IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY SECTION

Received 1 August 2024, accepted 26 August 2024, date of publication 4 September 2024, date of current version 13 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3454513

RESEARCH ARTICLE

Multitemporal SAR-to-Optical Image Translation Using Pix2Pix With Application to Vegetation Monitoring

DONATO AMITRANO, (Senior Member, IEEE)
Italian Aerospace Research Centre, 81043 Capua, Italy
e-mail: d.amitrano@cira.it

This work was supported by the Italian Ministry of University and Research under the Aegis of the Italian Aerospace Research Program "PRORA-Bioscience."

ABSTRACT Information representation enhancement in synthetic aperture radar images is highly debated in remote sensing literature. In the past, the most credited solutions were based on traditional image processing integrating microwave scattering principles. More recently, neural network-based solutions conquered the scene, allowing for the introduction of representations obtained via domain translation. In this context, generative adversarial networks rule the roost. However, after an initial improvement, the performance of the newly developed solutions reached a plateau. This paper proposes a change of perspective, moving the translation problem from the architecture to input data and training modality. It will be shown that state-of-the-art multitemporal synthetic aperture radar processing provides products that, due to their enhanced texture and colorimetric attributes, appear more similar to their optical correspondent. Moreover, training constrained by spatial and temporal features is beneficial to increase the phenomenological correspondence between the radar reflectivity function and the terrain reflectance. As a result, the obtained translated products show a significant increase of the standard image quality parameters. Finally, the exploitability of multimodal products is demonstrated with an application concerning the estimation of the normalized difference vegetation index, which shows the comparability of the synthetic index with that calculated from native optical data.

INDEX TERMS Synthetic aperture radar, multispectral, generative adversial networks, multimodal image translation, vegetation monitoring.

I. INTRODUCTION

Passive remote sensing is a powerful tool for large scale Earth monitoring. Thanks to recent advances in sensor technology, it can provide very high spatial resolution images. The presence of a number of satellites guarantees global coverage with high revisit time. However, the limitations of this technology are well-known. The dependence on illumination and weather conditions are limiting factors in some scenarios like emergencies or those needing continuous monitoring of a specific area of interest. A statistical analysis on Terra and Aqua sensors acquisitions revealed that the 67% of the Earth surface is, on average, covered by clouds [1].

In this regard, synthetic aperture radar (SAR) sensors provide a valid alternative. Being active instruments, they are able to generate their own power source. This makes them independent from the availability of an external illumination source. Moreover, operating at microwaves, they are almost independent from weather conditions [2]. These characteristics make SAR sensors very attractive from the viewpoint of the acquisition capability, which can be assumed as "all weather and all time" but represent, at the same time, a limitation to their full exploitation.

Passive remote sensing typically offers data acquired in the spectra of the visible, near-infrared (which is the domain

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni.

of the chlorophyll), and far infrared (which is related to the emissivity and thus to the surface temperature). Most of remote sensing scientists and end-users are familiar with these frequencies. SAR platforms, conversely, operate at microwave frequencies, typically included between 1 and 10 GHz. These acquisitions are strongly influenced by the surface roughness [3]. The peculiar image formation mechanism makes data difficult to interpret by standard users of satellite images. This makes the technology underexploited against its real potential [4].

The problem of the enlargement of the SAR community and, in general, of the interpretation of satellite images has been widely debated in literature [4], [5], [6], [7], [8], [9], [10]. Indeed, it assumes different meanings depending on the specific sensor at hand. As an example, dealing with hyperspectral data, the topic is how to reduce the dimensionality of data in order to display meaningful information on a standard tristimulus RGB device [11], [12], [13]. Working with SAR, proposed solutions aim at increasing the exploitability through improved pre-processing and/or multitemporal data fusion techniques [4], [8], [14], [15]. More recently, the progress in artificial intelligence made available new tools allowing for the enlargement of the available range of solutions for data enhancement, like super-resolution [16] or multimodal image translation [17].

The latter has been initially conceived for the transformation/modification of consumer images. For example, Reference [18] showed how to change the style of a photograph using deep convolutional neural networks. In this work, famous paintings are combined with everyday pictures to create original artworks. The concept of image-to-image translation can be applied to several of situations, from dawn to dusk scene conversion [19] to the generation of portraits from sketches [20].

In the past, each of these tasks has been tackled with a special-purpose algorithm. Today, the problem has been unified under a unique class of algorithms known as generative adversarial networks (GANs). Firstly introduced in [21], they are designed in such way to produce, ideally, a synthetic output indistinguishable from the original input [20]. GANs are constituted by a pair of networks trained simultaneously and in competition with each other, i.e., the generator and the discriminator. A common analogy applicable to visual data is thinking of the generator as an art forger and the discriminator as an art expert. The forger creates forgeries to make realistic images. The expert tries to separate real artworks from fake ones [22].

In the field of remote sensing, these concepts have been recently exploited to translate SAR images into the optical domain. Reference [23] proposed an improved cycle-consistent GAN framework in which the standard loss function is integrated with the mean square error (MSE) loss for better performance. A similar approach has been adopted by [24]. In this case, the standard loss function has been integrated with the structure similarity index measure (SSIM) [25]. Reference [26] introduced an improved conditional GAN method to enhance the contour sharpness, the texture rendering and the colour fidelity in translated images. Reference [27] proposed the exploitation of CycleGAN architectures to tackle with multimodal images registration issues. A comparative study of different methodologies for SAR-to-optical image translation has been proposed by [17]. In this work, the authors reported that, independently from the adopted loss function, the quality of the reconstruction tends to get worse as the texture content of the scene increases. As a result, the reconstruction of natural scenes is, on average, quite true to reality. Conversely, urban areas are more affected by artifacts, as testified by the decrease in the adopted image quality indicators.

In the last years, the research on the topic has been boosted by the release of the SEN1-2 dataset [28], which, providing couples of SAR and optical images, constitutes a fruitful environment for the development of new translation techniques. This work aims to introduce a new perspective in SAR to optical image translation. In fact, so far the literature pursued the performance improvement by developing new architectures and/or through fine-tuning of literature network loss functions using as input single-look SAR images. However, as reported in [17], the performance advance measured via standard indicators is quite limited. In other words, nevertheless the abundant literature published on the topic, the quality of the generated output seems to have reached a plateau.

This evidence makes some questions rising about the development of new nets derived from major architectures or the research of the best parametrization. It is reasonable to think that the improvement of generated images, which is a key to enable new applications based, for example, on hybrid change detection models [29], should be pursued following alternative ways. like a better management of the training phase and the ingestion of an input more suitable for translation, i.e., closer to the target. The shift of the focus from the architecture to such concepts is new in the literature and constitutes the principal novelty of the work.

To this end, SAR data are opportunely pre-processed using state-of-the-art multi-temporal concepts [4] to make the texture and the chromatic content of images more similar to their optical correspondent. The training is constrained by spatial and temporal features to increase the phenomenological correspondence between the radar reflectivity function and the terrain reflectance. Data are processed using a literature network, run with standard parameters, with no actions taken on the original loss function. It will be shown that this change of perspective significantly increases the performance metrics against typical literature values. Finally, the exploitability of multimodal products is demonstrated with an application concerning vegetation monitoring, in which it will be shown that a synthetic vegetation index can reasonably approximate the estimate obtained from native optical data.

