UNSUPERVISED DOMAIN ADAPTATION METHODS FOR LAND COVER MAPPING WITH OPTICAL SATELLITE IMAGE TIME SERIES

E. Capliez[1,3] *D. Ienco*[1] *R. Gaetano*[2] *N. Baghdadi*[1] *A. Hadj Salah*[3]

[1] INRAE, UMR TETIS, Univ. Montpellier, France
[2] CIRAD, UMR TETIS, Univ. Montpellier, France
[3] Airbus Defence and Space, Toulouse, France

ABSTRACT

Nowadays, Satellite Image Time Series (SITS) are employed as input to derive land cover maps (LCM) to support decision makers in several application domains like agriculture and biodiversity. The generation of LCM largely relies on available ground truth (GT) data to calibrate supervised machine learning models. Unfortunately, this data are not always accessible. In this scenario, the possibility to transfer a model learnt on a particular year (*source domain*) to another period of time (*target domain*) could be a valuable tool to deal with the previously mentioned restrictions.

In this paper, we provide an experimental evaluation of recent Unsupervised Domain Adaptation (UDA) methods in the specific context of temporal transfer learning for SITS-based LCM. The objective is to learn a classification model at a certain year (exploiting available GT data) and, successively, transfer such a model on a subsequent year where no labelled samples are accessible. The obtained findings reveal that UDA methods represent a promising research direction to cope with the problem of temporal transfer learning for LCM. While a model learnt on the source data and directly applied on target data achieves an weighted F1-score of 67.1, the best UDA method obtains an F1-score of 83.7 with more than 15 points of positive gap. Nevertheless, there is still room for improvement that should be explored in future works.

***Index Terms*—** Domain Adaptation, Satellite Image Time Series, Land Cover Map, Deep Learning.

1. INTRODUCTION

Nowadays, satellite imagery represents a fundamental source of information to monitor the dynamic of the Earth surface providing valuable knowledge to support decision makers in several application domains [**?**]. Recent spatial programmes (i.e. the European Union's Sentinel program) provide images acquisition with high revisit time period. This produces Satellite Image Time Series (SITS) data that can be leveraged to describe how natural and semi-natural areas evolve over time [**?**].

To this end, SITS data is regularly used as input to modern supervised machine learning approaches with the aim to provide up to date Land Cover Maps (LCM) over a specific region [**?**]. SITS data, conversely to mono-date imagery, contains information about the evolution of the signal depicting the Earth surface allowing, for instance, to distinguish land covers that evolve differently over time (i.e. soybeans vs. corn crops due to the fact that they exhibit a different dynamic in their radiometric signal over time).

Unfortunately, supervised machine learning methods require large amount of reference (or Ground Truth (GT)) data to be calibrated, hence posing serious challenges to their use in situations characterized by a reduced amount of or unavailable reference data. For instance, when LCM should be updated from an year to a successive one, costs related to human-effort and resources deployment prevent the possibility to regularly conduct field campaigns to collect new GT data [**?**].

Directly transfer a model learnt on a particular year to another period of time can be ineffective since the two time periods can be affected by different environmental, weather or climate conditions. This results in differences in the distributions of acquired remote sensing data (per year). In the general field of machine learning, the UDA framework has the objective to provide methods and strategies to cope with data distribution shifts between the data on which the model is calibrated (*source domain*) and the data on which the model is deployed (*target domain*) [**?**]. For instance, in the previous example, the goal is to have a model capable to deal with temporal transfer learning where the year on which the machine learning system is calibrated, where both remote sensing and GT data are available, is referred as *source domain* while the successive year on which the machine learning model should be deployed, where only remote sensing data is available, is referred as *target domain*. While a significant amount of research work exists on UDA for the general field of computer vision and pattern recognition [**?**], only few research studies were devoted to leverage such a learning setting in the context of time series of remote sensing data [**?**].

In this paper, with the aim to cope with the scenario of temporal transfer learning for SITS-based land cover mapping, we have revised, adapted and evaluated recent state-of-the-art UDA methods. To the best of authors knowledge, this is the first work that compares recent UDA methods for temporal transfer learning for SITS land cover mapping. As a benchmark for the evaluation study, we have considered two Sentinel-2 SITS data describing the same study area over two different years. The study site is located in the south of France and characterized by a set of classes spanning from natural to semi-natural land covers.

