ConvLSTM for Spatio-Temporal Feature Extraction in Time-Series Images

Gael Kamdem De Teyou

Abstract

Earth observation programs have provided highly useful information in global climate change research over the past few decades and greatly promoted its development, especially through providing biological, physical, and chemical parameters on a global scale. Programs such as Landsat, Sentinel, SPOT, and Pleiades can be used to acquire huge volume of medium to high resolution images every day. In this work, we organize these data in time series and we exploit both temporal and spatial information they provide to generate accurate and up-to-date land cover maps that can be used to monitor vulnerable areas threatened by the ongoing climatic and anthropogenic global changes. For this purpose, we combine a fully convolutional neural network with a convolutional long short-term memory. Implementation details of the proposed spatio-temporal neural network architecture are described. Examples are provided for the monitoring of roads and mangrove forests on the West African coast.

1 Introduction

Land cover can be defined as the physical features that cover the land, such as trees or pavement. A similar notion is land use that can be defined as human management and activities on land, such as mining or recreation. Climate can affect and be affected by changes in land cover. A forest, for instance, would likely include tree cover but could also include areas of recent tree removals currently covered by open grass areas. Land cover and use are inherently coupled: changes in land-use practices can change land cover, and land cover enables specific land uses. Changes in land cover can occur in response to both human and climate drivers. Therefore understanding how land cover, use, condition, and management vary in space and time is important to analyze vulnerable areas that are threatened by the ongoing climatic and anthropogenic global changes. With a low topography, a weak geological substratum, a poor fresh water supply, and a dense and rapidly expanding population, some littoral areas such as the West African coast for example are highly vulnerable to current sea level rise, extreme climatic phenomena, erosion, and modifications of the ecosystems and resources. Many Earth observation programs such as Landsat, Sentinel, SPOT and Pleiades produce huge volume of medium to high resolution multispectral images every day that can be organized in time series and used to produce accurate and up-to-date land cover and use maps that can monitor environmental changes at different places and time ranges.

In the past few years, Deep learning (DL), a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from complex data, has gained attention. Deep learning architectures, especially Fully Convolutional Networks (FCNs) show a great potential for application to various remote sensing problems such as land cover and use mapping. However, even if FCNs deal very well with spatial representation of features within an image, they are unable to learn the additional information provided by the multi-temporal structure of time series images, which can improve the land cover classification accuracy and efficiency. In addition, when applied on a single image, they lead systematically to a variance, that depends on the day when the image was acquired.

Tackling Climate Change with Machine Learning workshop at NeurIPS 2020.

To capture the temporal dependency of images, Recurrent Neural Networks (RNN), a class of deep learning architectures where connections between nodes form a directed graph along a temporal sequence, can be used. For example, (G. Sumbul, B. Demir, 2019) present a novel multi-attention driven system that jointly exploits Convolutional Neural Network (CNN) and RNNs in the context of multi-label remote sensing image classification. The article (Stoian et al., 2019) proposes a framework for working with Sentinel-2 L2A time-series image data, and an adaptation of the U-Net model for dealing with sparse annotation data while maintaining high resolution output. In (C. Pelletier et al., 2019), an analysis of RNN and Temporal Convolutional Neural Networks (TempCNNs) for the classification of Sentinel-2 image time series is provided. In (Emmanuel Maggiori, 2017), RNNs are used to correct satellite image classification maps.

In this work, we exploit both temporal and spatial information provided by multi-temporal Sentinel-2 images to generate accurate and up-to-date land cover and use maps. For this purpose, we combine a FCN with a RNN.