The works is organized as follows. The general methodology is introduced in Section II. Experimental results are presented and discussed in light of the recent literature in

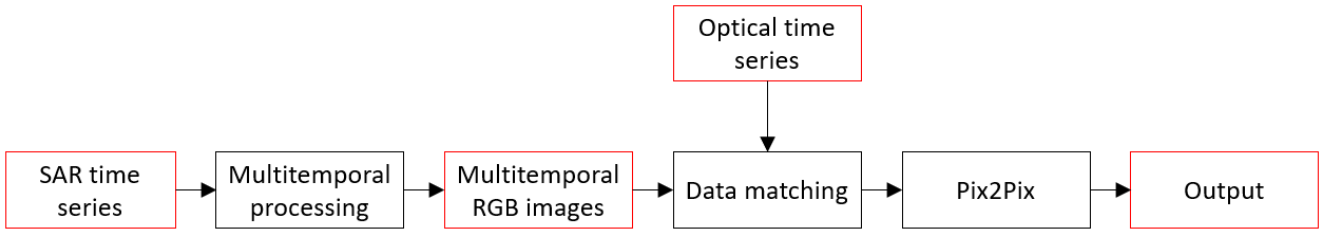


FIGURE 1. General workflow of the proposed multimodal image translation methodology. Blocks with red edges identify input/output data.

Section III. Vegetation monitoring using multimodal products is presented in Section IV. Conclusions are drawn at the end of the work.

II. METHODOLOGY

The general workflow of the proposed methodology is depicted in Figure 1. It mainly exploits techniques already known in literature, introducing novelties in the way they are fed and managed. As stated in the introduction, the focus is moved from the network architecture to input products, which are pre-processed to make them more similar to target data, thus favoring the translation process. Moreover, the training phase is optimized according to the spatio-temporal similarity of the input, which allows for tackling the differences between the SAR backscattering and the optical reflectance.

As shown in the diagram, the starting point is a time series of SAR data. They are pre-processed according to state-of-the-art multi-temporal concepts [4] to generate the data to be ingested within the GAN [20] through the extraction of geocoded patches. The second branch of the workflow involves optical/multispectral data acquired over the same area. They are pre-processed for atmospheric and radiometric correction and cut for the generation of patches suitable to be the targets for the training phase. The maximum time span allowed between SAR and optical patches of the same area is 5 days.

Such multimodal data constitute the input for the GAN, which is trained to learn the optical representation from SAR data to implement the domain translation. The output of the workflow is an optical image obtained from the original SAR one.

In the following, each processing block will be discussed in detail.

A. MULTI-TEMPORAL SAR PROCESSING

The purpose of the first part of the workflow is the generation of input products for image domain translation. To this end, a time series of SAR images is considered. It is subject to state-of-the-art pre-processing to make data suitable for ingestion in the training phase. In other words, the objective of the processing is to make the appearance of SAR data closer to their optical correspondent at both texture and chromatic content. This is achieved by exploiting change-detection principles allowing for the enhancement of the correspondence between the SAR reflectivity function and the optical reflectance [30].

Multi-temporal SAR pre-processing accounts for radiometric calibration, geometric co-registration, multi-temporal despeckling, time-series cross-calibration, multi-temporal composites generation and geocoding [4]. All the SAR data exploited in this work have been acquired by the Sentinel-1 constellation.

Radiometric calibration allows for compensating effects due to the orbit and/or the sensor [31]. It is performed by using calibration coefficients available in the product metadata. Radiometric calibration makes the data of the time series comparable. This means that, after calibration, the same object, imaged at different times, will exhibit the same reflectivity value, if not subject to changes.

Coregistration is the process of geometric alignment of images. It exploits both orbital information stored in the metadata and data-driven processing [32]. After coregistration, each pixel of each image of the time series will correspond to the same point on the Earth's surface, even if still in SAR geometry.

For the purpose of image translation, despeckling is probably the most important processing step. Speckle is a characteristic of SAR images. It is due sub-resolution elements within the resolution cell causing a random reflection [32]. This means that, given two areas macroscopically identical, they may exhibit a different reflectivity. This phenomenon causes the typical salt-and-pepper appearance of data, which corrupts the scene texture preventing the correct training of the neural net responsible for the translation.

Indeed, speckle is a phenomenon non-existent in optical images, which makes the texture of SAR images very different against their optical correspondence. However, it can be mitigated by the application of proper despeckling techniques [33]. Working with time series allows for the exploitation of spatio-temporal despeckling [34], like the algorithm developed in [35] used in this work.

The availability of multi-temporal data allows for building RGB composites, enhancing the information carried by the single acquisition in a change-detection framework [4], [8]. To this end, time-series cross-calibration must be implemented to re-quantize data in an 8-bit format suitable for display on standard tristimulus devices. The methodology proposed by [4] is implemented in this work. It consists of a histogram clip guided by a reference image, which is assumed

to be the one exhibiting the minimum of the maxima. The reflectivity value corresponding to the 98% of the cumulative histogram of the reference image is assumed as a clipping parameter. The choice of the 98th percentile is the best compromise between the percentage of saturated pixels and the preservation of the image entropy, which is a measure of the image information content [36].

Once the metrics of the time-series is established, data can be combined in order to obtain photo-realistic images. According to [5], all the composites belonging to the time series share the red channel, which plays as reference image. It is convenient to choose such an image when the vegetation is at its minimum. In such ways, loading one by one the other images on the green channel, an increasing vegetation content in the scene is associated with an increase of the green color contribution. This is due to the backscattering enhancement triggered by the growth of vegetation canopies [37]. Finally, the blue channel is reserved to a texture measure, i.e., the data range within a box of fixed size. This is useful to highlight high-texture areas like urban areas [5].

The last step of SAR pre-processing is geocoding. It is necessary to re-project data from the SAR geometry into a cartographic system [32]. This is crucial for the alignment with the optical data and the ingestion into the GAN training pipeline.

B. OPTICAL DATA PRE-PROCESSING AND PAIRING WITH THE SAR DATASET

The pre-processing of optical data accounts for atmospheric corrections. In this work, the exploited optical data have been acquired by the Sentinel-2 constellation. In particular, Level-2A images have been used. They are pre-processed products provided in bottom of atmosphere reflectance format thanks to the application of the Sen2Cor processor [38].

Optical images are paired and overlapped with the SAR dataset and cropped for the extraction of relevant patches for training the subsequent GAN. The pairing is made using images acquired with a maximum time span of 5 days to minimize the possibility of changes occurring in the scene. Similarly to SAR data, images are converted to 8-bit format via clipping at the 98% of the cumulative histogram and rescaled in the interval [0, 255].

