remote sensing

MDPI

Article

Crop Type Mapping from Optical and Radar Time Series Using Attention-Based Deep Learning

Stella Ofori-Ampofo [1,2], Charlotte Pelletier [1,*] and Stefan Lang [2]

[1] IRISA UMR CNRS 6074, Campus de Tohannic, Université Bretagne Sud, 56000 Vannes, France; stella.ofori-ampofo@stud.sbg.ac.at
[2] Christian Doppler Laboratory for Geospatial and EO-Based Humanitarian Technologies, Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria; stefan.lang@plus.ac.at
* Correspondence: charlotte.pelletier@univ-ubs.fr

Abstract: Crop maps are key inputs for crop inventory production and yield estimation and can inform the implementation of effective farm management practices. Producing these maps at detailed scales requires exhaustive field surveys that can be laborious, time-consuming, and expensive to replicate. With a growing archive of remote sensing data, there are enormous opportunities to exploit dense satellite image time series (SITS), temporal sequences of images over the same area. Generally, crop type mapping relies on single-sensor inputs and is solved with the help of traditional learning algorithms such as random forests or support vector machines. Nowadays, deep learning techniques have brought significant improvements by leveraging information in both spatial and temporal dimensions, which are relevant in crop studies. The concurrent availability of Sentinel-1 (synthetic aperture radar) and Sentinel-2 (optical) data offers a great opportunity to utilize them jointly; however, optimizing their synergy has been understudied with deep learning techniques. In this work, we analyze and compare three fusion strategies (input, layer, and decision levels) to identify the best strategy that optimizes optical-radar classification performance. They are applied to a recent architecture, notably, the pixel-set encoder–temporal attention encoder (PSE-TAE) developed specifically for object-based classification of SITS and based on self-attention mechanisms. Experiments are carried out in Brittany, in the northwest of France, with Sentinel-1 and Sentinel-2 time series. Input and layer-level fusion competitively achieved the best overall F-score surpassing decision-level fusion by 2%. On a per-class basis, decision-level fusion increased the accuracy of dominant classes, whereas layer-level fusion improves up to 13% for minority classes. Against single-sensor baseline, multi-sensor fusion strategies identified crop types more accurately: for example, input-level outperformed Sentinel-2 and Sentinel-1 by 3% and 9% in F-score, respectively. We have also conducted experiments that showed the importance of fusion for early time series classification and under high cloud cover condition.

Keywords: fusion; satellite image time series; Sentinel-1; Sentinel-2; pixel-set encoder; temporal attention encoder

check for updates

Citation: Ofori-Ampofo, S.; Pelletier, C.; Lang, S. Crop Type Mapping from Optical and Radar Time Series Using Attention-Based Deep Learning. *Remote Sens.* **2021**, *13*, 4668. https://doi.org/10.3390/rs13224668

Academic Editors: Mauro Dalla Mura, Lucas Drumetz and Guillaume Tochon

Received: 1 October 2021
Accepted: 16 November 2021
Published: 19 November 2021

1. Introduction

Causal factors such as climate change have a high likelihood to threaten food security at global, regional, and local levels [1]. Recent reports reveal that agriculture absorbs 26% of the economic impact of climate-induced disasters, which rises to more than 80% for drought in developing countries [2]. The agricultural sector is not only impacted by changing climates but contributes about 24% of greenhouse gas (GHG) emissions together with forestry and other land use [3]. Under certain conditions, warmer temperatures and carbon dioxide presence can stimulate crop growth [4], especially in temperate regions. Extreme thresholds, however, may have dire consequences on crop productivity [5]. Remote sensing has become an integral tool supporting the monitoring and management of agriculture as well as efforts to mitigate climate change.

The contribution of remote sensing is demonstrated in studies related to land cover mapping [6–8], urban studies [9], forest inventories [10], and burnt area mapping [11]. They are the backbone for diverse large-scale operational services (e.g., European Copernicus services), addressing issues concerning land, disasters, marine, and atmosphere. In agriculture, they have been used for crop classification [12], phenology studies [13,14], yield estimation [15,16], and insurance applications [17]. The availability and access to open satellite data archives (e.g., moderate resolution imaging spectroradiometer (MODIS), Landsat, and Sentinel) is advancing methodologies to solve environmental and societal challenges. One interesting context, especially in crop studies and land cover mapping, is the use of satellite image time series (SITS) where images of the same area are acquired at different dates. SITS presents opportunities for studying the seasonality or evolution of objects through time to aid their discrimination [14,18].

Until recently, land cover and crop type mapping tasks mainly focused on single-sensor data (e.g., optical SITS), which was used to feed traditional algorithms such as random forests or support vector machines [6,19]. Although successful, these learning algorithms do not harness temporal information, an essential dimension for vegetation studies [20]. Current advances in SITS classification are marked by the use of deep learning, which can make the most of the temporal structure of SITS data. They are designed to leverage high-end computational power and can attain similar or higher classification performance compared to classical machine learning algorithms [7,21,22].

Moreover, the diversity in sensor characteristics (spatial, spectral, or polarimetric) has triggered multi-sensor data fusion to enhance information for discriminating objects. For example, the combination of polarimetric radar and optical data provides information on different physical characteristics: the former describes structure and moisture content, whereas the latter provides spectral information across a wide range of spectra. Such synergy can complement information loss rising from cloud cover in optical data. Even in the absence of this limitation, their unison has resulted in increased performance compared to the use of each sensor [23–25].

Although promising, the combined use of optical and radar data is sparingly explored compared to the use of a single sensor, especially in the case of deep learning techniques for crop classification. There have been only a few attempts, to the extent of our literature review, to leverage optical and radar synergies through deep learning [7,26]. Our study extends these research works by investigating more forms of fusion along with an advanced deep learning architecture, namely pixel set encoder–temporal attention encoder (PSE-TAE), and proposes the optimal level of synergy in this setup between Sentinel-1 and Sentinel-2 for crop classification. Additionally, we explore (i) sparse SITS to mimic scenarios where the presence of clouds hinders the use of a complete Sentinel-2 SITS in optical-radar fusion, and (ii) incremental learning to assess the benefit of fusion for mapping crops before having a full year of data.

The rest of this paper is organized as follows: Section 2 presents related works with an overview of current state-of-the-art approaches for the classification of Earth observation (EO) time series in Section 2.1, and fusion strategies and application to radar and optical time series in Section 2.2. Section 3 describes the study area and the data used. In Section 4, we present the fusion strategies as well as the deep learning architecture employed in the study. Section 5 is the core section of this paper with qualitative and quantitative comparison of fusion strategies. Finally, we draw the conclusion in Section 6.

2. Related Work

2.1. Satellite Image Time Series Classification

The tremendous increase in the amount of free satellite data has triggered the development of robust methodologies to efficiently exploit high volumes and varieties of data for automated mapping at large scales. During the last decade, the classification of SITS data mainly relied on standard machine learning algorithms, notably decision-based (e.g., random forests [27]) and kernel-based (e.g., support vector machines [28]) for crop recognition [25,29,30] and land cover mapping [6,31–34]. Several studies have established that random forests perform better than support vector machines in winter land use mapping [24], in crop mapping [19], or in land cover mapping [6]. Presently, techniques based on deep learning have become prominent given their capacity to adaptively discover patterns from dense data, leveraging high-end computational power. In the context of SITS, they have attained similar or higher classification performance compared to classical machine learning techniques [7,21,22] and improved the separation of mixed or under-represented classes [33]. Deep learning has been applied using convolutional neural networks (CNNs) for handling the temporal dimension [21,35]; recurrent neural networks (RNNs)-like models [36,37], including long short term memory (LSTM) [33,38] or gated recurrent unit (GRU); and strategies that combine CNN with recurrent models [39,40], or ConvLSTM [41,42]. Recently, attention-based architectures have been proposed for the SITS classification in the context of crop type mapping [22,43]. The work presented in [22] shows that attention-based mechanisms outperformed CNNs but are at par with LSTM on unprocessed data, e.g., cloudy optical SITS; however, when extensive data pre-processing is applied, results are comparable to random forests. In [43], a modified self-attention-based mechanism architecture, namely pixel-set encoder–temporal-attention encoder (PSE-TAE), extracts more expressive features than CNNs and GRUs. PSE-TAE is an object-based classifier developed for crop type mapping, which is composed of a spatial encoder (PSE) to exploit the spatial context in SITS and a temporal encoder (TAE) to encode the temporal structure of SITS. A light-weight TAE also exists [44].

