MULTI-TEMPORAL CONVOLUTIONAL NEURAL NETWORKS FOR SATELLITE-DERIVED SOIL MOISTURE OBSERVATION ENHANCEMENT

Grigorios Tsagkatakis[1], *Mahta Moghaddam*[2], *Panagiotis Tsakalides*[1]

[1]Foundation for Research and Technology - Hellas (FORTH), Crete, Greece
[2]University of Southern California, Los Angeles, CA

ABSTRACT

In this work, we propose a novel Convolutional Neural Network architecture for increasing the low spatial resolution SMAP radiometer based soil moisture estimations from 36 km to 3 km resolution by using time-series of observations from both SMAP's radiometer and Sentinel-1 radar. By simultaneously extracting features from both current low-resolution input and residuals between high and low resolution at previous time instances, the proposed network is capable of accurately estimating soil moisture using coarse resolution observations. Experimental results on three different locations demonstrate that the proposed scheme is able to estimate soil moisture with accuracy in the range of the requirements set by the SMAP science team.

Index Terms— Soil Moisture; Deep Learning; Observation Fusion; SMAP; Sentinel-1.

1. INTRODUCTION

Accurate and timely soil moisture (SM) estimation is critical for a number of applications ranging from understanding the water cycle to predicting extreme climate events. The NASA's Soil Moisture Active Passive (SMAP) satellite, is tasked with the high temporal and spatial resolution estimation of global surface SM and freeze-thaw. To achieve its objective, SMAP was equipped with an active L-band radar, acquiring observation with 3 km spatial resolution, and a passive L-band radiometer acquiring 36 km spatial resolution observations. By fusing these two types of observations, Level-2 global surface SM products at 9 km spatial resolution were produced every three days [1]. Unfortunately, after a few months in operation, the radar module failed. To mitigate the impact of the radar loss, a collaboration between NASA and ESA considers the use of C-band radar measurements from Sentinel 1A and 1B satellites with SMAP observations [2]. The synergy between the observations from two missions has allowed the generation of highly accurate SM products at 3 km. However, such high-resolution products can only be generated when concurrent observations from both platforms are available, limiting the effective global coverage to 12 days, in contrast to 2-3 days revisit frequency for SMAP. The limited availability of high spatial resolution synergistically derived SM maps motivates this work, which introduces a novel spatio-temporal Convolutional Neural Network (CNN) architecture for increasing the spatial resolution (downscaling) of SMAP derived products, exploiting the limited availability of SMAP/Sentinel 1 derived SM.

This work was partially funded by the CALCHAS project (contract no. 842560) within the H2020 Framework Program of the European Commission.

2. STATE-OF-THE-ART

Estimation of SM from in-situ and/or satellite observations is an extensively investigated topic and numerous approaches have been proposed [3], while machine learning based approaches have gained considerable attention due to their flexibility and ability to process a large number of inputs [4, 5, 6]. More recently, the Deep Learning framework has gained considerable attention for the enhancement of remote sensing observations [7] and has been considered for the problem of SM estimation. In order to capture the information encoded in time-series (LSTM) networks were considered for SM estimation from Brightness Temperature measurements from SMAP, MODIS Vegetation Water Content and soil temperature in [8]. A work similar to the one reported here is the method proposed by Mao et al. [9] where the authors considered machine learning, random forests in particular, for estimating the high-resolution SMAP/Sentinel-1 estimation given low-resolution SMAP radiometry data. A earlier version considered CNNs for downscaling SMAP radiometer brightness temperature measurements, focusing only on the period when both SMAP radar and radiometer were operational [10].

3. SOIL MOISTURE ENHANCEMENT NETWORK (SMEN)

The objective of the SMEN is to encode time-series of low spatial resolution SM observations from SMAP and produce high spatial resolution SM, equivalent to those produced by

the fusion of Sentinel 1 and SMAP observations. To achieve this objective, the proposed CNN architecture must be capable of extracting appropriate features from the time-series, as well as utilizing training examples for increasing the spatial resolution. In order to accommodate both requirements, we proposed a network architecture consisting of a cascade of layers extracting features from previous time instances and subsequently combining these features with the current low-resolution observations. A high-level visual description of the proposed architecture is shown in Figure 1.