From each geocoded pair, a number N = 200 of patches is extracted for ingestion into the GAN training pipeline. In Figure 2, an example of the data used for the training is shown. They concern the city of Edmonton, Canada. SAR data have been extracted from a dataset of about 60 images. The compositions shown in Figure 2a and Figure 2c share the reference image, acquired in January 2019 and loaded on the red band. The green channel has been acquired in August 2020. As previously discussed, the purpose of SAR pre-processing is to make data as much as possible photo-realistic. However, they shall be evaluated from a change-detection perspective, being the result of the fusion of two acquisitions. For example, buildings are rendered in yellow or white colours depending on the low or high texture

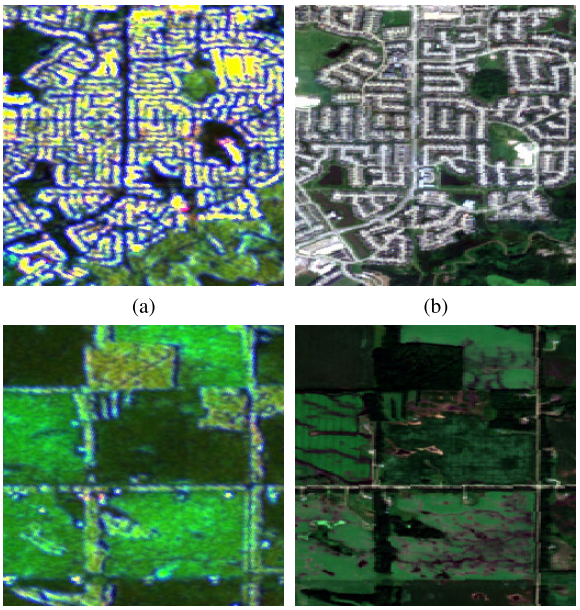


FIGURE 2. Examples of patches ingested within the GAN. Area with high urban texture: (a) Multi-temporal RGB SAR image and (b) correspondent optical image. Rural area: (c) Multi-temporal RGB SAR image and (d) correspondent optical image.

of the area due to the high contribution of both the reflectivity channels [39]. Areas with significant green contribution are characterized by increasing backscattering due to the growth of the vegetation against reference conditions. A balanced contribution of the two reflectivity channels is representative of stable land-cover.

The crucial aspects of the SAR pre-processing are the mitigation of the speckle and the introduction of a stable colour rendering of the land cover. As shown in the samples reported in Figure 2a and Figure 2b and in Figure 2c and Figure 2d, this makes SAR data closer to their optical correspondent. This is expected to enhance the training phase.

C. GAN BASICS AND TRAINING

In this Section, the fundamentals of GANs will be briefly recalled omitting most of the mathematical foundations of these architectures. The readers interested in such details are referred to dedicated works [20], [21], [40].

In recent years, the scientific community has been increasingly interested in GANs due to their ability to manage large quantities of unlabeled data. The classic architecture of a GAN is constituted by two competing nets [21]. As aforementioned, the two nets replicate the dualism between a forger and an art expert [22]. Referring to Figure 3, the forger, i.e., the generator G creates a copy y with the objective of making it as much as possible similar to the source x. The expert, i.e., the discriminator D, examines both the original x and the copy y trying to understand whether the artwork is fake. The two nets are trained simultaneously and compete with each other. The generator does not have direct access to real data. This means that the training of this net is

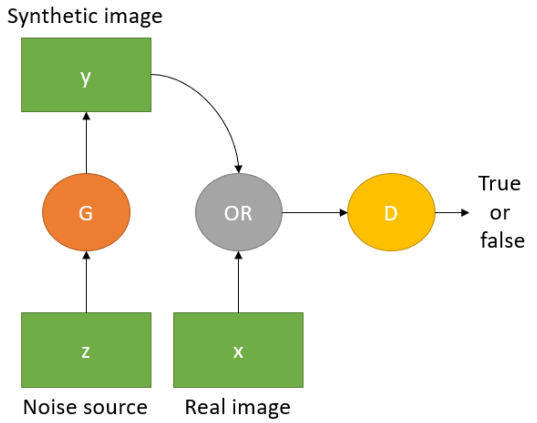


FIGURE 3. Basic architecture of a GAN.

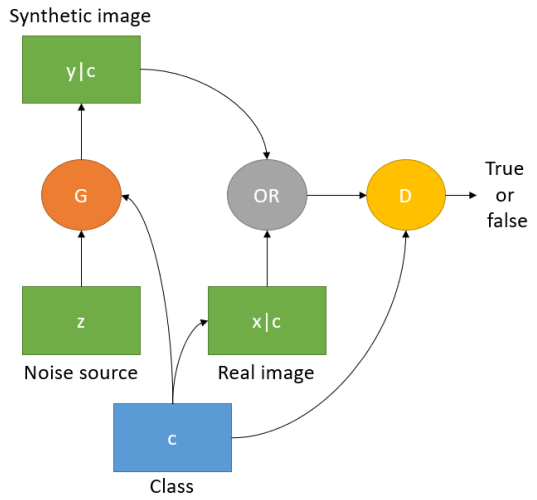


FIGURE 4. Basic architecture of a cGAN.

related only to interactions with the discriminator that, on the other hand, can access to both original and synthetic data. The error signal provided to the discriminator is represented by the truth about the correct decision and is used to train the generator, making it able to create synthetic data of better quality.

In this work, the translation from SAR to optical has been implemented through a conditional GAN (cGAN) [41]. As shown in FIGURE 4, in this setting, both the discriminator and the generator are class-conditional. According to [22], they provide better representations for multimodal data generation. Therefore, they are particularly well-suited for the problem at hand.

The architecture exploited is the one known as Pix2Pix, presented in [20]. It is a conditional image-to-image translation architecture using a conditional GAN objective combined with a reconstruction loss. The conditional GAN objective for observed images x, output images y, and the random noise vector z can be expressed as

$$\mathcal{L}_{GAN} = \mathbb{E}_{x,y}\left[\log D\left(x, y\right)\right] + \mathbb{E}_{x,z}\left[1 - D\left(x, G(x, z)\right)\right]. \quad (1)$$

The discriminator is trained like in a traditional GAN architectures, i.e., minimizing the negative log likelihood of identified real and fake images. The generator is trained using both the adversarial loss for the discriminator model and the L1 loss between the generated translation and the expected target image. In other words, G tries to minimize the objective expressed in Equation 1 against an adversarial D that tries to maximize it, i.e.

$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_{L1}(G). \quad (2)$$

where

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}\left[||y - G(x, z)||_1\right] \quad (3)$$

The combination of the L1 loss with the adversarial loss is controlled by the hyperparameter λ, which is set to 10. It determines the importance of the L1 loss against the adversarial loss during the training of the generator.

As stated in [20], both the generator and the discriminator use modules of the form convoluton-BatchNorm-ReLu [42]. The network optimization follows the standard approach proposed in [21]. The used solver is Adam [43] with a learning rate of 0.0002 and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$.

The network has been trained using SAR/optical pairs with a maximum time span of 5 days to minimize the possibility of abrupt changes within the scenes. Nevertheless, the major problem to tackle in the translation task is the ambiguity of the SAR response against optical land covers. Indeed, the SAR reflectivity function depends on a number of factors like the roughness, the dielectric constant, the orientation of the scatterers and the occurrence of volumetric scattering phenomena [2], [32]. They do not have a unique optical correspondent. In other words, the same SAR reflectivity value could correspond to different optical land covers and vice-versa, thus causing the the generator to fail.

