

Article

Mapping of Fluvial Morphological Units from Sentinel-1 Data Using a Deep Learning Approach

Massimiliano Gargiulo ^{1,*}, Carmela Cavallo ² and Maria Nicolina Papa ²

¹ Earth Observation Systems and Application (AOTD), Italian Aerospace Research Centre (CIRA), 81043 Capua, Italy

² Department of Civil Engineering, University of Salerno, 84084 Fisciano, Italy; ccavallo@unisa.it (C.C.); mnpapa@unisa.it (M.N.P.)

* Correspondence: m.gargiulo@cira.it; Tel.: +39-0823-623164

Abstract: The identification of ongoing evolutionary trajectories, the prediction of future changes in the functioning of riverine habitats, and the assessment of flood-related risks to human populations all depend on regular hydro-morphological monitoring of fluvial settings. This paper focuses on the satellite monitoring of river macro-morphological units (assemblages of water, sediment, and vegetation units) and their temporal evolution. In particular, we develop a deep-learning semantic segmentation method using Synthetic Aperture Radar (SAR) Sentinel-1 dual-polarized data. The methodology is executed and tested on the Po River, located in Italy. The training of a relatively deep convolutional neural network requires a large amount of ground-truth data, which is often limited and challenging to acquire. To address this limitation, the dataset is augmented using a random forest (RF) classification algorithm. RF parameters are trained with both Sentinel-1 (S1) and Sentinel-2 (S2) data. The RF classification algorithm is very robust and achieves excellent performance. To overcome the limitation linked with the scarce availability of contemporary acquisition by S1 and S2 sensors, the deep learning (DL) model is trained by using only the Sentinel-1 input data and the ground truth from the RF result. The proposed approach achieves promising results in the classification of water, sediments, and vegetation along rivers such as the Italian Po River with low computational costs and no concurrency constraints between S1 and S2.

Keywords: Sentinel-1 Synthetic Aperture Radar (SAR) data; Sentinel-2 data; fluvial satellite monitoring; image segmentation; convolutional neural network; google earth engine;



Academic Editor: Won-Ho Nam

Received: 26 November 2024

Revised: 10 January 2025

Accepted: 17 January 2025

Published: 22 January 2025

Citation: Gargiulo, M.; Cavallo, C.; Papa, M.N. Mapping of Fluvial Morphological Units from Sentinel-1 Data Using a Deep Learning Approach. *Remote Sens.* **2025**, *17*, 366. <https://doi.org/10.3390/rs17030366>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fluvial geomorphological investigations have gained increasing importance over the past decade, emerging as an essential tool for the sustainable management of fluvial environments [1]. These studies have proven particularly effective in assessing flood risks and providing crucial data to understand hydro-morphological variations, both natural and anthropogenic, including those induced by restoration interventions. There is a strong link between river physical forms and biological conditions [2]. Geomorphic units (e.g., riffles, pools, bars, islands) constitute essential habitats for fluvial biota. For example, they offer refuge from disturbance or predation, spawning spots, etc. Therefore, continuous monitoring of geomorphic units is essential for gathering insights into habitat availability and dynamics. This study focuses on mapping macro-morphological units, defined as aggregates of one or more units belonging to the same typology. These units, characterized by common textural features such as aquatic portions, sediments, and vegetation, provide a

preliminary level of characterization of the river environment [3,4]. Wet channel, sediment bars, and vegetated areas are the macro-morphological units identified in this work.

To ensure effective and long-lasting river management, it is essential to conduct continuous geomorphological observations over time. However, these investigations can become quite complex and costly if carried out in the field. Remote sensing (RS) technologies play a key role in addressing numerous environmental and societal challenges. In particular, one field of application is hydrology and river monitoring, which has become increasingly critical due to climate change. Rivers are highly sensitive to sudden variations, such as floods, as well as medium-term changes like droughts and the drying of non-perennial riverbeds. The fusion of remote sensing data from multiple sources has proven invaluable in this context. By leveraging the enhanced spatial, spectral, and temporal resolutions offered by data fusion, the monitoring of rapid and instantaneous phenomena has become more effective. Earth observation is nowadays supported by a huge number of satellites [5,6], providing precious data for various applications in fluvial hydro-morphology, such as monitoring and quantifying hydro-morphological changes [7–10], and analyzing the dynamics of riparian vegetation [11]. The extensive coverage provided by satellite systems ensures that even the most remote and inaccessible river regions can be monitored effectively. This global reach is crucial for creating comprehensive hydro-morphological maps and detecting changes over large areas. Secondly, the frequent revisit time (around 5 days for the Sentinel mission) enables the monitoring of rapid changes and events in river environments, such as floods, sediment transport, and alterations in river courses. This high temporal resolution allows for timely interventions and informed decision-making in water resource management and disaster response. Furthermore, the cost-effectiveness of satellite data is a significant advantage. While traditional field surveys and airborne missions can be expensive and logistically challenging, satellite observations offer a more economical and efficient alternative. The ability to repeatedly capture data without the need for extensive on-ground infrastructure reduces both the financial and environmental costs associated with river monitoring. Multispectral and Synthetic Aperture Radar (SAR) imagery are commonly used to provide information about vegetation, risk management, river water mapping, and much more. Multispectral imagery such as MODIS, SPOT, Landsat-8, and Sentinel-2 are readily available and easy to process, but are often disturbed by clouds and cloud shadows. In contrast, SAR imagery from satellites such as Envisat, RADARSAT, and Sentinel-1 SAR can penetrate clouds, thus providing great advantages such as, for example, the possibility of observing hydro-morphological variations immediately after flood events, when cloud cover often limits the use of multispectral data. Numerous studies have utilized multispectral imagery to monitor fluvial hydro-morphology. Multispectral images have already been used to study various river morphologies and dynamics [10,12]. For example, the Sentinel-2 images are used to delimit water, vegetation, and dry sediment, detect vegetation growth, and delimit active channels on four Italian rivers: the Po River, the Sesia River, the Paglia River, and the Bonamico River [13]. The Sentinel-2 images have been further used to study alternate sandbars movement on the Vistula River in Poland [14]. In [9], a combination of Landsat and Sentinel-2 imagery is used to assess the river dynamics and changes over time in the Po River. Regarding multispectral images, classification methods based on multispectral indices have been successfully applied to rivers with a width-to-resolution ratio in the order of 3:1 or greater [10,15–18]. According to our knowledge, studies that have utilized SAR data for monitoring fluvial hydro-morphology rather than flooding events are limited [19]. SAR data are more complex to interpret without a strong background in electromagnetics. The monitoring of rivers is also affected by difficulties linked with the possible presence of hill shadows and water surface ripples. Hill shadows are particularly problematic in mountainous regions, where terrain-induced shadows and complex water

dynamics complicate accurate water surface detection. Surface ripples generate higher backscatter that can cause confusion between water surfaces and non-water surfaces [20]. Only a few approaches have been developed for extracting geomorphic macro-units from SAR satellite data. For instance, SAR data are used to analyze river morphology [20] and to assess braided river-bed dynamics over time [19]. Among the most promising sensors for fluvial monitoring, the Sentinel-1 and Sentinel-2 satellites of the ESA Copernicus program, provide data with a higher revisit time than the other satellites at finer spatial resolution, and at a higher spatial resolution than other missions, such as Landsat, MODIS, and so on [21–24]. Sentinel-2 is an advanced multispectral sensor that captures data in 13 spectral bands with a temporal resolution of 5 days and a spatial resolution of up to 10 m. On the other hand, Sentinel-1, equipped with a C-band, is one of the most recent and advanced SAR sensors, with a high temporal and spatial resolution capability, providing data with temporal and spatial resolutions of up to 6 days and 10 m, respectively. The main limitation of relying on a single sensor lies in the unavoidable trade-off between spatial and temporal resolution. Image blending, achieved by combining multiple data sources, is a widely used method for enhancing temporal resolution [25,26]. In fact, in recent decades, data fusion algorithms based on the use of multispectral and SAR data have gained attention in the research community. Thus, a wide range of data fusion studies combine specific bands and polarizations of optical and SAR sensors to gain a more thorough understanding of fluvial processes [27–35]. Data fusion algorithms were developed to fully exploit the potential of S1 and S2 sensors and overcome their limitations [27,36–38]. In the last decade, deep learning (DL) with increasingly complex architectures [39–42] has attracted increasing interest in the computer vision community. Thus, an increasing number of works address data fusion using DL architectures, as they are very effective in global monitoring for a plethora of RS applications [35,43–46]. The growing interest and the increasing number of satellites (and thus data) have encouraged the use of methods based on Convolutional Neural Networks in fluvial monitoring. We present a new deep learning-based methodology that leverages the temporal coverage of Sentinel-1 and the segmentation accuracy of S1 and S2 data fusion to classify fluvial macro-morphological units [3], such as channels, wet areas, sediments, and vegetation. The suggested method focuses on two main challenges in dynamic fluvial environments: (i) the constraints caused by the non-contemporaneity of S1 and S2 datasets, and (ii) the dependency on large ground-truth datasets for deep learning model training. To overcome these obstacles, the suggested method uses Random Forest (RF)-generated semantic segmentation maps (obtained by the use of S1 and S2 data) as training labels for a deep learning architecture that has only been trained on S1 dual-polarized data. As a result, the method demonstrates high-frequency monitoring capabilities, limited only by Sentinel-1's six-day revisit time, which makes it especially appropriate for monitoring rapid hydro-morphological changes.