2 METHODOLOGY

Our acrhitecture combines a FCN that captures the spatial representation of features in the image with a RNN that learns the temporal variations of these features. The architecture of our FCN is based on the U-Net. It is a symmetric encoder-decoder structure consisting of a contracting branch that captures the context and an expanding branch that enables precise localization for the segmentation masks. The contracting path consists of the repeated application of two 3×3 convolutions, each followed by a ReLu and a 2×2 max pooling for downsampling. At each downsampling step, the number of feature channels is doubled. Every step in the expansive path consists of a transpose convolution that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path and one 3x3 convolution. This FCN enables us to generate a Probability Map (PMap) from an input image. The PMap is an image of the same size as the input but with l_c channels, where l_c is the number of land cover or use classes. For each pixel, it gives the probability to belong to a particular land cover or use class.

We use Sentinel-2 that acquires pictures of the Earth every five days at 10-m spatial resolution, with 3-7 days revisit frequency. These time-series data change slightly from one day to another. One factor for example, is the sun movement that changes the sun-target-sensor geometry constantly. As a consequence, this effect causes an additional alteration of the radiometric data on pixels with the same land cover and similar structure. Learning only relevant information from these temporal data and removing time-dependent variance can yield accurate and up-to-date land cover maps. To encode temporal dependencies in the PMap, we can use RNNs. In this paper, LSTM is used to capture the correlation among the PMap generated by the FCN for the same land cover, but with a sequence of images taken at different dates. Then we aim to reduce the variance related to the day of acquisition and produce a *single*, *accurate* and *up-to-date* probability map. For the purpose of keeping the spatial structure of the feature map, we use Convolutional LSTM (ConvLSTM) where matrix multiplication is replaced by convolution at each gate. ConvLSTM networks capture spatiotemporal correlations better than standard fully connected LSTM (Xingjian Shi et al., 2015).

We used a joint loss function L, combining cross entropy with a differentiable form of intersection over union (V. Iglovikov et al., 2017):

$$L = \alpha H - (1 - \alpha) \log(J), \tag{1}$$

with α an hyperparameter that we can tune, H the cross entropy and J the IoU.

The final architecture is shown on Fig. 1. It consists of a FCN, two ConvLSTM layers and one convolution layer. Gerenerally, two LSTM layers are enough to detect complex features. More layers can be better but also harder to train. The output of the FCN is sent to the first convLSTM layer. The input is a sequence of images from the same land cover but taken at different times. The images in the sequence are first encoded with the FCN to extract the probability maps. For each image, the output of this operation is an image of the same size as the original one but with l_c channels (number of land cover classes). The first convLSTM combines this transformed image (x_t in the equations) with its short-term h_{t-1} and long-term c_{t-1} memories. The output is another l_c-channels image that is sent to the second convLSTM layer. Finally, the short-term memory output, h_t , is linearly transformed per pixel with the last convolution layer and scaled with a sigmoid function to obtain the final probability map.

