Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 236; doi:10.3390/rs10020236

Article

A CNN-BASED FUSION METHOD FOR FEATURE EXTRACTION FROM SENTINEL DATA

Giuseppe Scarpa[1]*, Massimiliano Gargiulo[1], Antonio Mazza[1] and Raffaele Gaetano[2]

[1] DIETI, University Federico II, Via Claudio 21, 80125 Naples, Italy
[2] UMR-TETIS Laboratory, CIRAD, 34000 Montpellier, France
* Correspondence: giscarpa@unina.it; Tel.: +39 081 768 3768

Abstract: Sensitivity to weather conditions, and specially to clouds, is a severe limiting factor to the use of optical remote sensing for Earth monitoring applications. A possible alternative, is to resort to weather-insensitive synthetic aperture radar (SAR) images. However, in many real-world applications, critical decisions are made based on some informative spectral features, such as water, vegetation or soil indices, which cannot be extracted from SAR images. In the absence of optical sources, these data must be estimated. The current practice is to perform linear interpolation between data available at temporally close time instants. In this work, we propose to estimate missing spectral features through data fusion and deep-learning. Several sources of information are taken into account - optical sequences, SAR sequences, DEM - so as to exploit both temporal and cross-sensor dependencies. Based on these data, and a tiny cloud-free fraction of the target image, a compact convolutional neural network (CNN) is trained to perform the desired estimation. To validate the proposed approach, we focus on the estimation of the normalized difference vegetation index (NDVI), using coupled Sentinel-1 and Sentinel-2 time-series acquired over an agricultural region of Burkina Faso from May to November 2016. Several fusion schemes are considered, causal and non-causal, single-sensor or joint-sensor, corresponding to different operating conditions. Experimental results are very promising, showing a significant gain over baselines methods according to all performance indicators.

Keywords: Coregistration; pansharpening; multi-sensor fusion; multitemporal images; deep learning; normalized difference vegetation index (NDVI).

PACS: J0101

1. Introduction

The recent launch of coupled optical/SAR (synthetic aperture radar) Sentinel satellites, in the context of the Copernicus program, opens unprecedented opportunities for end users, both industrial and institutional, and poses new challenges to the remote sensing research community. The policy of free distribution of data allows large scale access to a very rich source of information. Besides this, the technical features of the Sentinel constellationmakes it a precious tool for a wide array of remote sensing applications. With revisit time ranging from two days to about a week, depending on the geographic location, spatial resolution from 5 to 60 meters, and wide coverage of the spectrum, from visible to short-wave infrared ($\sim 440 - 2200$ nm), Sentinel data may impact decisively on a number of Earth monitoring applications, such as climate change monitoring, map updating, agriculture and forestry planning, flood monitoring, ice monitoring, and so forth.

Especially precious is the diversity of information guaranteed by the coupled SAR and optical sensors, a key element for boosting the monitoring capability of the constellation. In fact, the information conveyed by the Sentinel-2 (S2) multi-resolution optical sensor depends on the spectral reflectivity of the target illuminated by sunlight, while the backscattered signal acquired by the Sentinel-1 (S1) SAR sensor depends on both target's characteristics and the illuminating signal. The joint processing of optical and radar temporal sequences offers the opportunity to extract the information of interest with an accuracy that could not be achieved based on only one these sources.

Of course, with this potential, comes the scientific challenge of how to exploit these complementary piece information in the most effective way.

In this work, we focus on the estimation of the Normalized Difference Vegetation Index (NDVI) in critical weather conditions, fusing the information provided by temporal sequences of S1 and S2 images. In fact, the typical processing pipelines of many land monitoring applications rely, among other features, on the NDVI for a single date or a whole temporal series. Unfortunately, the NDVI, as well as other spectral features, is unavailable under cloudy weather conditions. The commonly adopted solution consists in interpolating between temporally adjacent images where the target feature is present. However, given the availability of weather-insensitive SAR data of the scene, it make sense to pursue fusion-based solutions, exploiting SAR images that may be temporally very close to the target date. By so doing, we assume that SAR data, despite their very different nature, can provide precious information on NDVI. Even if this holds true, however, it is by no means obvious how to exploit such dependency. To address this problem we resort to deep learning, designing a simple three-layer convolutional neural network (CNN) for this task, and training it to account for both temporal and cross-sensor dependencies. Note that the same approach, with minimal adaptations, can be extended to estimate many other spectral indices, commonly used for water, soil, and so on. Therefore, besides solving the specific problem, we demonstrate the potential of deep learning for data fusion in remote sensing.

According to the taxonomy given in [1] data fusion methods, *i.e., processing dealing with data and information from multiple sources to achieve improved information for decision making*, can be grouped in three main categories:

- *pixel*-level: the pixel values of the sources to be fused are jointly processed [2–5];
- *feature*-level: features like lines, regions, keypoints, maps, and so on, are first extracted independently from each source image and subsequently combined to produce higer-level cross-source features which may represent the desired output or be further processed [6–12]; and
- *decision*-level: the high-level information extracted independently from each source is combined to provide the final outcome, for example resorting to fuzzy logic [13,14], decision trees [15], Bayesian inference [16], Dempster-Shafer theory [17], and so forth.

In the context of remote sensing, with reference to the sources to be fused, fusion methods can be roughly gathered for the most part in the following categories:

- multi-*resolution*: concerns a single sensor with multiple resolution bands. One of the most frequent application is pansharpening [2,18,19], although many other tasks can be solved under a multi-resolution paradigm, such as segmentation [20] or feature extraction [21], to mention a few.
- multi-*temporal*: is one of the most investigated forms of fusion in remote sensing due to the rich information content hidden in the temporal dimension. In particular it can be applied to strictly time-related tasks, like prediction [9], change detection [22,23], co-registration [24], and general-purpose tasks, like segmentation [3], despeckling [25], feature extraction [26–28], which do not necessarily need a joint processing of the temporal sequence but can benefit from it.
- multi-*sensor*: is gaining an ever growing importance due both to the recent deployment of many new satellites, and to the increasing tendency of the community to share data. It represents also the most challenging case because of the several sources of mismatch (temporal, geometrical, spectral, radiometrical) among involved data. Like for other categories, a number of typical remote sensing problems can fit this paradigm, such as classification [6,12,29–31], coregistration [11], change detection [32], feature estimation [33–36].
- *mixed*: the above cases may also occur jointly, generating mixed situations. For example hyperspectral and multiresolution images can be fused to produce a spatial-spectral full-resolution datacube [5,37]. Likewise, low-resolution temporally dense series can be fused

with high-resolution but temporally sparse ones to simulate a temporal-spatial full-resolution sequence [38]. The monitoring of forests [16], soil moisture [39], environmental hazards [8], and other processes, can be also carried out effectively by fusing SAR and optical time series. Finally, works that mix all three aspects, resolution, time, and sensor, can also be found in the literature [7,17,40].

Turning to multi-sensor SAR-optical fusion for the purpose of vegetation monitoring, a number of contributions can be found in the literature [7,12,16,34,41]. In [7] ALOS POLSAR and Landsat time-series were combined at feature level for forest mapping and monitoring. The same problem was addressed in [16] through a decision-level approach. In [41] the fusion of single-date S1 and simulated S2 was presented for the purpose of classification. In [34], instead, RADARSAT-2 and Landsat-7/8 images were fused, by means of an artificial neural network, to estimate soil moisture and leaf area index. The NDVI index obtained from the Landsat source was combined with different SAR polarization subsets for feeding *ad hoc* artificial networks. A similar feature-level approach, based on Sentinel data, was followed in [12] for the purpose of land cover mapping. To this end, the texture maps extracted from the SAR image were combined with several indices drawn from the optical bands.