2. RELATED WORK

The following section presents an overview of the state of the art methods in the field of unsupervised domain adaptation (UDA). In the UDA setting, we dispose of data from a source domain $\mathcal{D}_s$, with associated label information and unlabelled data from a target domain $\mathcal{D}_t$. The main assumptions are the follows: i) there is a shift in the data distribution between $\mathcal{D}_s$ and $\mathcal{D}_t$ and ii) $\mathcal{D}_s$ and $\mathcal{D}_t$ share the same label space (homogeneous domain adaptation). The goal is to build a classification model capable to exploit, simultaneously, the information associated to the source domain (data and labels) and the available information associated to the target domain (data) to perform prediction on the unlabelled data belonging to $\mathcal{D}_t$. For a general overview of unsupervised domain adaptation approaches please refer to [**?**].

2.1. State-of-the-art unsupervised DA methods

*Geodesic Flow Kernel (GFK) [**?**]:* This approach aligns the source and target data distributions by means of a geodesic flow kernel-based strategy. The method allows to project both source and target data into a shared, low-dimensional, space in which the distribution shift between the two domains is reduced. Successively, any standard supervised classification method can be trained on the source data and tested on the target ones.

*Domain-Adversarial Neural Networks (DANN) [**?**]:* This approach enrich a standard neural network-based supervised classification strategy with a domain classifier that may distinguish between source and target examples. The domain classifier is associated with a gradient reverse layer (GRL) that enforces the features extracted by the encoder to be invariant w.r.t. the distribution shift that can be present between $\mathcal{D}_s$ and $\mathcal{D}_t$.

*Adversarial Discriminative Domain Adaptation (ADDA) [**?**]:* Inspired by the concept of generative adversaial network (GAN), this approach set up a two players learning strategy where a discriminator network tries to distinguish between source and target data representation provided by the generator while the generator tries to fool the discriminator network. Also in this case, the objective is to extract data representations that are invariant w.r.t. possible distribution shifts that can occur between data coming from $\mathcal{D}_s$ and $\mathcal{D}_t$.

2.2. DA of Satellite Image Time Series

Although there is a significant amount of work on UDA, only a limited number of studies are devoted to cope with time series data [**?**] and, even less for satellite image time series [**?**]. For the latter, the work proposed in [**?**] clearly underlines that state-of-the-art UDA methods cannot directly deal with spatial transfer learning while no research study, to the best of our literature survey, has evaluated the quality of recent UDA methods in the context of temporal transfer learning for land cover mapping based on SITS data.

3. DATA

The study area is centered on the city of Balaruc-les-Bains in south-eastern France. It borders the Mediterranean Sea and it has a surface of 100 km^2. It is a small but dense urban area surrounded by agricultural crops - mostly vineyards - and natural vegetation. Figure **??** presents the GT map superposed on the Sentinel-2 image of the 23th January 2018.

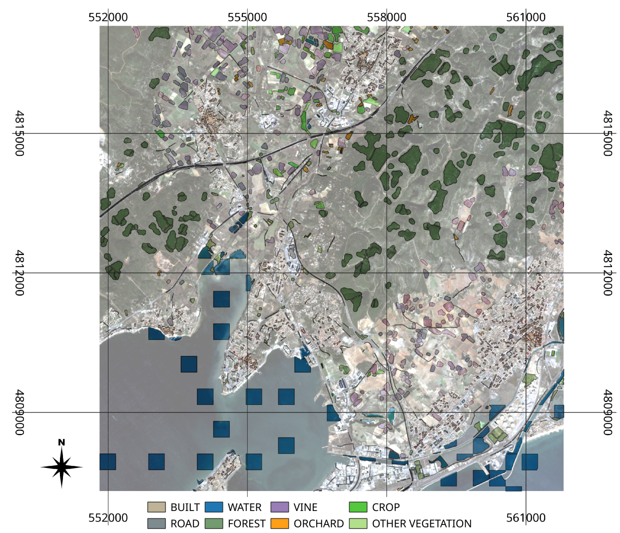

Fig. 1. Ground truth data superposed to a Sentinel-2 image.