The paper is organized as follows. Section 2.1 provides an overview of the case study, including the area of investigation, and details the S1 and S2 data used in this research. In Section 3, we explain the general workflow of the proposed method, and in particular the proposed deep learning architecture used for the semantic segmentation training. Further, we describe the definition of a specific loss function and the implementation of the training phase. In Section 4, the results are shown in detail with visual and numerical assessments. Section 5 focuses on evaluating the performance. Section 6 summarizes the main findings and outlines the future directions of this research.

2. Study Area and Satellite Imagery Used

2.1. Study Area Overview

To illustrate our approach, we chose two downstream sections of Italy's Po River to train our model and assess the outcomes obtained through the classification technique. The Po River is the longest watercourse in Italy (spanning approximately 652 km), is located in northeast Italy, and flows from the Cozie Alps to the Adriatic Sea, draining a catchment area of approximately 74,091 km² (Figure 1a). This research primarily concentrated on a 40 km reach between Boretto and Borgoforte (Figure 1b), with an average bank width of 200–500 m, and another section close to Boschina Island located downstream of the Revere railway bridge, near Ostiglia village (Figure 1c). Within the analyzed reaches, the river exhibits a predominantly single-thread channel with a planform ranging from straight to meandering. The most frequent geomorphic features include point bars, mid-channel bars, and chute channels. The alluvial bed consists of well-sorted coarse sand, with a d_{50} of approximately 0.4 mm [9]. Figure 1b shows the RGB representation of the Borgoforte area, used as a training dataset, together with a zoomed image (Figure 1c) of the test area (Ostiglia, Italy). The test area was not considered in the training phase because we wanted to validate the model in other areas with similar fluvial morphological conditions. The dates were selected for cloud-free conditions, and hence with available reference ground-truth for any possible optical feature.

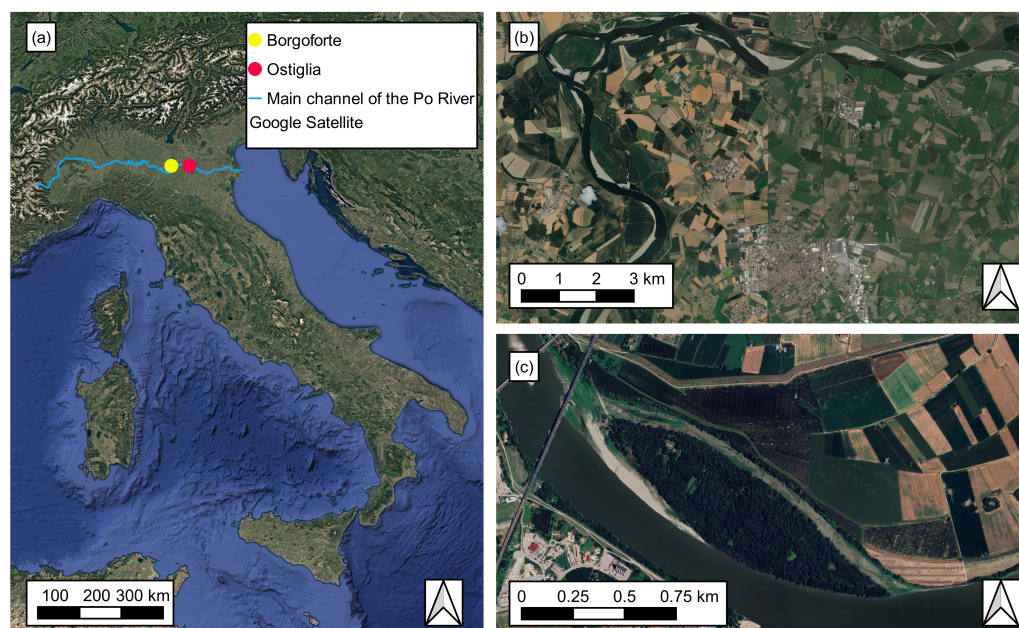


Figure 1. (a) Main channel of the Po River, highlighting the locations of the cases study. (b) View of Boretto-Borgoforte area. (c) View of Ostiglia island.

2.2. Sentinel-1 and Sentinel-2 Data

In recent years, the GEE has been increasingly used in the remote sensing community [47], thanks to its free access to satellite data and specific algorithms useful for various applications [48]. Its data collection over four decades around the world provides valuable opportunities for global-scale applications in hydrology [49], channel change detection [18], and other areas [50]. The data collections available on GEE are from the complete Landsat series, S1, S2, Sentinel-3, and others [51]. In this study, we used Copernicus Sentinel data, specifically the S1 and S2 datasets, which were downloaded and pre-processed using Google Earth Engine (GEE), as shown in Figure 2. The GEE presents the S2 at different levels: the Level-1C (L1C), provided in Top Of Atmosphere (TOA) reflectances, and the

used Level-2A (L2A), provided in Bottom Of Atmosphere (BOA) reflectances, obtained by the atmospheric correction from the L1C (see Figure 2). The GEE provided the Sentinel-1 data processed with the S1 Toolbox using the following processing chain: thermal noise removal, radiometric calibration, and terrain correction using SRTM 30 m. Finally, the terrain-corrected product was converted to decibels. Then, simple despeckling was applied to each S1 polarization based on the mean filter with a Gaussian kernel. After the described processing, the despeckled S1 backscatter in VV and VH polarizations was used as input to the Random Forest algorithm (as shown in Figure 2). Conversely, we did not consider any despeckling algorithms on the S1 input when we developed the deep-learning solution. As shown in [44], the usage of speckled S1 images allowed us to save time and still achieve excellent classification/segmentation results. The information about the data provided by the two missions used is reported in Tables 1 and 2.

Table 1. Overview of the Sentinel-1 SAR dual-polarized data.

| Characteristics | Sentinel-1A Data |
|--------------------|------------------|
| Acquisition orbit | Descending |
| Imaging mode | IW |
| Imaging frequency | C-band (5.4 GHz) |
| Polarization | VV, VH |
| Data Product | Level-1 GRDH |
| Spatial Resolution | 10-m |

Table 2. Overview of the Sentinel-2 Multispectral data.

| Spectral Bands (Bands Number) | Wavelength Range [μm] | Spatial Resolution [m] |
|--|-----------------------|------------------------|
| Blue (2), Green (3), Red (4), and NIR (8) | 0.490–0.842 | 10 |
| Vegetation Red Edge (5, 6, 7, 8A) and SWIR (11, 12) | 0.705–2.190 | 20 |
| Coastal Aerosol (1), Water Vapour (9), and SWIR (10) | 0.443–1.375 | 60 |

Depending on sediment size, the SAR signal returns a different response; if sediment size exceeds a threshold, the surface is rough, and vice versa smooth. We considered the threshold h defined in [52] and described below:

$$h < \frac{\lambda}{25 \cdot \theta_{inc}} \quad (1)$$

where λ is equal to 5.6 cm (C-band) and θ_{inc} is approximately equal to 38° in the considered area. In our case, the condition was $h < 2$ mm, that is, the sediments of the Italian Po River, at approximately 1 mm, satisfied the condition of a smooth surface. We conclude that in the smooth state, both the VV and the VH backscattering values are lower compared to the rough state (see Figure 3). However, the VH/VV backscattering ratio of the smooth state is higher than that of water, indicating a significant difference in the scattering behaviour between the two states [53].

The GEE provided the S2 data at the L2A and L1C levels [54]. We used the L2A level data, obtained by processing the S2 L1C with atmospheric correction through the Sen2cor toolbox, as shown in Figure 2. To create a dataset for supervised Random Forest algorithms, we manually selected a ground truth corresponding to irregular 100 polygons drawn in GEE. The 100 chosen polygons were manually selected to cover specific geomorphological

features, such as sediment bars, wet channel, and vegetated areas. The total area of the polygons was equal to 2.16 km². Specifically, each polygon was associated with one of three specific classes (water, sediment, and vegetation). These polygons were drawn for a day with acquisitions of S1 and cloudless S2, benefiting from expert knowledge that ensured the representativeness of the geomorphic diversity of the river.

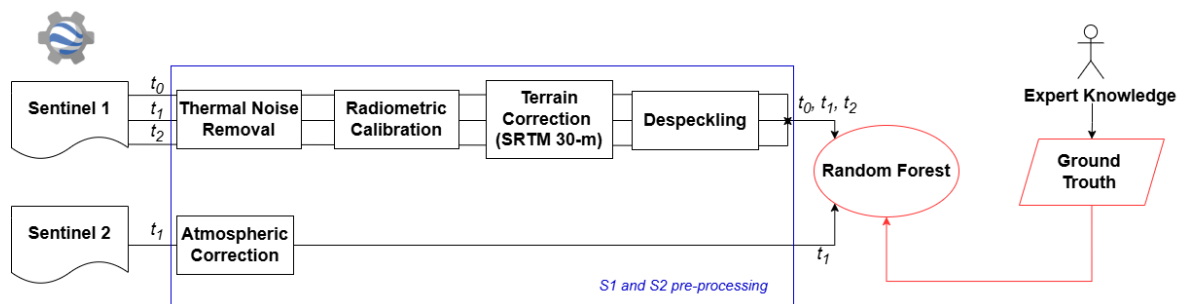


Figure 2. The processing chain on S1 and S2 data to train the Random Forest model using expert knowledge on fluvial morphological conditions.

