

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://SPIDigitalLibrary.org/conference-proceedings-of-spie)

## Multi-year crop type mapping using pre-trained deep long-short term memory and Sentinel 2 image time series

Giulio Weikmann, Claudia Paris, Lorenzo Bruzzone

Giulio Weikmann, Claudia Paris, Lorenzo Bruzzone, "Multi-year crop type mapping using pre-trained deep long-short term memory and Sentinel 2 image time series," Proc. SPIE 11862, Image and Signal Processing for Remote Sensing XXVII, 118620O (12 September 2021); doi: 10.1117/12.2600559

**SPIE.**

Event: SPIE Remote Sensing, 2021, Online Only

# Multi-year Crop Type Mapping using pre-trained deep Long-Short Term Memory and Sentinel 2 Image Time Series

Giulio Weikmann, Claudia Paris, and Lorenzo Bruzzone

Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 5 I-38123, Trento, Italy

## 1. ABSTRACT

This work presents a system for multi-year crop type mapping based on the multi-temporal Long Short-Term Memory (LSTM) Deep Learning (DL) model and Sentinel 2 image Time Series (TS). The method assumes the availability of a pre-trained LSTM model for a given year and aims to update the corresponding crop type map for a different year considering a small amount of recent reference data. To this end, the proposed approach combines Self-Paced Learning (SPL) and fine-tuning (FT) techniques. While the SPL technique gradually incorporates samples from crop types that can be classified with high-confidence by the pre-trained model, the FT strategy adapts the network to those classes having low-confidence accuracy. This condition allows us to reduce the labeled samples required to achieve accurate classification results. The experimental results obtained on three tiles of the Austrian country on TSs of Sentinel 2 data acquired in 2019 and 2020 (considering a model pre-trained on images of 2018) demonstrate the capability of the LSTM to adapt to TS of images with different temporal and radiometric characteristic with respect to the one used to pre-train the model, with a relatively small number of training samples. As expected, by directly applying the model without performing any adaptation, we obtain a mean F-score (F1%) of 64% and 62% compared to 76% and 70% achieved by the proposed technique with only 1500 samples for 2019 and 2020, respectively.

**Keywords:** Long Short Term Memory (LSTM), automatic classification, multi-temporal Deep Learning (DL), multi-temporal analysis, multi-year crop type mapping, remote sensing.

## 2. INTRODUCTION

The production of multi-year crop type maps is extremely important from the operational viewpoint since crop cultivation varies significantly from year to year.<sup>1</sup> In the literature, several multitemporal Deep Learning (DL) architectures have been defined to classify agricultural areas by taking advantage from the availability of long and dense Time Series (TS) of satellite multispectral images.<sup>2</sup> Both recurrent<sup>3</sup> and time-convolutional DL models<sup>4</sup> proved to be more effective than standard shallows classifier for this peculiar classification task.<sup>5</sup> In particular, LSTM performed well against the other techniques, demonstrating a good capability of modeling complex temporal dynamics and long-term dependencies. However, most of the architectures focus on a specific training for the production of a single-year crop type map, even though the crop rotation practice requires to frequently update the map obtained. Although a DL model successfully trained can achieve accurate crop type mapping for the year of the labeled samples used in the network training, its accuracy can drastically drop if another year is considered because of the crop rotation practice (that may introduce large changes in the priors of classes) and the different acquisition condition of TS of optical images. These factors may lead to a strong variation of the class statistical distributions of TSs acquired over different years.<sup>6</sup> In this scenario, it is unfeasible to assume that a large number of labeled samples can be collected every year to generate updated maps. From the operational viewpoint, a large number of labeled samples must be collected every year.

To address this issue, a well-explored solution in the Machine Learning (ML) literature is the fine-tuning (FT) approach that aims to adapt a pre-trained network to a target dataset.<sup>7</sup> In general, this approach offers better performance compared to random initialization of the DL model and sharply reduces the amount of data required to successfully train a deep architecture from scratch. The traditional implementation of the FT aims at optimizing the parameters of the pre-trained DL model using recent training data. However, this approach requires a number of recent labeled data that has to be at least comparable to the numbers parameters that have to be updated in order to avoid underfitting.<sup>8</sup> To address this issue, several FT approaches freeze the weights of

all the layers in the pre-trained network except for the latest, which is re-trained using the recent training data.<sup>9</sup> However, this partial adaptation of the network leads to sub-optimal results. In this context, even though the FT approach improves the classification results, its accuracy depends on the amount of available recent training data, which is difficult to collect and time consuming.

To mitigate the need of labeled data, Self-Paced Learning (SPL) strategies have been widely used in the Remote Sensing (RS) literature. SPL take inspiration from the humans learning process which is based on the gradual incorporation of samples having increasing complexity (from easy to complex). Similar to semi-supervised learning, SPL optimizes the initial training model by iteratively incorporating samples from the image to be classified.<sup>10</sup> In particular, at each iteration more samples of the recent RS data can be added by gradually increasing the learning pace parameter on the SPL regularizer in an unsupervised way. The effectiveness of the approach has been demonstrated in many computer vision and RS tasks.<sup>11</sup> For instance, the SPL learning strategy has been used to accelerate the learning convergence of the multilayer auto-encoders network trained on PolSAR images.<sup>12</sup> The results obtained demonstrate that the use of SPL allows for a stronger generalization capability. SPL has been incorporated into convolutional networks to support change detection in heterogeneous RS images.<sup>13</sup> The approach leverages on the SPL strategy to dynamically select reliable samples while modelling the relations between the two RS data. The possibility of integrating SPL and FT to achieve multi-year crop type mapping with a small amount of recent labeled data is extremely interesting from the operational viewpoint.

This work presents a system for multi-year crop type mapping based on the multi-temporal Long Short-Term Memory (LSTM) model and Sentinel 2 Image TSs. The method assumes available a pre-trained LSTM model for a given year and aims to update the corresponding crop type map for a different year, for which a small amount of reference data is available. This is achieved by a hybrid strategy based on FT and SPL. First, the SPL allows the adaptation of the pre-trained classification model to the recent TS of images by progressively enlarging the initial training set in an unsupervised way. In the considered SPL strategy, at each iteration an increasing number of training samples is extracted for each class, forcing the network to obtain a balanced training set. Then, a FT step is performed at the convergence of the SPL step. The number of samples required to achieve accurate classification results is much lower than the one needed by the standard FT approach due to the SPL strategy.