Although some fusion techniques have been proposed for spatio-temporal NDVI super-resolution [38] or prediction [9], they use exclusively optical data. None of these papers attempts to directly estimate a pure multispectral feature, NDVI or the likes, from SAR data. In most cases the fusion, occurring already at feature level, is intended to provide high-level information, like the classification or detection of some physical item. Conversely, we can register some notable example of indices directly related to physical items of interest, like soil moisture or the area leaf index, which have been estimated by fusing SAR and optical data [33,34].

In this work, we propose several CNN-based algorithms to estimate the NDVI through the fusion of optical and SAR Sentinel data. With reference to a specific case study, we acquired temporal sequences of S1 SAR data and S2 optical data, covering the same time lapse, with the latter partially covered by clouds. Both temporal and cross-sensor (S1-S2) dependencies are used to obtain the most effective estimation protocol. From the experimental analysis, very interesting results emerge. On one hand, when only optical data are used, CNN-based methods outperform consistently the conventional temporal interpolators. On the other hand, when also SAR data are considered, a further significant improvement of performance is observed, despite the very different nature of the involved signals. It is worth underlining that no peculiar property of NDVI was exploited, and therefore these results have a wider significance, suggesting that other image features can be better estimated by cross-sensor CNN-based fusion.

The rest of the paper is organized as follows. In Section 2, we present the dataset and describe the problem under investigation. In Section 3, the basics of the CNN methodology are recalled. Then, the specific prediction architectures are detailed in Section 4. Finally, in Section 5, we present and discuss the experimental results. Section 6 draws conclusions.

2. Dataset and Problem Statement

The objective of this work is to propose and test a set of solutions to estimate a target optical feature at a given date from images acquired at adjacent dates, or even from the temporally closest SAR image. Such different solutions reflect also the different operating conditions found in practice. The main application is the reconstruction of a feature of interest in a target image which is available but partially or totally cloudy. However, one may also consider the case in which the feature is built and used on a date for which no image is actually available.

In this work, we focus on the estimation of the Normalized Difference Vegetation Index, but it is straightforward to apply the same framework to other optical features. With reference to Sentinel

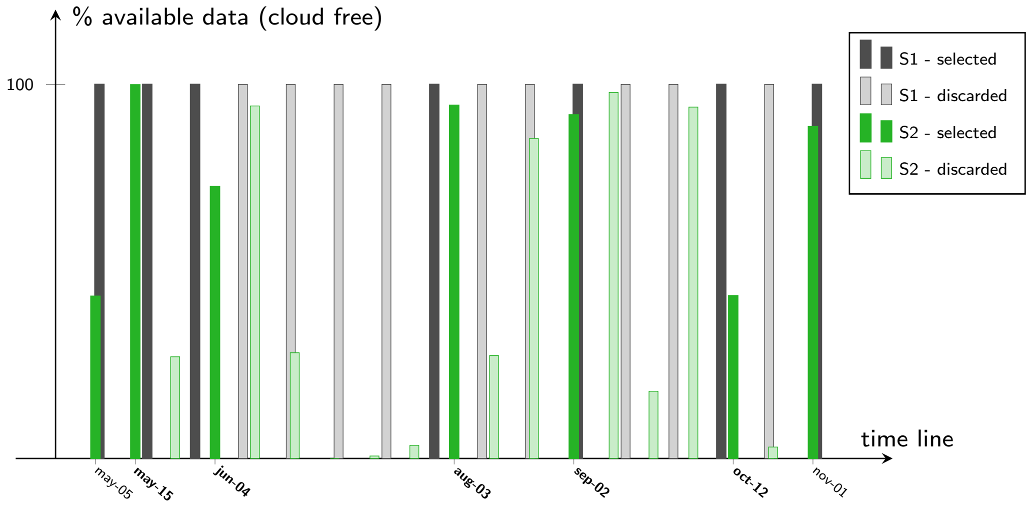

Figure 1. Available S1 (black) and S2 (green) images over the period of interest. The bar height indicates the fraction of usable data. Solid bars mark selected images, boldface date mark test images.

images, the NDVI is obtained at a resolution of 10 meters by combining pixel-by-pixel two bands, near infrared (NIR, 8th band), and red (Red, 4th band), as:

$$\text{NDVI} \triangleq \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \in [-1, 1] \tag{1}$$

The area under study is located in the province of Tuy, Burkina Faso, around the commune of Koumbia. This area is particularly representative of West African semiarid agricultural landscapes, for which the Sentinel missions offer new opportunities in monitoring vegetation, notably in the context of climate change adaptation and food security. The use of SAR data in conjunction with optical images is particularly appropriate in these areas, since most of the vegetation dynamics take place during the rainy season, especially over the cropland, as smallholder rainfed agriculture is dominant. This strongly reduces the availability of usable optical images in the critical phase of vegetation growth, due to the significant cloud coverage [42] from which SAR data are only loosely affected. The 5253x4797 pixels scene is monitored between May 5th and November 1st 2016, over a period that corresponds to a regular cultural season in the area.

Fig. 1 indicates the available S1 and S2 acquisitions in this period. In the case of S2 images, the bar height indicates the percentage of data which are not cloudy. It is clear that some dates provide little or no information. Note that, during the rainy season, the lack of sufficient cloud-free optical data may represent a major issue, preventing the extraction of spatio-temporal optical-based features, like time-series of vegetation, water or soil indices, and so on. S1 images, instead, are always completely available, as SAR data are insensitive to meteorological conditions.

For the purpose of training, validation and testing of the proposed methods, we kept only S2 images which were cloud-free, or such that the spatial distribution of clouds did not prevent the selection of sufficiently large test and training areas. For the selected S2 images (solid bars in Fig. 1) the corresponding dates are indicated on the x-axis. Our dataset was then completed by including also the S1 images (solid bars) which are temporally closest to the selected S2 counterparts. The general idea of the proposal is to use the closest cloud-free S2 and S1 images to estimate the desired feature on the target date of interest. Therefore, among the seven selected dates, only the five inner ones are used as targets. Observe, also, that the resulting temporal sampling is rather variable, with intervals ranging from ten days to a couple of months, allowing us to test our methods in different conditions.

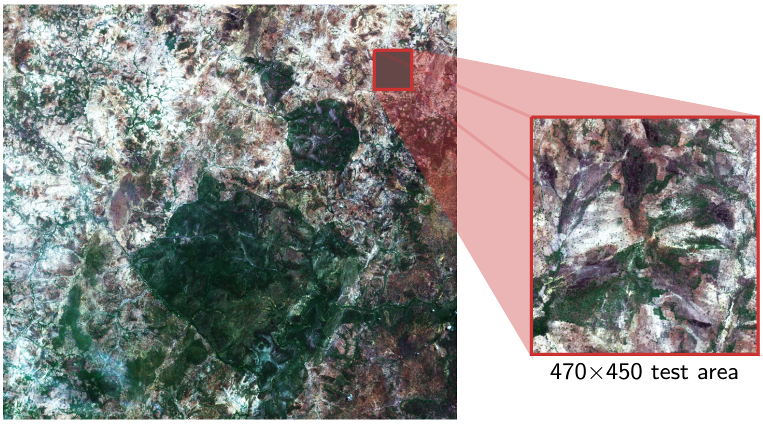

Figure 2. RGB false-color representation of the 5253×4797 S2-Koumbia dataset (May 15th, 2016), with a zoom on the area selected for testing.

To allow also temporal analyses, we chose a test area, of size 470×450, which is cloud-free in all the selected dates, and hence with available reference ground-truth for any possible optical feature. Fig. 2 shows a false-color representation of a complete image of the Koumbia dataset (May 15th, 2016), together with a zoom of the selected test area. Even after discarding the test area, quite a large usable area remains, from which a sufficiently large number of small (33×33) cloud-free patches are randomly extracted for training and validation.