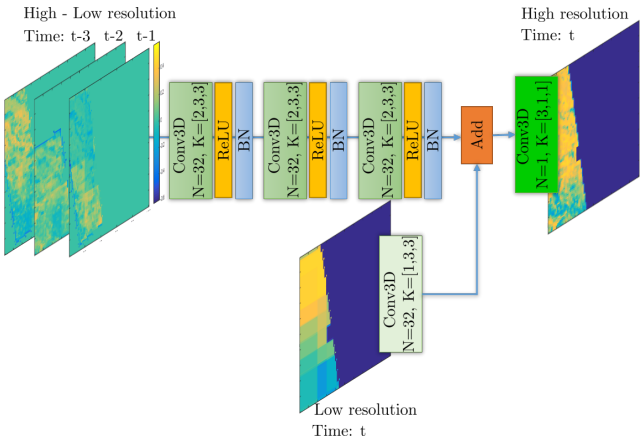

Fig. 1: Overview of resolution enhancement scheme.

Let $\mathcal{I}(t)$ and $\mathcal{J}(t)$ denote three-dimensional tensors encoding the low (SMAP only) and high (SMAP and Sentinel-1) resolution observations acquired at times $[t-k, t-k+1, \cdots, t-1]$ and $\mathcal{D}(t) = \mathcal{J}(t) - \mathcal{I}(t)$ a three-dimensional tensor encoding the residual, i.e., the difference between low and high-resolution observations. Given $I(t)$ and $J(t)$, two-dimensional matrices encoding the low and high-resolution observations at time instance t, respectively, the generating model employed in this work assumes that

$$J(t) = h(g(I(t)) + f(\mathcal{D}(t))) \tag{1}$$

where $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ are in general non-linear functions that represent the extractions of features by the associated layers of the CNN, from the low-resolution observations, the residuals, and the fused output, respectively.

More specifically, the architecture representing $f(\cdot)$ consists of three convolutional blocks where each convolutional block is composed of a 3D convolution with 32 kernels of size $(2 \times 3 \times 3)$, a (ReLU) activation, and a batch normalization layer, while the function $g(\cdot)$ operating on the low-resolution input consist of a single convolutional layer of 32 kernels of size $(1 \times 3 \times 3)$. The two streams of extracted features are added together following the assumption that the temporal-focused components will produce the texture, which will be incorporated to the input image. A last convolution layer, the h function, consisting of a single filter of size $(1 \times 3 \times 3)$, is introduced at the end for fusion of the 32 individual feature-extracted images into a single prediction $2D$ image.

In order to derive relevant soil moisture estimation, we introduce the unbiased root mean squared error (uRMSE) [11] as the loss function $\mathcal{L}$, which is minimized by the network. Formally, given the estimated $\hat{J}(t)$ and the true $J(t)$ high-resolution SM, the loss function is defined as

$$\mathcal{L} = M(t) \odot \sqrt{E[(\hat{J}(t) - E[\hat{J}(t)]) - (J(t) - E[J(t)])]^2} \tag{2}$$

where $M(t)$ is a binary mask, which forces the error to be estimated only on pixels where measurements are available.

4. PERFORMANCE EVALUATION

The objective of the following experiments is to understand the capabilities of the SMEN in terms of (i) high-resolution SM estimation quality; (ii) impact of each input data type; (iii) impact of the number of training examples on the estimation accuracy. To quantify the performance, surface SM prediction quality is reported in m^3/m^3 while the specification of SMAP's science team define a volumetric accuracy requirements of $0.04\ m^3/m^3$ measured in terms of the uRMSE.

4.1. Experimental setup

The first source of observations is Level-2 SM products from SMAP radiometer at 36 km resolution and the second input is a time-series of difference images generated by subtracting the high from the low-resolution images, at previous time instances, at 3 km resolution. The objective is to estimate "ground truth" SM of each pixel at any given time, equivalent to the accuracy of the 3 km resolution of the Level 2 SMAP/Sentinel Active-Passive SM product. In all cases, the observations are encoded in 100×100 pixel images while the SMEN is trained for 1000 epochs with the Adam optimizer and a learning rate of 10^{-4}.

We consider observations from three regions, two in the continental USA, namely regions around Oklahoma City, characterized by a humid subtropical climate, and Tucson Arizona, characterized by a hot desert climate, as well as a region around Geneva, Switzerland, characterized as moderately continental, with warm summers and cold winters. We collected observations from 2016, 2017 for training and 2018 for testing from NASA Worldview platform[1].

4.2. Experimental results

The estimation accuracy for three different locations is given for the training and testing data in Table I. The experimental results demonstrate that the proposed CNN architecture is

[1] https://worldview.earthdata.nasa.gov/

capable of producing high quality estimation on the test examples, in one case below and in the other two comparable to requirement set by the SMAP science team.