The idea is to reduce the variety of optical land covers by dividing the training problem into more sub-tasks with higher spatiotemporal consistency. In other words, the proposal is not to look for a ''general GAN'' able to translate images whenever and wherever acquired. Rather, the concept of ''zonal GAN'' is introduced, in which the net is trained with data acquired on a specific area, i.e., a specific city and in a limited time frame, like a specific month. Due to the intrinsic difference between the data at hand, this is useful to reduce the variety of texture and land covers to be translated. This means that the translation of the area of interest will be performed using several GANs, each one representative of a month. Due to the cyclic nature of land covers, the temporal consistency is preserved through the years.

D. DATA

The data used in this study have been acquired under the aegis of the Copernicus Programme by the constellations Sentinel-1 (SAR) and Sentinel-2 (optical). Input SAR data are in ground range detected (GRD) format. As suggested in

the past related literature [26], [44], [45], all the data have been acquired in the same configuration in terms of orbit and incidence angle in order to avoid backscattering differences on targets of the same nature due to the different imaging geometry.

Target multispectral data are in atmospherically corrected L2A format. All the images have 10-meter spatial resolution. As discussed above, they are converted to 8-bit format before ingestion in the neural network.

The locations for the test of the proposed methodology have been selected to diversify the land cover. To this end, two urban and two rural sites have been chosen. The formers are represented by two Canadian cities, i.e., Red Deer and Edmonton. The latters concern agricultural areas in Dodge City (US) and Wurzburg (Germany). In all cases, images were acquired during years 2018 to 2021.

III. EXPERIMENTAL RESULTS

A. PERFORMANCE METRICS

The assessment of the obtained results has been performed using literature image quality indicators, like the root mean square error (RMSE), the peak signal-to-noise ratio (PSNR), and the SSIM [16].

Indicating with x the luminance of the original image and with y that of the translated one, the first indicator exploited, the RMSE, is obtained from the square root of MSE, i.e.,

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-y)^2}, \tag{4}$$

where N is the total number of pixels. The luminance is calculated according to the following relation

$$I = 0.299\,R + 0.587G + 0.114B, \tag{5}$$

in which R, G and B are the intensities of the red, green and blue channels. The RMSE is a measure of the Euclidean distance between the original and the translated image evaluated pixel-wise and, therefore, an indicator of the fitting of the translated image with the original one. It has been preferred to the MSE as it provides values of the error measured in the same unit of the value to be predicted, i.e., digital numbers (DNs).

The second indicator, the PSNR, is defined as the ratio between the maximum signal power and the noise power. Mathematically, the following relation holds [16]

$$PSNR = 20\log_{10}\left(\frac{\max(y)}{\sqrt{MSE}}\right). \tag{6}$$

The PSNR is relatively simple to calculate and has a clear physical meaning. However, as the RMSE, it does not introduce human visual system characteristics into the image quality evaluation because the differences are analyzed purely from a mathematical perspective. In other words, although the PSNR and the RMSE remain accepted evaluation metrics, they cannot capture the differences in visual perception [16].

To this end, the SSIM index is exploited. It is defined as follows [25]

$$SSIM = \frac{4\sigma_{xy}\mu_x\mu_y}{\left(\sigma_x^2+\sigma_y^2\right)\left(\mu_x^2+\mu_y^2\right)}, SSIM \in [-1,1], \tag{7}$$

where

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{8}$$

$$\mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{9}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2, \tag{10}$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2, \tag{11}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)\left(y_i-\mu_y\right). \tag{12}$$

The SSIM is a perceptual metric and a comprehensive measure of similarity between images from three aspects, including structure, brightness, and contrast. Therefore, the SSIM can measure either the degree of distortion of a picture or the degree of similarity between two pictures. This makes the SSIM index more in line with human eye perception [16].

B. RESULTS AND DISCUSSION

The results of the application of the proposed workflow are reported in Figure 5 and Table 1.

In particular, the graphics reports patches arranged as follows. The rows are relevant to different sites. The columns reports input or output data. Specifically, the first row of Figure 5 concerns the Red Deer site. The second, the Edmonton site. The third, the Dodge City site. The fourth, the Wurzburg site.

The column organization of Figure 5 is as follows. The first one is reserved to sample SAR patches. The second reports the correspondent optical ones. The third column is the translated multimodal image. The last column depicts the RMSE map. For the latter, the colormap spans from cool colors to hot ones, so blue areas are characterized by an almost null RMSE, while red ones are those with exhibiting the higher values.

Qualitatively, it is remarkable that multimodal products preserve both shapes and color with a good degree of approximation. The textures of dense urban areas are respected after translation as well as the border of agricultural fields. Looking at the RMSE maps, there are limited areas with insurgence of significant errors due to incorrect color rendering. As a general comment, most of the errors are located in correspondence with sharp edges, which are slightly distorted by the domain translation.