In addition to the Sentinel data, we assume the availability of two additional features, the cloud masks for each S2 image, and a Digital Elevation Model (DEM). Cloud masks are obviously necessary to establish when the prediction is needed and which adjacent dates should be involved. The DEM is a complementary feature that integrates the information carried by SAR data, and may be useful to improve estimation.

For this work, we used Sentinel-1 data in the Ground Range Detected (GRD) format as provided by ESA. All images have been calibrated (VH/VV intensities to sigma nought) and terrain corrected using ancillary data, and co-registered to provide a 10 m resolution, spatially coherent time series, using the official ESA Sentinel Application Platform (SNAP) software [43]. No optical/SAR co-registration has been performed, assuming that the co-location precision provided by the independent orthorectification of each product is sufficient for the application. Sentinel-2 data are provided by the French Pole Thématique Surfaces Continentales (THEIA) [44] and preprocessed using the Multi-sensor Atmospheric Correction and Cloud Screening (MACCS) level 2A processor [45] developed at the French National Space Agency (CNES) to provide surface reflectance products as well as precise cloud masks. Finally, the DEM was gathered from the Shuttle Radar Topographic Mission (SRTM) 1 Arc-Second Global, with 30 m resolution resampled at 10 m to match the spatial resolution of Sentinel data.

3. Convolutional neural networks

Before moving to the specific solutions for NDVI estimation, in this Section we provide some basic notions and terminology about CNNs.

In the last few years, convolutional neural networks have been successfully applied to many classical image processing problems, such as denoising [46], super-resolution [47], pansharpening [4,19], segmentation [48], detection [49], classification [50,51]. The main strengths of CNNs are (i) an extreme versatility that allows them to approximate any sort of linear or non linear transformation, including scaling or hard thresholding; (ii) no need to design handcrafted filters, replaced by machine learning; (iii) high-speed processing, thanks to parallel computing. On the downside, for correct training, CNNs require the availability of a large amount of data with ground-truth (examples). In our specific case, data are not a problem, given the unlimited quantity of cloud-free Sentinel-2 time-series that can be downloaded from the web repositories. However, using large datasets has a cost in terms of complexity, and may lead to unreasonably long training times. Usually, a CNN is a chain[1] of different layers, like convolution, nonlinearities, pooling, deconvolution. For image processing tasks in which the desired output is an image at the same resolution of the input, as in this work, only convolutional layers interleaved with nonlinear activations are typically employed.

The generic l-th convolutional layer, with N-band input $\mathbf{x}^{(l)}$, yields an M-band output $\mathbf{z}^{(l)}$ computed as

$$\mathbf{z}^{(l)} = \mathbf{w}^{(l)} * \mathbf{x}^{(l)} + \mathbf{b}^{(l)},$$

whose m-th component can be written in terms ordinary 2D convolutions

$$\mathbf{z}^{(l)}(m,\cdot,\cdot) = \sum_{n=1}^{N} \mathbf{w}^{(l)}(m,n,\cdot,\cdot) * \mathbf{x}^{(l)}(n,\cdot,\cdot) + \mathbf{b}^{(l)}(m).$$

The tensor $\mathbf{w}$ is a set of M convolutional $N \times (K \times K)$ kernels, with a $K \times K$ spatial support (receptive field), while $\mathbf{b}$ is a M-vector bias. These parameters, compactly, $\Phi_l \triangleq \left(\mathbf{w}^{(l)}, \mathbf{b}^{(l)}\right)$, are learnt during the training phase. If the convolution is followed by a pointwise activation function $g_l(\cdot)$, then, the overall layer output is given by

$$\mathbf{y}^{(l)} = g_l(\mathbf{z}^{(l)}) = g_l(\mathbf{w}^{(l)} * \mathbf{x}^{(l)} + \mathbf{b}^{(l)}) \triangleq f_l(\mathbf{x}^{(l)}, \Phi_l). \tag{2}$$

Due to the good convergence properties it ensures [50], the Rectified Linear Unit (ReLU), defined as $g(\cdot) \triangleq \max(0,\ \cdot)$, is a typical activation function of choice for input or hidden layers.

Assuming a simple L-layer cascade architecture, the overall processing will be

$$f(\mathbf{x}, \Phi) = f_L(f_{L-1}(\ldots f_1(\mathbf{x}, \Phi_1), \ldots, \Phi_{L-1}), \Phi_L) \tag{3}$$

where $\Phi \triangleq (\Phi_1, \ldots, \Phi_L)$ is the whole set of parameters to learn. In this chain, each layer l provides a set of so-called *feature maps*, $\mathbf{y}^{(l)}$, which activate on local cues in the early stages (small l), to become more and more representative of abstract and global phenomena in subsequent ones (large l). In this work, all proposed solutions are based on a simple three-layer architecture, and differ only in the input layer, as different combinations of input bands are considered.

3.1. Learning

In order to learn the network parameters, a sufficiently large training set, say $\mathbf{T}$, of input-output examples $\mathbf{t}$ is needed:

$$\mathbf{T} \triangleq \{\mathbf{t}_1, \ldots, \mathbf{t}_Q\}, \quad \mathbf{t} \triangleq (\mathbf{x}, \mathbf{y}^{\mathrm{ref}})$$

In our specific case, $\mathbf{x}$ will be a sample of the combination of images from which we want to estimate the target NDVI map, with $\mathbf{y}^{\mathrm{ref}}$ the desired output. Of course, all involved optical images must be cloud-free over the selected patches.

Formally, the objective of the training phase is to find

$$\Phi = \arg\min_{\Phi} J(\mathbf{T}, \Phi) \triangleq \arg\min_{\Phi} \frac{1}{Q} \sum_{\mathbf{t} \in \mathbf{T}} L(\mathbf{t}, \Phi)$$

where $L(\mathbf{t}, \Phi)$ is a suitable loss function. Several losses can be found in the literature, like L_n norms, cross-entropy, negative log-likelihood. The choice depends on the domain of the output, and affects

[1] parallels, loops or other combinations are also possible.

Table 1. Hyper-parameters of the CNN architecture. Shape = # features × # channels × 2D support.

	ConvLayer 1	$g_1(\cdot)$	ConvLayer 2	$g_2(\cdot)$	ConvLayer 3
Shape	$48\times b_x\times 9\times 9$	ReLU	$32\times 48\times 5\times 5$	ReLU	$1\times 32\times 5\times 5$
Learning rate	10^{-4}		10^{-4}		10^{-5}
Momentum	0.9		0.9		0.9

the convergence properties of the networks [52]. Among the different solutions experimented, which will be later discussed, we have chosen the L_1-norm

$$L(\mathbf{t},\Phi) \propto ||f(\mathbf{x},\Phi) - \mathbf{y}^{\mathrm{ref}}||_1 \tag{4}$$

which proved effective in other generative problems in remote sensing [19]. As for minimization, the most widespread procedure, adopted also in this work, is the stochastic gradient descent (SGD) with momentum [53]. The training set is partitioned in batches of samples, $\mathbf{T} = \{\mathbf{B}_1, \ldots, \mathbf{B}_P\}$. At each iteration, a new batch is used to estimate the gradient and update parameters as

$$\nu^{(n+1)} \leftarrow \mu\nu^{(n)} + \alpha\nabla_\Phi J\left(\mathbf{B}_{j_n}, \Phi^{(n)}\right)$$
$$\Phi^{(n+1)} \leftarrow \Phi^{(n)} - \nu^{(n+1)}$$

A whole scan of the training set is called an *epoch*, and training a deep network may require from dozens of epochs, for simpler problems like handwritten character recognition [54], to thousands of epochs for complex classification tasks [50]. Accuracy and speed of training depend on both the initialization of Φ and the setting of hyperparameters like learning rate α and momentum μ, with α being to most critical, impacting heavily on stability and convergence time.

