real-Time Vegetation Anomaly Detection with MODIS NDVI: Timeliness vs. Accuracy and Effect of Anomaly Computation Options,” *Remote Sensing of Environment*, vol. 221, pp. 508–521, 2019.
- [5] J.-F. Pekel, C. Vancutsem, L. Bastin, M. Clerici, E. Vanbogaert, E. Bartholomé, and P. Defourny, “A Near Real-Time Water Surface Detection Method based on HSV Transformation of MODIS Multi-Spectral Time Series Data,” *Remote Sensing of Environment*, vol. 140, pp. 704–716, 2014.
- [6] X. Tang, E. L. Bullock, P. Olofsson, S. Estel, and C. E. Woodcock, “Near Real-Time Monitoring of Tropical Forest Disturbance: New Algorithms and Assessment Framework,” *Remote Sensing of Environment*, vol. 224, pp. 202–218, 2019.
- [7] M. Wulder, J. Masek, W. Cohen, T. Loveland, and C. Woodcock, “Opening the Archive: How Free Data has Enabled the Science and Monitoring Promise of Landsat,” *Remote Sensing of Environment*, vol. 122, pp. 2–10, 2012.
- [8] J. Aschbacher, “ESA’s Earth Observation Strategy and Copernicus,” in *Satellite Earth Observations and Their Impact on Society and Policy*, 2017, pp. 81–86.
- [9] P. Defourny, S. Bontemps, N. Bellemans, and et al., “Near Real-Time Agriculture Monitoring at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems Around the World,” *Remote Sensing of Environment*, vol. 221, pp. 551–568, 2019.
- [10] J. Li and D. P. Roy, “A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring,” *Remote Sensing*, vol. 9, no. 9, 2017.