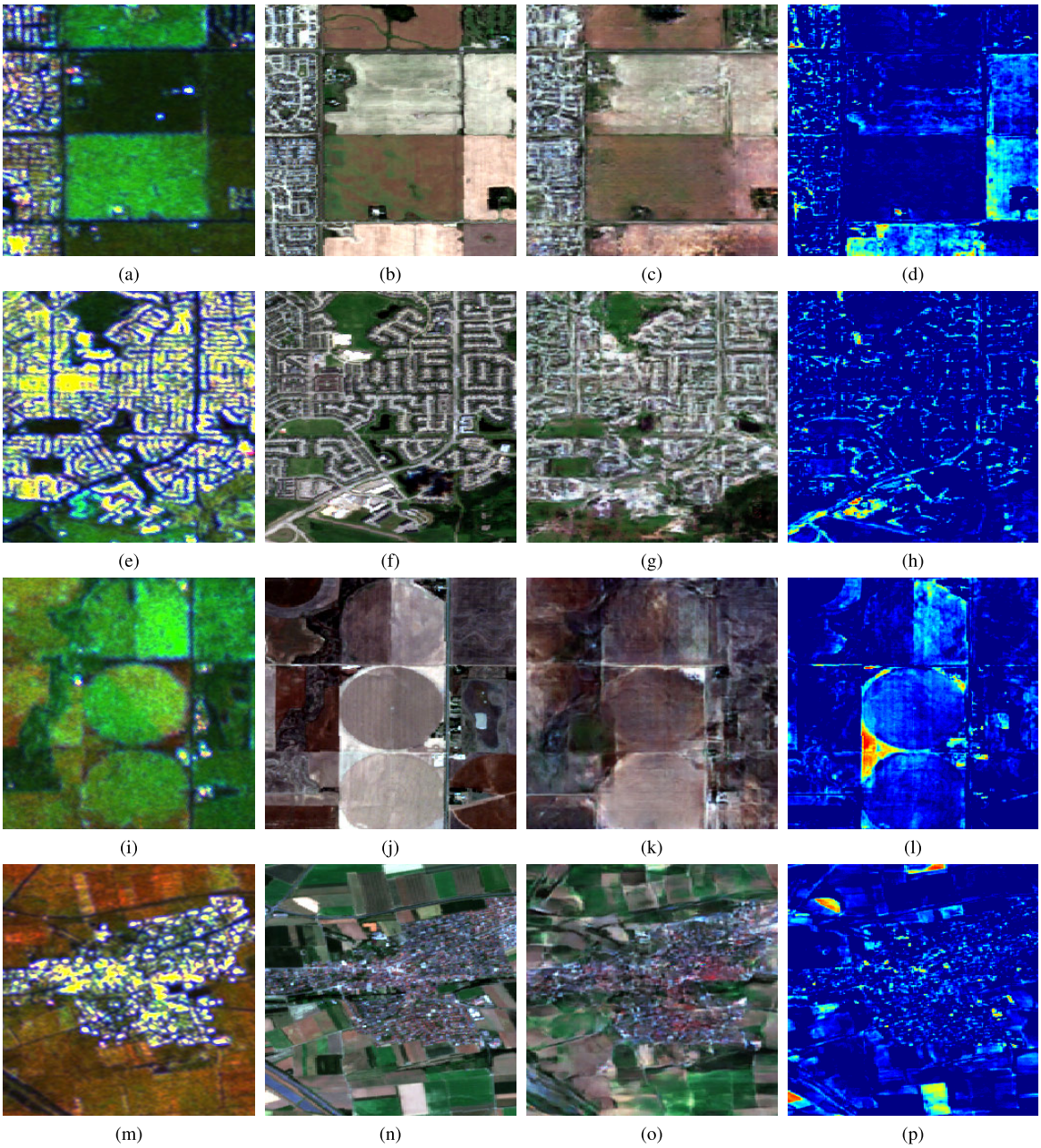


FIGURE 5. Examples of GAN output. First row: Red Deer dataset. Second row: Edmonton dataset. Third row: Dodge city dataset. Fourth row: Wurzburg dataset. First column: SAR patch. Second column: optical patch. Third column: translated image. Fourth column: RMSE.

Quantitative results of the application of the proposed methodology are reported in Table 1. As for the PSNR metric, it is quite stable, with a peak of 29.2 registered for the Wurzburg dataset (month of July) and an average value of 28.4.

The same holds for the RMSE measure, which ranges between 8.99 registered for the Wurzburg dataset (month of July) and 10.5 obtained for the Dodge City site (month of October). The registered average value is 9.73.

As for the SSIM index, higher fluctuations are observed. Its values range from 0.386, calculated for the Edmonton dataset (month of October), to 0.674, obtained using the Wurzburg dataset (month of August). The registered average value is 0.564.

In Table 1, the results of the application of the standard workflow [17] are also reported. In this case, the translation has been implemented using the Pix2Pix network without any substantial pre-processing on SAR data, apart 8-bit rescaling through histogram clip and geocoding. A general digradation of the image quality parameters is observed, although those concerning the PSNR and the RMSE are less significant. The substantial change in the performance concerns the

TABLE 1. Multimodal image translation: quantitative results for the different datasets and comparison with standard approach.

			Proposed			Standard		
Site	Month	Patches	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
Urban								
Red Deer	June	849	28.7	0.647	9.51			
Red Deer	August	672	28.6	0.631	9.50	28.5	0.157	9.77
Red Deer	September	471	28.6	0.637	9.55			
Edmonton	August	627	28.5	0.467	9.68	28.8	0.236	9.49
Edmonton	October	438	27.8	0.386	10.4			
Rural								
Dodge City	June	481	28.0	0.575	10.2	27.7	0.180	10.6
Dodge City	October	1196	27.7	0.466	10.5			
Wurzburg	July	1219	29.2	0.599	8.99	27.6	0.298	10.7
Wurzburg	August	852	28.9	0.674	9.24			
Mean			28.4	0.564	9.73	28.2	0.217	10.1

SSIM parameter, that is more than halved using the standard methodology. In fact, the MSE and the PSNR are measures of the absolute errors, while the SSIM is more sensitive to structural differences. As reported by [46], the MSE, and so the PSNR, show low performance in the discrimination of the structural content of images since diverse degradations applied to the same image can result in the same MSE. This means that, using the standard workflow, nevertheless similar value of absolute error metrics, the drop in SSIM values testify a significant texture degradation.

The obtained results can be interpreted also in light of the related recent literature. The results reported in some selected papers are summarized in Table 2. It is worthwhile to remark that this Table is not exhaustive, as the literature on the topic is wide. Consequently, only few papers have been selected for explicit comparison, in order to provided numbers to support the discussion. However, similar values can be found in many other works [23], [24], [47], [48], [49], [50], [51], [52]

From this Table, it arises that, although the change of the architecture used for the translation, the obtained values for the performance indicators are quite flat. For example, the reported PSNR is in most cases included in the range 16-19, thus about 30% lower than that obtained using the proposed approach. Using the Pix2Pix network, i.e., the same one used to implement the proposed workflow, the results obtained by the examined literature range between 17.4 and 17.8.

Considering the SSIM index, the same considerations can be replicated. The values reported by the examined papers range between 0.3 and 0.4 in most of cases, with a worsening of about 50% against the value obtainable using the proposed workflow.

The RMSE, indeed, is not adopted in the majority of the literature on the topic, although it provides an easy-to-understand measure of the distance between the original image and the translated one. In fact, it is worthwhile to note that this metric is expressed in the same unit of the data it refers to, in this case DNs. Nevertheless, in most of the reviewed papers, the MSE, whose unit of measure is the square of the unit of measure of the data, is used for the purpose. Moreover, it is observed a normalization of the data against their maximum [45], which makes the numbers difficultly interpretable. However, when expressed, the RMSE is about 30% than that obtained by applying the proposed methodology.

Concerning the loss function, some indications about its impact on the quality of the translated image can be found in [45] and [50]. In the first work, the authors tested two different loss functions within their proposed architecture. In particular, they proposed to use the L_1 loss, which is the one utilized in the standard Pix2Pix network, in combination with a perceptual loss [53] and with a perceptual loss and the SSIM loss. As a result, they obtained that the second option was the best performing and determined an increase of the PSNR and of the SSIM index from 13.2 to 13.5 and from 0.292 to 0.354, respectively. Similarly, Guo et al. [45] tested the architecture proposed in their work with different loss functions, obtaining variations of the PSNR and of the SSIM index between 18.1 and 19.4 and between 0.393 and 0.444, respectively.

The above numbers and considerations allow for arguing that i) the variation of the architecture exploited for translation, as well as modifications of the loss function, has a negligible impact on the standard translation quality indicators, especially concerning the PSNR; ii) the results obtained in literature are quite similar although different input datasets, in which the SAR part is provided in single-look format, are exploited.

In summary, the proposed methodology, using a standard architecture, allows for significantly improving the values of the translation quality indicators. This is obtained through i) optimization of input SAR data that, thanks to multitemporal processing, exhibit a textural and chromatic content closer to their corresponding optical representation and ii) a training constrained by spatial and temporal features, with the

TABLE 2. Multimodal image translation: quantitative results extracted from the recent literature.

