Monitoring of Agricultural Areas by using Sentinel 2 Image Time Series and Deep Learning Techniques

Claudia Paris, Giulio Weikmann, and Lorenzo Bruzzone

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

1. ABSTRACT

This paper presents a system that aims to exploit Sentinel 2 image Time Series (TSs) and deep learning techniques to perform agricultural monitoring in large areas. Although deep learning proved its effectiveness to classify optical multispectral data, this peculiar classification task shows many challenges: (1) the considered TSs are noisy due to the presence of clouds that corrupts the multitemporal spectral signature, thus affecting the classification results, (2) TSs of different tiles are made up of images acquired in different dates (different temporal sampling), and (3) a large training database of labeled samples is needed to train the deep model from scratch. To address these problems, this paper presents an approach based on three main steps: (1) a preprocessing step that aims to generate temporally homogeneous TSs of images across tiles that accurately represent the phenological behavior of the crops, (2) a training set extraction step that automatically establishes a large training database leveraging publicly available thematic products, and (3) a multitemporal deep learning classification where a Long Short Term Memory (LSTM) network is used for crop type mapping. Experimental results show that the proposed system architecture is promising for solving large-scale problems. The considered deep architecture achieved the most balanced classification results compared to existing state-of-the-art method, obtaining a mean Fscore (F1%) of 78.32% and an Overall Accuracy (OA%) of 85.86.

Keywords: Long-Short Term Memory (LSTM), Sentinel 2 Time Series (TSs), Crop Mapping, Deep Learning, Crop Type Mapping, Food Security, Remote Sensing.

2. INTRODUCTION

To generate accurate crop classification maps, it is necessary to exploit multitemporal approaches since cultivations change their spectral and textural appearance according to their crop type growth cycle.[1] In this context, the availability of dense Time Series (TSs) of multispectral images acquired at high spatial resolution such as Sentinel 2 data (10 m spatial resolution) represents an extremely useful information source. Deep learning models designed for handling sequential data have been recently demonstrated their effectiveness in generating accurate land cover maps from satellite image TSs to perform accurate crop type mapping due to their capability of extracting high-level abstract features extremely useful for complex image classification tasks. Through deep learning a better classification is achieved compared to the use of customized feature extraction approaches designed for specific tasks. However, most of the models in the literature focus the attention on the spatial information provided by the Remote Sensing (RS) data, neglecting the temporal information that is fundamental for the crop type classification task. Two main categories of multitemporal deep learning architectures have been tested on TSs of RS images for crop type mapping: (i) recurrence based deep learning models, and (ii) time convolutional based deep learning models.

Recurrence based deep learning models are suited for the crop classification task,[2] since they have been designed for handling sequential data as the network exploit the previous observations for the classification of the current one.[3–6] In particular, a special type of Recurrent Neural Networks (RNN), the Long Short Term Memory (LSTM), has been widely used for crop type mapping due to its long-term memory capabilities.[3,5] LSTM can store a huge amount of evidence to make decisions in that actual temporal context, giving a better solution compared to other recurrent deep learning models. Comparisons among the classification results obtained by using LSTM, RNN, the mono-temporal Convolutional Neural Network (CNN) and shallow Support Vector Machine (SVM) on a TSs of Sentinel 2 images for crop type mapping are presented in the literature.[3] LSTM outperforms the other techniques demonstrating its capability to accurately model temporal dynamics, even

though it requires a high computational burden for training. Recently, Turkoglu *et al* present the STAR recurrent neural network (StarRNN),[6] a network made up of STAckable Recurrent cells that requires less parameters compared to the LSTM cells. By stacking several STAR cells, the network avoids vanishing gradient issue when using deep architectures. Rußwurm *et al*[4] present a bidirectional sequential encoder network. The input sequence of observations is encoded to a representation which is then projected to softmax-normalized feature maps for each crop type using a convolutional layer. A cloudy TS of Sentinel 2 images is given as input to the deep architecture to verify the capability of the network to learn jointly the cloud masking and crop classification tasks. The results show that even though the network is able to handle the cloud coverage, some recurrent cells are sensitive to presence of cloudy pixels, thus affecting the classification result.

Time convolutional based deep learning models have also been tested to perform crop type mapping.[7] Such temporal 1D-CNN models extract features from the temporal profiles of the TSs by applying the convolutions in the temporal domain.[8,9] Different architectures have been proposed in the literature to classify sequential data,[10] by stacking the convolution layers with different combinations of non-linear activation, pooling, and normalization layers. Pelletier *et al*[9] propose a network for land cover classification made up of 3 convolutional layers having same filter size, a dense and softmax layers. Similarly, in the Omniscale Convolutional Neural Network (OmniscaleCNN)[10] the structure of the network is made up of 3 convolutional layers having different lengths, a global average pooling and a dense layer. Zhong *et al*[8] present the classification of TSs of Landsat images for distinguish summer crops using a Conv1D-based model. The model consists of a set of 1D-CNN layers and an inception module. The authors demonstrate that while lower convolutional layers allow to capture small scale temporal variations, upper layers model the whole seasonal trend.