2.2. Fusion Strategies for Satellite Image Time Series

The single use of optical or radar data has been beneficial to produce parcel-level crop maps at national and continental scales [45,46]. At times, data from a single sensor may not be enough to optimize target class separation. Given the different acquisition mechanisms of optical and radar sensors, they can enrich information for discriminating targets by respectively contributing information on reflectance and structural or moisture properties. This is evidenced in studies relating to land cover mapping [7,34], grassland monitoring [47], burnt area detection [11], urban mapping [34,48], crop mapping [12,18,26,29,49–51], soil moisture mapping [52], soil texture estimation [53], and crop phenological studies [14]. Today, the consistent acquisitions of ESA's Sentinel-1 and Sentinel-2 images has allowed their exploitation in unison.

According to [54], multi-modal image fusion can be grouped into pixel, feature, and decision levels. In deep learning, the pixel- and feature-level fusions have been redefined as input-level and layer-level, respectively [55].The input level fundamentally combines image bands from multi-source data (usually through resampling and concatenation, or image to image co-registration). This is the simplest and common form of fusion, but it might suffer some drawbacks due to the resulting high dimension of stacked data [56]. Layer fusion requires the extraction and concatenation of high-level features allowing each modality to learn individual feature representations. At the decision level, each modality is independently processed by a network to generate class confidence scores (in the case of classification), which are combined statistically (e.g., averaged) to yield a final fused decision. An interested reader can also refer to [57,58] for a general review of data fusion in remote sensing and its main challenges.

In multi-modal (typically optical and radar) SITS, fusion has been commonly applied using traditional machine learning algorithms. Most of the studies used an early fusion strategy with a random forest classifier, where optical and radar time series are stacked together in one data cube [7,11,12,24,29,34,49–51,53,59]. A late fusion strategy was also adopted in [7,56], where class probability vectors obtained by training two independent random forest models are combined together to obtain the final prediction. The combined use of Sentinel-1 and Sentinel-2 has been shown to improve accuracies compared to a single sensor but has mainly been explored with random forests. Although deep learning techniques achieved significant performance in crop and land cover mapping, only few studies [7,26], have leveraged optical and radar synergies through deep learning. The authors of [26] used 1D-CNN, whereas 2D-CNN and RNN are integrated in [7] to combine Sentinel-1 and Sentinel-2 time series. One of the best performances from [7] is yielded by feeding a random forest classifier with high-level spatio-temporal features extracted from two deep networks trained independently on each modality. SITS have also been fused with single-date high spatial resolution data using U-Net, CNN and/or GRU approaches [60,61]. Moreover, optical and radar data have also been used together without, strictly speaking, fusing them [47,62]. In these works, radar time series serve as the input of a deep learning model, while the target is a derived feature (e.g., normalized difference vegetation index (NDVI) time series [47] or a segmentation map [62]) extracted from optical data.

Our study extends these research works by investigating more forms of fusion along with an advanced deep learning architecture, namely pixel set encoder–temporal attention encoder (PSE-TAE), and proposes the optimal level of synergy in this setup between Sentinel-1 and Sentinel-2 for crop classification.

3. Data

3.1. Study Area

The study area, Finistère, is a *department* in the Western Brittany region, France (see Figure 1). It consists of four districts, namely Quimper, Châteaulin, Morlaix, and Brest, covering an area of 6733 km^2, which corresponds to about a quarter of Brittany. Bordered by the English Channel and the Atlantic Ocean on a 1200 km coastline, the department experiences a temperate oceanic climate characterized by mild temperatures in winter, temperate in summer, and rainfall distributed all throughout the year. Finistère features a high variety of crops, making it an interesting consideration for crop classification studies. The prevalent crops cultivated are *meadows* , *maize*, *vegetables*, *barley*, and *wheat* (see Figure 2). On average, croplands measure about two hectares in this area.

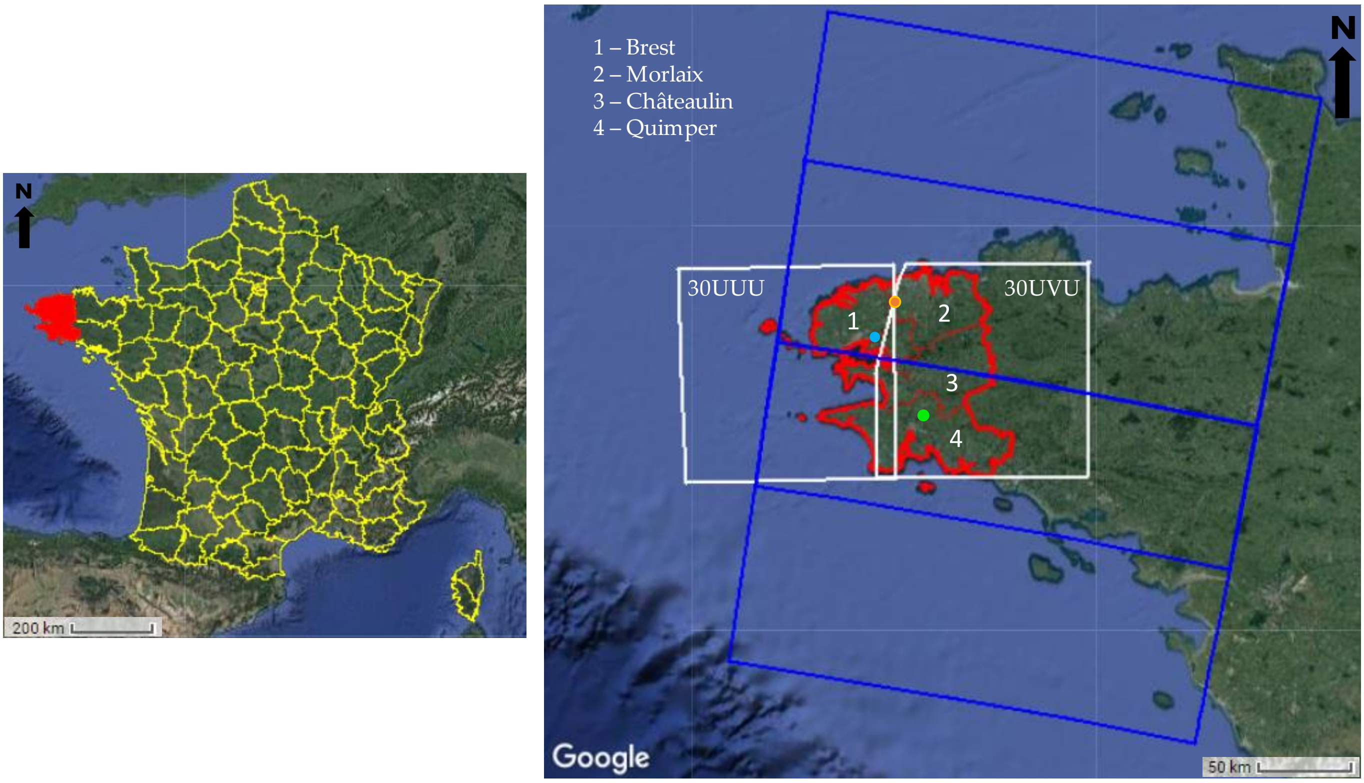


Figure 1. The study area, Finistère, is a French *department* in the Western Brittany region (in red). Finistère consists of four districts: Quimper, Châteaulin, Morlaix, and Brest. Sentinel-1 footprints and Sentinel-2 granule tiles are displayed in blue and white, respectively.

3.2. *Crop Type Labels*

Farmer declarations, which form the basis of subsidy disbursement under the European Union Common Agricultural Policy, have resulted in a huge archive of publicly available crop labels (Land Parcel Identification System). In France, this information is collated under the *Registre Parcellaire Graphique* (RPG), a comprehensive geographical depository allowing the identification of land use on agricultural parcels. RPG is obtained from the French open data platform (www.data.gouv.fr, accessed on 16 February 2021) for the year 2019. It is published, for each France region, with a hierarchical nomenclature of more than 300 categories for the lower level and 28 for the higher level. In Finistère, 20 out of the 28 categories of agricultural land use are identified. Although subject to some issues (e.g., mis-registration errors), we assume here that these data are free of errors.