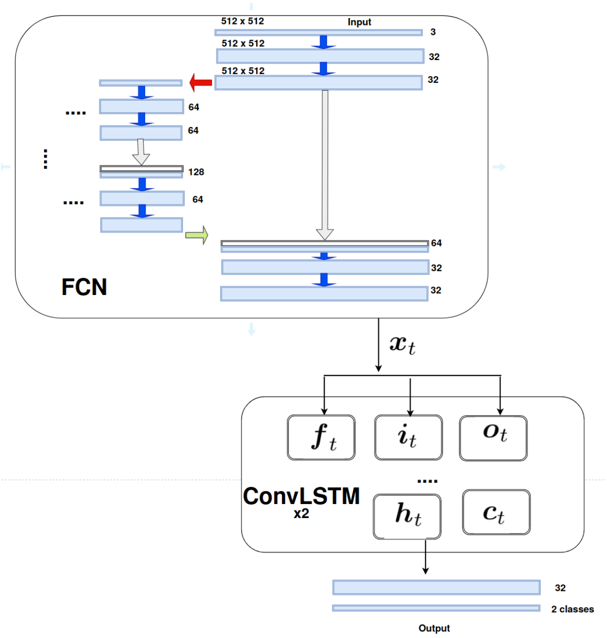

Figure 1: Proposed architecture for the model

3 Dataset and Experiments

For this work we focus on mangrove and roads classes. The corresponding Sentinel-2 images were downloaded from ESA website. The download folder contains the envelope of all resolutions including 10m, 20m and 60m. We used the True Color Image (TCI) at 10m resolution, built from the B02 (Blue), B03 (Green), and B04 (Red) bands. We set the sequence length of the RNN to 5 images. So for each city, 5 images corresponding to 5 different acquisition days were downloaded. Each image has a spatial resolution of 10m/pixel and a size of 10900 x 10900 pixels. We also used vertical/horizontal flips and 0/90/180/270 degrees rotations to augment data. Data augmentation helps in building a strong model which is less dependent on input image orientation. This is very helpful for our model to generalize to different regions other than regions in training set. To fit large images into GPU memory, we divided the input image into smaller patches $512 \times 512 \times 3$ pixels. First, the FCN is trained and then used to generate PMap sequences of length 5 to train the RNN. For both, we used the Adam optimization algorithm, with a base learning rate of 0.1 for 10 epochs, that we decreased to 0.01 for another 10 epochs and finally to 0.001 for the last 10 epochs.

3.1 Roads

In our dataset, roads are represented with a line in a shapefile, and very often this line is not properly located at the center of the road. Therefore we used morphological dilation to increase the width of the road after rasterization. We also tested dilation with a gaussian filter, but results were better with morphological dilation. Working with single images gave us an accuracy of 92.1% during training and 91.1% during validation. However, considering sequence of 5 images significantly increased the accuracy to 95.1% during for training and 93.4 % for validation. Fig. 2 shows roads extracted over the city of Saint-Louis in Senegal in 2019 (Left) and 2016 (Middle). At the right we derived new roads constructed during that time interval. This enable us to monitor easily vulnerable areas that are threatened by the construction of new roads.

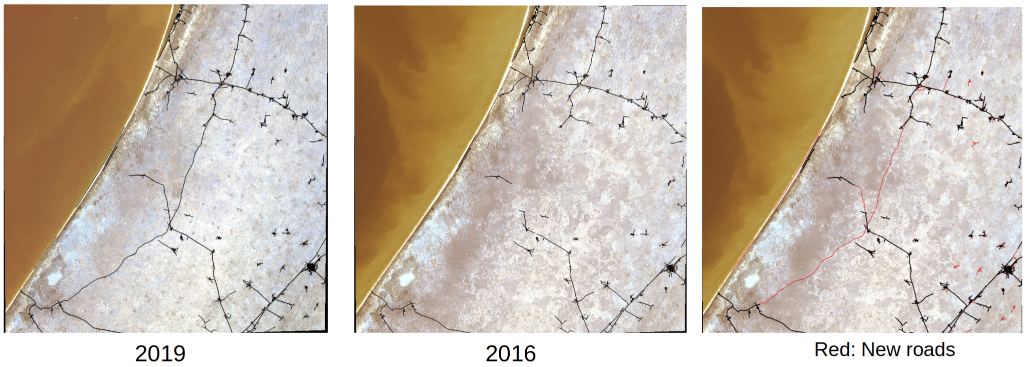

Figure 2: Road extraction over the city of Saint-Louis in Senegal using Sentinel-2 images

