



**HAL**  
open science

# Distilling Before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas Using Sentinel-1 Time Series

Hassan Bazzi, Dino Ienco, Nicolas Baghdadi, Mehrez Zribi, Valérie Demarez

► **To cite this version:**

Hassan Bazzi, Dino Ienco, Nicolas Baghdadi, Mehrez Zribi, Valérie Demarez. Distilling Before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas Using Sentinel-1 Time Series. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17 (11), pp.1909-1913. 10.1109/LGRS.2019.2960625 . hal-04610578

**HAL Id: hal-04610578**

**<https://hal.science/hal-04610578>**

Submitted on 26 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Distilling Before Refine: Spatio-Temporal Transfer Learning for Mapping Irrigated Areas Using Sentinel-1 Time Series

H. Bazzi<sup>1</sup>, D. Ienco<sup>1</sup>, N. Baghdadi, M. Zribi, and V. Demarez

**Abstract**—This letter proposes a deep learning model to deal with the spatial transfer challenge for the mapping of irrigated areas through the analysis of Sentinel-1 data. First, a convolutional neural network (CNN) model called “Teacher Model” is trained on a source geographical area characterized by a huge volume of samples. Then, this model is transferred from the source area to the target area characterized by a limited number of samples. The transfer learning framework is based on a *distill and refine* strategy, in which the teacher model is first distilled into a student model and, successively, refined by data samples coming from the target geographical area. The proposed strategy is compared with different approaches including a random forest (RF) classifier trained on the target data set and a CNN trained on the source data set and directly applied on the target area as well as several CNN classifiers trained on the target data set. The evaluation of the performed transfer strategy shows that the “distill and refine” framework obtains the best performance compared with other competing approaches. The obtained findings represent a first step toward the understanding of the spatial transferability of deep learning models in the Earth observation domain.

**Index Terms**—Deep learning, knowledge distillation, satellite image time series, Sentinel-1 (S1), transfer learning.

## I. INTRODUCTION

IRRIGATION, nowadays, plays a significant role in agricultural production in order to meet the global food requirement [2]. Due to this fact, a better management of irrigation policies is required to deal with the high demand of food with the increase in the global population [4]. To support such policies, accurate information on the irrigated area extent is essential to manage water resources or evaluate irrigation water requirements. Unfortunately, the extent and distribution of irrigated areas remain indefinite and the large-scale mapping of such a property remains a challenge for modern remote sensing analysis. Recent works have pointed out that the synthetic

aperture radar (SAR) signal seems to be more suitable to map different agricultural irrigated areas [3]. The main assumption is related to the fact that the radar signal is sensitive to the water content of the soil due to a positive correlation between the dielectric constant and the soil water content. Following this direction, time series of SAR information acquired via the recently Sentinel-1 (S1) SAR constellation provides an effective tool for large-scale area mapping and monitoring due to their high revisit period (six days revisit time). To perform such mapping, machine (and deep) learning-based techniques are becoming the standard tools, since they allow large-scale analysis and they provide acceptable results [8], [14]. One of the key questions about the adoption of machine learning-based solution is related to their ability, given a particular task, to be spatially transferred from a geographical area to another spatially uncorrelated one. Considering the task of mapping irrigated areas [1], the challenge is to build a predictive model from a rich set of labeled samples available on a particular area and, successively, adapt this model on another geographical area, with a limited set of labeled samples. This adaptation could be crucial to conceive large-scale monitoring systems in an operational scenario. With the aim to tackle this issue, we propose a deep learning [14] framework capable to adapt a predictive model trained from a source data set (characterized by a large volume of labeled samples) to deal with the irrigation mapping on a target data set (characterized by a limited number of labeled samples). To this end, we propose a *distill before refine* framework in which first a teacher model is learned on a source data set, then a smaller student model is distilled from the teacher model, and, finally, the student model is fine-tuned to deal with the classification task on the target data set. Here, the challenge behind our framework is that models with a huge number of parameters are difficult to fine-tune (refine) on classification tasks where a very limited number of labeled samples are available. The innovative component of our framework is associated with the use of knowledge distillation [5] to distill a new smaller model that can be easily fit with new limited labeled samples on the target study area. In the general context of fine-tuning approaches [11], [13], the model learned on the source data is directly fine-tuned on the target data without considering the issues related to scarce or limited information. To the best of our knowledge, we are not aware about any other research work that adopts a similar pipeline to deal with spatial transfer learning in the field of remote sensing. The proposed methodology is implemented on mapping irrigated