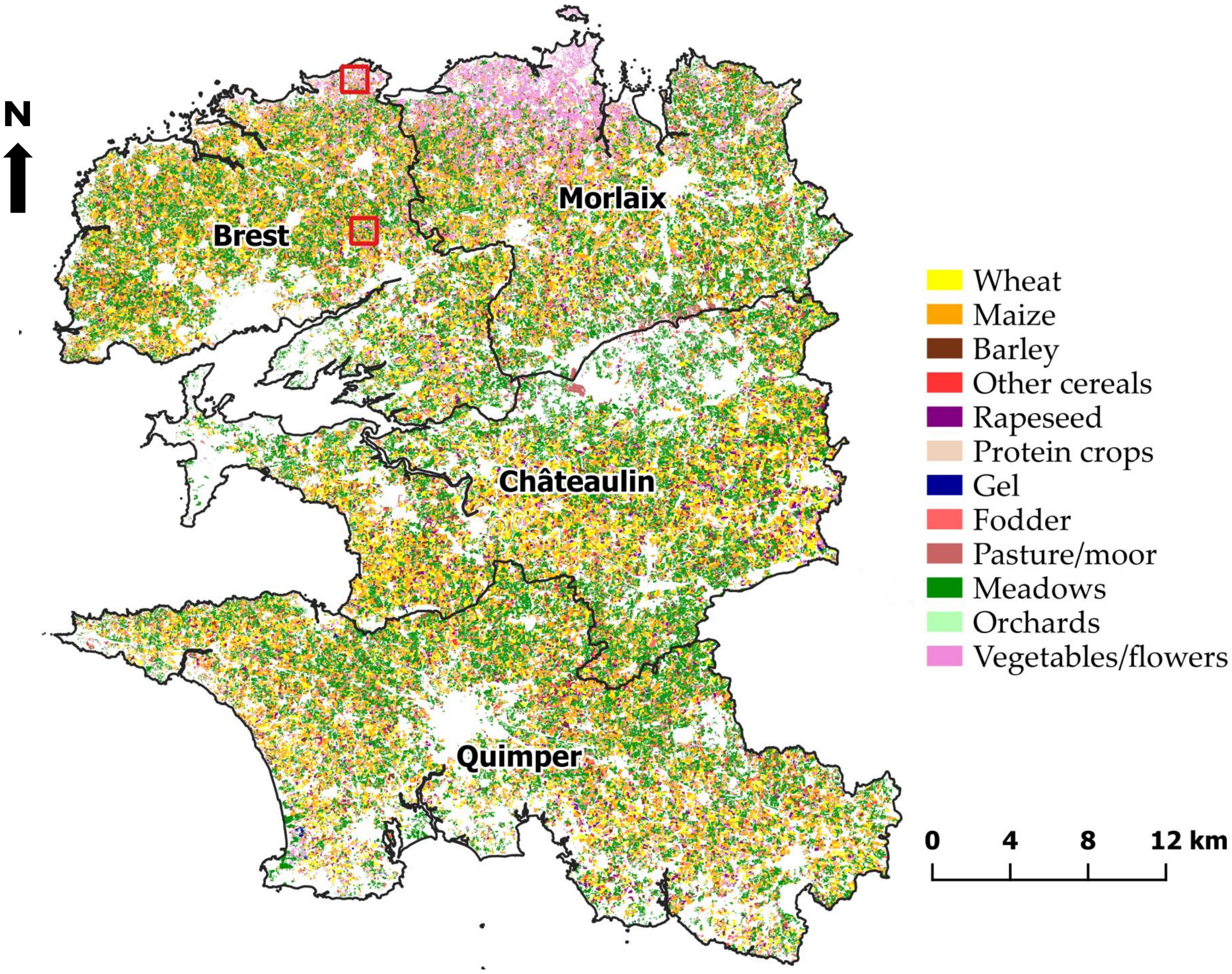


Figure 2. Reference crop map of Finistère. Red boxes are selected sites for qualitative evaluation. See Table 1 for the legend.

As an initial data preparation step, classes representing mixed crops or classes with relatively low occurrences (below 0.02% of total reference data) are discarded from the analysis. There is an exception for the class "*other cereals*", which includes *buckwheat*, a regional cereal. Permanent and temporary *meadows* are consolidated into one class as they cannot be distinguished within the time frame considered. In the end, a total of 12 classes are considered: *maize*, *wheat*, *barley*, *rapeseed*, *protein crops*, *gel (frozen surfaces)*, *fodder*, *pasture and moor*, *meadows*, *orchards*, *vegetables/flowers* and *other cereals*. The crop type distribution is highly imbalanced with the *meadow* class representing the highest frequency, nearly half of the total samples. The code for the different data preprocessing steps is publicly available at https://github.com/ellaampy/CropTypeMapping (accessed on 16 February 2021).

Table 1. Number of parcels per crop type and per district.

Legend		Quimper Training	Châteaulin Training	Morlaix Validation	Brest Testing	Total	%
Maize	🟧	11,237	7670	7830	10,934	37,671	20.87
Wheat	🟨	4820	4684	2583	3227	15,314	8.48
Barley	🟫	2378	2072	3417	3017	10,884	6.03
Rapeseed	🟪	899	1092	338	300	2629	1.46
Other cereals	🟥	1591	883	629	352	3455	1.91
Protein crops	⬜	169	104	58	46	377	0.21
Vegetables/flowers	🟪	1080	472	10,030	2502	14,084	7.8
Orchards	🟩	218	69	33	45	365	0.2
Meadows	🟩	25,020	22,602	18,414	21,151	87,187	48.3
Gel	🟦	1651	1059	579	354	3643	2.02
Fodder	🟥	1373	890	846	647	3756	2.08
Pasture/moor	🟥	153	369	466	176	1164	0.64
Total		50,589	41,966	45,223	42,751	180,529	100

Figure 2 displays the reference crop map for the whole studied area, and Table 1 gives the actual number of parcels per class and per district. In the experiments (presented in Section 4.2), the train–validation–test split is made at district level. Crop instances from Châteaulin and Quimper are combined to form the training set, whereas Morlaix crops form the validation set. The Brest district is used to report the final performance for the different baselines and fusion scenarios.

3.3. Satellite Data

The Copernicus mission is the European Earth observation program with a constellation of optical and radar satellites called Sentinel, providing free, accurate, continuous, global coverage for a better understanding and sustainable management of the environment. Sentinel-1 (radar) and Sentinel-2 (optical) image time series are acquired between October 2018 and December 2019 to obtain enough information to capture the phenology of different crops. Data from the end of 2018 are retained to provide additional information on soil structural changes during tilling in preparation for sowing [13]. We consider a multi-seasonal rather than a multi-annual time series. Figure 3 displays the acquisition dates for Sentinel-1 and Sentinel-2 time series.

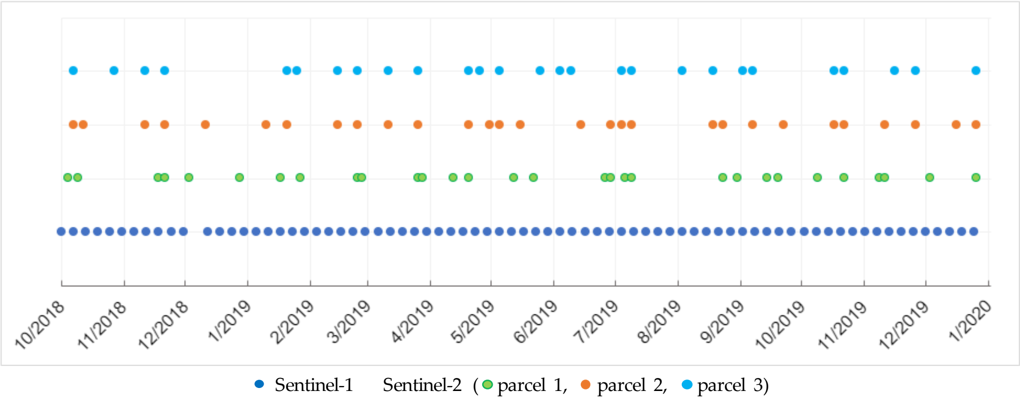


Figure 3. Acquisition dates of Sentinel-1 (in blue) and Sentinel-2 (in cyan, orange, and green) time series. Dates differ for Sentinel-2 depending on the parcel's location. Figure 1 displays colored dots (cyan, orange, and green) to infer the location of the three parcels.

Each mission is equipped with twin polar-orbiting satellites at 180° apart, enabling acquisitions at about 5 and 6 days repeat cycles for Sentinel-2 and Sentinel-1, respectively. Sentinel-2 provides multi-spectral images, whereas Sentinel-1 provides data from a dualpolarized C-band SAR enabling imaging through clouds. Both data are collected from Google earth engine (GEE), a cloud-based geospatial analysis platform hosting a rich catalog of publicly available data [63] including Sentinel, Landsat, and MODIS.

