Semantic Segmentation for Change Detection in Satellite Imaging

Kürşat Kömürcü, Linas Petkevicius

Vilnius University, Institute of Computer Science,
Didlaukio str. 47, LT-08303 Vilnius, Lithuania
kursat.komurcu@mif.stud.vu.lt, linas.petkevicius@mif.vu.lt

Abstract. Change detection is a common and actual problem in the field of remote sensing. The classical approaches using raw pixel information are very sensitive to noise. In this study we propose the usage of additional semantic information for change detection. We use the semantic segmentation methods like geospatial Segment Anything Model and encoder based U-Net to evaluate the predictions and tracing the semantic information as well as raw information in change detection. Later the multidimensional time series data is used via the Vector Autoregression model to predict the future changes in the landscape. The observations which fall out of the prediction interval are considered as the changes in the landscape. The proposed method is evaluated on the dataset of the random locations across the Baltic region. The research is accompanied by the data and reproducible code at Github repository[1].

Keywords: Deep learning, semantic segmentation, change detection, satellite imagery, Vector Autoregression

1 Introduction

Change detection is a problem to object variations by observing them over time. Satellite imagery is one of the domains where change detection is widely used. Remote sensing change detection is applied by using data from Earth-orbiting satellites and identifying the changes in the landscape structural permutations which have happened throughout a specific time track.

The change detection problem is analysed in classical mathematical modelling methods as algebraic analysis difference [1] or regression [2]. However, any methods based only on pixel intensity are very sensitive

[1] https://github.com/kursatkomurcu/semantic_segmentation_for-_change_detection_in_satellite_imaging

DOI: https://doi.org/10.15388/LMITT.2024.8

to noise. This common problem in computer vision, thus deep learning methods could be used to overcome this problem [3]. Deep learning methods for change detection could be divided into two groups based on different approaches. First approach uses 3-channel RGB images, while the second approach is based on multi-spectral imagery and contains more band information.

The RGB imagery approach is widely used in the field of remote sensing. The most popular method is based on U-Net [4] architecture. The U-Net architecture is used in the field of remote sensing for semantic segmentation [5] and change detection [6]. The U-Net architecture is also used in the field of remote sensing for change detection in combination with Siamese networks [7] demonstrating good results on cases like OSCD dataset [8]. Some recent methods apply deep learning models using bi-temporal images [9] or transformer based models [10].

The main challenges arise among approaches. The older satellite images do not enable them to form bi-temporal images. The bi-temporal images are formed by using the same place image taken at different times. The relief displacement is also a problem in remote sensing [6]. The relief displacement is a problem that occurs when the same object is imaged from different angles. Finally the seasonality of weather conditions is also a problem in remote sensing when climate conditions are changing, thus the same place could be imaged in different seasonality and weather conditions.

The addressing of challenges in remote sensing the deep learning methods are suitable choices. While some successful feature extraction methods could be used using end-to-end models [11] it raises the computational challenges. To address this we focus on the semantic segmentation methods. Very common approach is the usage of U-Net models modifications for semantic segmentation [12]. On the other hand, a recent successful model for semantic segmentation is Segment Anything Model [13] which was adopted to remote sensing imaging [14].

The most common non-commercial application of change detection falls under climate monitoring. For this current resolution of satellite imagery is rather sufficient. The 10 to 60 meter resolution is from Sentinel-2 [15]. The land site change detection for coastal zones is analysed [16].

The change detection is significantly impacted on noises and quality of images [8] or non stationary objects in the images like [17]. Thus it encounters the problems of joining datasets [18]. The class imbalance is also the common challenge [19, 20, 17, 21], mostly since background class in general is not changing [18]. Finally, clouds and noises are also a challenge [22] as increased saturation [17] or distorted colours [23].

The work is organised as follows. In Section 2, we discuss the proposed methodology. In Section 3, we discuss the dataset. In Section 4, we discuss the results. In Section 5, we discuss the conclusions.