The structure of the rest of the paper is as follows. Section 3 presents the proposed system architecture for multi-year crop type mapping. Section 4 describes the considered study area, while Section 5 presents the experimental setup defined to test the effectiveness of the proposed approach. Then, the experimental results obtained classifying TSs of Sentinel 2 images acquired 2019 and 2020 using a model pre-trained on 2018 are presented in Section 6. Finally, Section 7 concludes the paper.

### 3. PROPOSED MULTI-YEAR CROP TYPE MAPPING APPROACH

Fig. 1 shows the block scheme of the proposed approach, which consists of four main steps: (1) the multi-year TSs preprocessing step that aims to harmonize the TSs acquired in different years, (2) the training/deployment of the network pre-trained on a given target year, (3) the SPL strategy that aims to enlarge the initial training set with samples classified with high-confidence by the pre-trained network, and (4) the FT of the network with samples extracted from the considered year.

#### 3.1 Multi-Year TSs Harmonization

The first step of the proposed approach aims to harmonize the multi-year TSs from the temporal and radiometric viewpoint. Indeed, TSs acquired in different years may have different temporal sampling and lengths due to different cloud coverage that hampers the use of some images. To harmonize the data from the temporal viewpoint, the proposed method generates homogeneous TSs of 12 monthly composites per year. In the considered implementation of the method, we employed a statistics-based approach widely used to generate monthly, seasonal or annual composites.<sup>14</sup> The sequence of pixel observations acquired within a month are collapsed into the median value of these observations at band level.<sup>15</sup> Then, the obtained multi-year TSs of 12 monthly composites are harmonized from the radiometric viewpoint. In particular, the monthly composite of different years are pre-processed in order to match the radiometric characteristics of spectral bands of the TS used in the

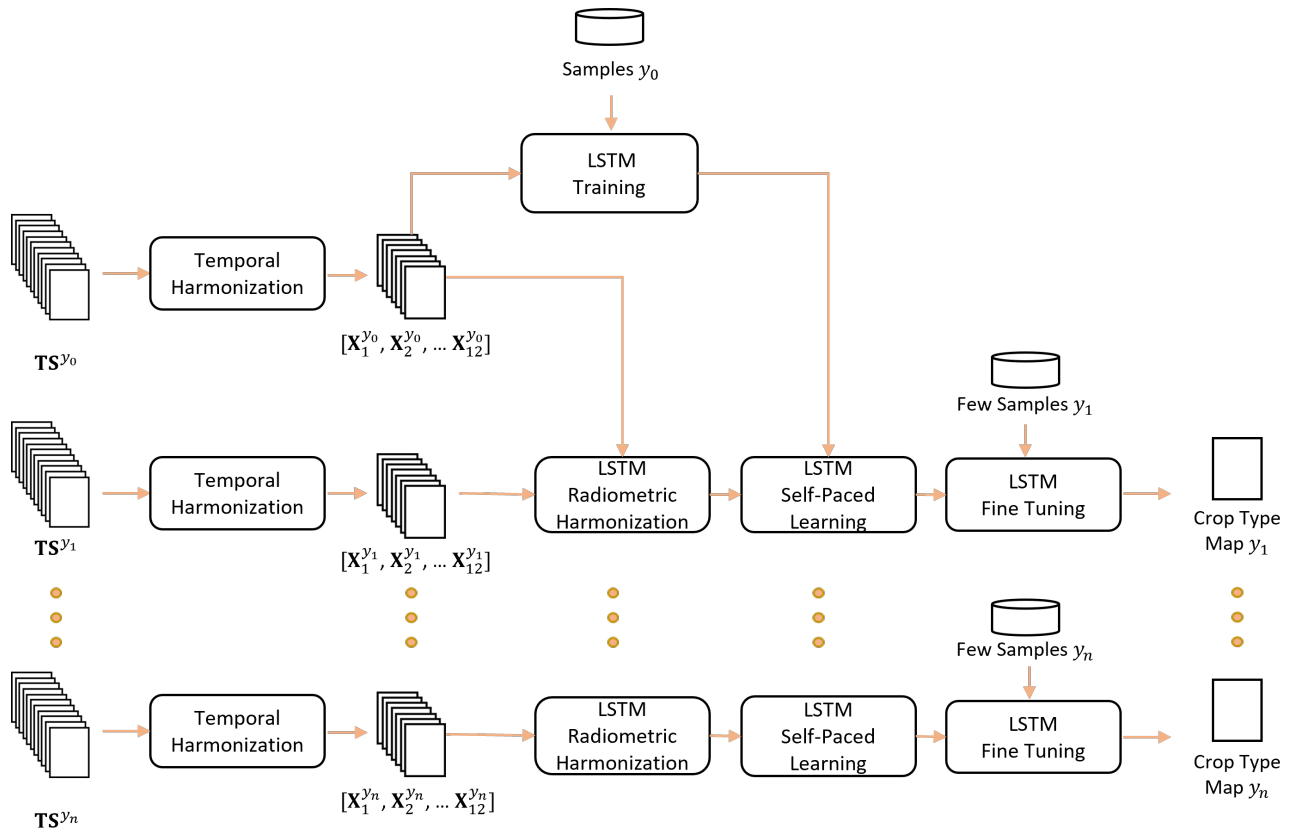


Figure 1. Architecture of the proposed system for the adaptation of a LSTM pre-trained on a different year to the considered TS of Sentinel 2 images.