Figure 4 shows NDVI extracted from Sentinel-2 images (left) and Sentinel-1 VV backscatter (right) in a time composite, highlighting different vegetation dynamics (which corresponds to winter and summer crops) observed through an optical and radar lens. Changes prominent between January and April, May and August, and September and December are shown as red, green, and blue, respectively. We describe in the following paragraphs the different preprocessing steps applied for each type of data.

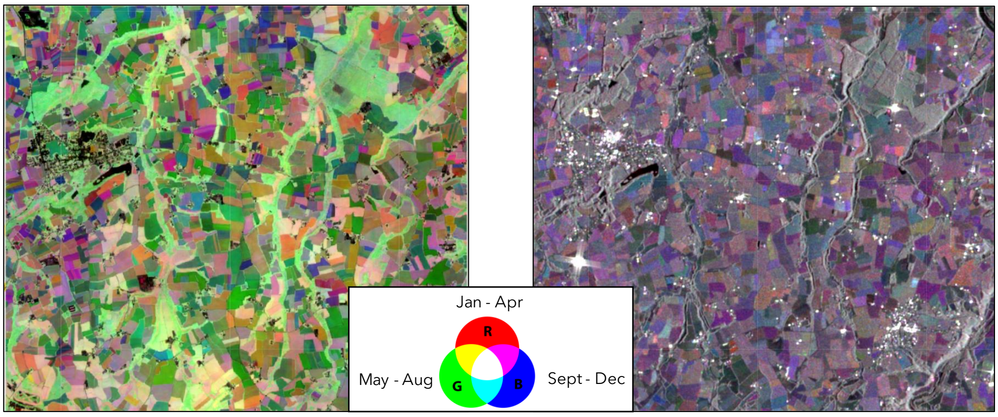


Figure 4. Multi-temporal RGB composite. Sentinel-2 NDVI (**left**). Sentinel-1 VV (**right**).

Sentinel-1: The Sentinel-1 collection comprises ground range detected (GRD) scenes in interferometric wide (IW) swath mode with dual polarization (VV and VH). The maximum number of acquisitions (in our given time frame) over a swath in Finistère is 75. Only data from one pass (descending) and the same orbit number (154) are used to ensure similar acquisition conditions. GEE adopts a standard processing workflow implemented in ESA's Sentinel-1 toolbox [64] to produce analysis-ready data. The workflow is applied in the order of: border noise removal, thermal noise removal, radiometric calibration, terrain correction, and conversion to decibels. Sentinel-1 images (provided at a 10 m spatial resolution in GEE) are co-registered to Sentinel-2. We further apply a ratio-based multi-temporal speckle filtering technique proposed by [65].

Sentinel-2: Sentinel-2 is a wide-swath, high-resolution, multi-spectral imaging mission. Sentinel-2 bands vary in spatial resolution at 10, 20, and 60 m. Surface reflectance products (level-2A) available in GEE are gathered. These have been generated using `Sen2Cor` [66], a processor for the atmospheric, terrain, and cirrus correction of top-of-atmosphere Level 1C data. Atmospheric bands (coastal aerosol, water vapor, and cirrus bands at 60 m) are excluded, and the remaining bands (in the visible and infra-red range) are considered. Bands at 20 m are resampled to 10 m. The collection is filtered for images with a cloudy cover percentage below 80% to limit the occurrence of cloudy observations. Then, for each month, two acquisitions with the least cloudy cover percentage are retained. This results in a total of 30 images or less (27) for few parcels, where available images could not meet the cloud criteria. As seen in Figure 3, temporal gaps in Sentinel-2 are irregular using this strategy, yet specific to parcels within the same scene footprints. Moreover, the strategy is hindered by its dependency on metadata, which is subject to quality issues [67]. Inglada et al. [12] have shown that using Sentinel-1 and Sentinel-2 without gap-filling cloudy areas can yield results equivalent to using Sentinel-2 only in a cloud-free setting; hence, the analysis proceeds without any pixel-based cloud masking or gap-filling approach.

3.4. Data Preparation

Feature normalization: For every image, the individual bands are normalized (per image, date, and channel) to transform pixel values into a common scale while preserving inherent similarities and variations. This process is known to accelerate convergence with machine learning algorithms and to avoid exerting importance on features with higher dynamic ranges. Normalization is applied using feature min-max normalization, which deduces minimum and maximum values from 2% and 98% percentile to be less sensitive to outlier values.

Input data organization: Besides features values, acquisition dates for each parcel are stored for both sensors. Acquisition dates for Sentinel-1 are constant, but vary for Sentinel-2 due to the cloud reduction strategy adopted. All parcels are cropped to the time series collection and stored per sensor as NumPy arrays of shape $T \times C \times N$, where T is the time series length, C is the number of channels, and N is the number of pixels within a parcel.

4. Methods

4.1. Overview of Pixel Set Encoder–Temporal Attention Encoder

To explore different fusion strategies, we adopted pixel set encoder–temporal attention encoder (PSE–TAE) [43] as the deep learning architecture over existing supervised learning algorithms dedicated to SITS classification. The PSE-TAE architecture is a spatio-temporal classifier for the classification of SITS at object-level. We thus assume that field geometries are accessible and known, which is the case for most fields in Europe [68]. Besides being the state-of-the-art algorithm for SITS classification, the choice of PSE-TAE lies in its ability to (i) address variable parcel size and account for irregular temporal sampling, (ii) learn long-term dependencies through self-attention mechanisms, and (iii) operate with lesser memory requirement, thus improving computational efficiency. The two main components of the architecture are (i) the spatial encoder (pixel set encoders) and (ii) the temporal attention encoder.

Pixel set encoder (PSE): In deep learning, textural information is usually extracted using CNNs. This information is lost when input images have a rather coarse resolution [69]. To overcome this limitation, the architecture uses pixel-set encoders, which computes learned statistical descriptors of the spectral distribution of the parcel's observations from a randomly selected number of pixels. The pixels are processed by shared consecutive MLPs to obtain a spatio-spectral embedding per date.

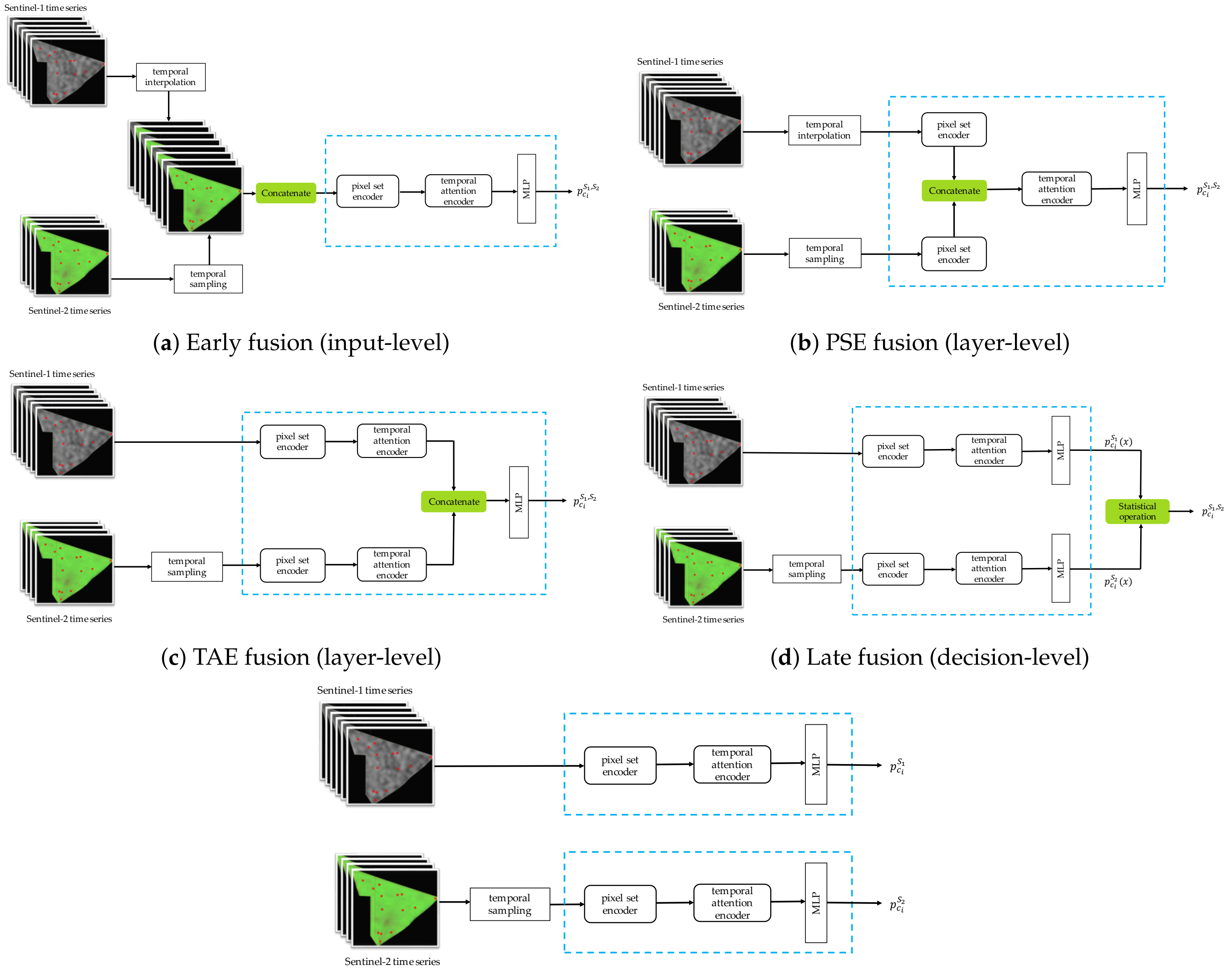
Temporal attention encoder (TAE): The TAE is founded on self-attention mechanisms [70]. Using this concept emphasizes the relationship among the different positions of an input sequence (here a time series) in order to compute a representation of the sequence. The relative positions of the sequences are preserved by using a positional encoder (based on sine and cosine functions), and this information is added to the PSE embeddings. The TAE takes the sum of both embeddings as direct inputs. The use of multi-head attention allows the model to jointly attend to information from different representation subspaces at different temporal positions while facilitating parallelization and improving computation as opposed to the sequential processing of RNNs.