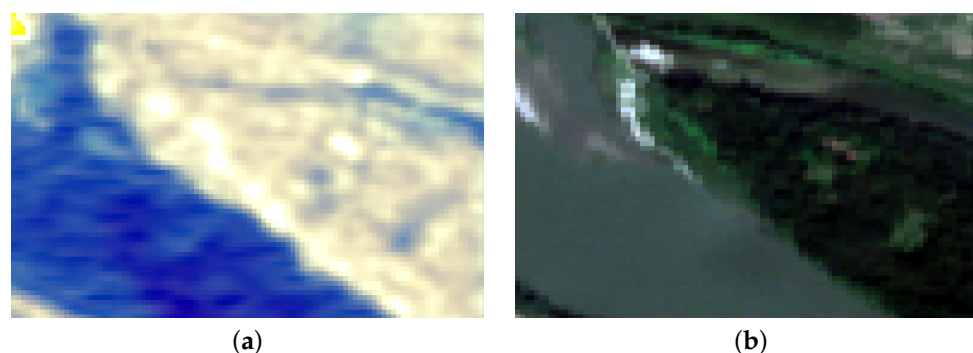


Figure 3. (a) false colour RGB visualization of Sentinel-1 (R: VH, G: VV, B: $\frac{VH}{VV}$) and (b) RGB of Sentinel-2 on site under investigation.

3. Proposed Method

In this paper, we propose an automatic semantic segmentation method for land cover classes along river corridors. The proposed solution, in particular, removes the constraint of concurrence between S1 and S2, which is too stringent a limitation. This is particularly relevant in tropical areas or when monitoring rapid phenomena, such as hydro-morphological changes in rivers, where cloud cover limits the availability of Sentinel-2 data. For this reason, we trained the DL architecture so that the presence of the S2 data was not required during the testing phase. However, we did apply the S2 data when using ground truth in Random Forest training and testing. This deep learning strategy has been used in previous studies [44] to extract information from lentic wetlands, whereas in the current work it was employed to extract data on water, vegetation, and sediments in dynamic environments such as rivers. The proposed workflow was composed of a three-step solution:

1. We created a manual ground truth (polygons shown in Figure 4) from the observation of Sentinel-2 data and the corresponding higher spatial resolution data (on the same day). These polygons were used to train the Random Forest model, for which the input was composed of the contemporary data from S1 and S2 (as shown in Figure 2);
2. We trained a deep neural network starting from only the Sentinel-1 input data and, as a reference, the segmentation obtained from Random Forest (refer to Figure 5);
3. We tested the trained deep neural network (see Figure 6).

Compared with our previous solution [44], the novelties introduced in this paper are:

- a novel deep learning architecture that means the U-Nets cascade;
- a three-terms loss function;
- a different segmentation map as reference, obtained by using the random forest algorithm;
- a case study that best matches the characteristics of the proposed solution.

Given the structure obtained as a cascade of three U-Nets and the presence of three terms of the loss function that carry an implicit denoising capability as in the work [55], known as Noise2Noise (N2N), the proposed network, hereafter, will be called CUN2Net (Cascade Unet Noise2Noise net).



Figure 4. Polygons, manually extracted in GEE, and evaluated in Random Forest Algorithm. In the image, green polygons is for the vegetation, yellow ones for the sediments, and blue for the water.

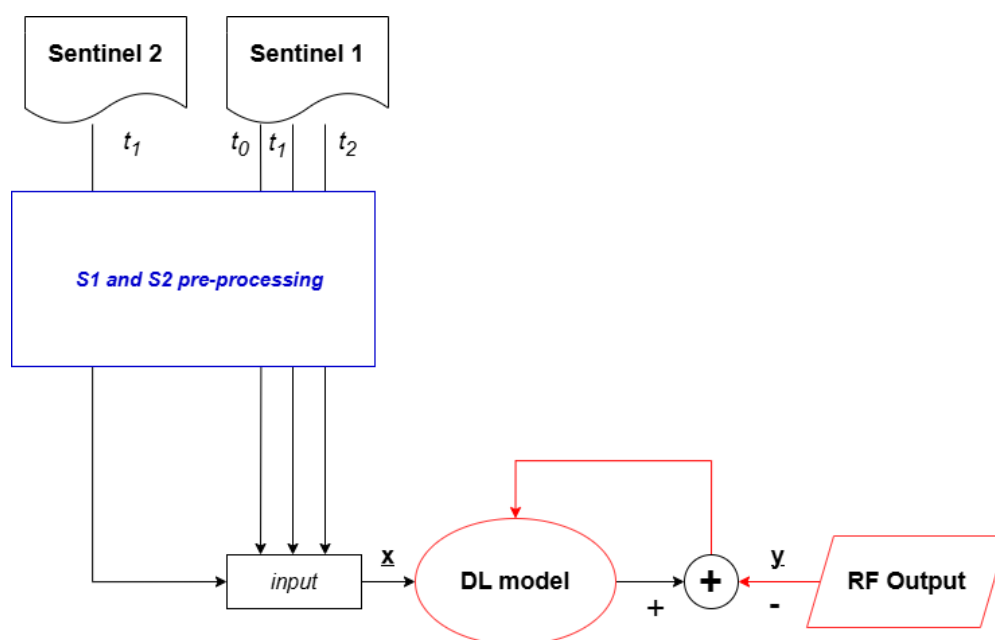


Figure 5. The general workflow of the proposed method to train the proposed CUN2Net architecture. The S1 and S2 pre-processing block is the same as in Figure 2.

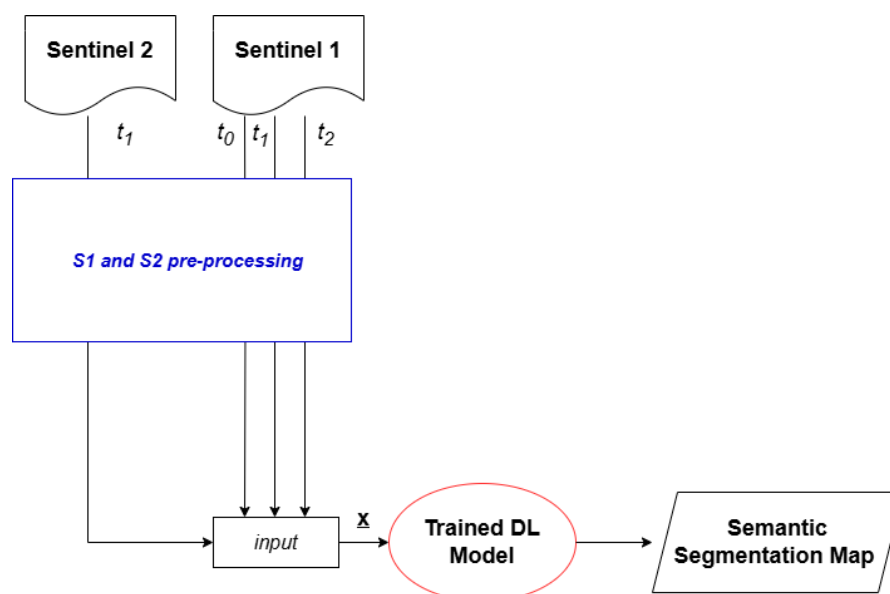


Figure 6. The general workflow to use and test the trained CUN2Net architecture. The S1 and S2 pre-processing block is the same as in Figure 2.

3.1. Dataset Generation with Random Forest Model (Step I)

The Random Forest (RF) algorithm is a common pixel classification approach [56,57]. To benefit from the S1 and S2 data, we used to train the RF in the contemporaneous presence of S1 and S2 for a specific date. However, the simultaneousness of the two datasets was a limitation. To overcome this drawback, we used the output of an RF-based segmentation map to move forward with a deep learning solution trained with only S1 dual-polarized data. However, this approach represented a supervised algorithm, and required many examples from which it was able to learn helpful information in remote sensing applications. In our context, we used S2 images to provide the necessary examples by drawing polygons on the Google Earth Engine, facilitated by expert knowledge. Just as for all supervised algorithms, the RF required examples from which to learn, so we built pairs (input, target) that allowed the Random Forest to reproduce similar classifications as in example pairs, but in areas never seen before. In our case, the input stack was composed of the Sentinel-1 dual-polarized data and the 20 m and 10 m Sentinel-2 bands. The S1 dual-polarized data were reprojected into the geographical raster of the S2 ones at 10 m. The central date (t_2) is reported in Table 3 for S2, while for S1 it was anticipated or postponed by a maximum of 2 days with respect to the S2 data. The dates t_1 and t_3 considered were thus 6 days from the S1 date. This input configuration gave better results than other configurations composed of a subset of this configuration. The output was obtained from manually extracted information, which means the polygons drawn in Google Earth Engine. For each polygon, we associated a specific class between the classes mentioned earlier (water, sediments, and vegetation), and thus a single specific class corresponded to each input pixel. The RF model was trained using 80% of the drawn polygons, while 20% were used for testing. Figure 2 illustrates the processing chain applied to S1 and S2 data used to train the Random Forest model, incorporating polygons drawn (ground truth). To better understand the diversity of the input data, some examples of polygons are shown in Figure 4. After the specific training, achieved by using 300 trees and no limit for the maximum number of leaf nodes in each tree, we tested the trained model on other areas. We considered a numerical inspection, as shown in Table 4, to determine the goodness and generalization capability of this method with respect to other machine learning algorithms: Support Vector Machine (SVM) and Classification And Regression Trees (CART) [58]. The overall accuracy (OA) of the RF (with the usage of S1 and S2) was equal to 0.9991, and was higher than other

supervised classification algorithms. Moreover, the RF was trained only on S1 (first line in the Table 4) with an overall accuracy of 0.9171, and only on S2 (second line in the Table 4) with an OA of 0.9367. The significance of training the RF to use S1 and S2 simultaneously to provide superior results is confirmed by this further analysis. Similar considerations can be made for the other algorithms.