Manuscript received October 29, 2019; revised November 28, 2019; accepted December 15, 2019. This work was supported by the French National Research Agency under the Investments for the Future Program under Grant ANR-16-CONV-0004 (DigitAg). (Corresponding author: H. Bazzi.)

H. Bazzi and N. Baghdadi are with the UMR-TETIS Laboratory, INRAE, University of Montpellier, 34090 Montpellier, France (e-mail: hassan.bazzi@irstea.fr; nicolas.baghdadi@teledetection.fr).

D. Ienco is with the LIRMM laboratory, 34090 Montpellier, France (e-mail: dino.ienco@irstea.fr).

M. Zribi and V. Demarez are with CESBIO, Université Paul Sabatier, 31330 Toulouse, France (e-mail: mehrez.zribi@ird.fr; valerie.demarez@univ-tlse3.fr).

Digital Object Identifier 10.1109/LGRS.2019.2960625

areas in which the classification problem involves *Irrigated* versus *Non Irrigated* area detection. The experiments are conducted considering a source study area from the *Catalonia* region of north east Spain where a time-consuming and cost-intensive field campaign was conducted collecting around 193 000 labeled samples and a target study area, from the *West Occitanie* region of South France, where the reference data are constituted of less than 500 examples.

## II. DATA

In this section, we introduce the geographical areas involved in this article: the source data set over the *Catalonia* region and the target data set over the *West Occitanie* (South of France) region. We also introduce the S1 SAR satellite images involved in this article and detail the different preprocessing steps performed to obtain the S1 time series backscattering coefficients at plot scale.

We point out that, in terms of climate, both studied zones (*Catalonia* and *West Occitanie*) are extremely different. In fact, the climate of the *Catalonia* region is typically Mediterranean, where the average annual precipitation is around 376 mm. On the other hand, the climate in *West Occitanie* is humid to oceanic with an average annual precipitation of 1200 mm. In both regions, irrigation mainly occurs in the summer season between May and October of each year. However, the summer season in *Catalonia* is very dry with rare rainfall events, whereas the summer season in *West Occitanie* is more humid with an average precipitation of 300 mm between May and October.

### A. Catalonia Data Set

Over the Catalonia region of north east Spain [Fig. 1(a)], the Geographic Information System for Agricultural Parcels (SIGPAC) data are provided by the General Direction of Rural Development of the Generalitat of Catalonia. The SIGPAC data represent the agricultural plots digitized using the orthophotograph images of 25-cm spatial resolution at scale 1:2000. Each plot in the provided data set is identified by a unique identification code, surface area, land use, and an irrigation coefficient (0 for nonirrigated and 100 for irrigated). Each year, an update of the database is provided based on an annual large field campaign in order to maintain the credibility of the data set (mainly irrigation information and crop type). In our study, a total of 193 000 plots covering an area of 3795 km<sup>2</sup> of different crop types and irrigation management have been used to develop the later called the teacher model. Among different land cover types, only agricultural crops (summer and winter crops) were considered for the irrigation classification. Forests, urban, and orchards plots were eliminated. In general, winter cereals such as wheat, oat, and barely are rarely irrigated with some exceptions. On the other hand, irrigated plots mainly include alfalfa, maize, grassland, beans, rapeseed, and rice. Among the total number of plots, 126 000 are nonirrigated, whereas 67 000 plots are irrigated.

