Monitoring Urban Changes with Ensemble of Neural Networks and Deep-Temporal Remote Sensing Data

Georg Zitzlsberger*
georg.zitzlsberger@vsb.cz
IT4Innovations, VŠB – Technical University of Ostrava
Ostrava-Poruba, Czechia

Michal Podhoranyi
michal.podhoranyi@vsb.cz
IT4Innovations, VŠB – Technical University of Ostrava
Ostrava-Poruba, Czechia

Václav Svatoň
vaclav.svaton@vsb.cz
IT4Innovations, VŠB – Technical University of Ostrava
Ostrava-Poruba, Czechia

Milan Lazecký
milan.lazecky@vsb.cz
University of Leeds
Leeds, UK

Jan Martinovič
jan.martinovic@vsb.cz
IT4Innovations, VŠB – Technical University of Ostrava
Ostrava-Poruba, Czechia

ABSTRACT

Urban change detection with remote sensing data offers a wide field of applications like understanding socio-economic impacts, identifying new settlements, or analyzing trends of urban sprawl. It is used for decades. However, analyses are usually carried out manually by selecting high-quality samples, restricted to small scale scenarios either temporarily limited or with low spatial or temporal resolution. To process a large amount of available remote sensing observations for a selected period, we propose a fully automated method to train an ensemble of neural networks, without the need to manually select samples. We consider two eras with three sites, each with at least 500km^2, and deep observation time series with hundreds and up to over a thousand of combined synthetic aperture radar (SAR) and multispectral optical observations. In order to train such large data sets, we apply data-parallel deep learning with Horovod and multiple NVIDIA Tesla V100 GPUs.

KEYWORDS

urban change detection, neural network, SAR, optical multispectral, deep-temporal

1 INTRODUCTION

Digital remote sensing-based land cover and land use (LCLU) change detection with multi-temporal satellite data has been studied for decades, dating back to the early 1970s [5]. Due to the reliance of repeat cycles and sensor technologies of satellite-based remote sensing it is sought after by urban change detection to understand socio-economic impacts and effects, identifying new settlements, monitoring urban growth, or analyzing trends of urban sprawl, just to name a few. For these fields, knowledge of how urban and man-made structures change over time is crucial and, depending on the exact application, can require high spatio-temporal resolution. In addition, automatic processing is required if a large scale analysis is desired to monitor larger areas over a longer time, and to make available the methods and results to a wider range of interdisciplinary users to thrive on remote sensing-based urban change detection and monitoring.

In our work [8], we consider four important properties: change detection method, remote sensing missions used, a large (deep) number of observations, and scaling of the method. Objectives of our work are the combination of both Synthetic Aperture Radar and multi-spectral optical for increased temporal resolution. Furthermore, multiple eras with different generations of remote sensing technologies are considered, such as 1991-2011 (ERS-1/2 and Landsat 5 TM) and 2017-2021 (Sentinel 1 & 2) to enable longer range monitoring. The training of the neural network model is done with synthetic but noisy labels to provide a fully automated process and to eliminate the need for expensive and time-consuming manual ground truth labeling. The use of publicly available level 1 remote sensing data provides maximum temporal and spatial coverage. Finally, resilient handling of partial and irregular observations is considered. Different from other work, we combine fully automated processes with high spatio-temporal observations from two different remote sensing types applied to two different technological generations (cf. [8]).

2 METHODOLOGY

We demonstrate two eras for which we obtained level 1 SAR and multi-spectral optical remote sensing data from ESA, USGS and SentinelHub. Almost all data from 1991-2011 and 2017-2021 is considered as long as co-registration is possible. In addition, optical remote sensing data is a subject of cloud coverage for which additionally available cloud masks are used.

Observations are further pre-processed to temporally stack, assemble, and tile. The former ensures that each real observation time point of all remote sensing types, such as multi-spectral optical and SAR, has an (approximate) representation, thus avoiding gaps. The assembling step combines bands from the different remote sensing types like optical and SAR polarization bands for every possible observation point. The final tiling breaks down the full Area of Interest (AoI) into manageable spatial samples (32x32 pixels).