pre-trained network, i.e., the target year  $y_0$ . Let us define the  $q$ th monthly composite of a generic TS of images as  $\mathbf{X}_q$  having size  $M \times L \times B$ , where the TS can be defined as  $[\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{12}]$ . The  $q$ th composite to be harmonized (i.e., belonging to the most recent TS) can be defined as  $\mathbf{X}_q^{y_1}$ , while the reference  $q$ th composite of the TS used for the pre-trained network is  $\mathbf{X}_q^{y_0}$ , with  $q = 1, 2, \dots, 12$ . First,  $\mathbf{X}_q^{y_1}$  is normalized in order to have zero mean and unit variance at band level. Let  $\mathbf{X}_{b,q}^{y_1}$  represents the  $b$ th band of the composite, with  $b = 1, 2, \dots, B$ . The normalized version of the  $b$ th band of the composite  $\dot{\mathbf{X}}_{b,q}^{y_1}$  is computed as follows:

$$\dot{\mathbf{X}}_{b,q}^{y_1} = \frac{\mathbf{X}_{b,q}^{y_1} - \mu_{b,q}^{y_1}}{\sigma_{b,q}^{y_1}} \quad \text{with } b = 1, 2, \dots, B \quad (1)$$

where,  $\mu_{b,q}^{y_1}$  and  $\sigma_{b,q}^{y_1}$  represent the mean and variance of the  $b$ th band of  $\mathbf{X}_q^{y_1}$ . Then, the normalized composite is processed to match the mean and the variance properties of the target image  $\mathbf{X}_q^{y_0}$ . Let  $\mu_{b,q}^{y_0}$  and  $\sigma_{b,q}^{y_0}$  be the mean variance of the  $b$ th band of  $\mathbf{X}_q^{y_0}$ . The harmonized source image is obtained as follows:

$$\hat{\mathbf{X}}_{b,q}^{y_1} = (\dot{\mathbf{X}}_{b,q}^{y_1} \cdot \sigma_{b,q}^{y_0}) + \mu_{b,q}^{y_0} \quad \text{with } b = 1, 2, \dots, B \quad (2)$$

This condition allows us to harmonize the TSs from both the temporal and spectral viewpoint, mitigating the shift between the different years. At the end of this step, we have the recent TS of harmonized composites  $[\hat{\mathbf{X}}_1^{y_1}, \hat{\mathbf{X}}_2^{y_1}, \dots, \hat{\mathbf{X}}_{12}^{y_1}]$ .

### 3.2 Deep Learning Architecture

The pre-trained network adopted is a multi-layer LSTM made up of three layers of 200, 125 and 100 hidden units respectively, a fully connected layer and a softmax layer. The pre-trained network used in<sup>16</sup> is particularly

suited for the crop classification task due to its ability to accurately model the phenological characteristics of the different crop types over the TS. Moreover, the network is robust to the presence of residual cloudy pixels in the scene, that may not have been correctly masked during the pre-processing step. Thanks to the pre-processing mentioned in the previous sub-section, the network can be adapted to the different years considered, since the TS length is harmonized. The network takes as input the harmonized TS, representing an agronomic year, and outputs from the softmax layer a pixel-wise posterior associated to its level of confidence in the classification results. This can be provided later in the SPL. In greater detail, the fully connected layer provides a number of output values matching the number of classes in the considered problem, which are then processed by the softmax layer to output a posterior probability. Since the problem is a multi-class classification problem, the loss function adopted is the categorical cross-entropy loss. The loss is back propagated through the network and the gradients are utilized by the RMSprop optimizer.<sup>17</sup>

### 3.3 Self-Paced Learning Strategy

To successfully train a DL network, a high number of training samples is typically required. However, this is not always feasible due to the scarcity of ground truth and reference samples. Indeed, the collection of annotated data is often labor-extensive and time-consuming. For this reason, it is not reasonable to assume that a DL architecture can be trained from scratch every time that a recent TS of satellite images is acquired. Although the adoption of pre-trained networks can solve this problem, the model has to be tuned in order to obtain satisfying accuracy on the new classification task. To ensure reliable results, recent labeled samples are still required. To reduce the amount of labeled data needed to achieve accurate classification results, the aim of this step is to exploit the capability of the pre-trained network to correctly classify some crop types of the recent TSs of images. To this end, an SPL strategy is used to progressively adapt the pre-trained network to the new dataset in an unsupervised way. Since the shift between the two years is limited, we can enlarge the initial training set with recent classified samples having a high posterior probability, i.e., high-confidence.

Let us define the model parameters of the pre-trained network as  $\theta$  and let  $\hat{\mathbf{x}}_i^{y_1} \in \mathbb{R}^d$  be the  $i$ th mutispectral pixel vector of the TS of harmonized composites  $[\hat{\mathbf{X}}_1^{y_1}, \hat{\mathbf{X}}_2^{y_1}, \dots, \hat{\mathbf{X}}_{12}^{y_1}]$ , where  $d = B \times 12$  and  $i = 1, 2, \dots, M \times L$ . The pre-trained network iteratively classifies the unlabeled recent samples  $\hat{\mathbf{x}}_i$ , giving as output of the softmax function the level of confidence, i.e., membership probability assigned by the classifier. This condition allows us to detect the high-confidence samples, which have the highest probability to be correctly labeled. Thanks to the SPL strategy, at each iteration in an unsupervised way we incorporate in the training set pseudo-labeled samples, gradually selecting samples from easy to complex by tuning a self-paced regularizer  $\lambda$ . In greater details, the rule can be defined as follows:

$$\min_{\theta, \omega} \mathcal{L}(\theta, \omega, \lambda) = \left\{ \sum_{i=1}^l \omega_i L(y_i, f(\hat{\mathbf{x}}_i^{y_1}, \theta)) + h(\lambda, \omega_i) \right\} \quad s.t. \omega_i \in 0, 1 \quad (3)$$

where  $\mathcal{L}(\cdot, \cdot)$  is the model loss function,  $\theta$  the parameters of the network,  $\omega$  the weights of the network where  $\omega_i$  is a weight associated the sample  $\hat{\mathbf{x}}_i^{y_1}$  that represents its degree of complexity and  $h(\lambda, \omega_i)$  is the self-paced function. In the considered implementation of the SPL strategy, the weight  $\omega_i$  is calculated as

$$\omega_i = \begin{cases} 1 & \text{if } L(y_i, f(\mathbf{x}_i, \theta)) < \lambda \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $\lambda$  controls the learning pace, thus affecting the size of the model. The unlabeled samples having loss smaller than  $\lambda$  are classified with high-confidence by the pre-trained network. Thus, these samples are included in the training set (i.e.,  $\omega_i = 1$ ). By gradually increasing the  $\lambda$  value, more and more complex samples can be selected since it is reasonable to assume, under the condition that the two dataset are comparable, that the classifier becomes more and more reliable at later learning iterations. If the condition is not satisfied, the network may diverge.