4. Proposed prediction architectures

In the following developments, with reference to a given target S2 image acquired at time t, we will consider the items defined below:

- t_-, t_+: times of previous and next useful S2 images;
- $\hat{t}, \hat{t}_-, \hat{t}_+$: times of closest S1 image (always $|t-\hat{t}| \le 5$) and of previous and next S1 images;
- F: unknown feature (NDVI in this work) at time t;
- F_- and F_+: feature F at times t_- and t_+, respectively;
- $\mathbf{S} \triangleq (S^{VV}, S^{VH})$: double polarized VV-VH SAR image at time $\hat{t}$;
- $\mathbf{S}_-$ and $\mathbf{S}_+$: SAR images at times $\hat{t}_-$ and $\hat{t}_+$, respectively;
- D: DEM.

The several models considered here differ in the composition of the input stack $\mathbf{x}$, while the output is always the NDVI at the target date, that is, $\mathbf{y} = F$. Apart from the input layer, the CNN architecture is always the same, depicted in Fig. 3, with hyper-parameters summarized in Tab. 1. This relatively shallow CNN is characterized by a rather small number of weights (as CNNs go), and hence can be trained with a small amount of data. On the other hand, this very same architecture has proven to achieve state-of-the-art performance in closely related applications, such as super-resolution [47] and data fusion [4,19].

The number b_x of input bands depends on the specific solution and will be made explicit below. In order to provide output values falling in the compact interval [-1,1], as required by the NDVI semantics (Eq. 1), one can include a suitable nonlinear activation, like $\tanh(\cdot)$, to complete the output layer. In such a case, it is customary to resort to a cross-entropy loss for training. As an alternative, one may remove the nonlinear output mapping altogether, and simply take the result of the convolution, which can be optimized using, for example, a L_n-norm. Obviously, in this case, a hard clipping of the output is still needed, but this additional transformation does not participate in the error back

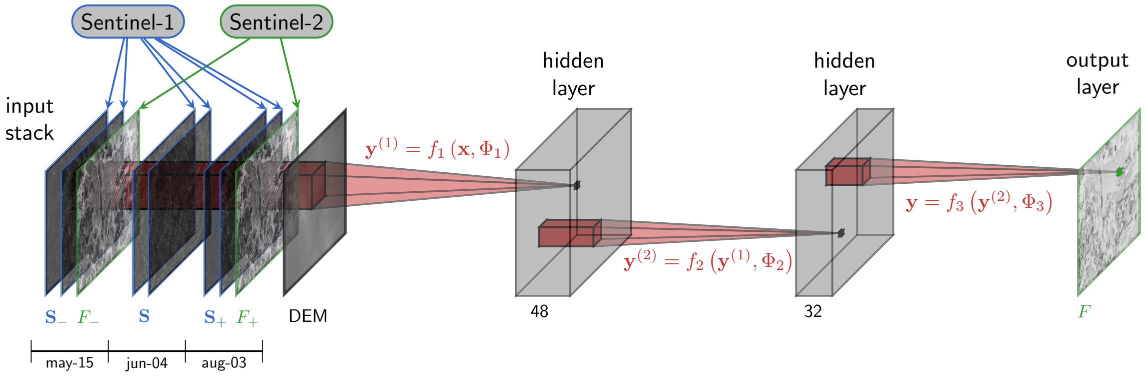

Figure 3. Proposed CNN architecture. The depicted input corresponds to the Optical-SAR+ case. Other cases use a reduced set of inputs.

Table 2. Proposed models.

		Input Bands		
Model name	b_x	Optical	SAR	DEM
SAR	2		$\mathbf{S}$	
SAR+	3		$\mathbf{S}$	D
Optical-C	1	F_-		
Optical-SAR/C	5	F_-	$\mathbf{S}_-, \mathbf{S}$	
Optical-SAR+/C	6	F_-	$\mathbf{S}_-, \mathbf{S}$	D
Optical	2	F_-, F_+		
Optical-SAR	8	F_-, F_+	$\mathbf{S}_-, \mathbf{S}, \mathbf{S}_+$	
Optical-SAR+	9	F_-, F_+	$\mathbf{S}_-, \mathbf{S}, \mathbf{S}_+$	D

propagation, hence should be considered external to the network. Through preliminary experiments, we have found this latter solution more effective than the former, for our task, and therefore we train the CNN considering a linear activation in the last layer, $g_3(\mathbf{z}^{(3)}) = \mathbf{z}^{(3)}$. Experiments also showed the L_1-norm (Eq. 4) to be more effective than the L_2-norm for training, and hence we opted for the former in our tests.

We now describe briefly the different solutions considered here, which depend on the available input data and the required response time.

Concerning data, we will consider estimation based on optical-only, SAR-only, and optical+SAR data. When using SAR images, we will also test the inclusion of the DEM, which may convey relevant information on them. Instead, the DEM is useless, and hence neglected, when only optical data are used. All these cases are of interest, for the following reasons.

- The *optical-only* case allows for a direct comparison, with the same input data, between the proposed CNN-based solution and the current baseline, which relies on temporal linear interpolation. Therefore, it will provide us with a measure the net performance gain guaranteed by deep learning over conventional processing.
- Although SAR and optical data provide complementary information, the occurrence of a given physical item, like water or vegetation, can be detected by means of both scattering properties and spectral signatures. The analysis of the *SAR-only* case will allow us to understand if significant dependencies exist between the NDVI and SAR images, and if a reasonable quality can be achieved even when only this source is used for estimation. To this aim we do not count on the temporal dependencies in this case trying to estimate a S2 feature from the closest S1 image only.
- The *optical-SAR* fusion, of course, is the case of highest interest for us. Given the most complete set of relevant input, and an adequate training set, the proposed CNN will synthesize expressive features, and is expected to provide a high-quality NDVI estimate.

Peer-reviewed version available at *Remote Sens.* **2018**, *10*, 236; doi:10.3390/rs10020236

Turning to response time, except for the SAR-only case, we will distinguish between causal estimation, in which only data already available at time t (or shortly later) can be used, and non-causal estimation, when the whole time series is supposed to be available.

- *Causal estimation* is of interest whenever the data must be used right away for the application of interest. This is the case, for example, of early warning systems for food security. We will include here also the case in which $\hat{t} > t$, namely the closest SAR image becomes available after time t, since the maximum delay is always very limited.
- On the other hand, in the absence of temporal constraints, all relevant data should be taken into account to obtain the best possible quality, resorting therefore to *non-causal estimation*.

Table.2 summarizes all these different solutions.

4.1. Training

In order to carry out an effective training of the networks, a large cloud-free dataset is necessary, with geophysical properties as close as possible to those of the target data. This is readily guaranteed whenever all images involved in the process, for example F_{-}, F and F_{+}, share a relatively large cloud-free area. Patches will be extracted from this area to train the network which, afterwards, will be used to estimate F also on the clouded area, obtaining a complete coverage at the target date.

For our relatively small network, a training+validation set of 18000 patches is sufficient for accurate training. With our patch extraction process, this number requires an overall cloud-free area of about 1000×1000 pixels, namely, about 4% of our 5253×4797 target scene (Fig. 2). If the unclouded regions are more scattered, this percentage may somewhat grow, but remains always quite limited. Therefore, a perfectly fit training set will be available most of the times (always, in our experiments). However, even if this is not the case, for example because the scene is almost completely covered by clouds at the target date, one can build a good training set by using data collected on other regions with geophysical features similar to those of the target scene (for example tropical-tropical). Subsequently, even a very small unclouded region of the target scene can be used to fine-tune the network parameters.