Table 3. The used datasets for training and testing phases, both for Random Forest and deep learning approaches. For S1, we also considered the t_1 and t_3 dates (6 days before and after the t_2 due to the temporal resolution of S1).

| Datasets | Year | S2 and S1 Date (t_2) | Considered Area | Data Type (Size) |
|-------------|--------------|-------------------------------------|-----------------|------------------------------------|
| RF Training | 2019 | 09-11 | Po | Polygons (80) |
| RF Testing | 2018 | 12-10 | Po | Polygons (20) |
| DL Training | 2018 2019 | 09-16; 12-10 01-04; 01-09; 09-11 | Borgoforte | Patches 128×128 (9.5k) |
| DL Testing | 2018 2019 | 09-16; 12-10 01-04; 09-11 | Ostiglia | Patches 128×128 (500) |

Table 4. Comparison of different classification algorithms in terms of accuracy.

| Methods | S1 Data | S2 Data | Accuracy | Reference |
|---------|---------|---------|----------|-----------|
| RF | ✓ | | 0.9171 | [56] |
| RF | | ✓ | 0.9367 | [56] |
| CART | ✓ | ✓ | 0.9980 | [59] |
| SVM | ✓ | ✓ | 0.9797 | [60] |
| RF | ✓ | ✓ | 0.9991 | [56] |

Convinced of the effectiveness of the training, we used the RF output to extensively train a deep neural network (Figure 5). We extended the results to a broader area, and we also included different dates with different conditions. For the RF training, the contemporaneous presence of S1 and S2 for a specific date was the main limitation, and so to overcome this drawback, we moved forward with a deep learning solution trained with only S1 dual-polarized data. It disengaged itself from the need to have simultaneous or contiguous dates of the two sensors because this condition is complicated to obtain due to the unavailability of S2 in the presence of clouds. Therefore, the deep learning method that we proposed exploited only Sentinel-1 dual-polarized data in the input, also available in cloud presence, and in the output, the result from the RF model.

3.2. Deep Learning Architecture (Step II)

The CUN2Net method we propose is built on a supervised deep learning architecture based on the concatenated use of convolutional layers and, in particular, on the use of a U-Net cascade similar to the W-Net strategy described in [44,61]. The core of the proposed architecture is the U-Net architecture [41]. In this work, we consider three sequential U-Nets, one after the other. The U-Net structure consists of contracting and expansive paths, as described in [41]. The blocks of the U-Nets are composed of: (i) the batch normalization layers, (ii) the 3×3 convolutional layer with a Rectified Linear Unit (ReLU), and (iii) the max pooling 2×2 layers. More details on the single parts of the U-Nets are included in [44]. At the end of the first and third U-Net, there is a convolutional layer that produces

a number of feature maps equal to the number of classes considered in our problem (i.e., 3) and, to produce the output, a softmax activation function. At the end of the second U-Net, there is a convolutional layer with a ReLU as an activation function to reconstruct the SAR information (VV and VH polarizations) in the central data (t_2). The output of the blocks at one level in a U-Net part is concatenated with the output of the max pooling layers (as shown in Figure 7) in the following U-Net, which is identical to the first. The first and third U-Nets have the same objective, i.e., image segmentation. On the other hand, the second UNet has the SAR image reconstruction objective. These outputs and objectives are differently weighted in the loss function, as described later. The output of the second UNet is an ancillary and helpful result that we show in visual inspection, but it is not the primary focus of this work. To support the supervised training required for the proposed CUN2Net, we first need to realize input-target (x,y) samples in order to start the training phase. The network is trained using six distinct input stacks, each representing a different combination of VV and VH polarizations, as reported in Table 5 and as already presented in [44]. In the multi-temporal input configuration (CUN2Net in Table 5), we consider the S1 dual-polarized data on three different dates: one is the closest to the target date (1), and the others are the next closest dates, before (0) and after (2) the target date. Hereafter, the closest date is also defined as the central date. The same considerations on the different dates are made for the C4 and C5 configurations in Table 5, but for these configurations, we only consider one polarization at a time. The other two subsequent U-Nets are fed with the output of the previous UNet. The first and third U-Net outputs consist of three-class segmentation maps. Conversely, the second UNet's output is the SAR data considered as input. The strategy of using the speckled data (input) in the output follows the denoising method described in [55].

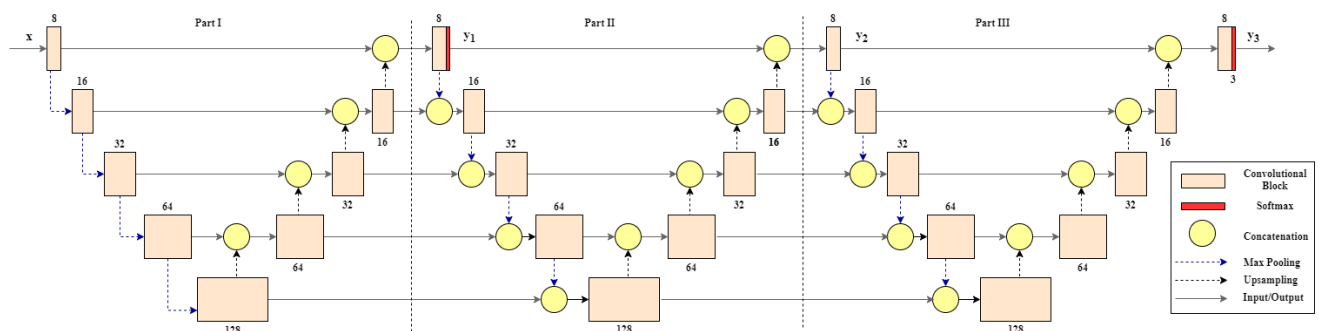


Figure 7. The proposed CUN2Net architecture.

Table 5. Different input stacks are considered in the training phase.

| Configurations | No. Input Bands | Description | Considered Times |
|----------------|-----------------|--------------|------------------|
| C1 | 1 | VH_i | 1 |
| C2 | 1 | VV_i | 1 |
| C3 | 2 | VV_i, VH_i | 1 |
| C4 | 3 | VH_i | 0, 1, 2 |
| C5 | 3 | VV_i | 0, 1, 2 |
| CUN2Net | 6 | VV_i, VH_i | 0, 1, 2 |

Training

After designing the architecture, it is necessary to define an appropriate loss function that must be minimized to enable the learning process. In supervised learning, we mainly distinguish two kinds of problems: the generative and the discriminative. In the discriminative (for instance, segmentation) context, IoU or Jaccard losses are the main choices, due to their simplicity and robustness, but a plethora of losses could be used, as reported in [62].

In the generative context, the L2 or L1 norms are typical choices [63] due to the effectiveness of speeding up training, as observed in [64]. However, in the proposed CUN2Net solution, we define a combined loss that accounts for a segmentation task, specific to our application, and a generative loss for an ancillary task. Specifically, we use an objective function that consists of three terms:

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_{IoU}^1 + \lambda_2 \cdot \mathcal{L}_1 + \lambda_3 \cdot \mathcal{L}_{IoU}^3 \quad (2)$$

where \mathcal{L}_{IoU}^k with $k = 1, 3$ are the terms of the loss based on the Intersection over Union for the first U-Net and the third U-Net, respectively, and the term \mathcal{L}_1 represents the terms of the loss for the second U-Net and is computed on the pixel-wise L1 norm basis. The terms \mathcal{L}_{IoU}^k with $k = 1, 3$ are based on the Intersection over Union (IoU) to be more effective in error backpropagation [65]. The IoU function can be defined as:

$$IoU = \frac{I}{U} = \frac{y \cap \hat{y}}{y \cup \hat{y}} = \frac{TP}{TP + FP + FN} \quad (3)$$

where I is the Intersection and U is the Union, and y and \hat{y} are the reference and the predicted map, respectively. Further, the IoU is also expressed as the combination of True Positive (TP), False Positive (FP), and False Negative (FN). Specifically, the IoU loss is computed on the objective function by averaging over the mini-batch samples at each updating step of the learning process:

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{1}{N} \sum_{n=1}^N \frac{y_n \cap \hat{y}_n}{y_n \cup \hat{y}_n} \quad (4)$$

where N is the batch size during the training phase (equal to 64 in this context), y_n is the n -th reference, and \hat{y}_n is the n -th predicted map, dependent on all the network trainable weights. In the following analyses, the three different terms are weighted with the following values: $\lambda_1 = 0.25$, $\lambda_2 = 0.05$, and $\lambda_3 = 0.7$. The weights were empirically chosen and the higher weight was assigned to the final segmentation task ($\lambda_3 = 0.7$) to prioritize accuracy, while lower weights ($\lambda_1 = 0.25$, $\lambda_2 = 0.05$) were assigned to intermediate tasks for stabilization. Furthermore, in the next section, the importance of each term is explained in more detail. In addition, we adopted the Adam optimizer, implemented in the Tensorflow Python package, and we considered a learning rate of $\eta = 0.002$, and the decay rate of the first and second moments to be $\beta_1 = 0.9$ and $\beta_2 = 0.999$, configured as in [66]. In the training phase, we started the training via Glorot initialization of the weights [67]. Because of the relative lightness of the considered network, we obtained considerable results despite the Glorot initialization of the network weights. In particular, we considered six input configurations (see Table 5) that differed from each other in the composition of the input stack x , while the output y is always the RF-based classification result obtained from Sentinel-1 and Sentinel-2 at the target date. The training phase was performed for just 10 epochs. We stopped training after 10 epochs due to overfitting that occurred when increasing the number of epochs. Each epoch passed over all mini-batches, composed of 64 128×128 input–output samples, over which the training set had been divided. The training dataset consisted of 10k patches (9.5k reserved for the training phase and 500 for the validation phase) derived from S1 images acquired on the reported in Table 3.