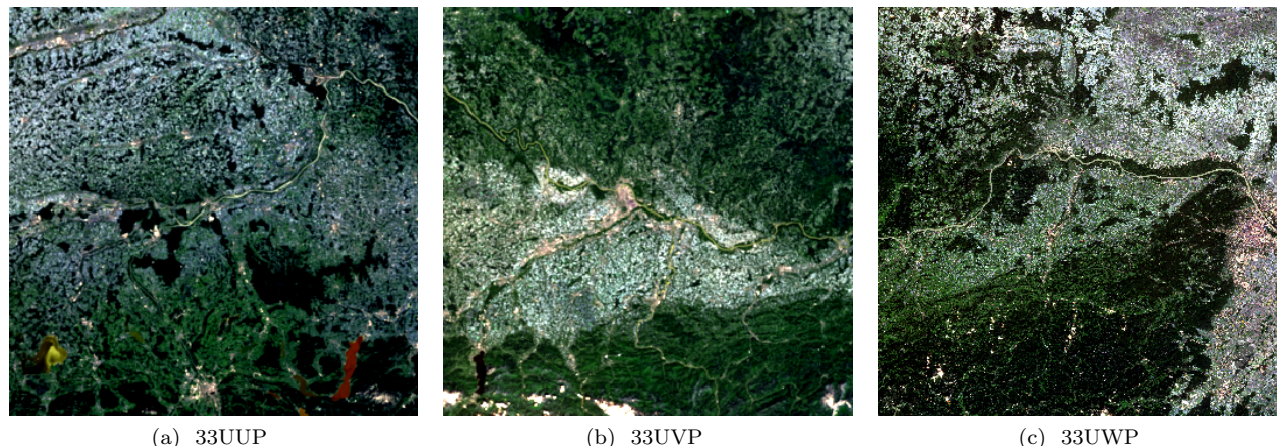


Figure 2. True color representation of the Sentinel-2 images (tiles “33UUP”, “33UVP”, and “33UWP”), acquired respectively on 27<sup>th</sup> August 2018, on 06<sup>th</sup> May 2018, and on 8<sup>th</sup> September 2017 over the study area.

### 3.4 Fine-Tuning Strategy

In the last step of the considered method, we adopt a FT strategy to accurately model the crop types that are classified with low confidence by the pre-trained network. Usually, the number of training samples required to fine tune the network must be larger or comparable to the number of parameters that have to be updated.<sup>18</sup> The fine-tuning of a pre-trained network is carried out in two steps: (i) training the fully connected layer at the end of the network with all the other layers frozen, and (ii) training the whole unfrozen network with a low learning rate. The subdivision in two steps allows the last class-specific layers to quickly reach convergence, while the first more generic layers are trained later in order to optimize computational time and avoid diverging from the original network.<sup>8</sup>

Differently from the standard approaches, due to the application of the SPL step, the number of samples required to perform an accurate classification is expected to be lower than the one usually needed. For this reason, the network obtained after the SPL is fully unfrozen and trained in a supervised manner using a low number of samples from each considered class.

## 4. DATASET DESCRIPTION

Fig. 2 shows the considered study area, which is located in Austria and is characterized by a spatial extent of 360 km<sup>2</sup>. The remote sensing data employed are TSs of Sentinel 2 images acquired over three neighboring tiles, namely tiles “33UUP”, “33UVP”, and “33UWP”, mainly covered by agricultural areas. The TS of Sentinel 2 images employed to train the network has been acquired from September 2017 to August 2018, in order to correctly represent the agronomic year (i.e., the period from one year’s harvest to the next). For this reason, the recent TSs have been acquired from September 2018 to August 2019 and September 2019 to August 2020. Such TSs allow us to accurately map the evolution of the crops present in the scene. The network has been trained on a dataset consisting of more than one million labeled samples covering the whole Austrian territory, i.e., 15 Sentinel 2 tiles covering a spatial extent of 36000 km<sup>2</sup>. The classification scheme of the pre-trained network consists of 15 classes, namely: “legumes”, “grassland”, “maize”, “potato”, “sunflower”, “soy”, “winter barley”, “winter caraway”, “rye”, “rapeseed”, “beet”, “spring cereals”, “winter wheat”, “winter triticale”, and “permanent plantations”. Both the labeled data used to train the pre-trained network and the ones used to perform the FT were extracted from the Austrian crop type maps of 2018, 2019 and 2020, respectively, which are publicly available.<sup>19</sup> These thematic products are generated in the context of the Common Agricultural Policy (CAP) of the European Union relying on the farmer declarations. The polygon field boundaries are provided by the Land Parcel Identification System (LIPS).