4 Mangrove

Mangrove forests are made up of a collection of saltwater tolerant tree species and that in different parts of the world, a single mangrove forest can be made up of numerous tree species. In addition, mangrove forests don't exist in isolation but are situated adjacent to other forest types (such as rain forests), arid areas, urban areas and areas that have been deforested that now contain land uses such as aquaculture and agriculture. In addition, mangroves can exist in different geomorphological settings including at the mouth of rivers, in coastal lagoons and in low lying areas that are under tidal influence. Given this, we defined Areas of Interest (AOIs) on the West African coast that captured as much of the possible variability in the appearance of mangrove forest as we could. We used a pre-existing data set (Global Mangrove Watch Baseline 2010) to guide the selection of AOIs. Working with single images gave us a validation accuracy of 80.1% on the West African coast region. However, considering sequence of 5 images significantly increased the accuracy to 86.3% accuracy. Since mangrove forests are difficult to distinguish with human eye from Sentinel-2 images, we make use of Google view to analyse the mangrove. Fig. 3 represents the google view of the city of Douala, Cameroon (Left) and the predicted mangrove (Middle). The ground truth is at the right. We can see that most of the mangrove forest are detected.

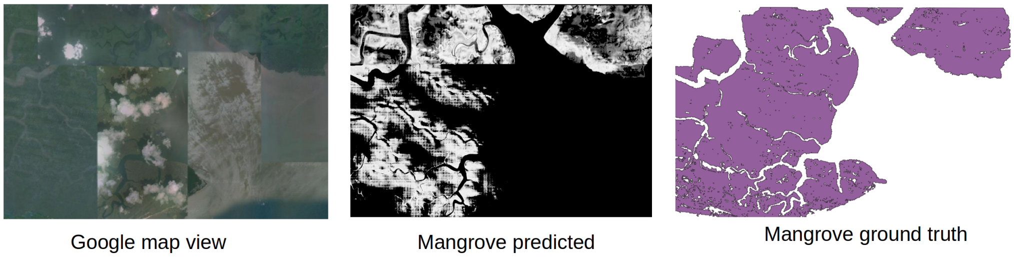

Figure 3: Mangrove in Douala, Cameroon

5 Conclusion

In this paper, accurate and up-to-date land cover and use maps are generated by applying a deep learning model that exploits both temporal and spatial information provided by 10-m resolution multi-temporal and multi-spectral satellite images. The deep learning model we designed is based on the combination of a fully convolutional neural network with skip connections, which takes into account spatial information, together with a convolutional LSTM layer, which exploits the temporal information. The proposed methodology is used to identify road networks and mangrove forests from Sentinel-2 images. Experimental results show that encoding the temporal information from the image time series into the LSTM layer memory cells improves significantly the segmentation performance.

References

[1] Abbas TAATI, et al., “Land Use Classification using Support Vector Machine and Maximum Likelihood Algorithms by Landsat 5 TM Images,” Engineering and Physical Sciences Journal, 2014

[2] Andrei Stoian, et al. "Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems" MDPI, Remote Sensing 2019.

[3] A. Ben Hamida et al. Amina Ben Hamida, Alexandre Benoit, P. Lambert, L Klein, Chokri Ben Amar, et al.. Deep Learning for Semantic Segmentation of remote sensing images with rich multispectral content. IEEE International Geoscience and Remote Sensing Symposium, Jul 2017,Fort Worth, United States

[4] C. Pelletier, G. I. Webb and F. Petitjean, "Deep Learning for the Classification of Sentinel-2 Image Time Series," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 461-464.

[5] Das, P., Pandey, V. Use of Logistic Regression in Land-Cover Classification with Moderate-Resolution Multispectral Data. J Indian Soc Remote Sens 47, 1443–1454 (2019). https://doi.org/10.1007/s12524-019-00986-8

[6] Emmanuel Maggiori, Guillaume Charpiat, Yuliya Tarabalka, Pierre Alliez. Recurrent Neural Net-works to Correct Satellite Image Classification Maps. IEEE Transactions on Geoscience andRemote Sensing, Institute of Electrical and Electronics Engineers, 2017, 55 (9), pp.4962-4971.

[7] Gael Kamdem De Teyou. Deep Learning Acceleration Techniques for Real Time Mobile Vision Applications, arXiv:1905.03418, 2019

[8] Gael Kamdem De Teyou. Junior Ziazet. Convolutional Neural Network for Intrusion Detection System In Cyber Physical Systems, arXiv:1905.03168

[9] Gael Kamdem De Teyou. Deep Residual Convolutional Neural Network for Face Segmentation,

[10] G. Sumbul, M. Charfuelan, B. Demir, V. Markl, "BigEarthNet: a large-scale benchmark archive for remote sensing image understanding," IGARSS 2019.