3.3. Testing the Model with Classification Metrics (Step III)

We employed metrics derived from the confusion matrix to assess our model's performance in terms of segmentation results. In particular, we considered the accuracy, precision,

recall, and F1-score, which are used to assess the correctness of multi-class classification and segmentation algorithms as reported and defined in [68].

These metrics rely on a ground truth (or reference), which, like the training data, was generated using the RF-based technique from the S1 and S2-L2A products. However, the RF-based classification could pose a limitation during the training phase.

4. Results

In this section, we describe the classification results obtained by the proposed CUN2Net method (Section 4.1). Then, a comparison is made between the proposed solution and some alternative methods from the literature (Sections 4.2–4.4). Performance is evaluated by visual inspection and by the previously defined metrics.

4.1. Numerical and Visual Results

The proposed CUN2Net solution allowed us to obtain encouraging results. Basically, the training data in the best-performing configuration (CUN2Net) consisted of three S1 couples (VV and VH on three different dates) in input, and a segmentation map from the previously described Random Forest algorithm in output. The general idea was to obtain from the standalone S1 dual-polarized data a result comparable to that obtained by combining the information from the S1 and S2 data. The comparison between the Random Forest result and the manually extracted ground truth showed an overall accuracy of 0.9915. Figure 1 shows the RGB representation of the Borgoforte area, used as a training dataset, together with a zoomed image of the test area (Ostiglia, Italy). The test area was completely disjointed from the area considered in the training phase because we wanted to validate the model in other areas with similar fluvial morphological conditions. The dates were selected for cloud-free condition, and hence with available reference ground-truth for any possible optical feature. The Random Forest algorithm in our configuration had the limitation of being able to work only in the presence of near contemporaneity between S1 and S2. However, it provided very robust results for the different couples of S1 and S2 considered as input of the algorithm. Ideally, we wanted the S2 data corresponding to the S1 central date, but clearly, this condition was not easily obtainable, so we trained and tested with the S2 image, which was at most 2 days away from the central date. This temporal shift did not lead to any kind of defection in the use of the Random Forest algorithm, although there was still the constraint of the patchy presence of S2. In fact, the results of Random Forest obtained very high accuracy with respect to ground truth data even when the S2 data was not coincident with the central data of S1. As already underlined, this does not restrict the implementation of the deep learning solution, as it is trained and tested only on S1 images. However, the constraint of the contemporary presence of S1 and S2 limits the application point of view and also the creation of the training dataset. We first compared the results obtained by using as inputs either of the two VV and VH polarizations singularly with the one obtained by the joint use of the two polarizations. The results shown in the first three rows of Table 6 and in Figure 8 clearly show that the use of both polarizations provided the best results, while VV is better than VH when used alone. This behaviour was also confirmed in multi-temporal configurations (see C4 and C5 in Table 6). We can conclude that with three dates and with the two polarizations, we obtained performances that were comparable to a Random Forest segmentation starting from the fusion of the multi-spectral data (S2) and the SAR data (S1). This confirms that reliable results are also obtained by the only use of the S1 dual-polarized data. The multi-temporal configurations, shown in Figure 8, highlight the importance of including temporal information from three dates to improve segmentation performance. This is clearly advantageous, especially in areas where the S2 data is more frequently corrupted by the presence of clouds. Moreover, even

in the absence of clouds, it also removes the problems associated with the non-constant time shifts between the S1 central date and the S2 date. In particular, in all configurations, both single-date and multi-temporal, the greatest ambiguities arise between the water and sediment classes. The multi-temporal version with both polarizations exhibited the best performances in terms of the F1-score for the single classes (Table 6).

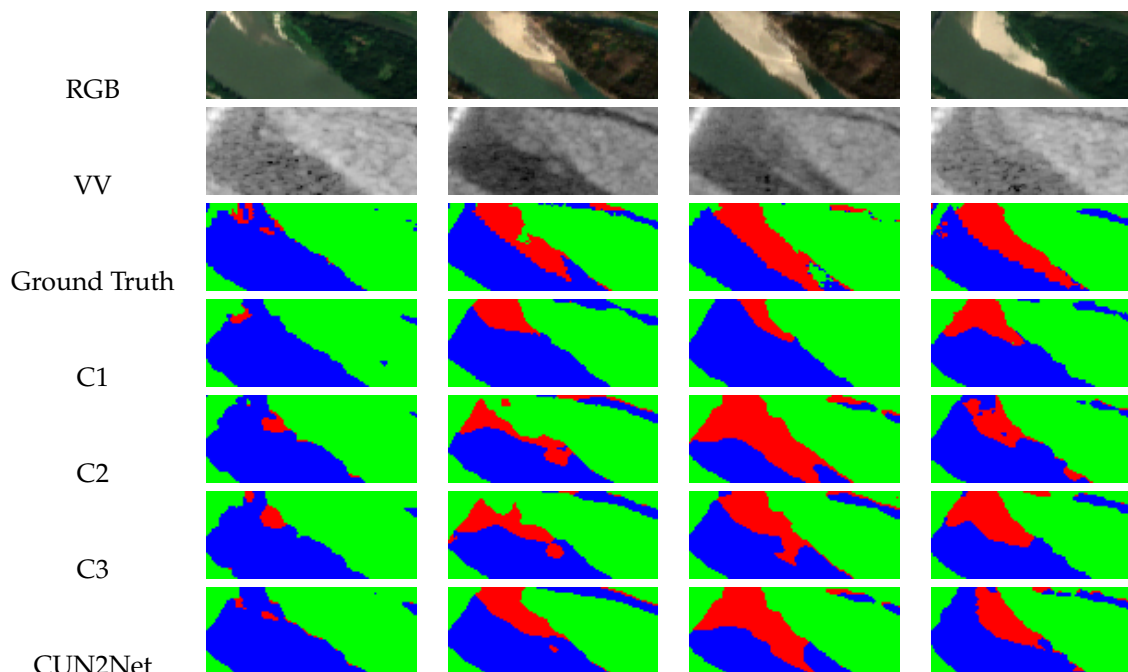


Figure 8. Comparison between three multi-temporal configurations of the proposed CUN2Net architecture in the testing area under investigation: blue is for the water class, red for the sediments, green for the vegetation.

Table 6. Comparison of results between solutions with different input configurations in terms of F1-score and accuracy.

| | Sediments | F1-Score Vegetation | Water | Accuracy |
|---------|-----------|------------------------|--------|----------|
| C1 | 0.3822 | 0.9833 | 0.8421 | 0.6839 |
| C2 | 0.5094 | 0.9817 | 0.8459 | 0.6782 |
| C3 | 0.5584 | 0.9811 | 0.8389 | 0.7078 |
| C4 | 0.2735 | 0.9760 | 0.2221 | 0.5786 |
| C5 | 0.5583 | 0.9849 | 0.7869 | 0.6432 |
| CUN2Net | 0.7296 | 0.9842 | 0.8736 | 0.7866 |

4.2. Comparison with Literature Algorithms

We compared the results of the proposed CUN2Net method with those shown in [44] and the $W - Net_+$ solution, which is an advanced solution of [44]. Specifically, the $W - Net_+$ solution has the same architecture with respect to the solution of two cascaded U-nets proposed in [44]. Still, the loss is composed of two IoU losses: one related to the first U-net and the other to the second U-net. In Table 7, we see that the $W - Net$ is poorly able to discriminate the sediment pixels. In fact, the F1-score is very low and lower than the advanced solution $W - Net_+$ and the proposed CUN2Net. $W - Net_+$ provides a better classification of sediment, but the classification of water is less accurate. Eventually, the proposed CUN2Net solution is better than the other two and, in particular, Figure 9 allows

for the analysis of the effectiveness of the result obtained by the proposed method in all the displayed images. We consistently focused on the same area, and thus the same sediment bar, precisely to understand whether the Sentinel-1 data were capable of monitoring the variations in the spatial arrangement of sediments.