- [11] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.
- [12] B. Huang and H. Song, "Spatiotemporal Reflectance Fusion via Sparse Representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3707–3716, 2012.
- [13] S. Skakun, E. Vermote, B. Franch, J.-C. Roger, N. Kussul, J. Ju, and J. Masek, "Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models," *Remote Sensing*, vol. 11, no. 15, 2019.
- [14] N. Pahlevan, S. K. Chittimalli, S. V. Balasubramanian, and V. Vellucci, "Sentinel-2/Landsat-8 Product Consistency and Implications for Monitoring Aquatic Systems," *Remote Sensing of Environment*, vol. 220, pp. 19–29, 2019.
- [15] R. Raj, B. Bayat, P. Lukeš, L. Šigut, and L. Homolová, "Analyzing Daily Estimation of Forest Gross Primary Production Based on Harmonized Landsat-8 and Sentinel-2 Product Using SCOPE Process-Based Model," *Remote Sensing*, vol. 12, no. 22, 2020.
- [16] K. T. Peterson, V. Sagan, and J. J. Sloan, "Deep Learning-based Water Quality Estimation and Anomaly Detection using Landsat-8/Sentinel-2 Virtual Constellation and Cloud Computing," *GIScience & Remote Sensing*, vol. 57, no. 4, pp. 510–525, 2020.
- [17] P. Griffiths, C. Nendel, J. Pickert, and P. Hostert, "Towards National-Scale Characterization of Grassland use Intensity from Integrated Sentinel-2 and Landsat Time Series," *Remote Sensing of Environment*, vol. 238, 2020.
- [18] N. Kussul, M. Lavreniuk, A. Kolotii, S. Skakun, O. Rakoid, and L. Shumilo, "A workflow for Sustainable Development Goals indicators assessment based on high-resolution satellite data," *International Journal of Digital Earth*, vol. 13, no. 2, pp. 309–321, 2020.
- [19] D. K. Bolton, J. M. Gray, E. K. Melaas, M. Moon, L. Eklundh, and M. A. Friedl, "Continental-Scale Land Surface Phenology from Harmonized Landsat 8 and Sentinel-2 Imagery," *Remote Sensing of Environment*, vol. 240, 2020.
- [20] T. Wu, Y. Zhao, S. Wang, H. Su, Y. Yang, and D. Jia, "Improving the Accuracy of Fractional Evergreen Forest Cover Estimation at Sub-pixel Scale in Cloudy and Rainy Areas by Harmonizing Landsat-8 and Sentinel-2 Time-Series Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3373–3385, 2021.
- [21] P. D'Odorico, A. Gonsamo, A. Damm, and M. E. Schaepman, "Experimental Evaluation of Sentinel-2 Spectral Response Functions for NDVI Time-Series Continuity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1336–1348, 2013.
- [22] E. Mandanici and G. Bitelli, "Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use," *Remote Sensing*, vol. 8, no. 12, 2016.
- [23] L. Korhonen, Hadi, P. Packalen, and M. Rautiainen, "Comparison of Sentinel-2 and Landsat 8 in the Estimation of Boreal Forest Canopy Cover and Leaf Area Index," *Remote Sensing of Environment*, vol. 195, pp. 259–274, 2017.
- [24] D. Roy, H. Zhang, J. Ju, J. Gomez-Dans, P. Lewis, C. Schaaf, Q. Sun, J. Li, H. Huang, and V. Kovalsky, "A General Method to Normalize Landsat Reflectance Data to Nadir BRDF Adjusted Reflectance," *Remote Sensing of Environment*, vol. 176, pp. 255–271, 2016.
- [25] N. Flood, "Comparing Sentinel-2A and Landsat 7 and 8 Using Surface Reflectance over Australia," *Remote Sensing*, vol. 9, no. 7, 2017.
- [26] H. K. Zhang, D. P. Roy, L. Yan, Z. Li, H. Huang, E. Vermote, S. Skakun, and J.-C. Roger, "Characterization of Sentinel-2A and Landsat-8 top of Atmosphere, Surface, and Nadir BRDF Adjusted Reflectance and NDVI Differences," *Remote Sensing of Environment*, vol. 215, pp. 482–494, 2018.
- [27] R. Chastain, I. Housman, J. Goldstein, M. Finco, and K. Tenneson, "Empirical Cross Sensor Comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ Top of Atmosphere Spectral Characteristics over the Conterminous United States," *Remote Sensing of Environment*, vol. 221, pp. 274–285, 2019.
- [28] M. Claverie, J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, and C. Justice, "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Det," *Remote Sensing of Environment*, vol. 219, pp. 145–161, 2018.
- [29] S. Saunier, J. Louis, V. Debaecker, T. Beaton, E. G. Cadau, V. Boccia, and F. Gascon, "Sen2like, A Tool To Generate Sentinel-2 Harmonised Surface Reflectance Products - First Results with Landsat-8," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 5650–5653.
- [30] R. Shang and Z. Zhu, "Harmonizing Landsat 8 and Sentinel-2: A Time-Series-based Reflectance Adjustment Approach," *Remote Sensing of Environment*, vol. 235, p. 111439, 2019.
- [31] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019.
- [32] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.
- [33] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.
- [34] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.
- [35] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2017.
- [36] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced gan for remote sensing image superresolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5799–5812, 2019.
- [37] Y. Xiong, S. Guo, J. Chen, X. Deng, L. Sun, X. Zheng, and W. Xu, "Improved srgan for remote sensing image super-resolution across locations and sensors," *Remote Sensing*, vol. 12, no. 8, p. 1263, Apr 2020. [Online]. Available: <http://dx.doi.org/10.3390/rs12081263>
- [38] R. Zhang, G. Cavallaro, and J. Jitsev, "Super-Resolution of Large Volumes of Sentinel-2 Images with High Performance Distributed Deep Learning," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020, pp. 617–620.
- [39] B. Chen, J. Li, and Y. Jin, "Deep learning for feature-level data fusion: Higher resolution reconstruction of historical landsat archive," *Remote Sensing*, vol. 13, no. 2, 2021. [Online]. Available: <https://www.mdpi.com/2072-4292/13/2/167>
- [40] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network," *Remote Sensing*, vol. 12, no. 9, p. 1432, May 2020. [Online]. Available: <http://dx.doi.org/10.3390/rs12091432>
- [41] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised multiple-change detection in vhr multisensor images via deep-learning based adaptation," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Jul 2019. [Online]. Available: <http://dx.doi.org/10.1109/IGARSS.2019.8900173>
- [42] M. Fuentes Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "Sar-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits," *Remote Sensing*, vol. 11, no. 17, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/17/2067>
- [43] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Jul 2018. [Online]. Available: <http://dx.doi.org/10.1109/IGARSS.2018.8519215>
- [44] "The scientific case for high performance computing in europe 2012–2020," 2020. [Online]. Available: https://exdci.eu/sites/all/themes/exdci_theme/images/prace_-_the_scientific_case_-_full_text_.pdf
- [45] C. A. Lee, S. D. Gasster, A. Plaza, C.-I. Chang, and B. Huang, "Recent developments in high performance computing for remote sensing: A review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 3, pp. 508–527, 2011.
- [46] A. Plaza, Q. Du, Y.-L. Chang, and R. King, "High performance computing for hyperspectral remote sensing," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 4, pp. 528 – 544, 10 2011.
- [47] D. Krause, "Juwels: Modular tier-0/1 supercomputer at the jülich supercomputing centre," *Journal of large-scale research facilities JLSRF*, vol. 5, p. 135, 2019.
- [48] R. Sedona, G. Cavallaro, J. Jitsev, A. Strube, M. Riedel, and J. Benediktsson, "Remote sensing big data classification with high performance distributed deep learning," *Remote Sensing*, vol. 11, no. 24, p. 3056, Dec 2019. [Online]. Available: <http://dx.doi.org/10.3390/rs11243056>

- [49] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.
- [50] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [51] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," 2017.
- [52] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018.
- [53] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 63–79.
- [54] A. Jolicœur-Martineau, "The relativistic discriminator: a key element missing from standard gan," 2018.
- [55] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable augmentation for data-efficient gan training," 2020.
- [56] T. Ben-Nun and T. Hoefler, "Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis," *ACM Computing Surveys*, 2019.
- [57] A. Sergeev and M. Del Balso, "Horovod: Fast and Easy Distributed Deep Learning in TensorFlow," *arXiv preprint arXiv:1802.05799*, 2018.
- [58] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterton, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. [Online]. Available: <http://proceedings.mlr.press/v9/glorot10a.html>
- [59] E. Suarez, N. Eicker, and T. Lippert, *Modular Supercomputing Architecture: from Idea to Production; 3rd*. CRC Press, 2019, vol. 3, pp. 223–251.
- [60] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, p. 133–142, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.aqpro.2015.02.019>
- [61] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, p. 3012–3021, Oct 2007. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2007.904923>
- [62] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [63] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on geoscience and remote sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.



models on multiple GPUs of High Performance Computing (HPC) systems.