Reference	Dataset	Technique	PSNR	SSIM	RMSE
Yang *et al.* [26]	Sen1-2	ICGAN	18.8	0.395	na
		Pix2Pix	17.4	0.319	na
		CycleGAN	12.6	0.175	na
		NiceGAN	11.6	0.187	na
Wei *et al.* [44]	QXS-SAROPT	Pix2Pix	18.4	0.360	31.4
		CycleGAN	17.1	0.367	37.6
		S-cycleGAN	18.2	0.399	32.1
		NiceGAN	14.6	0.347	47.9
		GANILLA	16.4	0.159	38.9
		CFRWD-GAN	19.0	0.468	33.7
Guo *et al.* [45]	Sen1-2	Pix2Pix	17.8	0.345	na
		CycleGAN	17.5	0.370	na
		S-cycleGAN	18.3	0.385	na
		EPCGAN	19.4	0.444	na
Wang *et al.* [54]	QXS-SAROPT	DSen2-CR	13.5	0.215	na
		CMD	14.4	0.223	na
		Pix2Pix	17.7	0.345	na
		PCycleGAN	16.5	0.300	na
		Piz2PixHD	16.6	0.351	na
		CRAN	17.9	0.346	na
		S-cycleGAN	17.2	0.287	na
		ICGAN	17.0	0.370	na
		EPCGAN	17.2	0.357	na
		HLF	18.0	0.365	na

aim to reduce the quantity of patterns to be rendered, as the correspondence between the SAR reflectivity function and the reflectance of the scene is not bijective.

The message that we want to convey is that neural networks, and specifically GANs for multimodal image translation, can provide better results if the intrinsic nature of input and target data and their phenomenological correspondence are respected during the implementation of the information process. It is intended that the training phase is computationally more expensive with the adoption of the proposed framework, as each area has its own GAN to be trained. The optimization of such hassle is not faced in this work, whose principal objective is to demonstrate that a different way to use GANs in SAR to optical image translation tasks is possible, and that provides results outperforming those so far presented in the literature. Possible workaround for training phase optimization is to group scene sharing similar characteristics. For example, cities of the same country can share the same model, as well as images acquired in different seasons like spring/summer and fall/winter.

IV. VEGETATION MONITORING

In applications, multimodal images are mainly exploited for removing clouds from optical acquisitions [55], [56], thus helping photo-interpretation tasks. In this Section, we explore the suitability of multimodal images with vegetation monitoring in areas prone to massive cloud coverage. The idea is to estimate the normalized difference vegetation index (NDVI) via partial least squares regression (PLSR) [57] of a collection of indicators derived from both translated and native optical data.

To this end, a total of 12 optical literature indices has been considered. They include the three reflectance R,G, B bands and the luminance expressed in Equation 5. Moreover, the triangular greeness index (TGI) [58], the green/red vegetation index (GRVI) [59], the green leaf index (GLI) [60], the visible atmospherically resistent index (VARI) [61], the normalized difference red-green redness index [62], the coloration index [63], and the simple ratios blue/green and blue/red [64] have been used for regression using an approach derived from [65] in which regression variables are screened using the variable importance in projection (VIP) parameter [66] to determine which ones are the more informative for the estimation problem.

Regression results are shown in FIGURE 6 and FIGURE 7 for Red Deer and Wurzburg datasets, respectively. In the first case, the target is the average NDVI calculated on 20 plots. In the second one, 42 plots have been considered.

The first experiment had the purpose to verify the existence of a relation between the NDVI and the selected indices. Therefore, all the available acquisitions have been exploited to train the regression algorithm. As for the Red Deer datasets, the obtained results are depicted in FIGURE 6a and

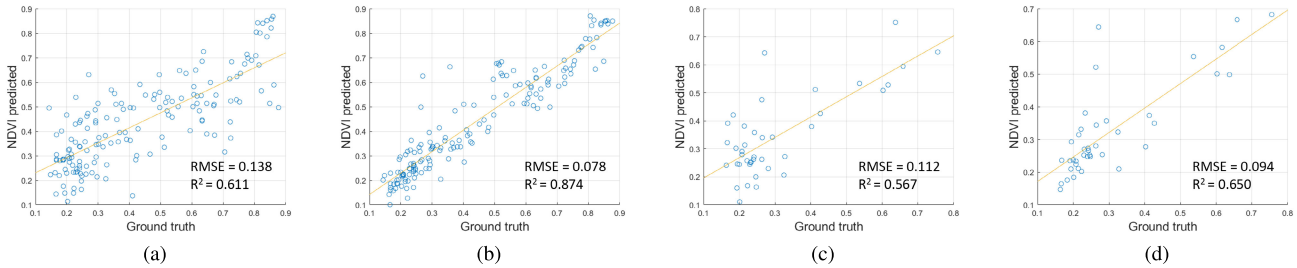


FIGURE 6. Regression results - Red Deer dataset. (a) All data, input multimodal ($RMSE = 0.138, R^2 = 0.611$). (b) All data, input optical ($RMSE = 0.078, R^2 = 0.874$). (c) Prediction, input multimodal ($RMSE = 0.112, R^2 = 0.567$). (d) Prediction, input optical ($RMSE = 0.094, R^2 = 0.650$).

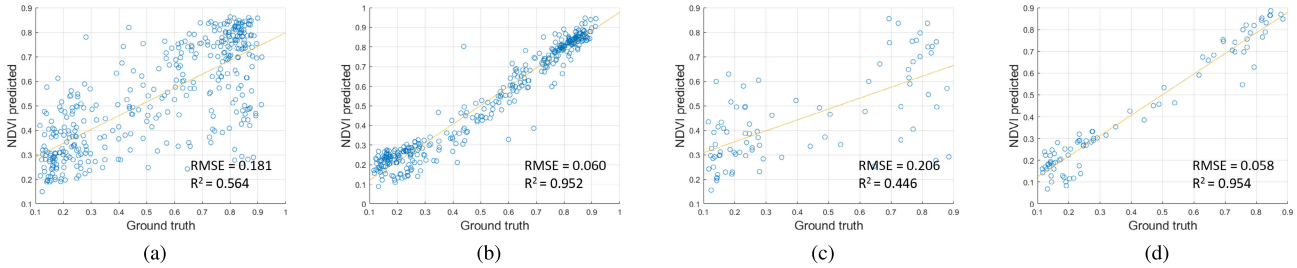


FIGURE 7. Regression results - Wurzburg dataset. (a) All data, input multimodal ($RMSE = 0.181, R^2 = 0.564$). (b) All data, input optical ($RMSE = 0.060, R^2 = 0.952$). (c) Prediction, input multimodal ($RMSE = 0.206, R^2 = 0.446$). (d) Prediction, input optical ($RMSE = 0.058, R^2 = 0.954$).

FIGURE 6b, for multimodal data and native optical data, respectively.

Using native data, a strong correlation is observed between optical indices and the target NDVI. The registered RMSE is 0.078, while the R^2 coefficient is 0.874. This is mostly confirmed in case of usage of multimodal data. The regression performance are slightly worse ($RMSE = 0.138, R^2 = 0.611$) but reasonably accurate considering the nature of the input.