Table 1: Estimation accuracy (uRMSE) for different locations. Numbers in parenthesis indicate the number of training and testing examples.

Location	Train error	Test error
Oklahoma City (80/50)	0.021	0.048
Tuscon (30/10)	0.024	0.034
Geneva (60/30)	0.057	0.046

Figure 2 presents an exemplary collection of high-resolution estimation, along with the low-resolution input and the high-resolution ground truth for the three regions under consideration. Given this collection, a number of observations can be made regarding the estimation process. First, one can observe that the estimation offers significantly more detail in the form of texture compared to the low-resolution input. Second, we observe that the severe blocking artifacts present in the low-resolution imagery are to a large extent suppressed in the estimation, capturing the temporal smoothness that characterizes real-world data. A third observation is that the proposed network estimates/completes the entire region, although the error is measured only in locations where high-resolution measurements are available.

4.3. Impact of inputs

The proposed CNN architecture seeks to simultaneously encode temporal information and spatial information by exploiting both past and present observations. In order to experimentally verify that exploiting both spatial and temporal information leads to better performance, Table 2 reports the estimation quality when using either low-resolution spatial only, temporal residuals only, and using both in the complete architecture in spatiotemporal case.

Table 2: Impact of different inputs on the Tuscon region

	uRMSE (train)	uRMSE (test)
Spatial	0.022	0.039
Temporal	0.026	0.049
Spatiotemporal	0.024	0.034

These results reveal that spatial information, i.e., the texture, is more important than temporal while using both sources of information leads to the best performance. This is manifested in the test error. We can also observe, that for the spatio-temporal case compared to the spatial case, the situation is reversed for the training data, which can be attributed to overfitting due to the limited amount of training data.

4.4. Impact of the availability of training data

Last, we explore the impact of the size of the training set on the estimation quality by varying the number of training examples while keeping the testing examples fixed to 10. Results on the estimation quality for the training and testing data are reported in Table 3. The results indicate that, as expected, using more training data has a significant impact, especially on the testing data, demonstrating that increasing the training set size can lead to better generalization.

Table 3: Impact of different training set sizes

#Training set	uRMSE (train)	uRMSE (test)
10	0.025	0.085
20	0.025	0.068
50	0.023	0.033
100	0.022	0.034

5. CONCLUSION

In this work we present a CNN architecture encoding spatial-temporal information, which is capable of estimating high spatial resolution SM equivalent to the one available when observations from two platforms, namely the SMAP and Sentinel-1 are available, using only observations from the high temporal, low spatial resolution SMAP.

6. REFERENCES

[1] N. N. Das, D. Entekhabi, R. S. Dunbar, A. Colliander, F. Chen, W. Crow, T. J. Jackson, A. Berg, D. D. Bosch, T. Caldwell *et al.*, "The smap mission combined active-passive soil moisture product at 9 km and 3 km spatial resolutions," *Remote sensing of environment*, vol. 211, pp. 204–217, 2018.

[2] N. N. Das, D. Entekhabi, S. Kim, T. Jagdhuber, S. Dunbar, S. Yueh, and A. Colliander, "High-resolution enhanced product based on smap active-passive approach using sentinel 1a and 1b sar data," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 2543–2545.

[3] E. Babaeian, M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller, "Ground, proximal and satellite remote sensing of soil moisture," *Reviews of Geophysics*.

[4] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data," *Remote Sensing*, vol. 7, no. 12, pp. 16 398–16 421, 2015.