In practice, for each date, a dataset composed of 15200 33×33 examples for training, plus 3800 more for validation, was created by sampling the target scene with a 8-pixel stride in both spatial directions, always skipping test area and cloudy regions. Then, the whole collection was shuffled to avoid biases when creating the 128-examples mini-batches used in the SGD algorithm.

5. Experimental results

The performance of the proposed CNN-based estimators is assessed in terms of correlation index (ρ), peak signal-to-noise ratio (PSNR), and structural similarity measure (SSIM).

We consider two reference methods, a deterministic linear interpolator (temporal gap-filling) which can be regarded as the baseline, and a simple affine regressor. Temporal gap filling was proposed in [42] in the context of the development of a national-scale crop mapping processor based on Sentinel-2 time series, and implemented as a remote module of the Orfeo Toolbox [55]. This is a practical solution used by analysts [42] to monitor vegetation processes through NDVI time-series. Besides being simple, it is also more generally applicable and robust than higher-order models which require a larger number of points to interpolate and may overfit the data. Since temporal gap filling is non-causal, we add a further causal interpolator for completeness, a simple zero-order hold. Of course, deterministic interpolation does not take into account the correlation between available and target data, which can help performing a better estimate and can be easily computed based on a tiny cloud-free fraction of the target image. Therefore, for a fairer comparison, we consider as a further reference the affine regressors, both causal and non-causal, computed based on such correlations. If suitable, post-processing may be included for spatial regularization, both for the reference and proposed methods. This option is not pursued here.

Table 3. Correlation index, $\rho \in [-1, 1]$.

		may-15	jun-04	aug-03	sep-02	oct-12	average
	gaps (before/after)	10/20	20/60	60/30	30/40	40/20	
Cross-sensor	**SAR**	0.8243	0.8161	0.5407	0.4219	0.4561	0.6118
	SAR+	0.8254	0.7423	0.3969	0.4963	0.6428	0.6207
Causal	**Interpolator/C**	0.9760	0.8925	0.6566	0.6704	0.6098	0.7611
	Regressor/C	0.9760	0.8925	0.6566	0.6704	0.6098	0.7611
	Optical/C	0.9811	0.9407	0.7245	0.7280	0.7302	0.8209
	Optical-SAR/C	0.9797	**0.9432**	0.7716	**0.7880**	0.7546	0.8474
	Optical-SAR+/C	**0.9818**	0.9424	**0.7738**	0.7855	**0.7792**	**0.8525**
Non-causal	**Interpolator**	0.9612	0.8915	0.7643	0.7288	0.8838	0.8459
	Regressor	0.9708	0.9004	0.7618	0.7294	0.8930	0.8511
	Optical	**0.9814**	0.9524	0.8334	0.758	0.9115	0.8874
	Optical-SAR	0.9775	**0.9557**	**0.8567**	0.8194	0.9002	0.9019
	Optical-SAR+	0.9781	0.9536	0.8550	**0.8220**	**0.9289**	**0.9075**

Let us first discuss the numerical results and then move to a subjective assessment by visual inspection of some meaningful sample images. In Tabb. 3-5 we report the correlation index, the PSNR, and the SSIM for all proposed and reference methods and for all dates. In the last column we also report the average performance of each method over all dates. The target dates are shown in the first row, while the second row gives the temporal gaps (days) between the target and the previous and next dates used for prediction, $(t - t_-)$ and $(t_+ - t)$, respectively. The following two lines show results for fully cross-sensor, that is, SAR-only, estimation, while in the rest of the table we group together all causal (top) and non-causal (bottom) models, highlighting the best performance in each group with blue text.

Let us focus for the time being on the ρ table, and in particular on the last column with average values, which accounts well for the main trends. First of all, the fully cross-sensor solutions, based on only-SAR or SAR+DEM data, respectively, are not competitive with methods exploiting optical data, with a correlation index barely exceeding 0.6. Nonetheless, they allow one to obtain a rough estimate of the NDVI in the absence of optical coverage, proving that even a pure spectral feature can be inferred from SAR images, thanks to the dependencies existing between the geometrical and spectral properties of the scene. Moreover, SAR images provide information on the target which is not available in optical images, and complementary to it. Hence, their inclusion can help boosting the performance of methods relying on optical data.

Turning to the latter, we observe, as expected, that non-causal models largely outperform the corresponding causal counterparts. As an example, for the baseline interpolator, ρ grows from 0.761 (causal) to 0.846 (non-causal), showing that the constraint of near real-time processing has a severe impact on estimation quality.

However, even with the constraint of causality, most of this gap can be filled by resorting to CNN-based methods. By using the very same data for prediction, that is, only F_-, the Optical/C model reaches already $\rho = 0.821$. This grows to 0.847 (like the non-causal interpolator) when also SAR data are used, and to 0.852 when also the DEM is included. Therefore, both the use CNN-based estimation and the inclusion of SAR data guarantee a clear improvement. On the contrary, using a simple statistical regressor is of little or no[2] help. Looking at the individual dates, a clear dependence on the time gaps emerges. For the causal baseline, in particular, the ρ varies wildly, from 0.610 to 0.976. Indeed, when the previous image is temporally close to the target, like for May-15, and hence strongly correlated with it, even this trivial method provides a very good estimation, and more sophisticated methods cannot give much of an improvement. However, things change radically when the previous

[2] The causal interpolator and regressor have identical ρ by definition.

Table 4. Peak signal-to-noise ratio (PSNR) [dB].

		may-15	jun-04	aug-03	sep-02	oct-12	average
	gaps (before/after)	10/20	20/60	60/30	30/40	40/20	
Cross-sensor	SAR	24.30	19.52	12.34	17.30	10.70	16.83
	SAR+	23.49	17.96	14.78	16.12	19.01	18.27
Causal	Interpolator/C	30.11	19.48	10.62	17.70	14.59	18.50
	Regressor/C	30.86	22.60	18.30	20.39	20.02	22.44
	Optical/C	30.85	24.92	18.74	21.01	21.22	23.35
	Optical-SAR/C	31.24	**25.07**	**19.96**	21.56	20.71	23.71
	Optical-SAR+/C	**32.81**	24.90	19.79	**21.76**	**21.91**	**24.24**
Non-causal	Interpolator	27.91	21.97	19.12	17.41	23.61	22.00
	Regressor	30.26	22.86	20.01	21.14	24.67	23.79
	Optical	**32.61**	26.09	21.41	21.53	24.74	25.28
	Optical-SAR	29.72	**26.29**	**22.01**	**22.48**	23.89	24.88
	Optical-SAR+	31.62	25.65	21.84	22.30	**25.24**	**25.33**

Table 5. Structural similarity measure (SSIM) [-1,1].

		may-15	jun-04	aug-03	sep-02	oct-12	average
	gaps (before/after)	10/20	20/60	60/30	30/40	40/20	
Cross-sensor	SAR	0.5565	0.4766	0.3071	0.3511	0.2797	0.3942
	SAR+	0.5758	0.4534	0.3389	0.3601	0.3808	0.4218
Causal	Interpolator/C	0.9128	0.7115	0.3481	0.6597	0.6335	0.6531
	Regressor/C	0.9168	0.7364	0.4161	0.6425	0.6001	0.6624
	Optical/C	0.9557	0.8583	0.6057	0.7265	0.6671	0.7627
	Optical-SAR/C	0.9543	0.8600	0.6280	0.7539	0.6918	0.7776
	Optical-SAR+/C	**0.9565**	**0.8602**	**0.6365**	**0.7545**	**0.6989**	**0.7813**
Non-causal	Interpolator	0.8801	0.6798	0.6696	0.7177	0.8249	0.7544
	Regressor	0.9067	0.7330	0.6693	0.7218	0.8032	0.7668
	Optical	**0.9589**	0.8788	0.7623	0.7618	0.8470	0.8418
	Optical-SAR	0.9541	**0.8835**	**0.7780**	**0.7841**	0.8339	0.8467
	Optical-SAR+	0.9571	0.8788	0.7757	0.7834	**0.8559**	**0.8502**

available image is acquired long before the target, like for the Aug-03 or Oct-12 dates. In these cases, the baseline does not provide acceptable estimates anymore, and CNN-based methods give a large performance gain, ensuring a ρ always close to 0.8 even in the worst cases.