Sentinel 2 images having cloud cover > 80% and partial acquisitions have been discarded. The data were downloaded from the Food Security Thematic Exploitation Platform (TEP),<sup>20</sup> where atmospherically corrected

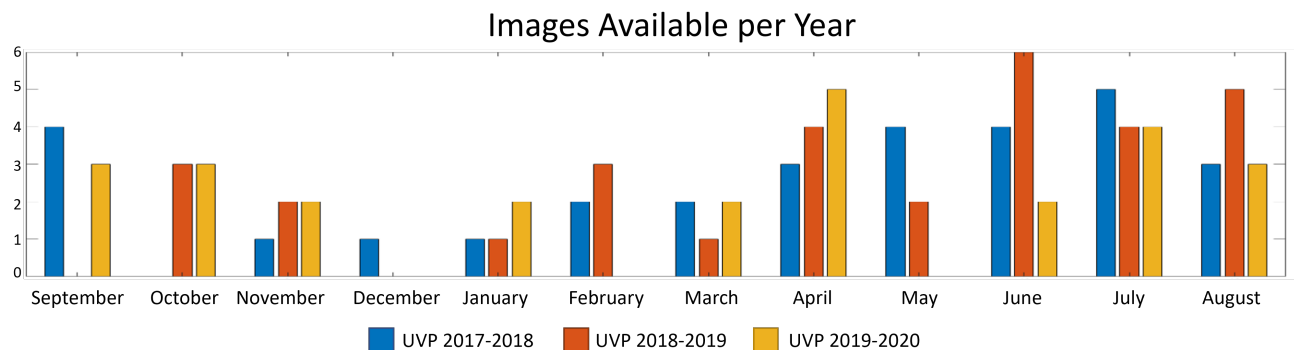


Figure 3. Distributions of number of images available per month in each year considered of the “33UVP” tile.

Sentinel 2 images are provided at 10 m spatial resolution having nine spectral bands, i.e., blue (B2 - 490 nm), green (B3 - 560 nm), red (B4 - 665 nm) the four vegetation red edge (B5 - 705nm, B6 - 740 nm, B7 - 0.783 nm and B8A - 865 nm) and the two short wave infrared (SWIR) (B11 - 1610 nm and B12 - 2190 nm) channels. Fig. 3 shows the number of images per month available for the different TSs of images considering the Sentinel 2 tile “33UVP”. Although this tile is located on overlapping orbits, thus having a much denser TS compared with the neighboring tiles, due to the heavy cloud coverage several images have to be discarded. As expected, this leads to TSs characterized by different temporal sampling. However, due to the harmonization step we are able to generate homogeneous TS, allowing the processing of all the data.
















## 5. DESIGN OF THE EXPERIMENTS

In the considered experimental setup, we assume available 100 samples per crop type, for a total number of 1500 samples used to perform the FT of the pre-trained network. The samples are randomly extracted from the three Sentinel 2 tiles under analysis. Regarding the SPL strategy, the  $\lambda$  parameter is tuned in order to have at each iteration a fixed number of samples to be included in the training set per class. In the first iteration, a very small number of high-confidence samples per crop type are included in the training set if and only if their posterior probability is higher than 0.9. In the experiments we fixed that number to 8. This condition allows us to adapt the weights of the pre-trained network to the recent TS of images by using samples belonging to the different crop types having the highest probability to be correctly classified. Since, under proper assumptions, the model is becoming more robust at each iteration, the number of samples selected by the SPL at a given iteration per crop type is twice as much as the previous iteration.

To assess the effectiveness of the proposed approach, we compare the results obtained on 2019 and 2020 agronomic year with the one obtained on the training year (which is 2018), considering: (i) the pre-trained LSTM network without adaptation (i.e., 2018 LSTM), and (ii) the unsupervised SPL strategy without FT. This condition allows us to compare the results obtained with no labeled samples for the new year, exploring the adaptation capability of the LSTM on TS with different characteristics than those of the TS considered for the training. Then, we compare the SPL+FT technique with the FT using only 1500 labeled samples. In this case, we want to assess the capability of the LSTM to adapt to a new target year, using the SPL+FT and the standard FT, with a small number of training samples. Finally, we tested the (i) SPL+FT and the standard (ii) FT using 15000 training samples. This condition allows us to check the classification upper bound that can be obtained by training the network considering a large number of samples on the target agronomic year. It is worth nothing that the 1500 and 15000 labeled samples used for the FT cannot be used properly for training from scratch the LSTM given the large number of parameters to estimate.

The experimental results obtained have been obtained considering a pool of test samples extracted from the tile “33UVP” in both the agronomic years. The tile has been randomly split into non overlapping patches of 1098 x 1098 pixels, allowing us to consider samples spatially disjoint from the ones used for FT. Experiments have been carried for both the agronomic years, i.e., September 2018 to August 2019 and September 2019 to August 2020, using the network pre-trained with the TS acquired from September 2017 to August 2018.

Table 1. Comparison between the Overall Accuracy (OA%), and the F-score (F1%) obtained on (1) the pre-trained model on the year used to perform the training, and the TSs of 2019 Sentinel 2 images using the: (2) pre-trained network without adaptation, (3) SPL without FT, (4) SPL + FT with 1500 labeled samples extracted from 2019, (5) FT with 1500 labeled samples extracted from 2019, (6) SPL + FT with 15000 labeled samples extracted from 2019, and (7) FT with 15000 samples extracted from 2019

		Ref. Year	No labeled data 2019		1500 samples 2019		15000 samples 2019	
Classes		2018	Pre-Trained	SPL	SPL+FT	FT	SPL+FT	FT
	Legumes	74.33	64.34	66.36	75.41	72.42	78.99	76.75
	Grassland	88.57	85.07	86.04	87.79	86.17	87.63	86.56
	Maize	92.91	91.89	92.80	91.67	91.95	94.66	94.92
	Potato	88.32	65.93	66.64	72.67	72.82	77.08	78.02
	Sunflower	79.64	37.93	39.84	56.10	54.54	57.34	58.87
	Soy	78.07	79.59	80.47	84.51	84.16	83.05	84.22
	Winter Barley	89.64	73.17	81.70	82.90	83.51	88.22	87.83
	Winter Caraway	88.66	37.88	55.83	86.68	85.50	86.18	86.40
	Rye	57.66	47.89	46.09	51.46	53.13	56.37	56.07
	Rapeseed	95.89	88.23	89.15	92.83	91.56	93.43	93.39
	Beet	95.77	84.89	83.87	88.14	85.13	91.39	90.54
	Spring Barley	78.40	77.12	81.13	83.31	81.18	84.58	82.81
	Winter Wheat	87.59	70.62	56.10	76.94	75.69	78.39	79.01
	Winter Triticale	54.92	44.75	44.63	50.13	51.57	50.67	52.35
	Perm. Plantation	77.66	14.26	30.46	68.70	63.38	71.92	70.12
OA%		84.82	74.96	75.81	80.58	79.25	82.46	81.83
Mean F1%		81.87	64.24	66.74	76.62	75.51	78.66	78.52
