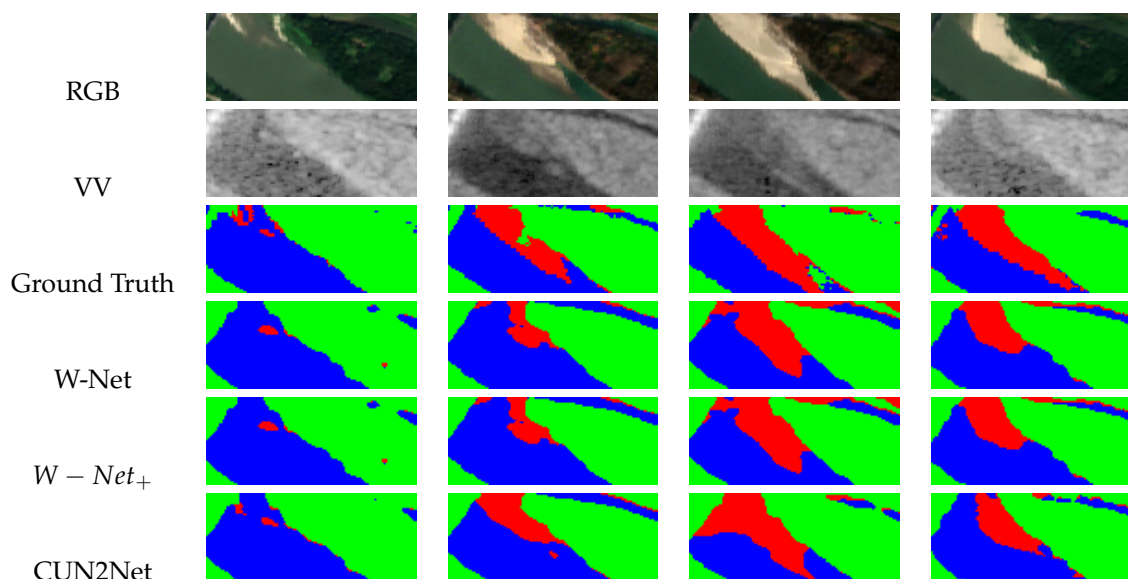


Figure 9. Comparison between the W-Net architecture, proposed in [44], the modified W-Net, and the proposed CUN2Net architecture in the testing area under investigation: blue is for the water class, red for the sediments, green for the vegetation.

Table 7. Comparison in terms of F1-score and accuracy between the proposed CUN2Net architecture and the W-Net architecture in two different configurations.

| | Sediments | F1-Score Vegetation | Water | Accuracy |
|----------------------|-----------|------------------------|--------|----------|
| W-Net | 0.2065 | 0.9381 | 0.8381 | 0.5686 |
| W – Net ₊ | 0.5578 | 0.9842 | 0.8014 | 0.6648 |
| CUN2Net | 0.7296 | 0.9842 | 0.8736 | 0.7866 |

4.3. Impact of Loss Terms

In order to understand the effect of the different components of the loss function (defined in Equation (2) in the Training Section), we evaluated a comparison between the proposed CUN2Net architecture and its variants that differed in the different combinations of weights in the loss function. In particular, we considered two other solutions, called: (i) IoU_1 , with $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_3 = 1$, i.e., a solution where only the first part was properly trained, and (ii) IoU_2 , with $\lambda_1 = 1$, $\lambda_2 = 0$, and $\lambda_3 = 1$, i.e., a solution where only the third part of the loss was properly weighted. As shown in Table 8, the solution IoU_2 (with only $\lambda_3 = 1$) presented a result that was improved by including the additional loss component ($\lambda_1 = 1$). Finally, the proposed loss (with an intermediate task not relevant for the segmentation purposes) presented the best result, improving in particular the sediment classification, as observed in Figure 10. This comparison allows us to understand that the different losses for each U-net contributed to achieving the final semantic segmentation objective.

Table 8. Comparison in terms of F1-score and accuracy between the proposed CUN2Net architecture and the other weight configurations of the same architecture. The check mark represents the presence in the loss of the related term.

| | λ_1 | λ_2 | λ_3 | Sediments | F1-Score Vegetation | Water | Accuracy |
|---------|-------------|-------------|-------------|-----------|------------------------|--------|----------|
| IoU_1 | | | ✓ | 0.5030 | 0.9842 | 0.7473 | 0.6106 |
| IoU_2 | ✓ | | ✓ | 0.6385 | 0.9846 | 0.8570 | 0.7544 |
| CUN2Net | ✓ | ✓ | ✓ | 0.7296 | 0.9842 | 0.8737 | 0.7866 |

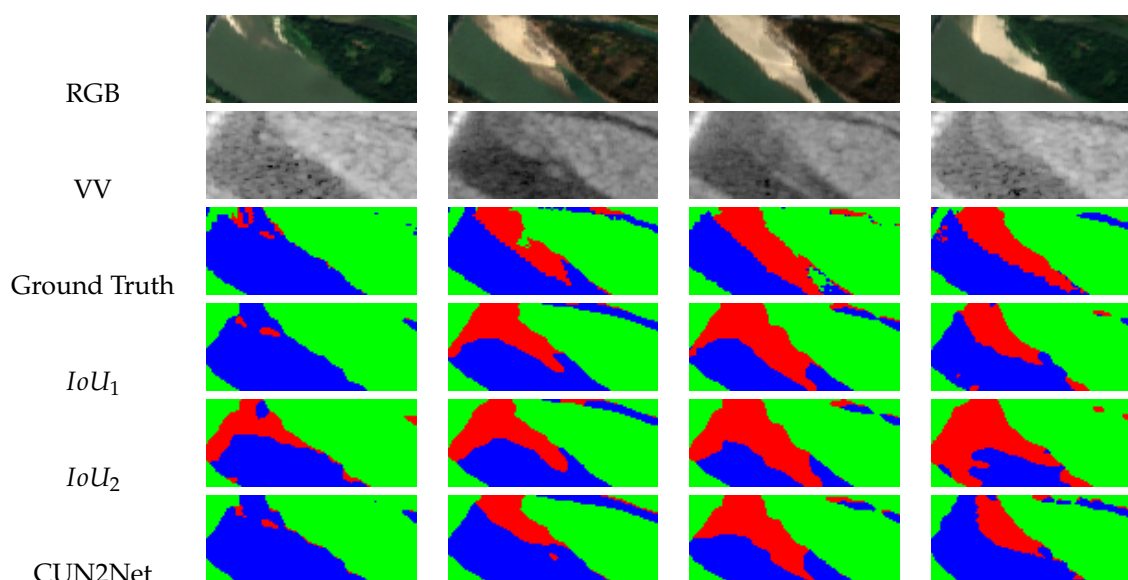


Figure 10. Comparison between three different weight configurations of the proposed CUN2Net architecture in the testing area under investigation: blue is for the water class, red for the sediments, green for the vegetation.

4.4. Ancillary Result

In this section, we showed the intermediate output of the proposed CUN2Net: the VV and VH backscattering of S1 data. This output can be seen as a despeckled version of the given VV and VH backscattering [69]. As shown in Figure 11, we can see that the VV from the intermediate output of the proposed CUN2Net network is effectively the despeckled version, which allows us to conclude that this architecture can realize a denoising version of the input data in the target date, as in [55]. Specifically, we obtain the reduction of the speckle noise, represented by an additive noise in the dB version of the SAR backscattering, without the use of a despeckled version of the VV and VH backscattering in the training phase. Furthermore, we have demonstrated the importance of this intermediate result in the final performances. In particular, we can see the improvement in the visualization of the sediment pixels. In these examples (Figure 11), we can recognize more clearly the sediment pixels as compared to water ones. The despeckled version of the input data clarifies the better performances in terms of semantic segmentation, in particular for sediments' pixels. Similar considerations in terms of despeckling visual results are obtained for VH backscattering.

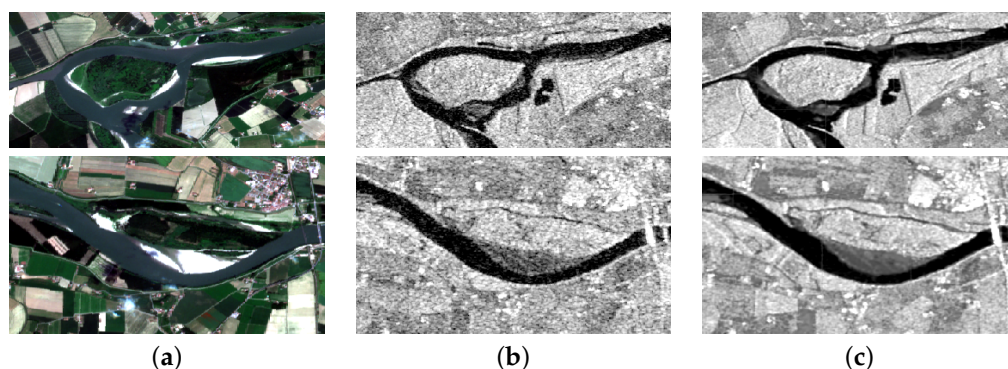


Figure 11. Two examples of intermediate VV output (c) that represented the despeckled version of the inputs (b). The areas are also shown in Sentinel-2 RGB representation (a).