Satellite Image Time Series: We collect satellite image time series of Sentinel-2 imagery spanning the years 2018 and 2019. For each year, we retain 24 images to guarantee a similar temporal sampling step. Images were also chosen with the aim to be representative of the temporal (annual) evolution of the land covers associated to the study area as well as to filter out images that were seriously impacted by cloud phenomena. Figure **??** depicts the acquisition dates of the two Sentinel-2 satellite image time series.

All images were provided by the THEIA pole platform [1] at level-2A in top of canopy reflectance values with associ-

[1]

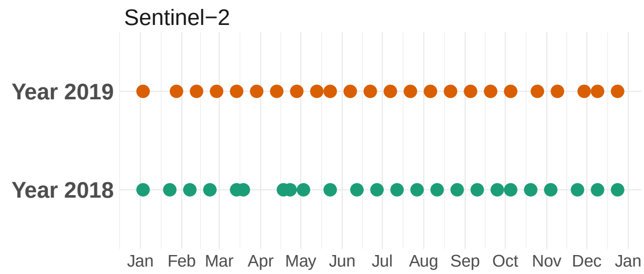

Fig. 2. Acquisition dates of each Satellite Image Time Series.

ated cloud masks. Only 10-m spatial resolution bands (Blue, Green, Red and Near infrared spectrum) were considered in this analysis. A preprocessing was performed over each band to replace cloudy pixel values as detected by the available cloud masks through a linear multi-temporal interpolation (cf. temporal gap-filling [**?**]). The entire study site is enclosed in the Sentinel-2 tile 31TEJ with relative orbit n°8.

Ground truth data: The GT data for 2018 was built from various sources: the *Registre Parcellaire Graphique* (RPG) reference data (the French land parcel identification system), the French National Geographic Institute *'BD-Topo & BD Forêt'* and the visual interpretation of a SPOT 6 image (to assess and enrich the GT data) as well. The GT was assembled in Geographic Information System (GIS) vector file, containing a collection of polygons, each attributed with a land cover category. Statistics about the GT data are reported in Table **??**. It was assumed that this GT data is also valid for 2019 because no significant change has occurred in this area between 2018 and 2019.

Class Name	Class ID.	# Polygons	# Pixels
BUILT	1	712	10411
ROAD	2	328	5165
WATER	3	82	32095
FOREST	4	172	49175
VINE	5	223	28630
ORCHARD	6	32	2075
CROP	7	39	5205
OTHER VEGETATION	8	56	4812
	Total	1644	137568

Table 1. Ground truth statistics.

4. EXPERIMENTS

In this section we firstly describe the settings associated to our experimental evaluation and, successively, we report the obtained results and the associated discussion considering the study area introduced in Section **??** and the competing approaches introduced in Section **??**.

4.1. Experimental settings

With the aim to assess the performances of recent UDA strategies in the context of temporal transfer learning for land cover mapping from SITS data, we have considered the state-of-the-art methods described in Section **??**.

About the evaluated methods, we couple the GFK approach with two different classification methods: a Random Forest (RF) and a Multi-layer perceptron (MLP). The latter has two fully connected layers of 512 neurons each (with ReLU activation function) and an output layer (with softmax) to perform the final classification. Both DANN and ADDA are equipped with a backbone especially tailored to manage satellite image time series data. To this end, we adopt the TempCNN model recently proposed in [**?**]. In addition, we also consider two baseline scenarios in which: i) a supervised classification model is trained with only source data and directly deployed on target data and (only $\mathcal{D}_s$) and ii) a supervised classification model is trained on labelled target data and deployed on the rest of the target examples (only $\mathcal{D}_t$). The former constitutes a straightforward baseline that does not take into account the necessity to deal with data distribution shift while, the latter represents the maximum performances we can (theoretically) achieve if we have label knowledge associated to $\mathcal{D}_t$. For both scenarios we consider as supervised classification methods the TempCNN and RF models.