### B. West Occitanie Data Set

Over the western part of the Occitanie region of South France [Fig. 1(b)], a terrain campaign was conducted over

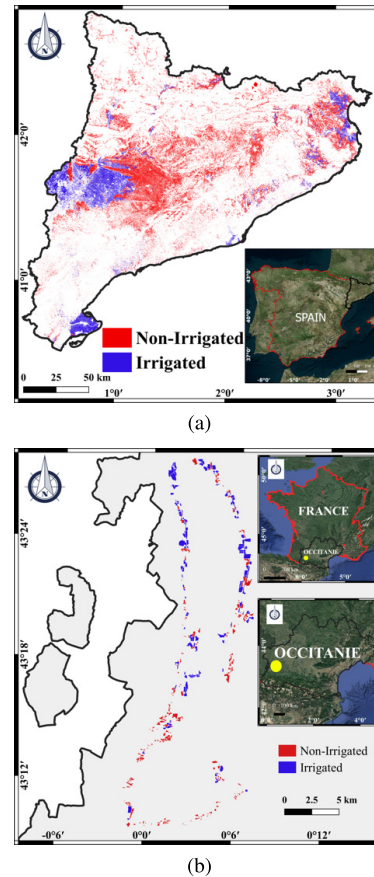


Fig. 1. Location of the source and the target study sites. (a) Catalonia region of North East Spain. (b) Western Occitanie region of South France.

different summer crop plots in 2017. In this field campaign, the existence or absence of irrigation activity over the summer crops was registered for each visited plot. The land cover of the plot was also recorded. Finally, a total of 451 plots including 300 nonirrigated and 151 irrigated plots were registered for the agricultural season of summer 2017. The dominant crop types of the collected plots are maize and soybeans.

### C. S1 Time Series

Over the *Catalonia* region, a total of 82 C-band (5.405 Ghz) SAR images acquired by S-1A and S-1B satellites were used for the period between September 2017 and December 2018. Over the *West Occitanie* region, the collected irrigation information corresponds to the year 2017. Thus, the same number of images (82 images) was collected over this region for the period between September 2016 and December 2017. The 164 images (82 for Catalonia and 82 for Occitanie) are acquired in the interferometric wide (IW) mode in both VV and VH polarizations. These S1 images are of Level-1 ground range detected (GRD) product with a pixel spacing of 10 m  $\times$  10 m and six days revisit time. The S1 toolbox<sup>1</sup> (S1TBX) developed by the European Spatial Agency (ESA) was used to calibrate the S1 images. This calibration converts the digital number of downloaded SAR images into backscattering coefficients in linear units and orthorectifies the

<sup>1</sup><https://step.esa.int/main/toolboxes/snap/>

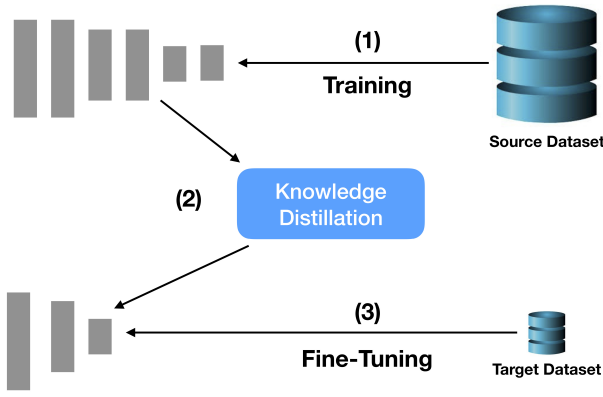


Fig. 2. Overview of our distill and refine pipeline. (1) Complex deep learning approach (teacher) is learned on a certain study area involving a large volume of labeled samples. (2) Smaller (student) model is distilled from the teacher one. (3) Student model is finally fine-tuned with the limited labeled samples available on the target study area.

images using a 30-m digital elevation model of the Shuttle Radar Topography Mission (SRTM). The temporal series of the S1 SAR backscattering coefficients over each agricultural plot in each region was then obtained by averaging the  $\sigma^0$  values of all pixels within each plot at each available date and at both VV and VH polarizations.

### III. CONTRIBUTIONS

In this section, we describe our *Distill before refine* framework devoted to deal with spatio-temporal transfer learning over different geographical areas. Converse to other scenarios in which the transfer is made between different classification problems or considering different types of data [6], [12], the process here focuses on the transfer between two distinct geographical areas fixing the classification problem (irrigation mapping) as well as the type of considered data (S1 time series data). Fig. 2 visually depicts the different steps of our proposed pipeline.