[11] G. Sumbul, B. Demir, "A novel multi-attention driven system for multi-label remote sensing image classification," IGARSS 2019.

[12] Itten, K.I.; Meyer, P. Geometric and radiometric correction of TM data of mountainous forested areas. IEEE Trans. Geosci. Remote Sens. 1993, 31, 764–770.

[13] Hochreiter, S.; Schmidhuber, J. Long Short–Term Memory. Neural Compution Journal. 1997, 9, 1735–1780.

[14] Justice, C.O.; Wharton, S.W.; Holben, B. Application of digital terrain data to quantify and reduce the topographic effect on Landsat data. Int. J. Remote Sens. 1981, 2, 213–230.

[15] Long, J., Shelhamer, E., Darrell, T., Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015 (pp. 3431–3440)

[16] M. Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, A Meygret, F Spoto, O Sy, F Marchese, and P Bargellini, “Sentinel-2: ESA’s optical high-resolution mission for GMES operational services,” Remote Sensing of Environment, vol. 120, pp. 25–36, 2012.

[17] N. Audebert, B. Le Saux, and S. Lef‘evre, “Semantic segmentation of earth observation data using multi-modal and multi-scale deep networks,” inAsian Confer-ence on Computer Vision, 2016.

[18] O. Ronneberger, P. Fischer, T. Brox, 2015., U-Net: Convolutional Networks for Biomedical Image Segmentation

[19] Priit Ulmasi1, Innar Liiv, Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification, arXiv 2020

[20] Robinson C, Hou L, Malkin K, Soobitsky R, Czawlytko J, Dilkina B, Jojic N. Large Scale High-Resolution Land Cover Mapping with Multi-Resolution Data. Proceedings of the 2019 Conference on Computer Vision and Pattern Recognition (CVPR 2019).

[21] S. Chantharaj et al., "Semantic Segmentation On Medium-Resolution Satellite Images Using Deep Convolutional Networks With Remote Sensing Derived Indices," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 1-6.

[22] Soenen, S.A.; Peddle, D.R.; Coburn, C.A. A Modified Sun-Canopy-Sensor Topographic Correction in Forested Terrain. IEEE Trans. Geosci. Remote Sens. 2005, 43, 2148–2159.

[23] Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 2018, 18, 18.

[24] Vasileios Syrris et al., Evaluation of the Potential of Convolutional Neural Networks and Random Forests for Multi-Class Segmentation of Sentinel-2 Imagery, MDPI, Remote Sensing 2019

[25] Vasileios Syrris.. Evaluation of the Potential of Convolutional Neural Networks and Random Forests for Multi-Class Segmentation of Sentinel-2 Imagery

[26] Vazquez-Jimenez, R.; Romero-Calcerrada, R.; Ramos-Bernal, R.N.; Arrogante-Funes, P.; Novillo, C.J. Topographic Correction to Landsat Imagery through Slope Classification by Applying the SCS + C Method in Mountainous Forest Areas. ISPRS Int. J. Geo-Inf. 2017, 6, 287.

[27] V. Iglovikov, S. Mushinskiy, and V. Osin, “Satellite imagery feature detection using deep convolutional neural network: A kaggle competition,” arXiv preprint arXiv:1706.06169, 2017.

[28] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS’15). MIT Press, Cambridge, MA, USA, 802–810.

[29] Yao X, Yang H, Wu Y, et al. Land Use Classification of the Deep Convolutional Neural Network Method Reducing the Loss of Spatial Features. Sensors (Basel). 2019;19(12):2792. Published 2019 Jun 21. doi:10.3390/s19122792