Finally, the resulting TAE embeddings are processed by a multi-layer perceptron (MLP) to generate class logits.

4.2. Fusion Strategies

Per the research objectives, the synergy between Sentinel-1 and Sentinel-2 is explored along different parts of the architecture to determine where they enhance classification performance. Figure 5 denotes the different forms of fusion termed (a) early fusion, (b) PSE fusion, (c) TAE fusion, and (d) statistical fusion of class probabilities. According to standard definitions in deep learning, the fusion strategies are classified as (a) input-level, (b–c) layer-level, and (d) decision-level.

Early fusion: In the early fusion scenario, the ten processed spectral bands of Sentinel-2 are combined along the channel dimension with the two polarizations (VV and VH) of Sentinel-1. This concatenation is only possible if both Sentinel-1 and Sentinel-2 time series have the same number of acquisitions. Hence, we decide to reduce the temporal dimension of Sentinel-1 to 27 from 75 to match the dimension of Sentinel-2. We explored two strategies for this temporal interpolation operation: (i) a temporal nearest neighbor resampling, where the Sentinel-1 observations having the same or near acquisition dates to Sentinel-2 are selected, and (ii) a linear temporal interpolation on the acquisition dates of Sentinel-2. While the second strategy should reduce the reconstruction error, especially if some big temporal gaps occur in Sentinel-2 time series, we observed that the first strategy is a good enough approximation for the classification task given.



(**a**) Early fusion (input-level)

(**b**) PSE fusion (layer-level)

(**c**) TAE fusion (layer-level)

(**d**) Late fusion (decision-level)

(**e**) No fusion

Figure 5. Levels of multi-modal fusion strategies with pixel set encoder (PSE)–temporal attention encoder (TAE): (**a**) early fusion, (**b**) PSE fusion, (**c**) TAE fusion, (**d**) late fusion, and (**e**) no fusion.

PSE fusion: In this scenario, the layer-level fusion is performed after the PSE module by using also a concatenation operation. The output of PSE is a time series embedding, whose length is the same as the input time series. The concatenation between Sentinel-1 and Sentinel-2 embeddings is thus possible if and only if Sentinel-1 and Sentinel-2 PSE embeddings have the same length. We adopt the same strategy as for the early fusion and resample Sentinel-1 input time series to match the length of Sentinel-2 time series. We could also directly resample the Sentinel-1 PSE embedding, but this would result in a PSE module for Sentinel-1 larger than necessary.

TAE fusion: For this second layer-level strategy, the fusion is performed after the TAE module by concatenating embeddings learned by two independent PSE-TAE networks. The concatenation occurs before the MLP classifier. It is a straightforward operation as both embeddings have the same size.

Late fusion: In this last scenario, we apply a decision-level fusion where the class probabilities generated by two independent PSE-TAE classifiers are combined. We explore two computation strategies: (i) a simple average between probabilities produced by both classifiers and (ii) the product of experts proposed by [56]. We observed that a simple average obtains higher performance, and thus report only these results.

In Section 5, two categories of experiments are designed. Firstly, data from one sensor (either Sentinel-1 or Sentinel-2) is fed into the architecture to generate class predictions (Figure 5e). An identical setup of the architecture is used in each case. This serves as the baseline. Secondly, their synergy is examined through the four fusion strategies previously presented. Please note that a temporal resampling of Sentinel-2 time series is required for all our experiments. Sentinel-2 sequence lengths vary between 27 and 30 acquisitions due to the cloud cover filtering strategy (see Figure 3); it thus creates feature vectors of variable length for each parcel. Such inconsistencies become an issue when training most machine and deep learning algorithms (e.g., distance computation or use of batches). Hence, we randomly sample 27 observations (minimum sequence length observed in Sentinel-2 time series) for each parcel.

Table 2 displays the number of trainable parameters for all the presented configurations. The numbers of parameters for single-sensor experiments are also displayed (Sentinel-1 only and Sentinel-2 only).

Table 2. Number of parameters for the different fusion strategies.

Experiments	No. Parameters
Sentinel-1 only	163,084
Sentinel-2 only	163,340
Early fusion	163,404
PSE fusion	798,956
TAE fusion	323,692
Late fusion	326,424

Sentinel-1 data have longer time series but only two (polarimetric) features, whereas Sentinel-2 data present shorter time series but a higher number of (spectral) features. This results in a similar number of trainable parameters for the two baseline models and slightly different training times (about 1 h for Sentinel-2 and 2 h for Sentinel-1). The early fusion also results in a similar number of trainable parameters such as the single sensor baselines, which is expected as only the amount of features slightly increases. The TAE and late fusion strategies double the number of trainable parameters as two PSE-TAE streams are used to process independently the two modalities. Finally, the PSE fusion implies the highest number of trainable parameters. This is due to the increase in the input embedding size for the TAE module that would require double the size of the weight matrix involved in the self-attention computations.

4.3. Experimental Setting

We adapt and extend the original GitHub implementation of the PSE-TAE [43] to (i) accommodate multi-sensor inputs and (ii) design the different fusion strategies. Our code is made publicly available at https://github.com/ellaampy/CropTypeMapping (accessed on 16 February 2021).

We use a similar configuration to the original PSE-TAE work. Networks are trained using the Adam optimizer with a learning rate of 0.001 for 100 epochs, which is, as in the original paper and in our experiments, enough for all models to achieve convergence. The loss function used is the focal loss ($\gamma = 2$ as suggested in [71]), which was proposed to address the class imbalance issue. As in [43], the number of randomly sampled pixels per parcel in the PSE module is set to 64, and the dropout rate (used in TAE) is set to 0.2. The model is trained using one RTX-2080-Ti GPU of RAM 11 GB. The influence of batch size is evaluated on the validation set, and a size of 512 is found to provide better performance. The best model is selected by monitoring the mean Intersection over Union (mIoU) on the validation set (see Section 3.2).

For each iteration, the parcel data, as well as the dates of acquisition (transformed into days of the year), are forwarded to the PSE-TAE architecture. The dates are used to index into a pre-computed positional encoding of maximum length 460 (maximum day of the year spanning October 2018 to December 2019) and added to the embeddings to preserve their order. The randomization processes, including the sampling of pixels per parcel is fixed for each run across all experiments. This ensures that the variations in results are independent of the input. We also ensure that the same pixels are selected per parcel for Sentinel-1 and Sentinel-2 data in fusion experiments.

4.4. Evaluation Metrics

Classification performance is assessed using established metrics including overall accuracy, average F-score, mean Intersection over Union (mIoU), and Kappa statistics. To produce reliable assessments, the result from each experiment is averaged over five runs. Prediction maps are also qualitatively compared to the ground truth labels over selected subsets of the test study area (red squares in Figure 2).