In[7] the authors present a benchmark dataset for crop type mapping for testing and comparing both recurrent and convolutional deep learning models. The experimental analysis demonstrates that recurrent methods in general obtained better results compared to time convolution approaches. However, all the approaches obtained relatively low average accuracy, due the complex classification task. The main reason is that, even though deep learning is effective, this peculiar classification task shows many challenges: (1) the model has to deal with noisy TSs due to the presence of clouds at irregular intervals, (2) the large reflectance of clouds introduces positive outliers in the multitemporal spectral signature of the crops, thus affecting the classification results, and (3) the TSs acquired in different tiles vary per length and area of acquisition (different temporal sampling). Another issue is that deep architectures for agricultural classification have to be trained from scratch, since pretrained architectures mainly rely on mono-temporal approaches. To successfully train the network, a large training set representative of the considered study area is required.

In this paper, we propose a system for monitoring agricultural areas by using TSs of Sentinel 2 images and a deep learning technique. To address the above-mentioned problems, the proposed method: (1) performs a preprocessing step which creates homogeneous TS of multispectral data across tiles and reduces the cloud coverage effect, (2) automatically extracts a reliable large training set representative of the considered study area leveraging publicly available crop type maps, and (3) trains a multitemporal deep architecture from scratch using the large training set. The preprocessing step allows us to effectively handle the cloud coverage problem while generating regular TSs of images at large scale. The large database of labeled samples can be extracted by using crop type maps available at country scale. These maps represent an extremely useful information source to generate large training databases to be used for training a deep architecture from scratch. We consider an LSTM multitemporal deep architecture, which is robust to the presence of cloudy pixels in the TS for small extents. Moreover, this type of network is able to better handle the imbalanced classification task compared to state-of-the-art methods.

3. PROPOSED APPROACH TO CROP MAPPING

Fig. 3.1 shows the block scheme of the proposed approach, which exploits the information provided by the long TSs of Sentinel 2 data and a large database of labeled samples extracted from existing thematic products for crop type mapping. The proposed approach is made up of three main steps: (1) the preprocessing for the harmonization of the TS of Sentinel 2 images, (2) the automatic and unsupervised extraction of a reference dataset representative of the considered study area, and (3) the production of crop type maps by using a multitemporal LSTM deep learning model trained from scratch.
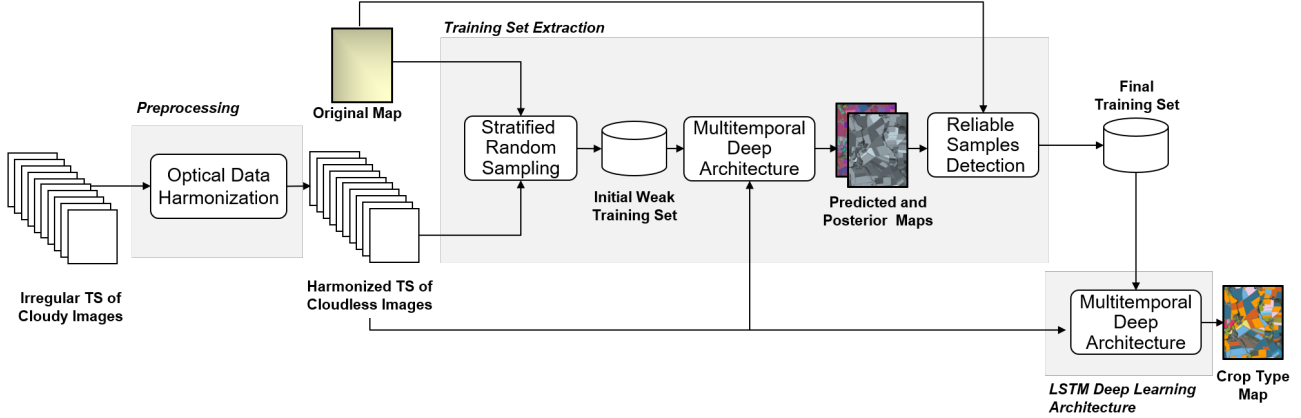
Figure 1. Architecture of the proposed method for monitoring agricultural areas by using TSs of Sentinel 2 images and deep learning techniques.