## 6. EXPERIMENTAL RESULTS

Tab. 1 shows the experimental results on the first agronomic year analyzed (i.e., September 2018 to August 2019) in terms of F-score (F1%) and Overall Accuracy (OA%). The table is divided in four columns, where the first column shows the results of the pre-trained architecture on the reference year (i.e. the year where it has been trained), while the others show the metrics obtained by different methods using the same number of labeled samples related to the considered agronomic year, i.e., 0 samples, 1500 samples and 15000 samples. From the results obtained one can notice that by applying the pre-trained network without adaptation, the values of F1% and OA% are lower with respect to the reference year, due to the shift of the land-cover classes distribution across years. Using the SPL method, we slightly increase the classification accuracy of the crop types in an unsupervised way. As expected, the LSTM is able to correctly map the phenological growth of the crops using 15000 samples from 2019 with both the SPL and FT and the standard FT approaches, that take advantage of the pre-trained network and adapt it to the target year. Thanks to the pre-processing step and the adaptation of the pre-trained network, the system is able to obtain similar accuracies by considering only 100 samples per class (i.e. 1500 labeled samples), with a mean F1% of 76.62% compared to an F1% of 78.66%. By considering only the the pre-processing step and the SPL, it is possible to perform unsupervised adaptation at the cost of a performance decrease. It is worth nothing that the considered classification task presents several challenges. The classes “sunflower”, “winter wheat”, “winter triticale”, and “rye” show lower value of accuracy when compared with the other crop types. This is due to the similarity between the “winter wheat”, “triticale” and “rye”, which are difficult to discriminate at a resolution of 10m. In contrast, the “sunflower” crop type is critical due to its small prior probability in the Austrian territory.

Similar results can be seen in Tab. 2, where the accuracy metrics increase as the number of samples increases. Note that the this multitemporal adaptation for the TS of September 2019 to August 2020 is more complicated than the previous one, since the TS of images considered has been acquired two years later than the one used to pre-train the network. As expected, the pre-trained network achieves the lowest accuracy, with an OA% of 70.76% and a mean F1% of 62.88%. Here, the SPL approach performs slightly better than the pre-trained network. In this case the LSTM with the SPL and FT approach using only 1500 labeled data is outperformed by the techniques considering 15000 samples, having a OA% of 75.11% and a mean F1% of 70.94%, compared to



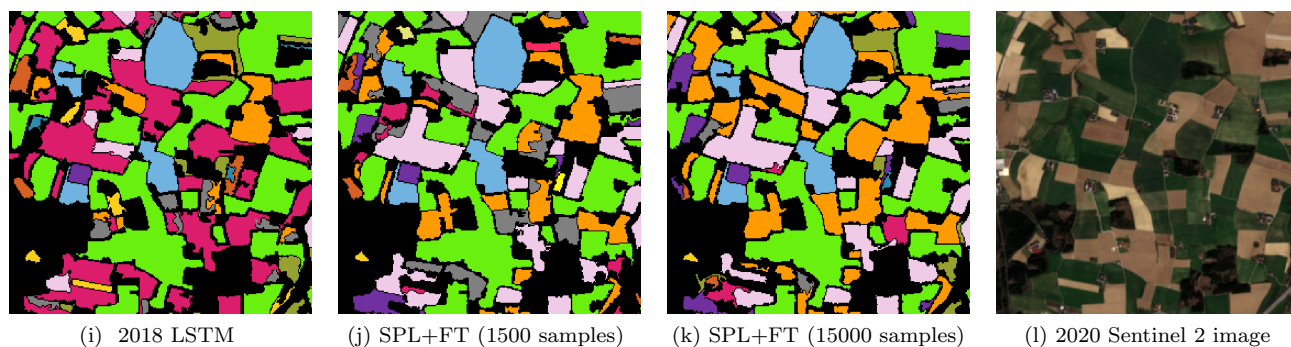
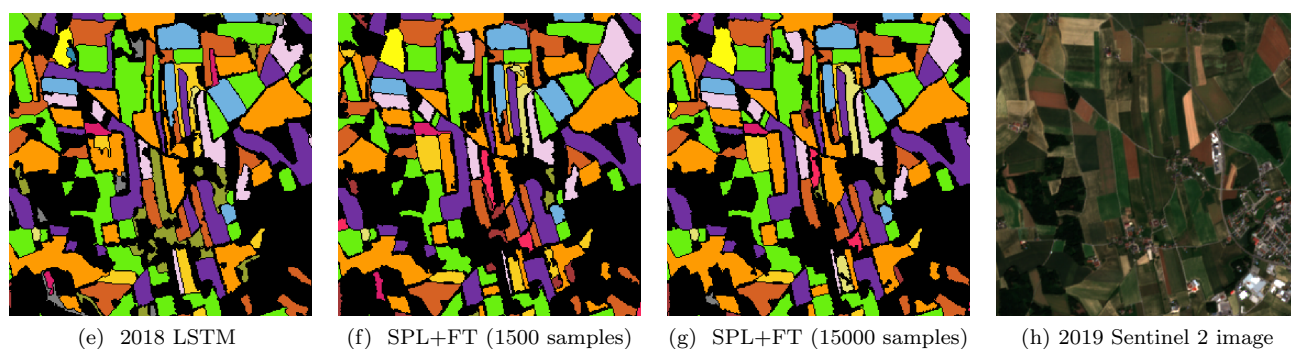
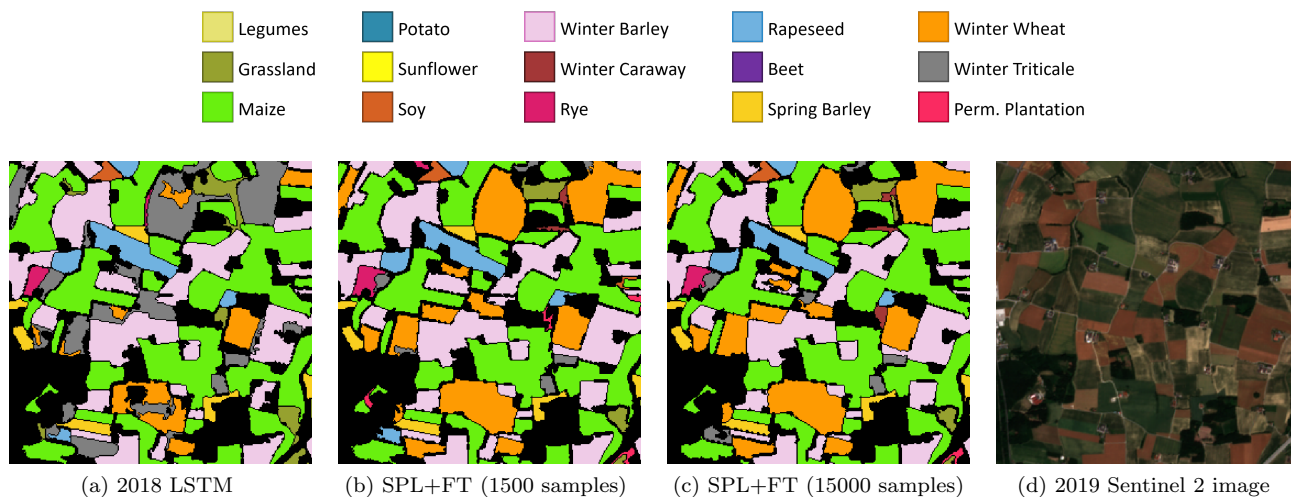
Table 2. Comparison between the Overall Accuracy (OA%), and the F-score (F1%) obtained on (1) the pre-trained model on the year used to perform the training, and the TSs of 2020 Sentinel 2 images using the: (2) pre-trained network without adaptation, (3) SPL without FT, (4) SPL + FT with 1500 labeled samples extracted from 2020, (5) FT with 1500 labeled samples extracted from 2020, (6) SPL + FT with 15000 labeled samples extracted from 2020, and (7) FT with 15000 labeled samples extracted from 2020