5. Experimental Results

5.1. Comparative Evaluation of Fusion Strategies

Table 3 summarizes the evaluation metrics for single and multi-sensor experiments averaged over five runs. Early and PSE fusions use a nearest-neighbor interpolation, whereas the late fusion strategy uses the average of class probabilities (see Section 4.2). We observe that for over 150,000 parcels, the PSE-TAE training and testing time averaged between 1–2 h for single-sensor and up to 3 h for multi-sensor experiments.

Table 3. Per-class accuracy and overall accuracy, F-score, and Kappa averaged (± one standard deviation) over 5 runs for each experiment. Bold values show the highest performance.

		S1	**S2**	**Early**	**PSE**	**TAE**	**Late**
Maize	🟧	0.943	0.968	0.967	0.973	0.972	**0.975**
Wheat	🟨	0.921	0.960	0.958	0.952	0.962	**0.971**
Barley	🟫	0.816	0.927	0.924	0.939	0.942	**0.946**
Rapeseed	🟪	0.911	0.900	0.891	0.935	**0.943**	0.942
Other cereals	🟥	0.260	0.341	0.375	0.411	**0.430**	0.335
Protein crops	🟫	0.213	0.213	0.284	**0.307**	0.271	0.218
Vegetables/flowers	🟪	0.652	0.736	**0.771**	0.738	0.736	0.761
Orchards	🟩	0.004	0.089	0.102	**0.138**	0.098	0.040
Meadows	🟩	0.961	0.952	0.957	0.937	0.951	**0.980**
Gel	🟦	0.003	0.047	0.040	**0.085**	0.042	0.001
Fodder	🟥	0.289	0.341	0.403	**0.442**	0.377	0.309
Pasture/moor	🟥	0.168	0.475	**0.546**	0.441	0.502	0.248
	OA	$0.896^{\pm.003}$	$0.916^{\pm.007}$	$0.922^{\pm.002}$	$0.913^{\pm.008}$	$0.913^{\pm.008}$	**0.934**$^{\pm.004}$
	Kappa	$0.842^{\pm.003}$	$0.875^{\pm.011}$	$0.883^{\pm.003}$	$0.872^{\pm.01}$	$0.881^{\pm.013}$	**0.901**$^{\pm.006}$
	mIoU	$0.444^{\pm.002}$	$0.502^{\pm.013}$	**0.525**$^{\pm.007}$	$0.519^{\pm.014}$	$0.519^{\pm.014}$	$0.515^{\pm.009}$
	F-score	$0.522^{\pm.003}$	$0.587^{\pm.015}$	**0.612**$^{\pm.007}$	$0.611^{\pm.015}$	$0.611^{\pm.015}$	$0.591^{\pm.008}$

Between single sensor baselines, Sentinel-2 outperforms Sentinel-1, as established in previous studies using random forests [12,14,24]. Sentinel-2 achieves 2% increase in overall accuracy, ≈7% units in F-score, and up to 30% gains in per-class accuracies. Both sensors have a comparable performance for *protein crops*, *rapeseed*, and *meadows*. The highest per-class accuracies and overall metrics are obtained by multi-sensor fusion. Fusion scenarios outperform Sentinel-1 with 2–4% gain in overall accuracy and up to 9% units in F-score. Lower gains in overall accuracy (below 2%) are observed over Sentinel-2 except for PSE and TAE where performance is comparable.

Interestingly, we also notice that under-represented classes (below 3%, see Table 1)—*rapeseed, other cereals, protein crops, orchards, gel, fodder*, and *pasture/moor*—are better predicted at input and layer levels (early, PSE, and TAE). Conversely, late fusion is well suited for dominant classes. Although *rapeseed* is an under-represented class, it was highly distinguishable in all scenarios with over 85% accuracy and small confusion with other winter crops (*wheat* and *barley*).

5.2. Qualitative Analysis

Figure 6 compares prediction maps for single (second and third column) and multi-sensor (last column) scenarios over subsets of Brest (see Table 1 for the legend). *Protein crops* and *maize* are rightly classified by Sentinel-1 in site A (black rectangle) but missed by early fusion. Subsequently, *vegetables* are confused with *other cereals* and *fodder* in early fusion (black square) but predicted accurately by Sentinel-2. In test site B (blue ellipses), *fodder* and *meadow* are missed by Sentinel-1 and Sentinel-2 but classified correctly through fusion. However, there are cases where neither single or multi-sensor predicts correctly (red box).

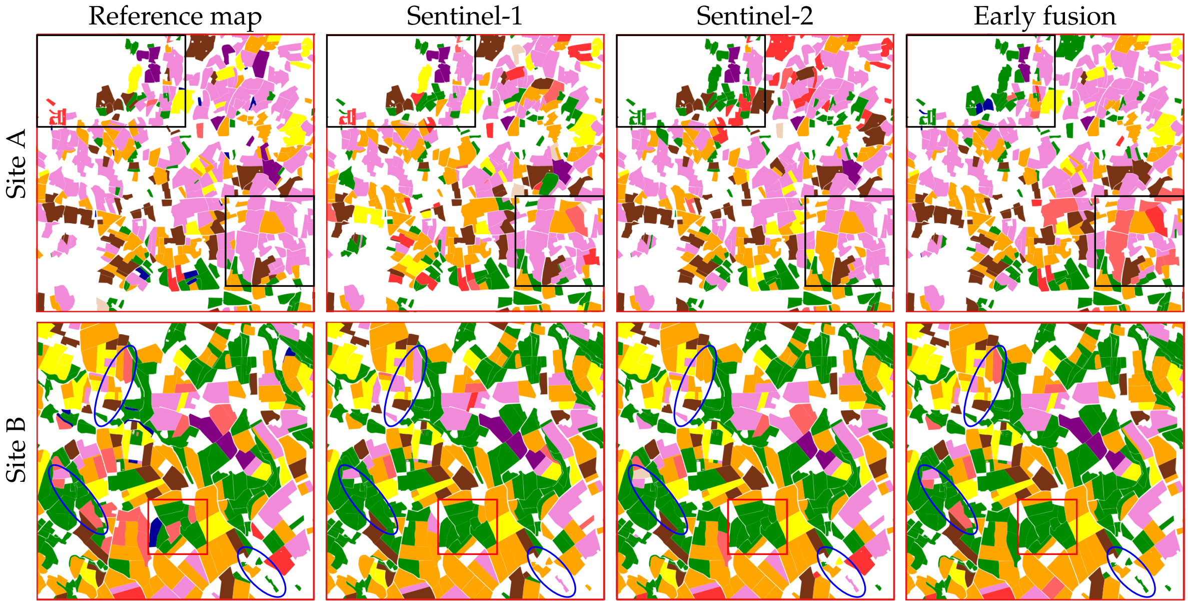


Figure 6. Qualitative comparison of prediction maps for two different sites. The first column displays the reference land cover map, the second and third column present predictions from Sentinel-1 and Sentinel-2, and the last column from early fusion. See Table 1 for the legend.

5.3. Sparser Time Series

We study the performance of PSE-TAE architecture on sparse time series by reducing the number of Sentinel-2 observations. The aim is to mimic cloudy scenarios, where the number of non-cloudy optical observations is limited. A certain percentage of the minimum length of Sentinel-2 (27) is randomly sampled at each trial of the five runs. We compare single sensor baseline (Sentinel-2 only and Sentinel-1 only) to three other forms of fusion: early, TAE, and late fusion with averaging. We exclude experiments using PSE fusion due to the relatively higher number of trainable parameters yet comparable results to other forms of fusion (see Table 2). Table 4 shows the average F-score. Except for early fusion where Sentinel-1 images are selected at nearest dates of Sentinel-2, we have used all Sentinel-1 time series.

Table 4. Sparse time series: F-score averaged (±one standard deviation) over 5 runs considering n% of Sentinel-2 data. Bold values show the highest performance for different quantities of missing data.

	20% **5 obs.**	**40%** **11 obs.**	**60%** **16 obs.**	**80%** **22 obs.**	**100%** **27 obs.**
Sentinel-2	$0.443^{\pm.012}$	$0.537^{\pm.015}$	$0.576^{\pm.012}$	$0.587^{\pm.012}$	$0.587^{\pm.017}$
Sentinel-1	$0.522^{\pm.003}$	$0.522^{\pm.003}$	$0.522^{\pm.003}$	$0.522^{\pm.003}$	$0.522^{\pm.003}$
Early	$0.478^{\pm.013}$	$0.564^{\pm.005}$	$\mathbf{0.595}^{\pm.004}$	$0.603^{\pm.014}$	$\mathbf{0.612}^{\pm.008}$
TAE	$\mathbf{0.558}^{\pm.010}$	$\mathbf{0.575}^{\pm.017}$	$0.590^{\pm.007}$	$\mathbf{0.605}^{\pm.007}$	$0.611^{\pm.017}$
Late	$0.540^{\pm.007}$	$0.564^{\pm.014}$	$0.584^{\pm.012}$	$0.583^{\pm.010}$	$0.591^{\pm.009}$

As more data are ingested, i.e., increasing the number of Sentinel-2 acquisitions, there is a general increase observed in F-score. Up until 100% sampling of optical data, most of the fusion strategy outperforms Sentinel-2. The TAE fusion, which uses all Sentinel-1 data, outperforms early fusion when the number of Sentinel-2 observations available is low. If a higher number of Sentinel-2 observations is available, early and TAE fusion strategies perform similarly. We also observe that the use of only five Sentinel-2 observations already allows a significant increase (about 0.036) in the F-Score compared to the use of only Sentinel-1 time series data.