3.1 Preprocessing

The aim of this step is the harmonization of the irregular TSs of cloudy images and the reduction of the cloud coverage effects. One of the challenges of crop type mapping at large scale is the use of TSs of images acquired over different tiles that have different lengths and are acquired at different times.[7] When training a deep learning model on local scale, accurate results can be achieved by using the whole TS of images available in the area. When scaling to a large area, TSs of images characterized by different sequence lengths are usually available on different tiles. This is mainly due to the irregular cloud coverage (which hampers the use of some images of the TS) and the different orbit acquisitions (different temporal sampling). This strongly affects the classification results obtained by the deep learning models. To solve this problem, in the proposed method we generate monthly composites. The idea is that we can represent the agronomic year (i.e., period from one year's harvest to the next one for agricultural commodity) by using 12 months in order to have enough information to represent the phenological trend of the crops. To this end, we consider a pixel composite approach that takes the collection of optical images acquired within each month and collapses them down to a single image. This is done by a statistic-based approach that computes the median value for each pixel, thus providing an harmonized TS from the temporal and radiometric view point across tiles. Hence, the pixel composite approach to mosaic generation provides consistent results at large scale in an automatic way by sharply reducing the spatial noise.[11] Fig. 3.2 shows a qualitative example of monthly composite obtained for September 2017 on an area located in the northern part of Austria, where 6 Sentinel 2 images are available. One can notice that this step also strongly mitigates the presence of clouds in the scene, thus reducing the probability of having high reflectance values in the phenological trends recorded by the optical sensor.

3.2 Training Set Extraction

To successfully train a deep learning model from scratch a large training set representative of the considered study area is required. However, labeling samples is a labor-extensive and time-consuming task. Moreover, it is not feasible to assume to collect in-situ measurements at such scale. To solve this problem, the aim of this step is to exploit publicly available crop type maps to extract a reliable training set representative of the considered study areas in an automatic and unsupervised way. Due to the *Common Agricultural Policy* of the European Union, several crop type maps based on farmers declarations are now available at country scale. Although such maps are an interesting source of information, the samples should be properly selected because the map may contain: (i) outdated information, (ii) the crop types present in the annual thematic products may not be the real ones for the whole year due to the rotation practice, and (iii) the map is made up of polygons which do not strictly match with to purely spectral pixels correctly associated to their crop types. Given the complexity of the considered problem, an automatic training set extraction step based on the considered recurrent deep learning model has been defined. The step aims to detect the labeled samples having the highest probability of being correctly associated with their labels in order to generate a large training database of weak labeled samples in an unsupervised way at large scale.
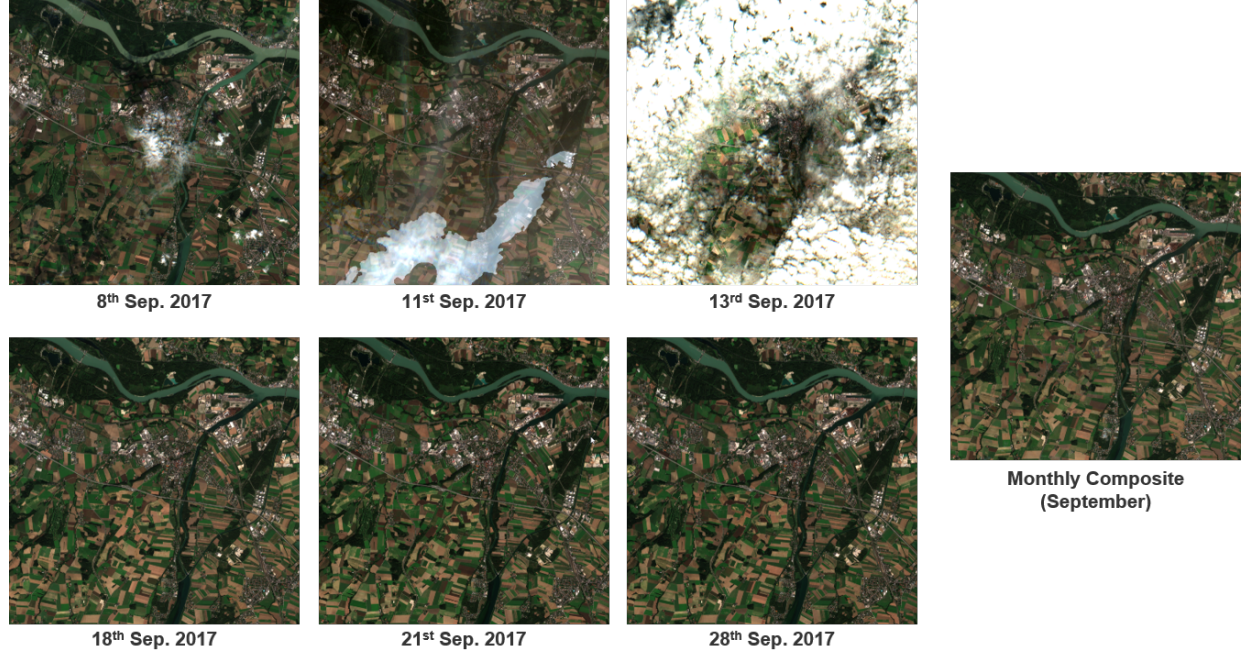
Figure 2. Qualitative examples of monthly composite result for September 2017 on an area located in the northern part of Austria. The composite is obtained by using all the images acquired in the month.