5. Discussion

The proposed CUN2Net approach for fluvial morphological mapping represents a significant step forward, as it allows for the extraction of valuable information exclusively from Sentinel-1 (S1) data. Unlike previous work [44], which focused on the analysis of the Albufera wetland area, the further novelty of this study lies in applying the approach to rivers, environments characterized by continuously changing conditions, in contrast to wetlands, which tend to have more stagnant conditions. In such dynamic environments, the use of high-frequency data, such as those from Sentinel-1 satellites, provides a crucial contribution to monitoring rapid and variable phenomena, such as the dynamics of watercourses.

The results indicate a high level of accuracy and very good generalization ability when applied to images not included in the training phase. In fact, in the proposed multi-temporal and cross-polarized configuration of the input, it is possible to obtain considerable performance in terms of the F1-score. In particular, we evaluated the performances separately for each considered class in order to have a more in-depth analysis that gave specific insight into the considered application. The experimental results showed that the main difficulties in the classification task were in the discrimination between water and sediments, for which classes we obtained F1-scores equal to 0.7296 and 0.8737, respectively. The vegetation class was the most distinguishable and obtained a significantly higher F1-score of 0.9842. However, the effectiveness of the deep learning algorithm was tested on an area with similar river morphological characteristics, which attests to the model's ability to independently discern and learn the intrinsic patterns and traits of river reaches from S1 dual-polarized data. We assessed the limits linked with the spatial resolution of S1 by visual analysis that showed an excellent ability of the S1 dual-polarized data to determine the exact position of the sediments (as in Figures 8–10). In the deep learning solution, we did not use the S2 data as input, so we freed our results from simultaneity between S1 and S2. Such contemporaneity is difficult to achieve in areas affected by frequent cloud coverage. Therefore, this deep learning method has shown the potential for high-frequency monitoring, constrained only by the six-day revisit time of Sentinel-1 dual-polarized data. This high-frequency monitoring capability is particularly helpful for capturing geomorphic changes in rapid dynamic riverine environments. Furthermore, we performed training on the speckled S1 dual-polarized data, and this efficiency not only shortened processing times but also simplified the overall workflow for data collection and analysis. The GEE usage for dataset creation introduced an important reduction in time consumption. Of course, the method is affected by some limitations, which pose new challenges and the basis for future works. One of the main limitations is the workload required for manually generating ground truth, although the use of the Random Forest algorithm for data

augmentation is helpful to train a deep learning algorithm. It is clear that the reliability of the training data depends largely on the amount of manually extracted information, and therefore a larger dataset can improve performance and generality, allowing for the implementation of deeper architectures than the ones used in this work. However, the manual creation of an ample archive of ground truth is extremely time-consuming. For smaller streams monitoring (especially in mountainous regions with geometric distortions, including foreshortening, layover, and shadow), another bottleneck can be the limited spatial resolution of the S1 dual-polarized data. Addressing these main limitations and improving the model's effectiveness in these contexts requires future research efforts, including the use of SAR satellite data with better spatial resolution (such as the X-band Cosmo-Skymed or Tandem-X constellations). Further analyses could also be conducted using various datasets with better spatial (for instance, Planet, WorldView, and so on) and/or spectral resolutions (for example, hyperspectral sensors).

6. Conclusions

This study presents a deep learning semantic segmentation method for land cover classes along river environments. In particular, the proposed CUN2Net solution removes the concurrency constraint on Sentinel-1 and Sentinel-2 in the testing phase with respect to the Random Forest algorithm because the input is composed of only Sentinel-1 dual-polarized data. In this work, we used Google Earth Engine (GEE) to create a three-class semantic segmentation using Sentinel-1 and Sentinel-2 data as input to the Random Forest algorithm and a manually selected ground truth (thanks to expert knowledge) on medium-resolution images (at 10 m, thanks to S2 data) from GEE. The advantage of the RF algorithm is its reliability and robustness, but it implies the condition that the images of S1 and S2 must be acquired almost simultaneously. This condition is very difficult to satisfy, especially during rainy periods when Sentinel-2 images are affected by the presence of clouds. We overcame this limitation by using the highly reliable RF-based solution as the ground truth for a deep learning method that only considers Sentinel-1 as input. The results showed significant performance for the classification of water, sediment, and vegetation classes. Moreover, it has been shown that we obtained a despeckled version of the input as an intermediate result of the proposed CUN2Net architecture, which confirms why the segmentation result in the proposed configuration is better than in the case where the intermediate result is not considered in the loss. The approach is sufficiently adaptable and can also be used for other datasets. The use of new datasets characterized by higher spatial resolution will allow the method to be applied to less wide riverbeds and thus extend its use to a larger portion of the hydrographic network.

Author Contributions: Conceptualization, M.G., C.C. and M.N.P.; methodology, M.G.; software, M.G.; validation, M.G. and C.C.; writing—original draft preparation, M.G.; writing—review and editing, M.N.P., M.G. and C.C.; supervision, M.N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ghosh, K.G.; Mukhopadhyay, S. Introductory Chapter: Current Practice in Fluvial Geomorphology: Research Frontiers, Issues and Challenges. In *Current Practice in Fluvial Geomorphology-Dynamics and Diversity*; IntechOpen: London, UK, 2020.
2. Poole, G.C. Stream hydrogeomorphology as a physical science basis for advances in stream ecology. *J. N. Am. Benthol. Soc.* **2010**, *29*, 12–25. [[CrossRef](#)]
3. Rinaldi, M.; Belletti, B.; Comiti, F.; Nardi, L.; Mao, L.; Bussettini, M. Sistema di rilevamento e classificazione delle Unità Morfologiche dei corsi d'acqua (SUM). *Versione Aggiornata* **2016**.
4. Belletti, B.; Rinaldi, M.; Bussettini, M.; Comiti, F.; Gurnell, A.M.; Mao, L.; Nardi, L.; Vezza, P. Characterising physical habitats and fluvial hydromorphology: A new system for the survey and classification of river geomorphic units. *Geomorphology* **2017**, *283*, 143–157. [[CrossRef](#)]
5. Schumann, G.; Giustarini, L.; Tarpanelli, A.; Jarihani, B.; Martinis, S. Flood modeling and prediction using earth observation data. *Surv. Geophys.* **2023**, *44*, 1553–1578. [[CrossRef](#)]
6. Gao, H.; Birkett, C.; Lettenmaier, D.P. Global monitoring of large reservoir storage from satellite remote sensing. *Water Resour. Res.* **2012**, *48*, 1–12. [[CrossRef](#)]
7. Baki, A.B.M.; Gan, T.Y. Riverbank migration and island dynamics of the braided Jamuna River of the Ganges–Brahmaputra basin using multi-temporal Landsat images. *Quat. Int.* **2012**, *263*, 148–161. [[CrossRef](#)]
8. Spada, D.; Molinari, P.; Bertoldi, W.; Vitti, A.; Zolezzi, G. Multi-temporal image analysis for fluvial morphological characterization with application to Albanian rivers. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 314. [[CrossRef](#)]
9. Cavallo, C.; Nones, M.; Papa, M.N.; Gargiulo, M.; Ruello, G. Monitoring the morphological evolution of a reach of the Italian Po River using multispectral satellite imagery and stage data. *Geocarto Int.* **2022**, *37*, 8579–8601. [[CrossRef](#)]
10. Cavallo, C.; Papa, M.N.; Gargiulo, M.; Ruello, G. Characterization of the flow regime of temporary rivers using Sentinel-2 satellite data. In *International Association for Hydro-Environment Engineering and Research (IAHR)*; Prof. Miguel Ortega-Sánchez, University of Granada: Granada, Spain, 2022; Volume 39, pp. 5413–5417.
11. Rivas-Fandiño, P.; Acuña-Alonso, C.; Novo, A.; Pacheco, F.A.L.; Álvarez, X. Assessment of high spatial resolution satellite imagery for monitoring riparian vegetation: Riverine management in the smallholding. *Environ. Monit. Assess.* **2023**, *195*, 81. [[CrossRef](#)]
12. Marchetti, G.; Bizzi, S.; Belletti, B.; Lastoria, B.; Comiti, F.; Carbonneau, P.E. Mapping riverbed sediment size from Sentinel-2 satellite data. *Earth Surf. Process. Landforms* **2022**, *47*, 2544–2559. [[CrossRef](#)]
13. Carbonneau, P.E.; Belletti, B.; Micotti, M.; Lastoria, B.; Casaioli, M.; Mariani, S.; Marchetti, G.; Bizzi, S. UAV-based training for fully fuzzy classification of Sentinel-2 fluvial scenes. *Earth Surf. Process. Landforms* **2020**, *45*, 3120–3140. [[CrossRef](#)]
14. Kryniecka, K.; Magnuszewski, A. Application of satellite sentinel-2 images to study alternate sandbars movement at Lower Vistula River (Poland). *Remote Sens.* **2021**, *13*, 1505. [[CrossRef](#)]
15. Shahrood, A.J.; Menberu, M.W.; Darabi, H.; Rahmati, O.; Rossi, P.M.; Kløve, B.; Haghighi, A.T. RiMARS: An automated river morphodynamics analysis method based on remote sensing multispectral datasets. *Sci. Total Environ.* **2020**, *719*, 137336. [[CrossRef](#)] [[PubMed](#)]
16. Seaton, D.; Dube, T.; Mazvimavi, D. Use of multi-temporal satellite data for monitoring pool surface areas occurring in non-perennial rivers in semi-arid environments of the Western Cape, South Africa. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 375–384. [[CrossRef](#)]
17. Cavallo, C.; Papa, M.N.; Gargiulo, M.; Palau-Salvador, G.; Vezza, P.; Ruello, G. Continuous monitoring of the flooding dynamics in the Albufera Wetland (Spain) by Landsat-8 and Sentinel-2 datasets. *Remote Sens.* **2021**, *13*, 3525. [[CrossRef](#)]
18. Boothroyd, R.J.; Williams, R.D.; Hoey, T.B.; Barrett, B.; Prasojo, O.A. Applications of Google Earth Engine in fluvial geomorphology for detecting river channel change. *Wiley Interdiscip. Rev. Water* **2021**, *8*, e21496. [[CrossRef](#)]
19. Rossi, D.; Zolezzi, G.; Bertoldi, W.; Vitti, A. Monitoring braided river-bed dynamics at the sub-event time scale using time series of Sentinel-1 SAR imagery. *Remote Sens.* **2023**, *15*, 3622. [[CrossRef](#)]
20. Mitidieri, F.; Papa, M.N.; Amitrano, D.; Ruello, G. River morphology monitoring using multitemporal SAR data: Preliminary results. *Eur. J. Remote Sens.* **2016**, *49*, 889–898. [[CrossRef](#)]
21. Barbouchi, M.; Abdelfattah, R.; Chokmani, K.; Aissa, N.B.; Mhammed, C.H. Sentinel 1 response to cereal leaf area index (LAI): Study case for central Tunisia. In *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, China, 10–15 July 2016; pp. 7125–7128. [[CrossRef](#)]
22. Zhang, K.; Zhang, F.; Wan, W.; Yu, H.; Sun, J.; Del Ser, J.; Elyan, E.; Hussain, A. Panchromatic and multispectral image fusion for remote sensing and earth observation: Concepts, taxonomy, literature review, evaluation methodologies and challenges ahead. *Inf. Fusion* **2023**, *93*, 227–242. [[CrossRef](#)]
23. Gargiulo, M.; Mazza, A.; Gaetano, R.; Ruello, G.; Scarpa, G. A CNN-based fusion method for super-resolution of Sentinel-2 data. In *Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 22–27 July 2018; pp. 4713–4716.