5.4. Incremental Learning

The goal of this experiment is to study the benefit of fusion for mapping crops before having a full year of data. For this purpose, we rerun our experiments by quarterly increments in data from October 2018. Figure 7 displays the overall F-Score averaged over five runs for single sensors and the best fusion strategy (early fusion).

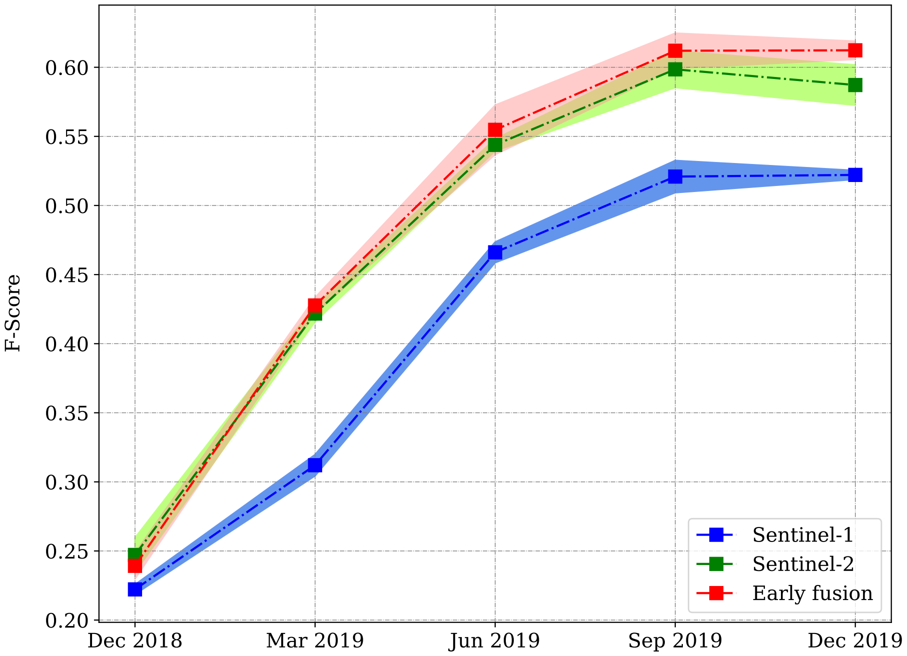


Figure 7. Average F-Score for quarterly classification for three scenarios: Sentinel-1 only (in blue), Sentinel-2 only (in green), and early fusion (in red).

We observe similar incremental patterns in F-score across all experiments as the season progresses. Sentinel-1 time series yields the lower performance, which is expected as optical data are a better estimator of phenological activity that is crucial to distinguish crops. Moreover, the early fusion dominates from March 2019 and onward. The performance tends to saturate after September 2019, which corresponds to the harvesting period of the summer crops. The use of both optical and radar modalities leads thus to a slight performance improvement for early crop type mapping.

6. Conclusions

The interplay between publicly available satellite data and current state-of-the-art remote sensing techniques can provide cost-effective, accurate, and timely information on crop extent and dynamics. In this study, Sentinel-1 and Sentinel-2 time series are harmonized for crop type mapping in Finistère, France using PSE-TAE, a deep learning architecture that leverages both spatial and temporal dimensions of SITS data. The computational efficiency offered by PSE-TAE allows the rapid assessment of different model configurations. Key findings from the study are summarized below:

- Combined Sentinel-1 and Sentinel-2 modalities are beneficial to increase classification performance [23–25] in majority and minority classes. Depending on the availability of class samples, different forms of fusion are suggested. When classes of interest are under-represented, it is better to use input or layer level fusion. In the case of well-represented classes, any form of fusion is sufficient, but decision-level fusion introduces more gains.
- The fusion of high-level temporal features obtained from Sentinel-1 and Sentinel-2 time series produces reliable results when optical data are very limited.
- Our experiments also confirm main results in the existing literature [12,14,24]: the use of optical time series results in higher performance than the use of radar time series.

Due to the limited research in optical and radar SITS fusion using deep learning, future work is required to investigate the contributions of radiometric indices, e.g., vegetation indices and polarimetric ratios [13] as well as augmentation techniques to improve the instances for under-represented classes.

Author Contributions: All the authors—S.O.-A., C.P. and S.L.—were involved in the writing of the manuscript and the design of experiments. Conceptualization, S.O.-A. and C.P.; Investigation, S.O.-A. and C.P.; Methodology, S.O.-A., C.P. and S.L.; Writing—original draft preparation, S.O.-A., C.P. and S.L.; Writing—review and editing, S.O.-A., C.P. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: The first author, Stella Ofori-Ampofo, received a European scholarship to engage in Master Copernicus in Digital Earth, an Erasmus Mundus Joint Master Degree (EMJMD). This research work was conducted during her master's thesis.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: We have made available the scripts to download the data in Github: https://github.com/ellaampy/GEE-to-NPY (accessed on 16 February 2021).

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

References

1. Brown, M.; Antle, J.; Backlund, P.; Carr, E.; Easterling, B.; Walsh, M.; Ammann, C.; Attavanich, W.; Barrett, C.; Bellemare, M.; et al. *Climate Change, Global Food Security and the US Food System*; Technical Report; University Library of Munich: Munich, Germany, 2015.
2. Food and Agriculture Organisation (FAO). *FAO's Work on Climate Change*; Technical Report, United Nations Climate Change Conference; FAO: Rome, Italy, 2019.
3. Smith, P.; Clark, H.; Dong, H.; Elsiddig, E.; Haberl, H.; Harper, R.; House, J.; Jafari, M.; Masera, O.; Mbow, C.; et al. *Agriculture, Forestry and Other Land Use (AFOLU)*; Technical Report; Cambridge University Press: Cambridge, UK, 2014.