The same test has been repeated with the Wurzburg dataset. Its results are reported in FIGURE 7a and FIGURE 7b for multimodal and native data, respectively. They confirm the above considerations. Using native data, a strong correlation ($RMSE = 0.060, R^2 = 0.952$) is observed between optical indices and the NDVI. The usage of multimodal data weakens such a relation ($RMSE = 0.181, R^2 = 0.564$), although the regression holds, for the majority of the plots, a sufficient degree of approximation.

The second experiment was focused on the building of a predictive model. In this case, a subset of the available images was used to train the regression model, which was used to predict the NDVI of the last two of the series. For the Red Deer dataset, the obtained results are shown in FIGURE 6c and FIGURE 6d for multimodal and native data, respectively. For the Wurzburg dataset, they are reported in FIGURE 7c and FIGURE 7d for multimodal and native data, respectively.

As for the Red Deer dataset, the regression is quite insensitive against the usage of multimodal or native optical data. In fact, in the first case the registered RMSE and R^2 are 0.112 and 0.567, respectively. In the second one, their values are 0.094 and 0.650. This holds only partially using the Wurzburg dataset, for which the regression using native optical data resulted in a very accurate prediction, with $RMSE = 0.058$ and $R^2 = 0.954$. The exploitation of multimodal data provided a lower quality prediction having $RMSE = 0.206$ and $R^2 = 0.446$.

V. CONCLUSION

Generative adversarial networks constitute a powerful tool for image domain translation, provided that input and target data share characteristics allowing for a believable translation. In the literature, the translation of synthetic aperture radar images to the visible domain has been mainly tackled with the same approach used for the treatment of consumer images, thus ignoring the peculiar structure of images and the lack of a bijective correspondence between the reflectivity function and the land cover.

In this paper, we demonstrated that acting on input products and network training allows for the improvement of output multimodal products according to standard image quality indicators. In particular, the input is pre-processed based on multi-temporal principles for the mitigation of speckle and the introduction of RGB composites enhancing the separability of the different land covers in a change-detection perspective. The training phase has been constrained by spatial and temporal features to account for land cover variability against the radar reflectivity function.

The higher quality of the obtained multimodal products allows for extending their exploitability to vegetation monitoring, which has been tested through the estimation of a synthetic normalized difference vegetation index derived from translated optical bands with promising results, thus

opening the possibility to the integration of such products in further remote sensing applications beyond visual interpretation support so far proposed in the literature.

The principal drawback of the proposed workflow is the computational burden required to train the spatio-temporal networks for domain translation. The workaround for the optimization of the training phase is object of current research and represents the way forward for the diffusion of the methodology.

REFERENCES

[1] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, Jul. 2013.

[2] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[3] F. T. Ulaby, P. P. Batlivala, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I—Bare soil," *IEEE Trans. Geosci. Electron.*, vol. 16, no. 4, pp. 286–295, Oct. 1978.

[4] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A new framework for SAR multitemporal data RGB representation: Rationale and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117–133, Jan. 2015.

[5] D. Amitrano, R. Guida, and G. Ruello, "Multitemporal SAR RGB processing for Sentinel-1 GRD products: Methodology and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1497–1507, May 2019.

[6] M. Datcu and K. Seidel, "Human-centered concepts for exploration and understanding of Earth observation images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 601–609, Mar. 2005.

[7] V. Madhok and D. A. Landgrebe, "A process model for remote sensing data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 3, pp. 680–686, Mar. 2002.

[8] D. Amitrano, F. Cecinati, G. D. Martino, A. Iodice, P.-P. Mathieu, D. Riccio, and G. Ruello, "Multitemporal level-1β products: Definitions, interpretation, and applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6545–6562, Nov. 2016.

[9] S. G. Dellepiane and E. Angiati, "A new method for cross-normalization and multitemporal visualization of SAR images for the detection of flooded areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2765–2779, Jul. 2012.

[10] C. O. Dumitru, S. Cui, G. Schwarz, and M. Datcu, "Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1635–1650, Apr. 2015.

[11] H. Su, Q. Du, and P. Du, "Hyperspectral image visualization using band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2647–2658, Jun. 2014.

[12] Z. Mahmood and P. Scheunders, "Enhanced visualization of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 869–873, Sep. 2011.

[13] N. P. Jacobson and M. R. Gupta, "Design goals and solutions for display of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2684–2692, Nov. 2005.

[14] E. Colin-Koeniguer, A. Boulch, P. Trouve-Peloux, and F. Janez, "Colored visualization of multitemporal SAR data for change detection: Issues and methods," in *Proc. 12th Eur. Conf. Synth. Aperture Radar*, Jun. 2018, pp. 1–4.

[15] E. I. Alves, A. I. A. S. S. Andrade, and D. A. Vaz, "A better view over Titan drainage networks through RGB fusion of Cassini SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 414–418, Mar. 2018.

[16] X. Wang, J. Yi, J. Guo, Y. Song, J. Lyu, J. Xu, W. Yan, J. Zhao, Q. Cai, and H. Min, "A review of image super-resolution approaches based on deep learning and applications in remote sensing," *Remote Sens.*, vol. 14, no. 21, p. 5423, Oct. 2022.

[17] J. Zhang, J. Zhou, M. Li, H. Zhou, and T. Yu, "Quality assessment of SAR-to-optical image translation," *Remote Sens.*, vol. 12, no. 21, p. 3472, Oct. 2020.

[18] L. Gatys, A. Ecker, and M. Bethge, "A neural algorithm of artistic style," *J. Vis.*, vol. 16, no. 12, p. 326, Sep. 2016.

[19] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," 2017, *arXiv:1703.00848*.

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*.

[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[22] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[23] L. Wang, X. Xu, Y. Yu, R. Yang, R. Gui, Z. Xu, and F. Pu, "SAR-to-optical image translation using supervised cycle-consistent adversarial networks," *IEEE Access*, vol. 7, pp. 129136–129149, 2019.

[24] Y. Li, R. Fu, X. Meng, W. Jin, and F. Shao, "A SAR-to-optical image translation method based on conditional generation adversarial network (cGAN)," *IEEE Access*, vol. 8, pp. 60338–60343, 2020.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[26] X. Yang, J. Zhao, Z. Wei, N. Wang, and X. Gao, "SAR-to-optical image translation based on improved CGAN," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108208.

[27] L. Liu and B. Lei, "Can SAR images and optical images transfer with each other?" in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 7019–7022.

[28] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The Sen1–2 dataset for deep learning in SAR-optical data fusion," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 141–146, Sep. 2018.

[29] S. Saha, M. Shahzad, P. Ebel, and X. X. Zhu, "Supervised change detection using prechange optical-SAR and postchange SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8170–8178, 2022.

[30] D. Amitrano, G. Di Martino, R. Guida, P. Iervolino, A. Iodice, M. N. Papa, D. Riccio, and G. Ruello, "Earth environmental monitoring using multi-temporal synthetic aperture radar: A critical review of selected applications," *Remote Sens.*, vol. 13, no. 4, p. 604, Feb. 2021.

[31] A. Freeman, "SAR calibration: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 6, pp. 1107–1121, Jun. 1992.