First, the prior probabilities of the crop types present in the scene are derived from the considered thematic product. This condition allows us to generate a statistically balanced training set by considering a stratified random sampling approach.[12] This sampling strategy allows the selection of a number samples per crop type proportional to the number of crops present in the considered study area. The erosion morphological operation is first performed to remove the boundaries of the crops to facilitate the selection of samples having almost pure spectral signature, i.e., representative of the phenological properties of the crop type. This condition allows us to generate an initial "weak" training set where misclassified samples may be present (i.e., samples with wrong labels associated). Then, we run the classifier to generate a preliminary land cover map by using the considered LSTM deep learning model trained on the harmonized TS of Sentinel 2 images. This map is compared with the original one to select only the samples located in the areas where the two maps agree. Moreover, we can use the pixel wise posterior probabilities provided by the network, as a confidence measure of the classification result for each samples. Hence, to increase the probability of detecting reliable samples we can tune the confidence level of the classifier used for the samples selection.

3.3 LSTM Deep Learning Architecture

In the last step of the proposed method, we train the deep learning model from scratch using the large training dataset. In the proposed system architecture, we consider a recurrent deep learning architecture to accurately model the phenological characteristics of the different crop types. In particular, we consider an LSTM network since it is robust to the presence of a small amount of cloudy pixels in the TSs. The LSTM model takes as input the TSs of harmonized Sentinel 2 images representing the agricultural year of the considered study area and the large training database.

We consider a multi-layer LSTM which provides the classification result at pixel level. Hence, comparing with single-layer LSTM, the multi-layer LSTM has stronger ability to model the information provided by TSs data, even though it requires a high training computational time due to the complex structure of the LSTM memory cells. The model used is an LSTM made up of three layers, a fully connected layer and a softmax layer. Since the classification results benefit from an initialization of each layer with a different number of cells, we set the first, second and third layer of the proposed system architecture with 200, 125 and 100 hidden units, respectively. Fig. 3.3 shows the considered deep learning model.
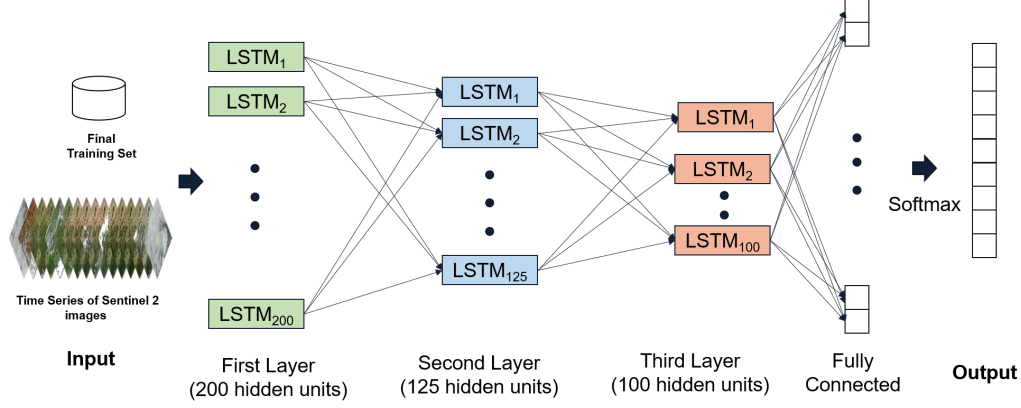
Figure 3. The proposed multi-layer LSTM network is made up of 3 layers, a fully connected layer and a softmax layer which provides the classification result at pixel level.

4. DATASET DESCRIPTION AND EXPERIMENTAL RESULTS

In this section we present: (i) the dataset employed for the experimental analysis, (ii) the experimental setup used to assess the effectiveness of the proposed system architecture, and (iii) the evaluation of the obtained crop type maps from the quantitative and qualitative view point.

4.1 Dataset Description

The considered study areas is located in Austria and presents a spatial extent of $\sim$36000 km^2. To extract the training set we consider the publicly available 2018 Austrian crop type map[13] based on farmer declarations (for the crop types) and the Geographic Information System (GIS) (for the field geometry). To extract reliable labeled samples from the considered thematic product, the crop type map is first carefully revised in order to: (i) discard ambiguous classes, (ii) discard semantically aggregated classes, and (iii) select a set of crop types that can be discriminated by using the spectral, spatial and temporal resolution of TSs of Sentinel 2 images. The considered classification scheme is made up of 12 classes, namely: "grassland," "maize," "legumes," "winter cereals," "rapeseed," "potato," "beet," "spring cereals," "soy," "sunflower," "permanent plantations" and "flowering legumes".

The RS data employed are TSs of Sentinel 2 images of an agronomic year acquired between September 2017 and August 2018. These images allow us to accurately characterize the phenological trends of the crop types present in the scene. Fig. 4.1 shows the considered TSs of Sentinel 2 images acquired on three neighboring tiles, namely tiles T33UUP, T33UVP and T33UWP. Images having cloud cover $> 80\%$ were discarded. Only the spectral channels acquired at 10 and 20 m spatial resolution of the Sentinel 2 images are considered. As expected the TSs are characterized by very different number of images due to the cloud coverage problem. Moreover, the TS acquired on tile T33UVP is much denser than the ones acquired on tiles T33UUP and T33UWP since the former is located on overlapping orbits. However, due the optical harmonization process, we are in the condition of generating homogeneous TSs across tiles, and thus, of extracting the training samples from all the tiles to train the deep learning model on the whole study area.