First, a model is trained from scratch on the source data set that is characterized by a large volume of labeled samples [Step (1) in Fig. 2]. Such a model, considering knowledge distillation literature [5], [9], is commonly denoted as the teacher model. Second, a lighter model (commonly named student) is distilled from the teacher model [Step (2) in Fig. 2]. The distillation step allows transferring the knowledge from the teacher to the student model. More in detail, during this step, we are trying to synthesize the teacher behavior in a usually smaller network that should behave similar to the bigger one. Once the knowledge is distilled in the student network, such a model is finally fine-tuned considering the limited labeled samples coming from the target study area [Step (3) in Fig. 2].

The main ratio behind our framework is as follows: models involving a huge number of parameters such as the teacher model can hardly be fine-tuned on a target data set characterized by a scarce number of labeled samples. Conversely, smaller and lighter models such as the student one, which behave similar to bigger ones, can be adapted more easily in the presence of scarce training data, since they involve a smaller number of parameters to modify.

TABLE I  
ARCHITECTURES OF THE TEACHER AND STUDENT CNNs

Teacher Model	Student Model
Conv(nf=64, k=7x1, s=1, act=ReLU)	Conv(nf=96, k=7x1, s=1, act=ReLU)
Conv(nf=64, k=5x1, s=2, act=ReLU)	Conv(nf=96, k=5x1, s=2, act=ReLU)
Conv(nf=128, k=5x1, s=1, act=ReLU)	Conv(nf=192, k=5x1, s=1, act=ReLU)
Conv(nf=128, k=5x1, s=2, act=ReLU)	Conv(nf=192, k=5x1, s=1, act=ReLU)
Conv(nf=256, k=5x1, s=1, act=ReLU)	Conv(nf=384, k=5x1, s=1, act=ReLU)
Conv(nf=256, k=5x1, s=1, act=ReLU)	GlobalAveragePooling
Conv(nf=256, k=3x1, s=1, act=ReLU)	FC(nn=2, act=None)
Conv(nf=256, k=1x1, s=1, act=ReLU)	
Flatten	
FC(nn=512, act=ReLU)	
FC(nn=512, act=ReLU)	
FC(nn=2, act=Linear)	
<b>Number Parameters</b>	<b>Number Parameters</b>
1 693 570	697 538

Considering the classical framework of knowledge distillation [5], given a teacher  $T$  and a student  $S$  networks, the main objective is to distill the knowledge of  $T$  inside the network  $S$ . Usually, the common assumption is that the network  $T$  is much bigger than the network  $S$  in terms of parameters. To deal with the knowledge distillation task, [5] proposes to learn the student network considering the following loss function  $\mathcal{L}$ :

$$\mathcal{L} = \frac{1}{|X|} \sum_{x_i \in X} \alpha * \mathcal{L}_1 + (1 - \alpha) * \mathcal{L}_2$$

$$\mathcal{L}_1 = L_{CE}(\text{Softmax}(f_S(x_i)), y_i)$$

$$\mathcal{L}_2 = KL \left( \text{Softmax} \left( \frac{f_S(x_i)}{\tau} \right), \text{softmax} \left( \frac{f_T(x_i)}{\tau} \right) \right) \quad (1)$$

where the main loss ( $\mathcal{L}$ ) is a linear combination of two other task-specific losses ( $\mathcal{L}_1$  and  $\mathcal{L}_2$ ) and the linear combination is weighted by the parameter  $\alpha$ . The first task-specific loss ( $\mathcal{L}_1$ ) is the classical cross-entropy loss usually employed in the multi-class classification task. The second loss ( $\mathcal{L}_2$ ) is the Kullback-Leibler divergence between the predicted output distribution of the teacher and the student.  $x_i \in X$  (resp.  $y_i$ ) is an example of the data set (resp. the associated label). Hinton *et al.* [5] employed a temperature scaling factor  $\tau$  to smooth the last output layer of the neural models before performing the softmax normalization. We remind that  $f_T(\cdot)$  and  $f_S(\cdot)$  are the presoftmax outputs for the teacher and student model, respectively (the  $\text{Softmax}^2$  operator is a standard component of the modern neural network classification model). The objective of the loss  $\mathcal{L}_2$  is to force the student model to simulate the output of the teacher model with the aim to distill the teacher behavior into the student network.