[32] G. Franceschetti and R. Lanari, *Synthetic Aperture Radar Processing*. Boca Raton, FL, USA: CRC Press, 1999.

[33] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1596–1615, Mar. 2014.

[34] G. Di Martino, A. Di Simone, A. Iodice, and D. Riccio, "Benchmarking framework for multitemporal SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5207826.

[35] G. F. De Grandi, M. Leysen, J. S. Lee, and D. Schuler, "Radar reflectivity estimation using multiple SAR scenes of the same target: Technique and applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, Aug. 1997, pp. 1047–1050.

[36] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[37] A. K. Fung, "Scattering from a vegetation layer," *IEEE Trans. Geosci. Electron.*, vol. GE-17, no. 1, pp. 1–6, Jan. 1979.

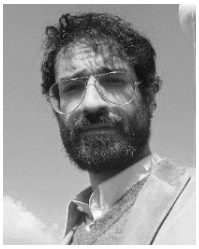
[38] J. Louis, V. Debaecker, B. Pflug, M. Main-Knorn, J. Bieniarz, U. Mueller-Wilm, E. Cadau, and F. Gascon, "Sentinel-2 Sen2Cor: L2A processor for users," in *Proc. Living Planet Symp.*, vol. SP-740, 2016, pp. 9–13.

[39] G. Franceschetti, A. Iodice, and D. Riccio, "A canonical problem in electromagnetic backscattering from buildings," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1787–1801, Aug. 2002.

[40] S. Jozdani, D. Chen, D. Pouliot, and B. A. Johnson, "A review and meta-analysis of generative adversarial networks and their applications in remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, Apr. 2022, Art. no. 102734.

[41] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[44] J. Wei, H. Zou, L. Sun, X. Cao, S. He, S. Liu, and Y. Zhang, "CFRWD-GAN for SAR-to-optical image translation," *Remote Sens.*, vol. 15, no. 10, p. 2547, May 2023.
[45] J. Guo, C. He, M. Zhang, Y. Li, X. Gao, and B. Song, "Edge-preserving convolutional generative adversarial networks for SAR-to-optical image translation," *Remote Sens.*, vol. 13, no. 18, p. 3575, Sep. 2021.
[46] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[47] K. Doi, K. Sakurada, M. Onishi, and A. Iwasaki, "GAN-based SAR-to-optical image translation with region information," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2069–2072.
[48] J. Zhang, J. Zhou, and X. Lu, "Feature-guided SAR-to-optical image translation," *IEEE Access*, vol. 8, pp. 70925–70937, 2020.
[49] M. Zhang, P. Zhang, Y. Zhang, M. Yang, X. Li, X. Dong, and L. Yang, "SAR-to-optical image translation via an interpretable network," *Remote Sens.*, vol. 16, no. 2, p. 242, Jan. 2024.
[50] J. Hwang, C. Yu, and Y. Shin, "SAR-to-optical image translation using SSIM and perceptual loss based cycle-consistent GAN," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2020, pp. 191–194.
[51] H. Shi, Z. Cui, L. Chen, J. He, and J. Yang, "A brain-inspired approach for SAR-to-optical image translation based on diffusion models," *Frontiers Neurosci.*, vol. 18, Jan. 2024.
[52] X. Bai, X. Pu, and F. Xu, "Conditional diffusion for SAR to optical image translation," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
[53] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, 2016, pp. 694–711.
[54] H. Wang, Z. Zhang, Z. Hu, and Q. Dong, "SAR-to-optical image translation with hierarchical latent features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5233812.
[55] S. Zhang, X. Li, X. Zhou, Y. Wang, and Y. Hu, "Cloud removal using SAR and optical images via attention mechanism-based GAN," *Pattern Recognit. Lett.*, vol. 175, pp. 8–15, Nov. 2023.
[56] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5222414.
[57] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometric Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, Oct. 2001.
[58] E. R. Hunt, P. C. Doraiswamy, J. E. Mcmurtrey, C. S. T. Daughtry, E. M. Perry, and B. Akhmedov, "A visible band index for remote sensing leaf chlorophyll content at the canopy scale," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 21, pp. 103–112, Apr. 2013.
[59] T. Motohka, K. N. Nasahara, H. Oguma, and S. Tsuchida, "Applicability of green-red vegetation index for remote sensing of vegetation phenology," *Remote Sens.*, vol. 2, no. 10, pp. 2369–2387, Oct. 2010.
[60] M. Louhaichi, M. M. Borman, and D. E. Johnson, "Spatially located platform and aerial photography for documentation of grazing impacts on wheat," *Geocarto Int.*, vol. 16, no. 1, pp. 65–70, Mar. 2001.
[61] A. A. Gitelson, R. Stark, U. Grits, D. Rundquist, Y. Kaufman, and D. Derry, "Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2537–2562, Jan. 2002.
[62] R. Escadafal and A. Huete, "Etude des proprietes spectrales des sols arides appliquee a l'amelioration des indices de vegetation obtenus par teledection," *Comp. Rendus de l'Academie des Sci.*, vol. 312, no. 2, pp. 1385–1391, 1991.
[63] R. Escadafal, A. Belghit, and A. Ben-Moussa, "Indices spectraux pour la teledetection de la degradation des milieux naturels en tunisie aride," in *Proc. Actes du 6th Symp. Int. Sur Les Mesures Physiques et Signatures en Télédétection*, 1994, pp. 253–259.
[64] P. Zarcotejada, A. Berjon, R. Lopezlozano, J. Miller, P. Martin, V. Cachorro, M. Gonzalez, and A. Defrutos, "Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy," *Remote Sens. Environ.*, vol. 99, no. 3, pp. 271–287, Nov. 2005.
[65] D. Amitrano, G. Candiani, L. Cicala, M. De Mizio, and F. Tufano, "Vegetation biomass and nitrogen content estimation using ensembles: Regression and active learning strategies for field sampling reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 13473–13482, 2024.
[66] M. Farrés, S. Platikanov, S. Tsakovski, and R. Tauler, "Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation," *J. Chemometrics*, vol. 29, no. 10, pp. 528–536, Oct. 2015.



DONATO AMITRANO (Senior Member, IEEE) was born in Naples, Italy, in December 1985. He received the bachelor's degree in aerospace engineering, the master's degree in aerospace and astronautical engineering, and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples "Federico II," Italy, in 2009, 2012, and 2016, respectively. From 2016 to 2018, he was a Postdoctoral Researcher with the Department of Electrical Engineering and Information Technology, University of Naples "Federico II." During this period, he was the Co-Founder and the CEO of Latitudo40, a high-tech start-up working on artificial intelligence solutions for Earth observation applications. In 2019, he joined as a Research Associate with Surrey Space Centre, University of Surrey, U.K. Since 2021, he has been with the Aerospace Systems and Earth Observation Laboratory, Italian Aerospace Research Centre. He is currently a Technical Consultant and an Expert Reviewer of public authorities, scientific journals, and private companies. His research interests include microwave, multispectral, and hyperspectral remote sensing, artificial intelligence, and computer vision. Since 2024, he has been an Associate Editor of IEEE Access.