		Ref. Year	No labeled data 2020		1500 samples 2020		15000 samples 2020	
Classes		2018	Pre-Trained	SPL	SPL+FT	FT	SPL+FT	FT
	Legumes	74.33	55.83	51.62	52.92	55.95	57.23	60.71
	Grassland	88.57	84.09	85.62	85.84	82.86	85.44	82.79
	Maize	92.91	87.57	87.52	84.92	86.39	89.46	88.45
	Potato	88.32	76.13	78.32	81.02	78.68	84.98	86.03
	Sunflower	79.64	33.73	58.56	62.39	67.24	72.16	74.19
	Soy	78.07	58.03	61.89	67.80	70.87	75.65	73.82
	Winter Barley	89.64	68.93	68.22	80.43	77.95	83.90	86.95
	Winter Caraway	88.66	59.96	80.69	79.99	85.44	84.88	83.23
	Rye	57.66	42.12	42.78	36.43	34.55	41.18	44.82
	Rapeseed	95.89	96.12	95.32	93.97	94.29	95.58	94.56
	Beet	95.77	88.30	90.58	92.42	92.20	92.83	92.72
	Spring Barley	78.40	55.76	60.35	66.97	68.30	74.08	73.58
	Winter Wheat	87.59	51.61	62.55	61.52	64.15	75.24	74.31
	Winter Triticale	54.92	41.17	40.32	46.66	45.36	53.44	52.33
	Perm. Plantation	77.66	43.79	63.85	70.75	66.07	75.29	71.29
OA%		84.82	70.76	73.30	75.11	74.32	78.65	78.44
Mean F1%		81.87	62.88	68.55	70.94	71.35	76.09	75.99

the OA% of 78.65% and a mean F1% of 76.09% of the method using more samples. Differently from the previous year, the performance of the LSTM using different training sample size varies greatly, due to the differences between the two agronomic years considered. Indeed, due to the increase in the problem complexity, the 2020 target year shows a decrease of the overall accuracies compared to the 2019. Similarly to the previous agronomic year, the critical classes are “sunflower”, “winter wheat”, “winter wheat”, and “rye”.

An example of qualitative results are shown in Fig. 4, where the crop type maps obtained with the different methods are reported. In particular, we compare the results obtained with the: (i) pre-trained model without adaptation, (ii) proposed SPL+ FT strategy using 1500 recent samples and (iii) proposed SPL+ FT using 15000 recent samples. The crop type map shown have been post-processed with basic morphological operators, attenuating the pixel-level noise typically present on the crop’s boundaries. The results obtained from the quantitative analysis are confirmed by the qualitative viewpoint. In particular, one can see that the network without adaptation provides less accurate maps than those with adaptation. The SPL method improves the pre-trained network if there are no available labeled samples and slightly improves the accuracy combined with the FT given a fixed number of labeled samples.