In a second processing stage, these observations are windowed and labelled. In the windowing process, an observation period of one year (ERS-1/2 & Landsat 5 TM) or six months (Sentinel 1 & 2) is constructed, containing all observation points which happen within. Window starting times are randomly selected per tile, allowing overlapping windows but following different observation patterns —

*First and corresponding author

no two tiles have the same series of windows, which mitigates memorization effects.

All windows $\boldsymbol{w}^t_{i,j}$ are annotated with synthetically created labels $\hat{\boldsymbol{y}}^t_{i,j}$ used for supervised training. The synthetic labels for the supervised training are defined as (cf. [8])

$$\hat{\boldsymbol{y}}^t_{i,j} := s_{cm}(\boldsymbol{w}^t_{i,j}[b^{asc}_{SAR}], \boldsymbol{w}^t_{i,j}[b^{dsc}_{SAR}]) \cdot o_{cm}(\boldsymbol{w}^{t-\Delta}_{i,j}[b_{OPT}], \boldsymbol{w}^{t+\Delta}_{i,j}[b_{OPT}])$$

with

$$s_{cm}(\boldsymbol{p}^{asc}, \boldsymbol{p}^{dsc}) = \frac{O(\boldsymbol{p}^{asc}, \sigma, \eta) + O(\boldsymbol{p}^{dsc}, \sigma, \eta)}{2}$$

and

$$o_{cm}(\boldsymbol{p}^-, \boldsymbol{p}^+) = abs\Bigg(E\Big(\big(\boldsymbol{p}^-\big)_{mean^t}, \alpha, \gamma\Big) - E\Big(\big(\boldsymbol{p}^+\big)_{mean^t}, \alpha, \gamma\Big)\Bigg).$$

This synthetic label creation is realized with a combination of two state-of-the-art methods. $O(\ldots)$ is the Omnibus test statistic to identify changes in SAR data (Conradsen et al. [3]). $E(\ldots)$ is our extended version [8] based on the enhanced normalized difference impervious surfaces index (ENDISI, Chen et al. [2]).

We define a new architecture suitable for our needs with an ensemble of neural networks, where each sub-network is specialized for SAR or optical data (figure 1).

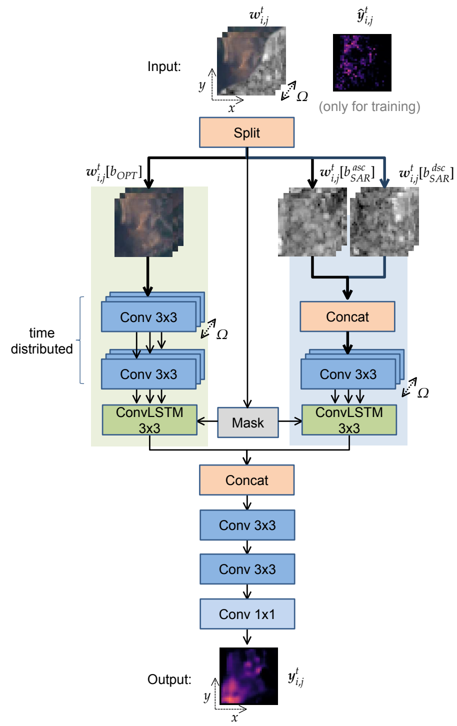

Figure 1: Ensemble neural network model architecture. Arrows indicate data flow of windows $\boldsymbol{w}^t_{i,j}$ for different data types (background green: multispectral optical, blue: SAR).

3 RESULTS

The two models, one for every era, are trained with different hyperparameters — mostly to accommodate the different remote sensing resolutions. The training was carried out on one compute node with four NVIDIA Tesla V100-SXM2 GPUs on the Barbora cluster at IT4Innovations[1]. Tensorflow 2.4 with Keras and Horovod [4] was used for data-parallel training on all GPUs. We used a batch size of 32 per GPU (effective batch size of 128), a learning rate of 0.004 (fixed over all epochs) and the distributed synchronous minibatch stochastic gradient descent (Sync-SGD) solver [1, 7] with a momentum of 0.8. The training data contained two different AoIs which are geographically different (Rotterdam and Limassol).

For inference, a third AoI (Liège), was selected to generate the urban change predictions. Some examples are shown in figure 2 for every era. These predictions either show a maximum of predictions over the entire era (center examples) or a series of six selected windows of the same area for every era.