4. Iizumi, T.; Furuya, J.; Shen, Z.; Kim, W.; Okada, M.; Fujimori, S.; Hasegawa, T.; Nishimori, M. Responses of crop yield growth to global temperature and socioeconomic changes. *Sci. Rep.* **2017**, *7*, 7800. [CrossRef] [PubMed]
5. Food and Agriculture Organisation (FAO). *The State of Agricultural Commodity Markets 2018*; Technical Report, Agricultural Trade, Climate Change and Food Security; FAO: Rome, Italy, 2018.
6. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [CrossRef]
7. Ienco, D.; Interdonato, R.; Gaetano, R.; Minh, D.H.T. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 11–22. [CrossRef]
8. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
9. Bhatta, B.; Saraswati, S.; Bandyopadhyay, D. Urban sprawl measurement from remote sensing data. *Appl. Geogr.* **2010**, *30*, 731–740. [CrossRef]
10. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [CrossRef]
11. Tanase, M.A.; Belenguer-Plomer, M.A.; Roteta, E.; Bastarrika, A.; Wheeler, J.; Fernández-Carrillo, Á.; Tansey, K.; Wiedemann, W.; Navratil, P.; Lohberger, S.; et al. Burned area detection and mapping: Intercomparison of Sentinel-1 and Sentinel-2 based algorithms over tropical Africa. *Remote Sens.* **2020**, *12*, 334. [CrossRef]
12. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* **2016**, *8*, 362. [CrossRef]
13. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [CrossRef]
14. Stendardi, L.; Karlsen, S.R.; Niedrist, G.; Gerdol, R.; Zebisch, M.; Rossi, M.; Notarnicola, C. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sens.* **2019**, *11*, 542. [CrossRef]
15. Doraiswamy, P.C.; Moulin, S.; Cook, P.W.; Stern, A. Crop yield assessment from remote sensing. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 665–674. [CrossRef]
16. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [CrossRef]
17. Mabalay, M.R.; Nelson, A.; Setiyono, T.; Quilang, E.J.; Maunahan, A.; Abonete, P.; Rala, A.; Raviz, J.; Skorzus, R.; Loro, J.; et al. Remote Sensing-Based Information and Insurance for Crops in Emerging Economies (Riice): The Philippine's Experience. In Proceedings of the 34th Asian Conference on Remote Sensing, ACRS, Bali, Indonesia, 20–24 October 2013.
18. Wang, J.; Xiao, X.; Liu, L.; Wu, X.; Qin, Y.; Steiner, J.L.; Dong, J. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sens. Environ.* **2020**, *247*, 111951. [CrossRef]
19. Son, N.T.; Chen, C.F.; Chen, C.R.; Minh, V.Q. Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto Int.* **2018**, *33*, 587–601. [CrossRef]
20. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [CrossRef]
21. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal Convolutional Neural Network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
22. Rußwurm, M.; Körner, M. Self-attention for raw optical satellite time series classification. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 421–435. [CrossRef]
23. Forkuor, G.; Conrad, C.; Thiel, M.; Ullmann, T.; Zoungrana, E. Integration of optical and Synthetic Aperture Radar imagery for improving crop mapping in Northwestern Benin, West Africa. *Remote Sens.* **2014**, *6*, 6472–6499. [CrossRef]
24. Denize, J.; Hubert-Moy, L.; Betbeder, J.; Corgne, S.; Baudry, J.; Pottier, E. Evaluation of using Sentinel-1 and-2 time-series to identify winter land use in agricultural landscapes. *Remote Sens.* **2019**, *11*, 37. [CrossRef]
25. Song, X.P.; Huang, W.; Hansen, M.C.; Potapov, P. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* **2021**, *3*, 100018. [CrossRef]
26. Liao, C.; Wang, J.; Xie, Q.; Baz, A.A.; Huang, X.; Shang, J.; He, Y. Synergistic Use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sens.* **2020**, *12*, 832. [CrossRef]
27. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
28. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
29. Campos-Taberner, M.; García-Haro, F.J.; Martínez, B.; Sánchez-Ruíz, S.; Gilabert, M.A. A Copernicus Sentinel-1 and Sentinel-2 classification framework for the 2020+ European common agricultural policy: A case study in València (Spain). *Agronomy* **2019**, *9*, 556. [CrossRef]
30. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sens.* **2015**, *7*, 12356–12379. [CrossRef]

31. Rodríguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [CrossRef]
32. 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]
33. Ienco, D.; Gaetano, R.; Dupaquier, C.; Maurel, P. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1685–1689. [CrossRef]
34. Tavares, P.A.; Beltrão, N.E.S.; Guimarães, U.S.; Teodoro, A.C. Integration of Sentinel-1 and Sentinel-2 for classification and LULC mapping in the urban area of Belém, eastern Brazilian Amazon. *Sensors* **2019**, *19*, 1140. [CrossRef] [PubMed]
35. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]
36. Ndikumana, E.; Ho Tong Minh, D.; Baghdadi, N.; Courault, D.; Hossard, L. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sens.* **2018**, *10*, 1217. [CrossRef]
37. Minh, D.H.T.; Ienco, D.; Gaetano, R.; Lalande, N.; Ndikumana, E.; Osman, F.; Maurel, P. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR Sentinel-1. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 464–468. [CrossRef]
38. Rußwurm, M.; Korner, M. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2016; pp. 11–19.
39. Khaki, S.; Wang, L.; Archontoulis, S.V. A cnn-rnn framework for crop yield prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [CrossRef]
40. Interdonato, R.; Ienco, D.; Gaetano, R.; Ose, K. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 91–104. [CrossRef]
41. Rußwurm, M.; Körner, M. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [CrossRef]
42. Martinez, J.A.C.; La Rosa, L.E.C.; Feitosa, R.Q.; Sanches, I.D.; Happ, P.N. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 188–201. [CrossRef]
43. Sainte Fare Garnot, V.; Landrieu, L.; Giordano, S.; Chehata, N. Satellite image time series classification with pixel-set encoders and temporal self-attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 12325–12334.
44. Sainte Fare Garnot, V.; Landrieu, L. Lightweight Temporal Self-attention for Classifying Satellite Images Time Series. In *International Workshop on Advanced Analytics and Learning on Temporal Data*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 171–181.
45. Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* **2019**, *221*, 551–568. [CrossRef]
46. d'Andrimont, R.; Verhegghen, A.; Lemoine, G.; Kempeneers, P.; Meroni, M.; van der Velde, M. From parcel to continental scale–A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *arXiv* **2021**, arXiv:2105.09261.
47. Garioud, A.; Valero, S.; Giordano, S.; Mallet, C. On the joint exploitation of optical and SAR imagery for grassland monitoring. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B3-2020*; ISPRS: Paris, France, 2020.
48. Hu, B.; Xu, Y.; Huang, X.; Cheng, Q.; Ding, Q.; Bai, L.; Li, Y. Improving Urban Land Cover Classification with Combined Use of Sentinel-2 and Sentinel-1 Imagery. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 533. [CrossRef]
49. Ferrant, S.; Selles, A.; Le Page, M.; Herrault, P.A.; Pelletier, C.; Al-Bitar, A.; Mermoz, S.; Gascoin, S.; Bouvet, A.; Saqalli, M.; et al. Detection of irrigated crops from Sentinel-1 and Sentinel-2 data to estimate seasonal groundwater use in South India. *Remote Sens.* **2017**, *9*, 1119. [CrossRef]
50. Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sens.* **2018**, *10*, 1642. [CrossRef]
51. Sun, L.; Chen, J.; Guo, S.; Deng, X.; Han, Y. Integration of Time Series Sentinel-1 and Sentinel-2 Imagery for Crop Type Mapping over Oasis Agricultural Areas. *Remote Sens.* **2020**, *12*, 158. [CrossRef]
52. El Hajj, M.; Baghdadi, N.; Zribi, M.; Bazzi, H. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sens.* **2017**, *9*, 1292. [CrossRef]
53. Bousbih, S.; Zribi, M.; Pelletier, C.; Gorrab, A.; Lili-Chabaane, Z.; Baghdadi, N.; Ben Aissa, N.; Mougenot, B. Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2. *Remote Sens.* **2019**, *11*, 1520. [CrossRef]
54. Poh, C.; Genderen, J. Multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854.
55. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3*, 100004. [CrossRef]
56. Valero, S.; Arnaud, L.; Planells, M.; Ceschia, E.; Dedieu, G. Sentinel's Classifier Fusion System for Seasonal Crop Mapping. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6243–6246.
57. Gómez-Chova, L.; Tuia, D.; Moser, G.; Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **2015**, *103*, 1560–1584. [CrossRef]

58. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proc. IEEE* **2015**, *103*, 1585–1601. [CrossRef]
59. Lopes, M.; Frison, P.L.; Durant, S.M.; Schulte to Bühne, H.; Ipavec, A.; Lapeyre, V.; Pettorelli, N. Combining optical and radar satellite image time series to map natural vegetation: Savannas as an example. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 316–326. [CrossRef]
60. Benedetti, P.; Ienco, D.; Gaetano, R.; Ose, K.; Pensa, R.G.; Dupuy, S. M3-Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4939–4949. [CrossRef]
61. Rustowicz, R.M.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; Lobell, D. Semantic segmentation of crop type in Africa: A novel dataset and analysis of deep learning methods. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 75–82.
62. 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]
63. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
64. Veci, L.; Prats-Iraola, P.; Scheiber, R.; Collard, F.; Fomferra, N.; Engdahl, M. The Sentinel-1 toolbox. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 1–3.
65. Zhao, W.; Deledalle, C.A.; Denis, L.; Maître, H.; Nicolas, J.M.; Tupin, F. Ratio-based multitemporal SAR images denoising: RABASAR. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3552–3565. [CrossRef]
66. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 Sen2Cor: L2A processor for users. In Proceedings of the Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016; pp. 1–8.
67. Sudmanns, M.; Tiede, D.; Augustin, H.; Lang, S. Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. *Int. J. Digit. Earth* **2020**, *13*, 768–784. [CrossRef] [PubMed]
68. Léo, O.; Lemoine, G. *Land Parcel Identification System in the Frame of Regulation (EC) 1593/2000 Version 1.4*; Discussion Paper; European Commission Directorate General Joint Research Centre (JRC)—ISPRA Space Application Institute Agriculture and Regional Information Systems Unit: Ispra, Italy, 2001.
69. Sainte Fare Garnot, V.; Landrieu, L.; Giordano, S.; Chehata, N. Time-space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6247–6250.
70. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
71. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.