Moving now to non-causal estimation we observe a similar trend. Both reference methods are significantly outperformed by the CNN-based solutions working on the same data, and further improvements are obtained by including SAR and DEM. The overall average gain, from 0.851 to 0.907 is not as large as before, since we start from a much better baseline, but still quite significant. Examining the individual dates, similar considerations as before arise, with the difference that now two time gaps must be taken into account, with previous and next images. As expected, the CNN-based methods provide the largest improvements when both gaps are rather large, that is, 30 days or more, like for the Aug-03 and Sep-02 images.

The very same trends outlined for the ρ are observed also with reference to the PSNR and SSIM data, shown in Tab. 4 and Tab.5. Note that, unlike ρ and SSIM, the PSNR is quite sensitive to biases on the mean, which is why, in this case, the statistical affine regressor provides significant gains over the linear interpolator. In any case, the best performance is always obtained using CNN-based methods relying on both optical and SAR data, with large improvements with respect to the reference methods.

Further insight into the behavior of the compared methods can be gained by visual inspection of some sample results. To this end we consider two target dates, June 4th and Aug 3rd, characterized by significant temporal changes in spectral features with respect to the closest available dates. In the first case, a high correlation exists with the previous date $\rho = 0.8925$ but not with the next $\rho = 0.6566$. In the second, both correlation indexes are quite low, 0.6566 and 0.6704, respectively.

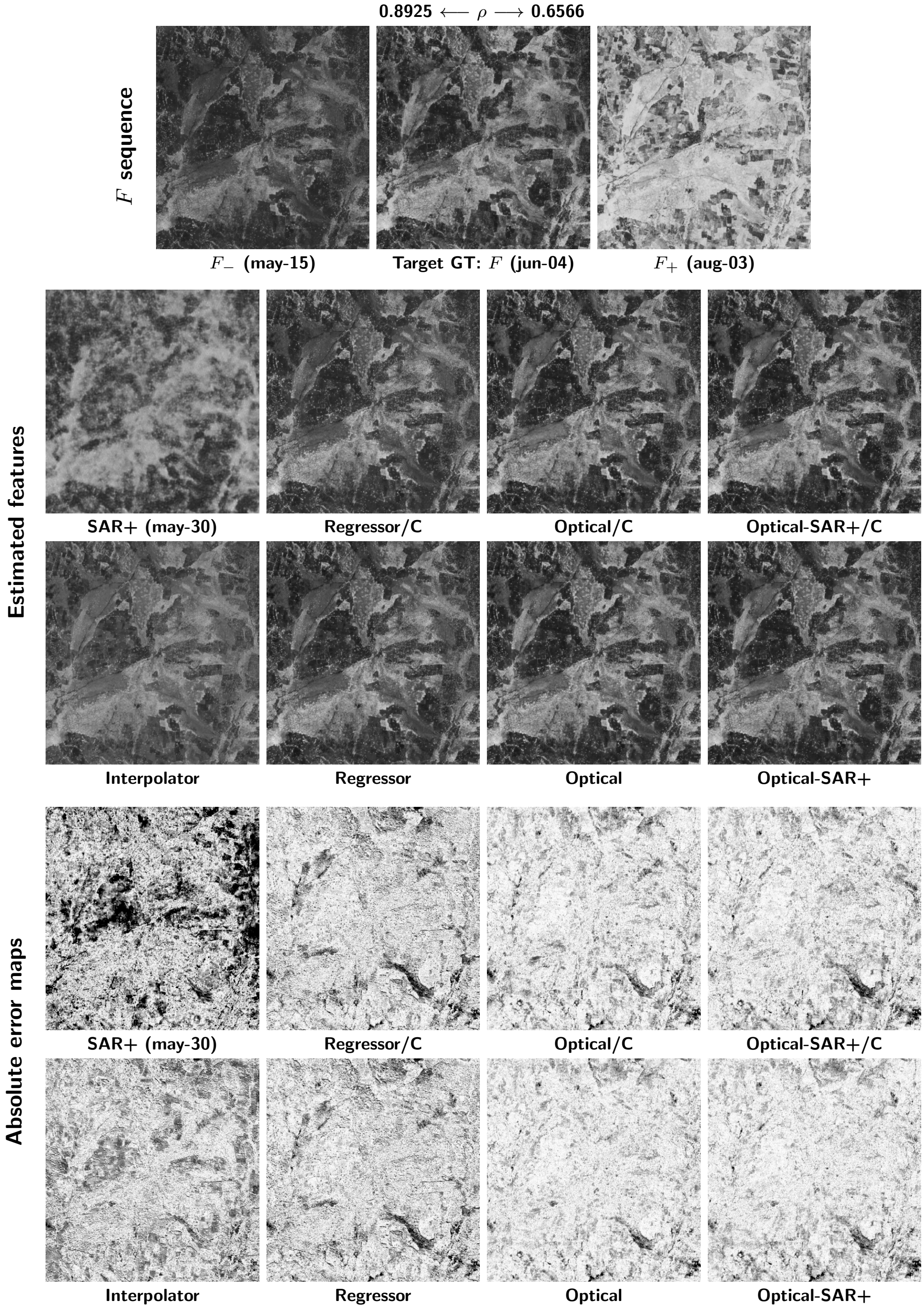

Figure 4. Sample results for the jun-04 target date. Top row: previous, target, and next NDVI maps of the crop selected for testing. Second/third rows: NDVI maps estimated by causal/non-causal methods. Last two rows: corresponding absolute error images.