4.2 Experimental Setup

Experiments have been carried out on tile T33UVP having a TS of 70 Sentinel 2 images in the considered agronomic year. To have a robust statistical analysis, the tile is randomly split into non overlapping mini tiles of 1098×1098 pixels ($\sim$ 120 km^2) in order to have spatially disjoint training, test and validation samples. Three kinds of experiments have been carried out.

In the first experiment, we tested the impact of the preprocessing step on the classification performances by comparing the results obtained training the LSTM on both the whole TS of 70 images and the 12 monthly composite. To perform this experiment, we extract the training set from tile T33UVP (1st Trset). This condition
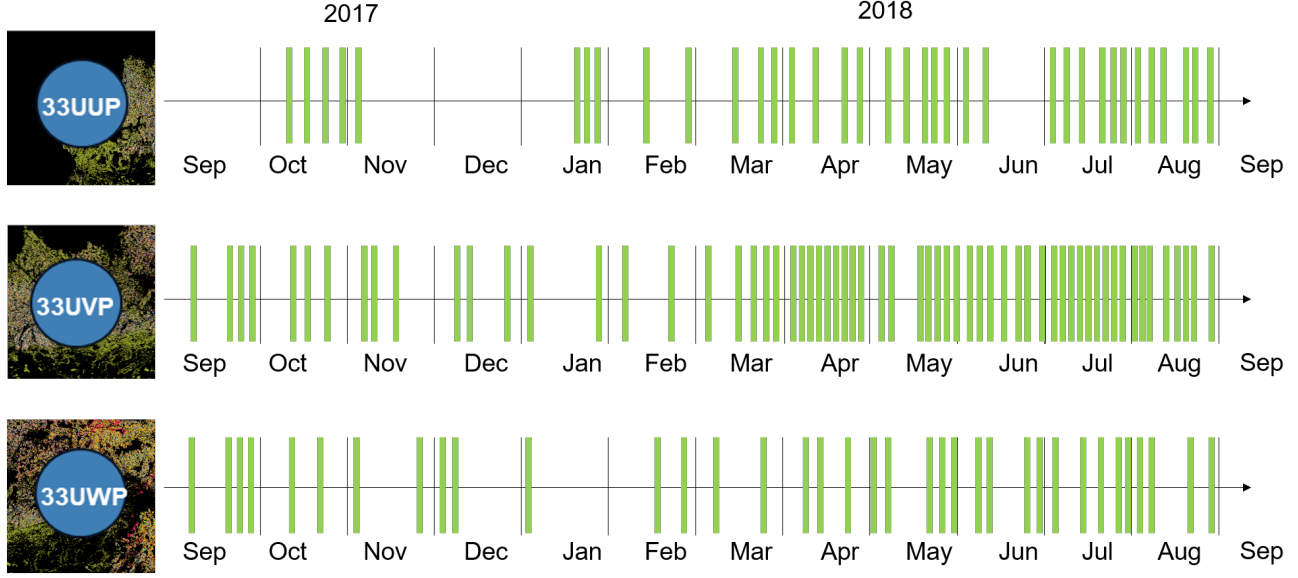
Figure 4. Images in the TSs having cloud cover $< 80\%$ on the three neighboring Sentinel 2 tiles used in the experiments. The acquisition dates of the TS acquired on tile T33UVP are much denser than the ones taken on T33UUP and T33UWP since the former is located on overlapping orbits.

allows us to check the classification upper bound that can be obtained with the LSTM when considering the whole TS of images. Note that, when working at single tile level no harmonization process is required to homogenize the multitemporal spectral signatures of the training samples with the ones of the test sample (i.e., samples to be classified). In the original TS of Sentinel 2 images, the clouds identified by the automatic atmospheric correction tool[14] are set to zero to regulate the input gates of the LSTM in order to dismiss the cloud information.[15]

In the second experiment, we assessed the capability of the large training to be representative of the whole study area by comparing the results obtained by training the deep learning model with: (1) samples extracted from the T33UVP (1st TrSet), (2) samples extracted from the T33UUP and T33UWP (2nd TrSet), and (3) samples extracted from the whole study area (3rd TrSet). Finally, the results obtained with the LSTM are compared with other recent state-of-the-art methods . The same harmonized TSs of cloudless Sentinel 2 images and the same training set (3rd TrSet) are used to feed all the deep architectures to have a fair comparison.

The hyper-parameters of the LSTM were selected through a grid search approach by testing all combinations of the number of network layers L $\in \{2, 3, 4\}$ and the number of cells per layer $\in \{100, 125, 200, 225, 300\}$. At each training step, the cross-entropy loss is calculated by comparing the predicted and the ground truth class probabilities. To adjust the model weights, the loss is then back propagated through the network layers as gradients, which in turn are utilized by RMSProp optimizer.[16] The initial learning rate has been set equal to 10^{-3} and the weight decay to 0.4.