#### A. Teacher and Student Model Implementation

In our case, both the teacher and student models are implemented as convolutional neural networks (CNNs). Table I resumes the architecture of the teacher and student CNNs. Each convolutional operation (*Conv*) is associated with a successive batch normalization and dropout layer. The convolution takes as input the number of filters (*nf*), the kernel size (*k*), the stride (*s*), and the activation (*act*). Our CNN follows the idea of temporal CNN introduced in [8], where the convolution is performed on the time dimension. Considering the

<sup>2</sup>[https://en.wikipedia.org/wiki/Softmax\\_function](https://en.wikipedia.org/wiki/Softmax_function)



fully connected (*FC*) layer, we apply batch normalization and dropout except for the output layer. The *FC* layer takes as inputs the number of neurons (*nn*) and the activation function (*act*). For all the layers, we adopt the rectifier linear unit [7] except for the last output layer that is associated with a linear function followed by the Softmax operator. Comparing the teacher and student models, we can observe that the former is deeper than the latter (11 versus 6 layers) and the teacher model involves more than the double of the parameters of the student model (about 1.69M versus 0.69M). Considering the student network, we eliminate all the *FC* layers. Moreover, the global average pooling (*GAP*) layer, at the end of the convolutional layers, is directly connected with the output layer. The *GAP* layer aggregates each feature maps via the average operator, producing a layer with as many neurons as the number of the feature maps at the precedent step.

#### IV. EXPERIMENTS

In this section, we evaluate our *distill before refine* framework in the context of irrigation area mapping using S1 time series data and the data sets introduced in Section II. We refer to the Catalonia data set as the source data set and the West Occitanie data set as the target data set. We compare the performance of our proposal with respect to several competing methods to pinpoint the benefits of compressing the deep learning model before performing fine-tuning on a study area characterized by a limited number of samples.

*Experimental Settings:* To assess the performances of the different methods, we consider two standard metrics: Accuracy and F-measure [10]. While the former is a standard metric in remote sensing, the latter corresponds to the harmonic mean between precision and recall, and it is well suited to evaluate classification performances in an unbalanced scenario. Considering the Teacher CNN model (trained on the Catalonia region characterized by a large volume of labeled samples), we consider 80% of the data set as the training data, while the rest as the validation set to choose the best model. More in detail, the model that achieves the best accuracy on the validation set is the one that is retained for the subsequent steps. Considering the fine-tuned step (trained on the French area characterized by a limited number of labeled samples), we adopt the following settings: we employ 33% (150 samples) of the data as test set, 20% (60 samples) of the data as the validation set, while we use the rest of the samples as training data (240 samples). Regarding the training set, given the validation and test set, we evaluate different models considering different amounts of training samples: {60, 120, 180, 240}. In this way, we can study the behavior of the different approaches considering a varying amount of training data to feed the learning process. For each method, we repeat the procedure ten times and we report the averaged value for each metric. For the different competing approaches, we involve a Random Forest (*RF*) classifier learned directly on the target study area (Occitanie) and we name such method *RF*; a CNN model with the same configuration of the teacher network (resp. the same configuration of the student network) learned from scratch directly on the target study area named