24. Gargiulo, M.; Mazza, A.; Gaetano, R.; Ruello, G.; Scarpa, G. Fast Super-Resolution of 20 m Sentinel-2 Bands Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2635. [[CrossRef](#)]
25. Pal, M.K.; Rasmussen, T.M.; Abdolmaleki, M. Multiple Multi-Spectral Remote Sensing Data Fusion and Integration for Geological Mapping. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019; pp. 1–5.
26. Wald, L. Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions. In *Les Presses de l'École des Mines*; Presses des MINES: Paris, France, 2002.
27. Clerici, N.; Valbuena Calderón, C.A.; Posada, J.M. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: A case study in the lower Magdalena region, Colombia. *J. Maps* **2017**, *13*, 718–726. [[CrossRef](#)]
28. Hall, D.L.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **1997**, *85*, 6–23. [[CrossRef](#)]
29. Zhang, J. Multi-source remote sensing data fusion: Status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24. [[CrossRef](#)]
30. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.; et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **2016**, *8*, 70. [[CrossRef](#)]
31. Castanedo, F. A Review of Data Fusion Techniques. *Sci. World J.* **2013**, *2013*, 19. [[CrossRef](#)]
32. 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. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5252–5255. [[CrossRef](#)]
33. Sukawattanavijit, C.; Chen, J.; Zhang, H. GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 284–288. [[CrossRef](#)]
34. Fan, M.; Ma, D.; Huang, X.; An, R. Adaptability Evaluation of the Spatiotemporal Fusion Model of Sentinel-2 and MODIS Data in a Typical Area of the Three-River Headwater Region. *Sustainability* **2023**, *15*, 8697. [[CrossRef](#)]
35. Scarpa, G.; Gargiulo, M.; Mazza, A.; Gaetano, R. A CNN-based fusion method for feature extraction from sentinel data. *Remote Sens.* **2018**, *10*, 236. [[CrossRef](#)]
36. Haas, J.; Ban, Y. Sentinel-1A SAR and sentinel-2A MSI data fusion for urban ecosystem service mapping. *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 41–53. [[CrossRef](#)]
37. Popescu, A.; Vaduva, C.; Faur, D.; Datcu, M. Enhanced Classification of Land Cover through Joint Analysis of Sentinel-1 and Sentinel-2 Data. In Proceedings of the ESA Living Planet Symp., Prague, Czech Republic, 9–13 May 2016; p. 9.
38. Prodromou, M.; Theocharidis, C.; Fotiou, K.; Argyriou, A.; Polydorou, T.; Hadjimitsis, D.; Tzouvaras, M. Fusion of Sentinel-1 and Sentinel-2 satellite imagery to rapidly detect landslides through Google Earth Engine. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 23–28 April 2023; p. EGU-12618.
39. Ouyang, W.; Wang, X. Joint Deep Learning for Pedestrian Detection. In Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 2056–2063. [[CrossRef](#)]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Marrakesh, Morocco, 6–10 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
41. Zhao, X.; Yuan, Y.; Song, M.; Ding, Y.; Lin, F.; Liang, D.; Zhang, D. Use of Unmanned Aerial Vehicle Imagery and Deep Learning UNet to Extract Rice Lodging. *Sensors* **2019**, *19*, 3859. [[CrossRef](#)]
42. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
43. Gargiulo, M.; Dell'Aglio, D.A.; Iodice, A.; Riccio, D.; Ruello, G. Semantic Segmentation using Deep Learning: A case of study in Albufera Park, Valencia. In Proceedings of the 2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Portici, Italy, 24–26 October 2019; pp. 134–138.
44. Gargiulo, M.; Dell'Aglio, D.A.; Iodice, A.; Riccio, D.; Ruello, G. Integration of Sentinel-1 and Sentinel-2 Data for Land Cover Mapping Using W-Net. *Sensors* **2020**, *20*, 2969. [[CrossRef](#)] [[PubMed](#)]
45. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
46. Stoian, A.; Poulain, V.; Inglada, J.; Poughon, V.; Derksen, D. Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sens.* **2019**, *11*, 1986. [[CrossRef](#)]
47. Moore, R.; Hansen, M. Google Earth Engine: A new cloud-computing platform for global-scale earth observation data and analysis. In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, CA, United States, 4–8 December 2011; Volume 2011, p. IN43C-02.
48. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]

49. Yang, X.; Pavelsky, T.M.; Allen, G.H.; Donchyts, G. RivWidthCloud: An automated Google Earth Engine algorithm for river width extraction from remotely sensed imagery. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 217–221. [[CrossRef](#)]
50. Mutanga, O.; Kumar, L. Google earth engine applications. *Remote Sens.* **2019**, *11*, 591. [[CrossRef](#)]
51. Malenovsky, Z.; Rott, H.; Cihlar, J.; Schaepman, M.E.; García-Santos, G.; Fernandes, R.; Berger, M. Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* **2012**, *120*, 91–101. [[CrossRef](#)]
52. Gaber, A.; Soliman, F.; Koch, M.; El-Baz, F. Using full-polarimetric SAR data to characterize the surface sediments in desert areas: A case study in El-Gallaba Plain, Egypt. *Remote Sens. Environ.* **2015**, *162*, 11–28. [[CrossRef](#)]
53. Dirgahayu, D.; Parsa, I.M. Detection Phase Growth of Paddy Crop Using SAR Sentinel-1 Data. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *280*, 012020. [[CrossRef](#)]
54. Main-Knorn, M.; Pflug, B.; Debaecker, V.; Louis, J. Calibration and validation plan for the l2a processor and products of the Sentinel-2 mission. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 1249–1255. [[CrossRef](#)]
55. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. *arXiv* **2018**, arXiv:1803.04189.
56. Jin, Y.; Liu, X.; Chen, Y.; Liang, X. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: A case study of central Shandong. *Int. J. Remote Sens.* **2018**, *39*, 8703–8723. [[CrossRef](#)]
57. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [[CrossRef](#)]
58. Kamal, M.; Jamaluddin, I.; Perala, A.; Farda, N.M. Comparison of Google Earth Engine (GEE)-based machine learning classifiers for mangrove mapping. In Proceedings of the 40th Asian Conference Remote Sensing, ACRS, Daejeon, Korea, 14–18 October 2019; pp. 1–8.
59. Jiang, L.; Wang, W.; Yang, X.; Xie, N.; Cheng, Y. Classification methods of remote sensing image based on decision tree technologies. In Proceedings of the Computer and Computing Technologies in Agriculture IV: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China, 22–25 October 2010; Selected Papers, Part I 4; Springer: Berlin/Heidelberg, Germany, 2011; pp. 353–358.
60. Huang, C.; Davis, L.; Townshend, J. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [[CrossRef](#)]
61. Xia, X.; Kulis, B. W-net: A deep model for fully unsupervised image segmentation. *arXiv* **2017**, arXiv:1711.08506.
62. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020; pp. 1–7.
63. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* **2016**, *8*, 594. [[CrossRef](#)]
64. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)]
65. Rahman, M.A.; Wang, Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 12–14 December 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 234–244.
66. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
67. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
68. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756.
69. Dalsasso, E.; Denis, L.; Tupin, F. SAR2SAR: A self-supervised despeckling algorithm for SAR images. *J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 4321–4329. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.