An advantage of our method is the support of "real-time" monitoring where the most recently available observations can be used for change detection, omitting averaging or lead time. Even though our method is a fully automated process, exposed parameters enable control of the synthetic label generation and training. Furthermore, the combination of SAR and optical observations helps to increase temporal resolution, especially in areas where atmospheric effects hinder analysis. Finally, an ensemble of networks offers more control of each remote sensing type as well as use for just one type.

Some limitations of our current method apply. High reflective surfaces (e.g., white or metallic objects) induce large gradients, and incomplete cloud removal can lead to artifacts in predictions. Also moving objects (e.g., cars, containers) lead to false positives. Farming land can be wrongly detected as an impervious surface in case of droughts. Finally, a slight bias towards optical observations exists.

4 CONCLUSION

The supervised training with synthetic but noisy labels has shown good utility, which avoids the need to manually create ground truth labels to train the neural networks. The ensemble of mixing two data types showed synergies, but with a tendency towards relying more on optical observations. Our fully automated and parametric environment allows training and inference with little manual work. Especially in combination with services like Sentinel Hub, the necessary data is directly available and can be used for training or inference, omitting further processing. We make available our work on Github[2] including the training environments, trained models and data samples.

5 ACKNOWLEDGEMENT

This research was funded by ESA via the *Blockchain ENabled DEep Learning for Space Data* (BLENDED) project [6] (SpaceApps Subcontract No. 4000129481/19/I-IT4I), and by the Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPS II) project IT4Innovations excellence in science - LQ1602 and by the IT4Innovations infrastructure which is supported from the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140) via Open Access Grant Competition (OPEN-21-31).