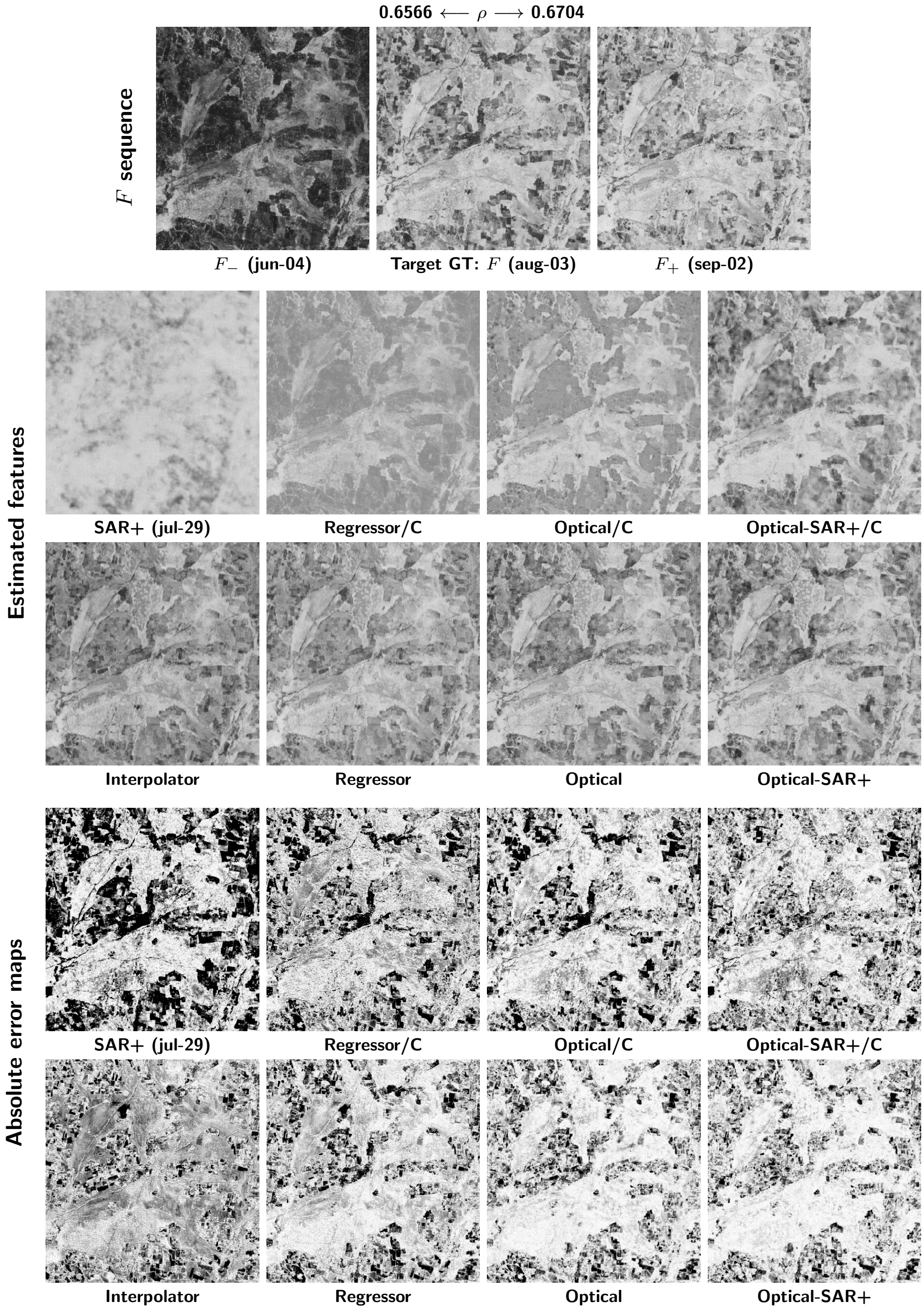

Figure 5. Sample results for the aug-03 target date. Top row: previous, target, and next NDVI maps of the crop selected for testing. Second/third rows: NDVI maps estimated by causal/non-causal methods. Last two rows: corresponding absolute error images.

These changes can be easily appreciated in the images, shown in the top row of Fig.4 and Fig.5, respectively. In both figures, the results of most of the methods described before are reported, omitting less informative cases for the sake of clarity. To allow easy interpretation of results, images are organized for increasing complexity from left to right, with causal and non-causal versions shown in the second and third row, respectively. As only exception, the first column shows results for SAR+ and non-causal interpolator. Moreover, in the last two rows, the corresponding absolute error images are shown, suitably magnified, with the same stretching and reverse scale (white means no error) for better visibility.

For jun-04, the estimation task is much simplified by the availability of the highly correlated may-15 image. Since this precedes the target, causal estimators work almost as well as non-causal ones. Moderate gradual improvements are observed going from left to right. Nonetheless, by comparing the first (interpolator) and last (Optical-SAR+) non-causal solutions, a significant accumulated improvement can be perceived, which becomes obvious in the error images. In this case, even the SAR-only estimate is quite good, and the inclusion of SAR data (third to fourth column) provides some improvements.

For the aug-03 image, the task is much harder, no good predictor images are available, especially the previous image, 60 days old. In these conditions, there is clear improvement when going from causal to non-causal methods, even more visible in the error images. Likewise, the left-to-right improvements are very clear, both in the predicted images (compare for example the sharp estimate of Optical-SAR+ with the much smoother output of the regressor) and in the error images, which become generally brighter (smaller errors) and with fewer black patches. In this case, the SAR-only estimate is very poor. Still, the inclusion of SAR data in the CNN-based estimators provides visible improvements.

6. Conclusions

We have proposed and analyzed CNN-based methods for the estimation of spectral features when optical data are missing. Several models have been considered, causal and non-causal, single-sensor and joint-sensor, to take into account various situations of practical interest. Validation has been conducted with reference to NDVI maps, using Sentinel-1 and Sentinel-2 time-series, but the proposed framework is quite general, and can be readily extended to the estimation of other spectral features. In all cases, the proposed methods outperform largely the conventional references, especially in the presence of large temporal gaps. Besides proving the potential of deep learning for remote sensing, experiments have shown that SAR images can be used to obtain a meaningful estimate of spectral indexes when other sources of information are not available,

Such encouraging results suggest further investigation on these topics. First of all, very deep CNN architectures should be tested, as they proved extremely successful in other fields. However, this requires the creation of a large representative dataset for training. In addition, more advanced deep learning solutions for generative problems should be considered, such as the recently proposed Generative Adversarial Networks [56]. Finally, cross-sensor estimation from SAR data is a stimulating research theme, and certainly deserves further study.

Supplementary Materials: Supplementary Materials: Software and implementation details will be made available online at http://www.grip.unina.it/ to ensure full reproducibility.

Author Contributions: Author Contributions: G. Scarpa has proposed the research topic, written the paper, and coordinated activities. M. Gargiulo and A. Mazza have equally contributed to develop and implement the proposed solutions, and validate them experimentally. R. Gaetano has provided and preprocessed the dataset, and contributed ideas from an application-oriented perspective.

Bibliography

1. Pohl, C.; Genderen, J.L.V. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing* **1998**, *19*, 823–854.

2. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200.
3. Gaetano, R.; Amitrano, D.; Masi, G.; Poggi, G.; Ruello, G.; Verdoliva, L.; Scarpa, G. Exploration of Multitemporal COSMO-SkyMed Data via Interactive Tree-Structured MRF Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 2763–2775.
4. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by Convolutional Neural Networks. *Remote Sensing* **2016**, *8*, 594.
5. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 639–643.
6. Gaetano, R.; Moser, G.; Poggi, G.; Scarpa, G.; Serpico, S.B. Region-Based Classification of Multisensor Optical-SAR Images. IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, 2008, Vol. 4, pp. IV – 81–IV – 84.
7. Reiche, J.; Souza, C.M.; Hoekman, D.H.; Verbesselt, J.; Persaud, H.; Herold, M. Feature Level Fusion of Multi-Temporal ALOS PALSAR and Landsat Data for Mapping and Monitoring of Tropical Deforestation and Forest Degradation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2013**, *6*, 2159–2173.
8. Errico, A.; Angelino, C.V.; Cicala, L.; Persechino, G.; Ferrara, C.; Lega, M.; Vallario, A.; Parente, C.; Masi, G.; Gaetano, R.; Scarpa, G.; Amitrano, D.; Ruello, G.; Verdoliva, L.; Poggi, G. Detection of environmental hazards through the feature-based fusion of optical and SAR data: a case study in southern Italy. *International Journal of Remote Sensing* **2015**, *36*, 3345–3367.
9. Das, M.; Ghosh, S.K. Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data. *IEEE Geosci. Remote Sensing Lett.* **2016**, *13*, 1984–1988.
10. Sukawattanavijit, C.; Chen, J.; Zhang, H. GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 284–288.
11. Ma, W.; Wen, Z.; Wu, Y.; Jiao, L.; Gong, M.; Zheng, Y.; Liu, L. Remote Sensing Image Registration With Modified SIFT and Enhanced Feature Matching. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 3–7.
12. Clerici, N.; Calderón, C.A.V.; Posada, J.M. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *Journal of Maps* **2017**, *13*, 718–726.
13. Fauvel, M.; Chanussot, J.; Benediktsson, J.A. Decision Fusion for the Classification of Urban Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* **2006**, *44*, 2828–2838.
14. Márquez, C.; López, M.I.; Ruisánchez, I.; Callao, M.P. FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. *Talanta* **2016**, *161*, 80 – 86.
15. Waske, B.; van der Linden, S. Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion. *IEEE Transactions on Geoscience and Remote Sensing* **2008**, *46*, 1457–1466.
16. Reiche, J.; de Bruin, S.; Hoekman, D.; Verbesselt, J.; Herold, M. A Bayesian approach to combine Landsat and ALOS PALSAR time series for near real-time deforestation detection. *Remote Sensing* **2015**, *7*, 4973–4996.
17. Du, P.; Liu, S.; Xia, J.; Zhao, Y. Information fusion techniques for change detection from multi-temporal remote sensing images. *Information Fusion* **2013**, *14*, 19 – 27.
18. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. CNN-based Pansharpening of Multi-Resolution Remote-Sensing Images. Joint Urban Remote Sensing Event 2017; , 2017.
19. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-adaptive CNN-based pansharpening. *ArXiv e-prints* **2017**, [arXiv:cs.CV/1709.06054].
20. Gaetano, R.; Masi, G.; Poggi, G.; Verdoliva, L.; G., S. Marker controlled watershed based segmentation of multi-resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1987–3004.
21. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. *Remote Sensing* **2016**, *8*, 354.
22. Zanetti, M.; Bruzzone, L. A Theoretical Framework for Change Detection Based on a Compound Multiclass Statistical Model of the Difference Image. *IEEE Transactions on Geoscience and Remote Sensing* **2017**.
23. Liu, W.; Yang, J.; Zhao, J.; Yang, L. A Novel Method of Unsupervised Change Detection Using Multi-Temporal PolSAR Images. *Remote Sensing* **2017**, *9*, 1135.