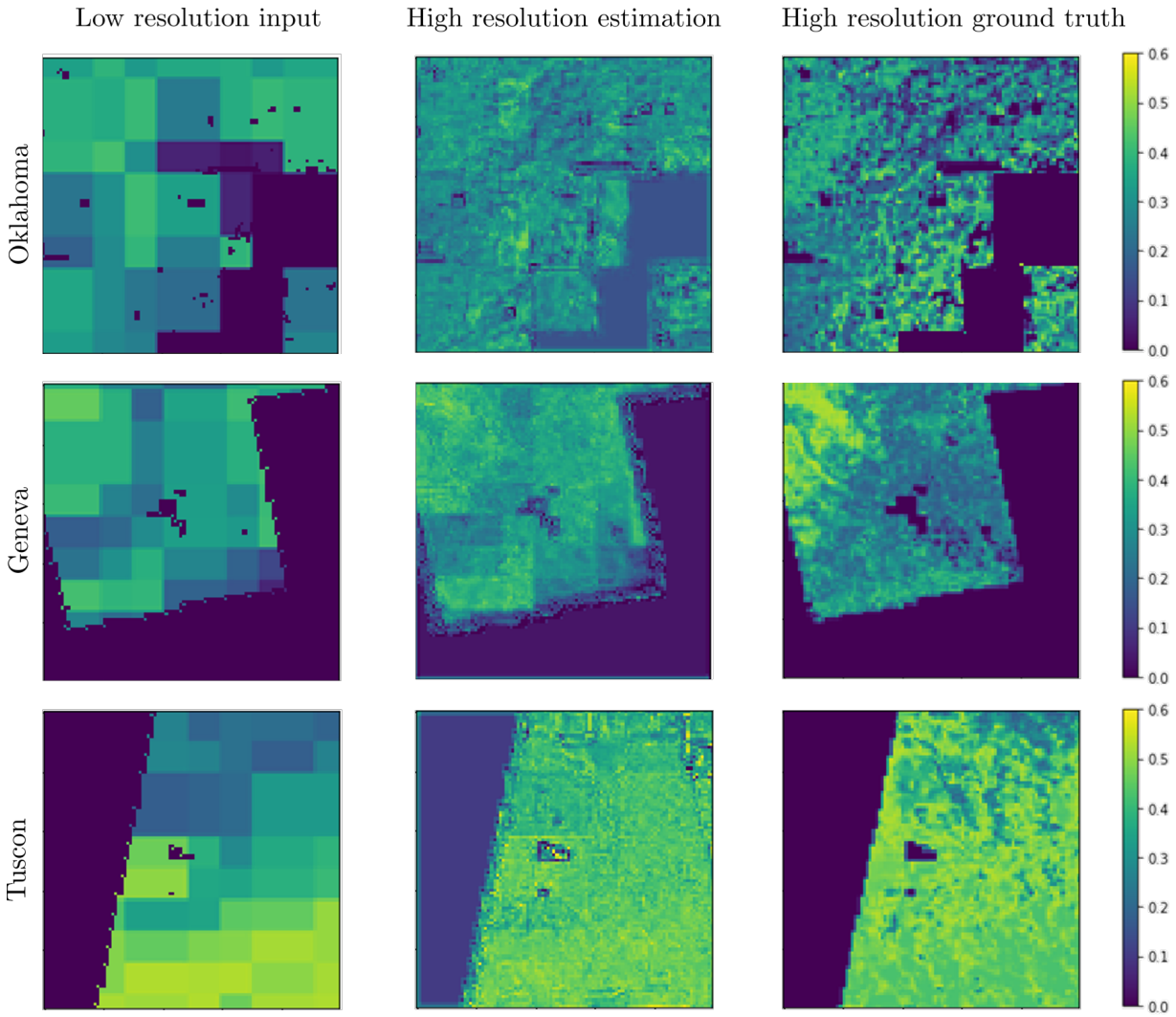

Fig. 2: Examples of SM input, estimation and "ground truth" (columns) for three different locations (rows).

[5] K. Fotiadou, G. Tsagkatakis, M. Moghaddam, and P. Tsakalides, "Recovery of soil moisture active passive (smap) instrument's active measurements via coupled dictionary learning," *Electronic Imaging*, vol. 2017, no. 17, pp. 185–190, 2017.

[6] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Advances in Water Resources*, vol. 33, no. 1, pp. 69–80, 2010.

[7] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep-learning approaches for remote sensing observation enhancement," *Sensors*, vol. 19, no. 18, p. 3929, 2019.

[8] A. B. Abbes, R. Magagi, and K. Goita, "Soil moisture estimation from smap observations using long short-term memory (lstm)," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1590–1593.

[9] H. Mao, D. Kathuria, N. Duffield, and B. P. Mohanty, "Gap filling of high-resolution soil moisture for smap/sentinel-1: A two-layer machine learning-based framework," *Water Resources Research*, 2019.

[10] M. Moghaddam, G. Tsagkatakis, and P. Tsakalides, "Convolutional neural networks for downscaling smap radiometer brightness temperature observations with high-resolution radar," in *AGU Fall Meeting Abstracts*, vol. 2018, Dec 2018.

[11] D. Entekhabi, R. H. Reichle, R. D. Koster, and W. T. Crow, "Performance metrics for soil moisture retrievals and application requirements," *Journal of Hydrometeorology*, vol. 11, no. 3, pp. 832–840, 2010.