[1] https://docs.it4i.cz/barbora/hardware-overview/

[2] https://github.com/It4innovations/ERCNN-DRS_urban_change_monitoring

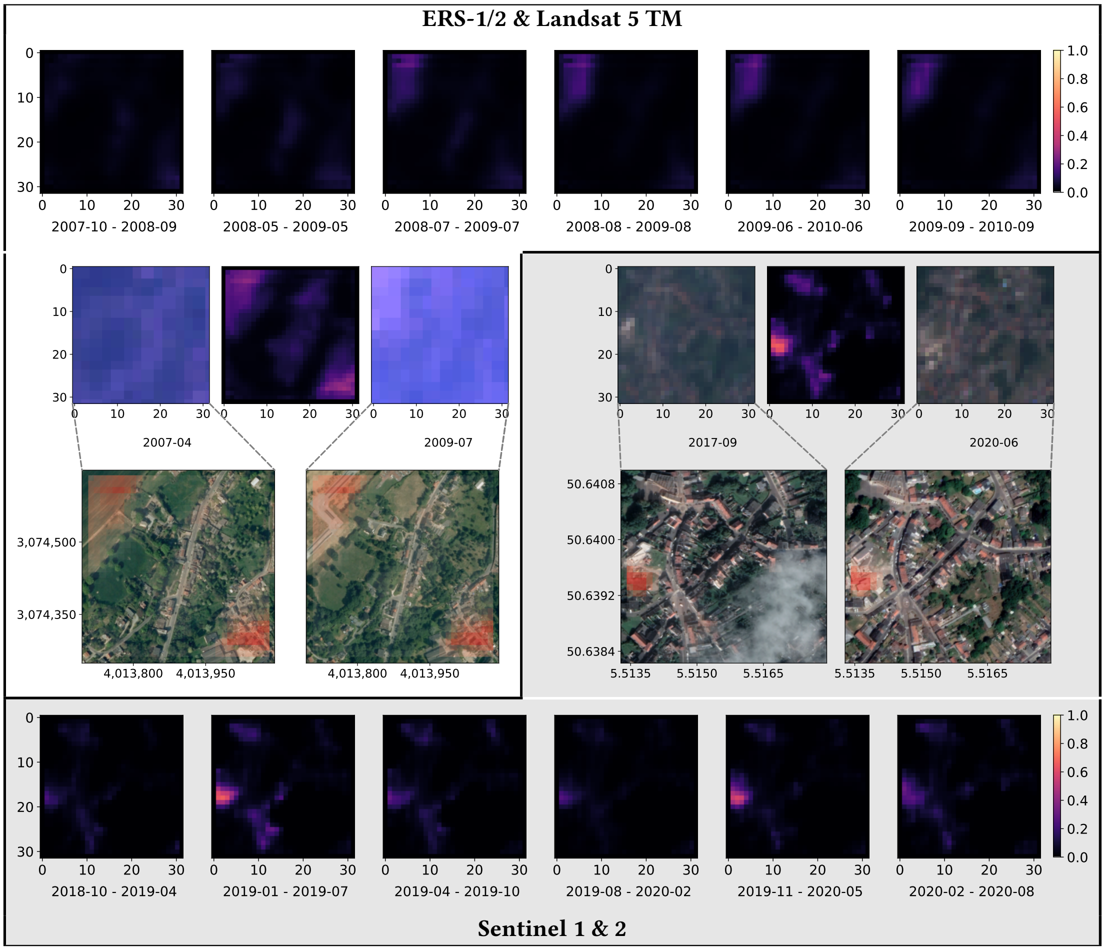

Figure 2: Detection examples for ERS-1/2 & Landsat 5 TM (top) and Sentinel 1 & 2 eras (bottom; gray background). For each, a detection series of six windows are shown on top and bottom, respectively. The center shows for both eras the true color optical observations at two points in time (left, right) with the maximum of all detected changes over the entire era (middle). Zoomed in are corresponding very-high-resolution imagery from Google Earth, © 2021 Maxar Technologies with detections superimposed in red. Higher values in detection outputs $y_{i,j}^t$ indicate a higher intensity of change.

REFERENCES

[1] Jianmin Chen, Rajat Monga, Samy Bengio, and Rafal Jozefowicz. 2016. Revisiting Distributed Synchronous SGD. In *Workshop Track of the International Conference on Learning Representations, ICLR*. publications/ps/chen_2016_iclr.ps.gz

[2] Junyi Chen, Kun Yang, Suozhong Chen, Chao Yang, Shaohua Zhang, and Liang He. 2019. Enhanced normalized difference index for impervious surface area estimation at the plateau basin scale. *Journal of Applied Remote Sensing* 13, 1 (2019), 1 – 19. https://doi.org/10.1117/1.JRS.13.016502

[3] Knut Conradsen, Allan Aasbjerg Nielsen, and Henning Skriver. 2016. Determining the Points of Change in Time Series of Polarimetric SAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 54, 5 (2016), 3007–3024. https://doi.org/10.1109/TGRS.2015.2510160

[4] Alexander Sergeev and Mike Del Balso. 2018. Horovod: fast and easy distributed deep learning in TensorFlow. *arXiv preprint arXiv:1802.05799* (2018).

[5] ASHBINDU SINGH. 1989. Review Article Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10, 6 (1989), 989–1003. https://doi.org/10.1080/01431168908903939 arXiv:https://doi.org/10.1080/01431168908903939

[6] Bernard Valentin, Leslie Gale, Hakim Boulahya, Betty Charalampopoulou, Charalampopoulou Christos K., Dimitris Poursanidis, Nektarios Chrysoulakis, Václav Svatoň, Georg Zitzlsberger, Michal Podhoranyi, Dušan Kolář, Vladimir Veselý, Ondrej Lichtner, Michal Koutenský, Dominika Regéciová, and Matús Múčka. 2021. BLENDED - USING BLOCKCHAIN AND DEEP LEARNING FOR SPACE DATA PROCESSING. In *Proceedings of the 2021 conference on Big Data from Space (JRC125131)*, Pierre Soille, Sveinung Loekken, and Sergio Albani (Eds.). Publications Office of the European Union, 97–100. https://doi.org/10.2760/125905

[7] Martin Zinkevich, Markus Weimer, Lihong Li, and Alex Smola. 2010. Parallelized Stochastic Gradient Descent. In *Advances in Neural Information Processing Systems*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta (Eds.), Vol. 23. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2010/file/abea47ba24142ed16b7d8fbf2c740e0d-Paper.pdf

[8] Georg Zitzlsberger, Michal Podhorányi, Václav Svato, Milan Lazecký, and Jan Martinovi. 2021. Neural Network-Based Urban Change Monitoring with Deep-Temporal Multispectral and SAR Remote Sensing Data. *Remote Sensing* 13, 15 (2021). https://doi.org/10.3390/rs13153000