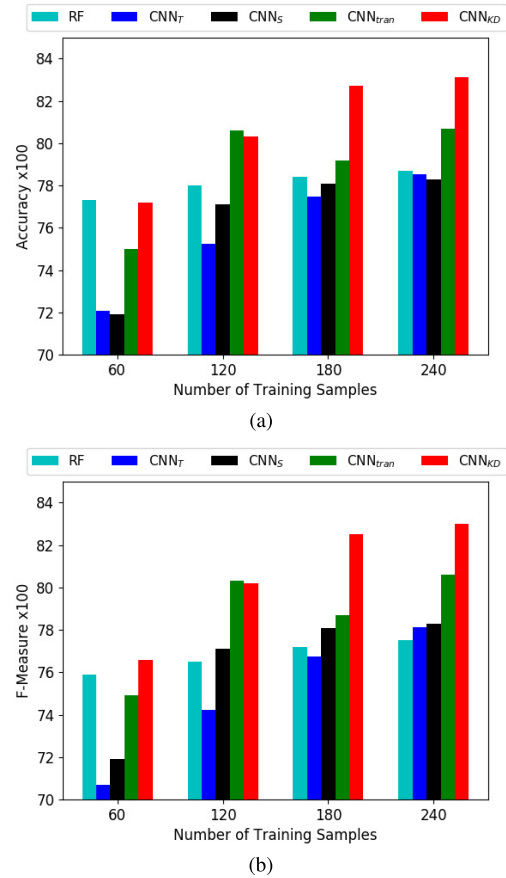


Fig. 3. (a) Accuracy and (b) F-Measure results considering the different competing methods varying the amount of (target) train samples to learn the model on the *West Occitanie* data set.

CNN<sub>T</sub> (resp. CNN<sub>S</sub>); and the teacher model directly fine-tuned on the target study area and we name such method CNN<sub>tran</sub>. The proposed distill before the refine framework is named CNN<sub>KD</sub>. For the *RF* models, we optimize the model via the maximum depth of each tree (in the range {20, 40, 60, 80, 100}) and the number of trees in the forest (in the set {100, 200, 300, 400, 500}). For the CNN models, we use Adam to optimize the parameters weights with a learning rate of  $1 \times 10^{-4}$ . CNN<sub>S</sub> learned on the source data set (Catalonia) are trained for 1000 epochs with a batch size equal to 256, while the CNN models trained (or fine-tuned) on the target data set (West Occitanie) are trained for 500 epochs with a batch size equal to 2. For the CNN<sub>KD</sub> method, we empirically set the  $\alpha$  and  $\tau$  values to 0.5 and 1, respectively.

#### A. Results

Fig. 3(a) and (b) resumes the Accuracy and F-Measure results, respectively. Considering both metrics, the *RF* methods are quite stable over the considered range. No bigger difference can be noted between the smallest amount of training samples (60) and the biggest one (240). Regarding the CNN<sub>T</sub> and CNN<sub>S</sub> methods trained from scratch on the target data set, as we can expect, they fail to be competitive considering small amount of training data. We can observe that when 60 labeled samples are considered, the model performances are

really poor compared with the competitors. On the other hand, they achieve similar performances to the *RF* method when bigger amounts of training samples are considered. Inspecting the performance of the teacher model fine-tuned on the target study area ( $CNN_{\text{tran}}$ ), it exhibits good performances starting from a training size of the labeled examples that is equal to 120, but it does not show any successive improvement for larger training size. This is probably due to the fact that this deep learning model would need more labeled information to effectively modify the huge number of parameters it involves. Such a behavior is highlighted regarding both evaluation metrics.

In general, we can observe that the  $CNN_{\text{KD}}$  obtains best (or comparable) results for all the considered amount of training samples regarding both Accuracy and F-Measure. Such a difference is clearly visible when the number of training samples is greater than or equal to 180. This behavior supports our *distill before refine* framework in which a smaller network distilled from a bigger one will be easier to adapt to a target domain characterized by a very limited number of labeled samples. This result constitutes a preliminary experiment in which the benefit of knowledge distillation is assessed to perform transfer learning from a source data set (a geographical area) to a target data set (another geographical area).