24. Han, Y.; Bovolo, F.; Bruzzone, L. Segmentation-Based Fine Registration of Very High Resolution Multitemporal Images. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 2884–2897.
25. Chierchia, G.; Gheche, M.E.; Scarpa, G.; Verdoliva, L. Multitemporal SAR Image Despeckling Based on Block-Matching and Collaborative Filtering. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 5467–5480.
26. Maity, S.; Patnaik, C.; Chakraborty, M.; Panigrahy, S. Analysis of temporal backscattering of cotton crops using a semiempirical model. *IEEE Transactions on Geoscience and Remote Sensing* **2004**, *42*, 577–587.
27. Manninen, T.; Stenberg, P.; Rautiainen, M.; Voipio, P. Leaf Area Index Estimation of Boreal and Subarctic Forests Using VV/HH ENVISAT/ASAR Data of Various Swaths. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 3899–3909.
28. Borges, E.F.; Sano, E.E.; Medrado, E. Radiometric quality and performance of TIMESAT for smoothing moderate resolution imaging spectroradiometer enhanced vegetation index time series from western Bahia State, Brazil. *Journal of Applied Remote Sensing* **2014**, *8*, 083580–083580.
29. Zhang, H.; Lin, H.; Li, Y. Impacts of Feature Normalization on Optical and SAR Data Fusion for Land Use/Land Cover Classification. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 1061–1065.
30. Man, Q.; Dong, P.; Guo, H. Pixel-and feature-level fusion of hyperspectral and lidar data for urban land-use classification. *International Journal of Remote Sensing* **2015**, *36*, 1618–1644.
31. Lu, M.; Chen, B.; Liao, X.; Yue, T.; Yue, H.; Ren, S.; Li, X.; Nie, Z.; Xu, B. Forest Types Classification Based on Multi-Source Data Fusion. *Remote Sensing* **2017**, *9*, 1153.
32. Pal, S.K.; Majumdar, T.J.; Bhattacharya, A.K. ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation. *ISPRS Journal of Photogrammetry and Remote Sensing* **2007**, *61*, 281–297.
33. Bolten, J.D.; Lakshmi, V.; Njoku, E.G. Soil moisture retrieval using the passive/active L- and S-band radar/radiometer. *IEEE Transactions on Geoscience and Remote Sensing* **2003**, *41*, 2792–2801.
34. Baghdadi, N.N.; Hajj, M.E.; Zribi, M.; Fayad, I. Coupling SAR C-Band and Optical Data for Soil Moisture and Leaf Area Index Retrieval Over Irrigated Grasslands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 1229–1243.
35. Santi, E.; Paloscia, S.; Pettinato, S.; Entekhabi, D.; Alemohammad, S.H.; Konings, A.G. Integration of passive and active microwave data from SMAP, AMSR2 and Sentinel-1 for Soil Moisture monitoring. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 5252–5255.
36. Addabbo, P.; Focareta, M.; Marcuccio, S.; Votto, C.; Ullo, S.L. Land cover classification and monitoring through multisensor image and data combination. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 902–905.
37. Jelének, J.; Kopačková, V.; Koucká, L.; Mišurec, J. Testing a Modified PCA-Based Sharpening Approach for Image Fusion. *Remote Sensing* **2016**, *8*, 794.
38. Bisquert, M.; Bordogna, G.; Boschetti, M.; Poncelet, P.; Teisseire, M. Soft Fusion of heterogeneous image time series. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, 2014, pp. 67–76.
39. Moran, M.S.; Hymer, D.C.; Qi, J.; Sano, E.E. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. *Agricultural and Forest Meteorology* **2000**, *105*, 69 – 80.
40. Wang, Q.; Blackburn, G.A.; Onojeghuo, A.O.; Dash, J.; Zhou, L.; Zhang, Y.; Atkinson, P.M. Fusion of Landsat 8 OLI and Sentinel-2 MSI Data. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 3885–3899.
41. Haas, J.; Ban, Y. Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping. *Remote Sensing Applications: Society and Environment* **2017**, *8*, 41 – 53.
42. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; Koetz, B. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing* **2015**, *7*, 12356–12379.
43. ESA. ESA Sentinel Application Platform (SNAP) software. http://step.esa.int/main/toolboxes/snap, (accessed on 13 December 2017).
44. THEIA home page. http://www.theia-land.fr, (accessed on 13 December 2017).

45. Hagolle, O.; Huc, M.; Villa Pascual, D.; Dedieu, G. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENμS and Sentinel-2 Images. *Remote Sensing* **2015**, *7*, 2668–2691.

46. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* **2017**, *26*, 3142–3155.

47. Dong, C.; Loy, C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2016**, *38*, 295–307.

48. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 2015, pp. 3431–3440.

49. Zhang, N.; Donahue, J.; Girshick, R.; Darrell, T. Part-Based R-CNNs for Fine-Grained Category Detection. Proceedings of European Conference on Computer Vision, 2014.

50. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012, pp. 1106–1114.

51. Jiao, L.; Liang, M.; Chen, H.; Yang, S.; Liu, H.; Cao, X. Deep Fully Convolutional Network-Based Spatial Distribution Prediction for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 5585–5599.

52. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016. http://www.deeplearningbook.org.

53. Sutskever, I.; Martens, J.; Dahl, G.E.; Hinton, G.E. On the importance of initialization and momentum in deep learning. *ICML (3)* **2013**, *28*, 1139–1147.

54. Cireşan, D.C.; Gambardella, L.M.; Giusti, A.; Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In NIPS, 2012, pp. 2852–2860.

55. Orfeo Toolbox: Temporal gap-filling. http://tully.ups-tlse.fr/jordi/temporalgapfilling, (accessed on 13 December 2017).

56. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*; 2014; pp. 2672–2680.