4.3 Experimental Results

Tab. 1 shows the experimental results obtained with the 1st TrSet (i.e., samples from the T33UVP) by using the whole TS of images and the 12 monthly composites. Moreover, it presents the results obtained by using the 2nd TrSet (i.e., samples from the T33UUP and T33UWP) and the 3rd TrSet (samples from the whole study area) by using the harmonized TSs of composites. As expected, the best results obtained with the 1st TrSet are the ones achieved with the whole TS of images. However, both the OA% and F1% score metrics obtained with the 12 composites are still comparable with the classification upperbound obtained by exploiting all available multitemporal information of the whole TS. The harmonization process allows us to obtained TSs comparable from the radiometric and temporal view point for the whole study area. The best results are achieved by using the training set having samples collected in three considered Sentinel 2 tiles (3rd TrSet), thus generating a larger and more representative set with respect to the one that can be extracted using only tile T33UVP. Moreover,

the results obtained on the UVP by training the network with samples extracted from the UUP and UWP demonstrate that the training set can be used at large scale and applied to classify regions where no training samples are available.

Tab. 2 reports the comparison with some recent state-of-the-art methods used for crop type mapping. In particular, the LSTM has been compared to the MSResNet,[17] the InceptionTime;[18] and the OmniScale[10] multitemporal deep networks. The prior probability of each crop type is reported. The considered classification problem is highly imbalanced having minority classes such as Soy of Flowering Legumes, which have prior probabilities of 1%. The considered deep learning model provides overall accuracy in line with the baseline approaches by having a mean F1% score of 78% and an OA% of almost 86%. However, it achieved the most balanced classification results, with a minimum F1% score much higher than the baselines approaches, i.e., 46 % for the flowering legumes class compared to 33.60%, 37.59%, 25.94% of MSResNet, InceptionTime and OmniScale, respectively.

Table 1. Crop type classification results obtained on the tile T33UVP. The Overall Accuracy (OA%) and the Fscore (F1%) are reported for the LSTM trained on: (1) the 12 monthly composites using the 1st TrSet extracted from the UVP tiles, (2) the whole TS using the 1st TrSet extracted from the UVP tiles, (3) the 12 monthly composites using the 2nd TrSet extracted from the UUP and UWP tiles, and (4) the 12 monthly composites using the 3rd TrSet extracted from the whole study area. For each crop type the prior probability is reported.

Crop Types	Prior (%)	1st TrSet 12 Composites F1%	1st TrSet Whole TS F1%	2nd TrSet 12 Composites F1%	3rd TrSet 12 Composites F1%
grassland	26.42	88.69	88.10	84.99	91.98
maize	14.13	91.30	94.13	91.16	93.62
legumes	1.41	62.43	68.27	57.68	69.53
winter cereals	9.03	86.00	87.74	84.86	87.95
rapeseed	4.66	55.90	58.68	40.35	57.67
potato	7.03	85.06	87.57	83.60	90.65
beet	6.03	94.25	95.85	88.49	96.68
spring cereals	7.43	70.07	71.71	66.62	74.86
soy	3.24	64.87	77.73	65.66	69.84
sunflower	1.41	53.18	68.53	55.78	78.26
p. plantation	16.39	74.64	71.44	59.51	82.48
f. legumes	1.40	14.52	26.44	41.14	46.37
OA%		81.53	83.47	77.62	85.87

Fig. 5, Fig. 6 and Fig. 7 show qualitative examples of crop type maps obtained in the considered study area. The results obtained from the quantitative view point are confirmed by the qualitative analysis that points out the capability of the proposed system architecture to generate accurate classification results. It is worth noting that the proposed system provides a classification map at pixel level. However, the obtained maps demonstrate the capability of the approach of properly detecting the fields associated with different crop types present in the scene due to a proper representation of the phenological trends of the cultivations present in the scene.