Both Accuracy and F-Measure depict almost the same behaviors among the different competing methods. The only point that changes between the evaluation of the two metrics is related to the comparison between the two *CNN* models trained from scratch ( $CNN_T$  and  $CNN_S$ ) and the *RF* method on the range 120-240. Here, we can observe opposite behavior. Regarding the accuracy, it seems that the *RF* model slightly outperforms the two *CNN* models, while the opposite happens when F-Measure is considered. This is due to the fact that the test data set, as the whole target data set, is unbalanced with respect to the involved classes. The *RF* method is biased toward the over-represented classes, and this is why it achieves best (resp. worst) performances in terms of Accuracy (resp. F-Measure). On the other hand, the two *CNN* models trained from scratch deal better with the unbalanced scenario achieving slightly better performance in terms of F-Measure, since they are less biased toward the over-represented classes. As an additional test, we also evaluate the performance of the teacher model without any additional fine-tuning directly on the test set of the target data set. In this experiment, we obtain a value of Accuracy and F-Measure equal to 27.5% and 36.86%, respectively. Considering that we are dealing with a binary task, such performances are lower than the average random performances we can obtain on such a task. This poor behavior indicates that the data distribution associated with the source (Catalonia area) and target (Occitanie area) data sets is really different and a heavy distribution shift exists. Such a result confirms once more the necessity to study and develop new machine learning approaches to deal with shifts in data distribution (due to spatial or temporal autocorrelation) to cope with the spatio-temporal model transfer for remote sensing data. With the increasing availability of remote sensing data, coming from large-scale monitoring systems (i.e., Copernicus data),

issues related to transfer models learned from a particular area (resp. in a particular time period) to deal with data coming from another geographical area (resp. another time period) will get more and more attention in a near future and, probably, greater effort will be made in this direction.

## V. CONCLUSION

We have introduced a new *distill before refine* framework to deal with spatial transfer between two geographical areas coming from different countries. We have deployed the proposed framework in the context of irrigation mapping leveraging S1 time series data. Experimental evaluations have underlined the effectiveness of the proposed framework in the context of irrigated area mapping. As future work, we plan to investigate more precisely the influence of the  $\alpha$  and  $\tau$  hyperparameters in the knowledge distillation phase of our framework.

## ACKNOWLEDGMENT

The authors would like to thank the National Center for Space Studies (CNES) for financing this project in the frame of the TOSCA program. They would like to thank M. J. Escorihuel from the isardSAT, Catalunya, for providing the SIGPAC data. They also would like to thank F. Dauriac of the Chamber of Agriculture of Hautes-Pyrenees, France.

## REFERENCES

- [1] H. Bazzi *et al.*, "Mapping irrigated areas using sentinel-1 time series in catalonia, spain," *Remote Sens.*, vol. 11, no. 15, p. 1836, 2019.
- [2] X. Cai and M. W. Rosegrant, "Global water demand and supply projections: Part 1. A modeling approach," *Water Int.*, vol. 27, no. 2, pp. 159–169, 2002.
- [3] Q. Gao, M. Zribi, M. Escorihuela, N. Baghdadi, and P. Segui, "Irrigation mapping using sentinel-1 time series at field scale," *Remote Sens.*, vol. 10, no. 9, p. 1495, 2018.
- [4] H. C. J. Godfray *et al.*, "Food security: The challenge of feeding 9 billion people," *Science*, vol. 327, no. 5967, pp. 812–818, 2010.
- [5] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*. [Online]. Available: <https://arxiv.org/abs/1503.02531>
- [6] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, p. 907, 2017.
- [7] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
- [8] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sens.*, vol. 11, no. 5, p. 523, 2019.
- [9] Y. Shi, M.-Y. Hwang, X. Lei, and H. Sheng, "Knowledge distillation for recurrent neural network language modeling with trust regularization," in *Proc. ICASSP*, May 2019, pp. 7230–7234.
- [10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Reading, MA, USA: Addison-Wesley, May 2005.
- [11] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Feature transfer learning for face recognition with under-represented data," in *Proc. CVPR*, 2019, pp. 5704–5713.
- [12] B. Zhao, B. Huang, and Y. Zhong, "Transfer learning with fully pretrained deep convolution networks for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1436–1440, Sep. 2017.
- [13] Z. Zhou, J. Y. Shin, L. Zhang, S. R. Gurudu, M. B. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in *Proc. CVPR*, 2017, pp. 4761–4772.
- [14] X. X. Zhu *et al.*, "Deep learning in remote sensing: A review," 2017, *arXiv:1710.03959*. [Online]. Available: <https://arxiv.org/abs/1710.03959>