2019



2020

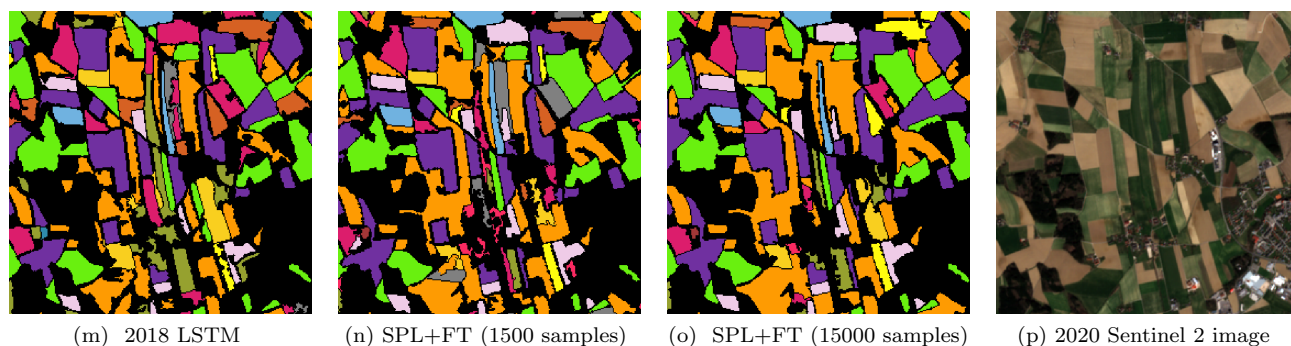


Figure 4. Crop Type maps obtained for 2019 and 2020 in small portions of the same considered test area considering the: (a),(e),(i),(m) pre-trained model without adaptation; (b),(f),(j),(n) proposed SPL+ FT strategy using 1500 recent samples; (c),(g),(k),(o) SPL+FT using 15000 recent samples. The last column shows the true color representation of one Sentinel 2 image acquired on the 30<sup>th</sup> June 2019 (d),(h) and the true color representation of one Sentinel 2 image acquired on the 5<sup>th</sup> April 2020 (l),(p).

## 7. CONCLUSIONS

In this paper we presented a novel system for multi-year crop type mapping based on a pre-trained LSTM network adapted by combining SPL and FT strategies. The system: (1) harmonizes the multi-year optical data to create homogeneous TSs from the temporal and spectral viewpoint, (2) applies an unsupervised SPL strategy to adapt the pre-trained network to the recent TS of images in an unsupervised way, and (3) performs a supervised FT considering a relatively small number of training samples. The experimental results obtained on two agronomic years demonstrate the adaptation capability of the LSTM architecture integrated in the proposed system. In the proposed system, the pre-processing step allows one to perform unsupervised adaptation by making the classification possible with no labeled samples at the cost of a performance decrease. If few new labeled samples are available, the LSTM increases its accuracy on the target year. This condition allows us to sharply reduce the amount of labeled samples required to FT the network. Indeed, the results obtained considering the availability of 1500 recent labeled samples are comparable to the results achieved considering a training set ten times larger. As expected, the accuracy obtained by the system after the adaptation is correlated to the difference between the target year and the year used to pre-train the architecture. As the shift increases, the performance of the pre-trained network decreases and more samples are required to obtain higher accuracy metrics.

As future developments, we aim to improve the SPL technique presented, by exploring a specific strategy able to handling imbalanced classification tasks. Indeed the minor classes, which are classified by the pre-trained network with more uncertainty, may not be accurately modeled by the standard SPL strategy. Finally, we aim to study the possibility of using a cascade classification approach to consecutively adapt the pre-trained network to successive target years, acquired more than 2 years after the one used for the pre-trained network.

## 8. ACKNOWLEDGMENTS

This work has been developed in the framework of the ExtremeEarth project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825258.

## REFERENCES

1. Solano-Correa, Y. T., Bovolo, F., Bruzzone, L., and Fernández-Prieto, D., "A method for the analysis of small crop fields in sentinel-2 dense time series," *IEEE Transactions on Geoscience and Remote Sensing* 58(3), 2150–2164 (2020).
2. Zhong, L., Hu, L., and Zhou, H., "Deep learning based multi-temporal crop classification," *Remote sensing of environment* 221, 430–443 (2019).
3. Rußwurm, M. and 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*], 11–19 (2017).
4. Pelletier, C., Webb, G. I., and Petitjean, F., "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing* 11(5), 523 (2019).
5. Rußwurm, M., Lefèvre, S., and Körner, M., "Breizhcrops: A satellite time series dataset for crop type identification," in [*Proceedings of the International Conference on Machine Learning Time Series Workshop*], 3 (2019).
6. Sonobe, R., Tani, H., Wang, X., Kobayashi, N., and Shimamura, H., "Parameter tuning in the support vector machine and random forest and their performances in cross-and same-year crop classification using terrasar-x," *International Journal of Remote Sensing* 35(23), 7898–7909 (2014).
7. Liu, X., Chi, M., Zhang, Y., and Qin, Y., "Classifying high resolution remote sensing images by fine-tuned vgg deep networks," in [*IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*], 7137–7140, IEEE (2018).
8. Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?," *arXiv preprint arXiv:1411.1792* (2014).
9. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J., "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging* 35(5), 1299–1312 (2016).

10. Chi, M. and Bruzzone, L., “A semilabeled-sample-driven bagging technique for ill-posed classification problems,” *IEEE Geoscience and Remote Sensing Letters* 2(1), 69–73 (2005).
11. Meng, D., Zhao, Q., and Jiang, L., “A theoretical understanding of self-paced learning,” *Information Sciences* 414, 319–328 (2017).
12. Chen, W., Gou, S., Wang, X., Li, X., and Jiao, L., “Classification of polsar images using multilayer autoencoders and a self-paced learning approach,” *Remote Sensing* 10(1), 110 (2018).
13. Li, H., Gong, M., Zhang, M., and Wu, Y., “Spatially self-paced convolutional networks for change detection in heterogeneous images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 4966–4979 (2021).
14. Flood, N., “Seasonal composite landsat tm/etm+ images using the medoid (a multi-dimensional median),” *Remote Sensing* 5(12), 6481–6500 (2013).
15. Roberts, D., Mueller, N., and McIntyre, A., “High-dimensional pixel composites from earth observation time series,” *IEEE Transactions on Geoscience and Remote Sensing* 55(11), 6254–6264 (2017).
16. Weikmann, G., Paris, C., and Bruzzone, L., “Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 4699–4708 (2021).
17. Tieleman, T. and Hinton, G., “Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude.” COURSERA: Neural Networks for Machine Learning (2012).
18. Mohanty, S. P., Hughes, D. P., and Salathé, M., “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science* 7, 1419 (2016).
19. <https://www.data.gv.at/?s=invekos>. Accessed: 2021-08-18.
20. Muerth, M., Migdall, S., Hodrius, M., Niggemann, F., Holzapfel, M., Bach, H., Gilliams, S., Van Roey, T., Cuomo, A., Harwood, P., et al., “Food security tep-supporting sustainable intensification of food production from space,” in [*IOP Conference Series: Earth and Environmental Science*], 509(1), 012038, IOP Publishing (2020).