For neural network approaches, the training stage has been conducted for 300 epochs, with a learning rate of 10^{-4} and a batch size of 64. When no domain adaptation is involved, the data is splitted into training, validation and test set considering 50%, 20% and 30% of the original data, respectively. The data was split to avoid possible spatial biases [**?**]. Weighted F1-score is adopted to evaluate the behaviors of the different competing approaches.

Experiments are carried out on a workstation with a dual Intel (R) Xeon (R) CPU E5-2667v4 with 256 GB of RAM and four TITAN X (Pascal) GPU. All the deep learning methods are implemented using the Python TensorFlow library and run on a single GPU.

4.2. Results and discussion

The study was conducted considering two possible configurations: ($\mathcal{D}_s$=2018,$\mathcal{D}_t$=2019) and ($\mathcal{D}_s$=2019,$\mathcal{D}_t$=2018). Nevertheless, only the evaluation for the former configuration is reported and discussed since similar results were obtained considering both configurations.

Table **??** reports the results, in terms of weighted F1-score, achieved by the different competing methods. For scenario (only $\mathcal{D}_s$), we can observe that deploying models learnt on SITS data from 2018 ($\mathcal{D}_s$) to SITS data coming from 2019 ($\mathcal{D}_t$) perform poorly. Generally, we can note more than 15 points of differences between the models trained w.r.t. this scenario and the models trained on $\mathcal{D}_t$ (scenario only $\mathcal{D}_t$). This fact clearly underlines that a serious shift in the data distribution occurs between 2018 and 2019.

Regarding transfer learning with the UDA methods (scenario w/ UDA), we can see that the majority of the strategies

Scenario	Method	F1-score	Training Time
2*Only $\mathcal{D}_s$	TempCNN	67.1	27 mins.
	RF	67.1	2 mins.
2*w/ UDA	GFK + MLP	65.8	146 mins.
	GFK + RF	76.7	14 mins.
	ADDA	80.6	650 mins.
	DANN	**83.7**	337 mins.
2*Only $\mathcal{D}_t$	TempCNN	<u>88.2</u>	26 mins.
	RF	85.7	2 mins.

Table 2. Weighted F1-score and training time of the competing approaches. The best F1-score score, for UDA methods, is highlighted in bold; the best score for supervised classification method calibrated on $\mathcal{D}_t$ is underlined.

ameliorate the performances on the target domain. The best results are achieved by the DANN strategy that, exploiting its ability to extract domain invariant features, reduces the gap performances up to less than 5 points w.r.t. the best model (TempCNN) trained directly on the target domain.

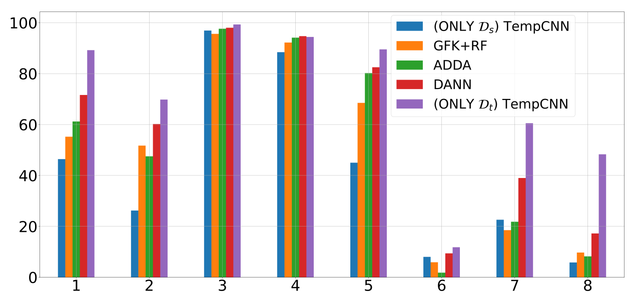

Fig. 3. Per class F1-score, ($\mathcal{D}_s$,$\mathcal{D}_t$) = (2018,2019).

Figure **??** depicts the per class results, in terms of weighted F1-score and computing time, of the different competing approaches. Firstly, we can note that some classes (*3-Water*, *4-Forest* and *6-Orchard*) seem to exhibit unchanged radiometric distributions between 2018 and 2019 while, the rest of the classes pose some challenges when the model is transferred (i.e. *1-Built*, *7-Crop* and *8-Other Vegegation*). Secondly, we can observe that DANN systematically achieves better performances than all the other competing approaches on the more challenging land cover classes. Finally, we can also advance the hypothesis that the best transfer (at class level) seems to be related to the amount of labelled examples we can access to in the source domain. For instance, UDA methods provide a notable improvement on class *5-Vine* that is the third more represented land cover class in the source domain.

5. CONCLUSION

The conducted experiments underline that state-of-the-art UDA methods can be applied in the context of temporal transfer learning for land cover mapping from SITS data. Despite the encouraging results we have obtained, our findings are still preliminary and, therefore, room for improvement still exists.