5. CONCLUSIONS

In this paper we presented an approach to monitor agricultural areas by using TSs of Sentinel 2 images and a deep learning technique. The approach: (1) harmonizes the optical data to create homogeneous TS across tiles while reducing the cloud coverage problem, (2) automatically extracts a large training database representative of the considered study area using publicly available crop type maps, (3) trains an LSTM model from scratch with the large training database, and (4) properly exploits the temporal and spectral properties of the TS of Sentinel 2 data to characterized the phenological trends of the crop types. The proposed method is promising for large-scale crop type mapping, generating consistent classification products on very large study areas. Although
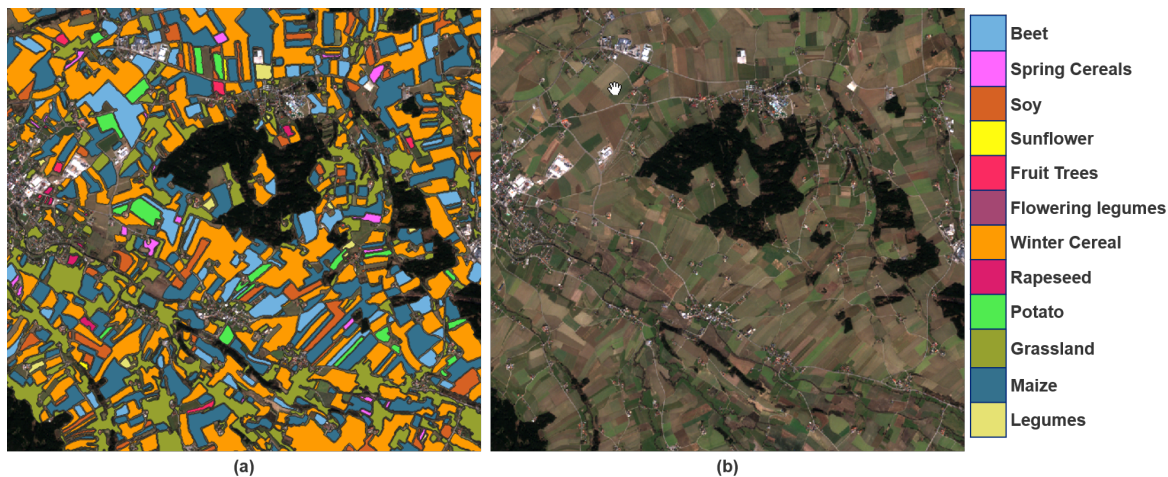
Figure 5. Qualitative example of the crop type map obtained with the proposed system: (a) predicted map, and (b) Sentinel 2 image acquired on the 25th March 2018.

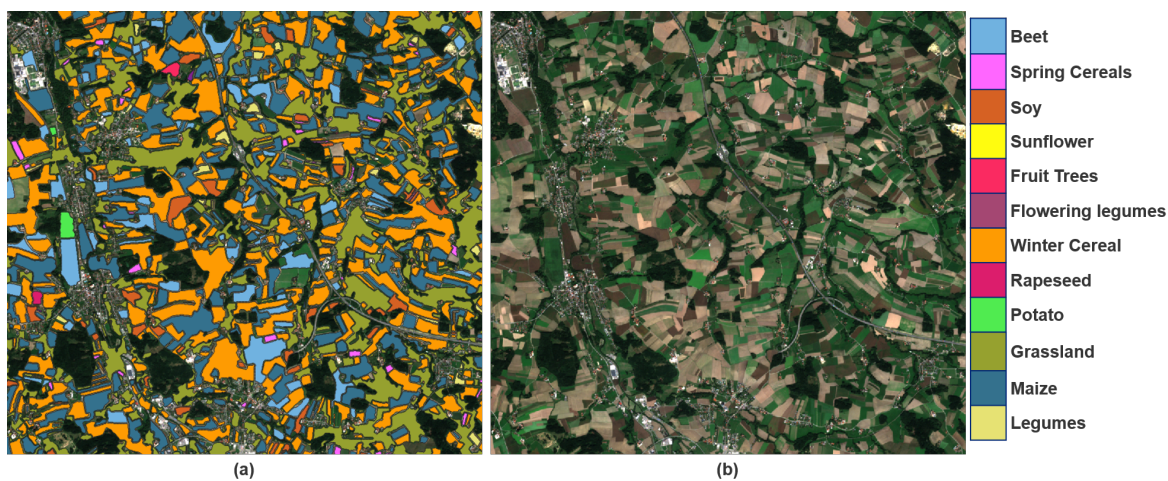Figure 6. Qualitative example of the crop type map obtained with the proposed system: (a) predicted map, and (b) Sentinel 2 image acquired on the 6th August 2018.
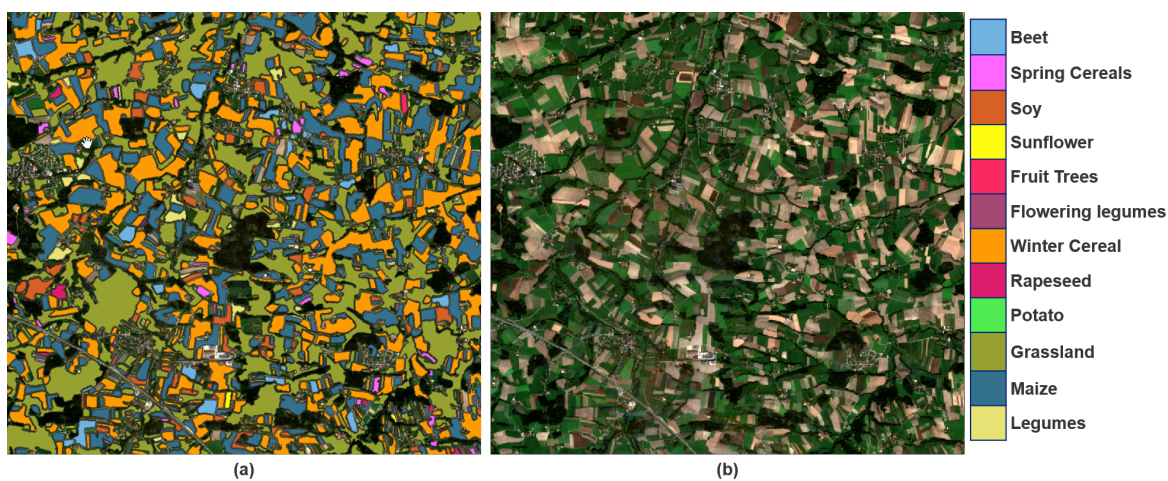

Figure 7. Qualitative example of the crop type map obtained with the proposed system: (a) predicted map, and (b) Sentinel 2 image acquired on the 26th August 2018.