Rocco Sedona (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in information and communications engineering from the University of Trento in 2016 and 2019, respectively. He is member of the "High Productivity Data Processing" (HPDP) research group at the Jülich Supercomputing Centre, Germany. He is currently pursuing the Ph.D. degree in computational engineering at the University of Iceland. His research interest is mainly in machine learning methods for remote sensing applications, with a particular focus on distributing deep learning



Claudia Paris (M'16) received the B.S. degree and M.S. (summa cum laude) degree in Telecommunication Engineering and the Ph.D. in Information and Communication Technology from the University of Trento, Italy, in 2010, 2012, 2016, respectively. She accomplished the Honors Master Program in Research within the Master Degree in Telecommunication Engineering in 2012. Since 2014 she is a teaching assistant at the Department of Information Engineering and Computer Science of the University of Trento, Italy, where she is currently Assistant

Professor. Her main research includes image and signal processing, machine learning and deep learning with applications to remote sensing image analysis. The main research interests include remote sensing single date and TSs image classification, land cover map update and fusion of multisource remote sensing data for the estimation of biophysical parameters. She conducts research on these topics within the frameworks of national and international projects. Dr. Paris won the very prestigious Symposium Prize Paper Award (SPPA) at the 2016 International Symposium on Geoscience and Remote Sensing (Beijing, China, 2016) and at the 2017 International Symposium on Geoscience and Remote Sensing (Fort Worth, Texas, USA, 2017). She has been a member of the program and scientific committee of several international conferences and workshops.



Gabriele Cavallaro (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Trento, Italy, in 2011 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Iceland, in 2016. He is currently the deputy head of the "High Productivity Data Processing" (HPDP) research group at the Jülich Supercomputing Centre, Germany. He is also the chair of the High-Performance and Disruptive Computing in Remote Sensing (HDCRS) Working

Group of the IEEE GRSS ESI Technical Committee.

He was the recipient of the IEEE GRSS Third Prize in the Student Paper Competition of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015 (Milan - Italy). His research interests cover remote sensing data processing with parallel machine learning algorithms that scale on high performance and distributed systems. He serves on the scientific committees of several international conferences and he is a referee for numerous international journals. Since 2019 he gives lectures on scalable machine learning for remote sensing big data at the Institute of Geodesy and Geoinformation, University of Bonn, Germany.



Lorenzo Bruzzone (Fellow Member, IEEE) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications. Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory (<https://rslab.disi.unitn.it/>) in the Department of Information Engineering and

Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, signal processing, machine learning and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is currently the Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the Jupiter ICy moons Explorer (JUICE) mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He is the author (or coauthor) of 294 scientific publications in referred international journals (221 in IEEE journals), more than 340 papers in conference proceedings, and 22 book chapters. He is editor/co-editor of 18 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 37000) and the value of the h-index (89) (source: Google Scholar). He was invited as keynote speaker in more than 40 international conferences and workshops. Since 2009 he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS), where since 2019 he is Vice-President for Professional Activities. Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards and the 2019 WHISPER Outstanding Paper Award. Dr. Bruzzone was a Guest Co-Editor of many Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003 he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the founder of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE for which he has been Editor-in-Chief between 2013-2017. Currently he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society between 2012-2016. He is a Fellow of IEEE.



Morris Riedel (Member, IEEE) received his PhD from the Karlsruhe Institute of Technology (KIT) and worked in data-intensive parallel and distributed systems since 2004. He is currently a Full Professor of High-Performance Computing with an emphasis on Parallel and Scalable Machine Learning at the School of Natural Sciences and Engineering of the University of Iceland. Since 2004, Prof. Dr. - Ing. Morris Riedel held various positions at the Juelich Supercomputing Centre of Forschungszentrum Juelich in Germany. In addition, he is the Head

of the joint High Productivity Data Processing research group between the Juelich Supercomputing Centre and the University of Iceland. Since 2020, he is also the EuroHPC Joint Undertaking governing board member for Iceland. His research interests include high-performance computing, remote sensing applications, medicine and health applications, pattern recognition, image processing, and data sciences, and he has authored extensively in those fields. Prof. Dr. - Ing. Morris Riedel online YouTube and university lectures include High-Performance Computing – Advanced Scientific Computing, Cloud Computing and Big Data – Parallel and Scalable Machine and Deep Learning, as well as Statistical Data Mining. In addition, he has performed numerous hands-on training events in parallel and scalable machine and deep learning techniques on cutting-edge HPC systems.