Table 2. Crop type classification results obtained on tile T33UVP. The Overall Accuracy (OA%) and Fscore (F1%) are reported for: 1) the MSResNet;[17] 2) the InceptionTime;[18] 3) the OmniScale;[10] and 4) the considered LSTM. All deep architectures are trained by using the harmonized Sentinel 2 TSs of cloudless images and the 3rd TrSet extracted from the whole study area. The prior probability of each crop type is reported.

	Crop Types	Prior	Baselines			Considered Model
			MSResNet	InceptionTime	OmniScale	LSTM
			F1%	F1%	F1%	F1%
	grassland	26.42	91.50	90.57	89.13	91.98
	maize	14.13	94.33	93.31	95.23	93.62
	legumes	1.41	72.96	72.15	66.29	69.53
	winter cereals	9.03	86.92	88.77	87.43	87.95
	rapeseed	4.66	59.19	64.78	60.97	57.67
	potato	7.03	92.29	91.90	92.80	90.65
	beet	6.03	96.28	96.46	95.93	96.68
	spring cereals	7.43	74.08	75.90	76.11	74.86
	soy	3.24	85.76	73.08	79.00	69.84
	sunflower	1.41	78.63	80.08	80.80	78.26
	p. plantation	16.39	82.39	80.03	77.35	82.48
	f. legumes	1.4	33.60	37.59	25.94	46.37
	OA%		86.32	85.98	85.60	85.87

LSTM proved its capability to mitigate the highly imbalanced classification problem compared to other state-of-the-art methods, we aim to further investigate possible solutions to handle this issue which severely affects real-world crop-type-mapping datasets. Moreover, we aim to test the proposed system architecture on country scale study areas by using the extracted large training set.

ACKNOWLEDGMENTS

REFERENCES

1. Solano-Correa, Y. T., Bovolo, F., Bruzzone, L., and Fernández-Prieto, D., "Spatio-temporal evolution of crop fields in sentinel-2 satellite image time series," in [*2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*], 1–4, IEEE (2017).
2. Mou, L., Bruzzone, L., and Zhu, X. X., "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **57**(2), 924–935 (2018).
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. Rußwurm, M. and Körner, M., "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS International Journal of Geo-Information* **7**(4), 129 (2018).
5. Rußwurm, M. and Körner, M., "Multi-temporal land cover classification with long short-term memory neural networks," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**, 551 (2017).
6. Turkoglu, M. O., D'Aronco, S., Wegner, J. D., and Schindler, K., "Gating revisited: Deep multi-layer rnns that can be trained," (2019).
7. Rußwurm, M., Lefèvre, S., and Körner, M., "Breizhcrops: A satellite time series dataset for crop type identification," *CoRR* **abs/1905.11893** (2019).
8. Zhong, L., Hu, L., and Zhou, H., "Deep learning based multi-temporal crop classification," *Remote sensing of environment* **221**, 430–443 (2019).
9. 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).
10. Tang, W., Long, G., Liu, L., Zhou, T., Jiang, J., and Blumenstein, M., "Rethinking 1d-cnn for time series classification: A stronger baseline," *arXiv preprint arXiv:2002.10061* (2020).
11. 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).

12. Paris, C. and Bruzzone, L., "A novel approach to the unsupervised extraction of reliable training samples from thematic products," *IEEE Transactions on Geoscience and Remote Sensing* , 1–19 (2020).

13. https://www.data.gv.at/katalog/dataset/e21a731f-9e08-4dd3-b9e5-cd460438a5d9 (2019). Accessed: 2020-09-10.

14. Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., and Gascon, F., "Sentinel-2 sen2cor: L2a processor for users," in [*Proceedings Living Planet Symposium 2016*], 1–8, Spacebooks Online (2016).

15. Sharma, A., Liu, X., and Yang, X., "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Networks* **105**, 346–355 (2018).

16. Zou, F., Shen, L., Jie, Z., Zhang, W., and Liu, W., "A sufficient condition for convergences of adam and rmsprop," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 11127–11135 (2019).

17. Wang, F., Han, J., Zhang, S., He, X., and Huang, D., "Csi-net: Unified body characterization and action recognition," *arXiv preprint arXiv:1810.03064* (2018).

18. Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., and Petitjean, F., "Inceptiontime: Finding alexnet for time series classification," *arXiv preprint arXiv:1909.04939* (2019).