2 Methodology

2.1 Semantic Segmentation

Semantic segmentation is a computer vision problem of assigning a class label to each pixel in an image from a predefined set of classes. Let's assume the input of format $X \in R^{c \times w \times h}$ of image consistent of tensor X with c - number channels, and width/height w, h respectively. The semantic segmentation mask $X \in R^{c \times w \times h}$ contain L number of classes, where each pixel is assigned to one of the classes. Such models could predict the class of each pixel in the image [12]. In our experiments we used the U-Net like model[2]. The pre-trained model had Building, Land, Road, Vegetation, Water and Unlabeled classes. For the generic segmentation models like Segment Anything Model [13] provide object mask prediction confidence score.

Upon completion of the segmentation process, the class probabilities for each pixel are aggregated to calculate average class probabilities for each image, forming a summarised representation of the segmentation outputs. To integrate these segmentation results into the VAR model, we construct feature vectors for each temporal pair of images by computing differences in the aggregated class probabilities between the two time points, thus capturing the changes in class distributions over time. These feature vectors are then weighted by the confidence scores derived from the segmentation phase, ensuring that the VAR model input emphasises data with higher predictive reliability.

2.2 Vector Autoregression

The vector autoregression (VAR) is a model used to capture the linear interdependencies among multiple time series data. The VAR model is a

2 https://github.com/ayushdabra/dubai-satellite-imagery-segmentation

generalisation of the univariate autoregressive model (AR) [24]. The VAR model is used to forecast tasks. The VAR model is defined as follows:

$$y_t = \beta + \sum_{i=1}^{p} \Omega_t Y_{t-i} + \epsilon_t$$

where y_t is a $k \times 1$ vector of endogenous variables at time t, β is a $k \times 1$ vector of bias, Ω is a $k \times k$ matrix of coefficients for i-th lag, p is the order of the VAR process, and ϵ_t is a $k \times 1$ vector of error terms at time t. The confidence interval of VAR models could be used either dynamic, or fixed. In our case use t distribution confidence interval which is the same for each time step. The critical value for the confidence level α, in our experiments we used α = 0.05. The VAR model was used using *Python* package *statsmodel*.

By transforming the class probabilities and confidence scores into a time-series format, we enable the VAR model to utilise these inputs effectively for forecasting landscape changes, ensuring that the transition from image data to predictive modelling is both smooth and logical.

3 Dataset

The investigation of change detections covered a wide range of diverse cases. We randomly chose 100 coordinates over the Baltic region (53,53100 - 59,69747 latitude values and 20,49722 - 28,22760 longitude) using uniform distribution. After that, we used COPERNICUS/S2 satellite in Google Earth Engine API to collect images of random chosen coordinates over the 2022 - 2023 time period. In our experiments, we used pixel intensities of B4, B3 and B2 bands which represent red, green and blue colours. For each coordinate, we made predictions using *geospatial* Segment Anything Model [12] and collected IOU and score values. Class probabilities are collected using the U-Net model. Cloud Probabilities collected using Google Earth Engine API. In such the dataset for each coordinate consist of 11 features of Raw Pixel Intensity of B4, Pixel Intensity of B3, Pixel Intensity of B2, IOU, Scores, Probabilities of 6 classes and Cloud Probabilities.

Table 1. Summary table

Index	Lat	Lon	RMSE	AIC	Fall In CI
0	57.9822	27.5759	0.085	-92.993	%0
1	54.8303	21.8945	0.118	-123.880	%0
2	59.1785	24.5851	0.058	-92.618	%0
3	57.2123	24.1739	<0.001	None	%99.363
4	55.4973	23.1317	0.089	-105.224	%0
5	59.5876	25.7885	0.053	-96.585	%0.746
6	57.2948	22.5929	0.093	-98.745	%0
7	53.6124	27.2380	0.080	-102.598	%0
8	54.6356	22.8023	0.096	-113.109	%0
9	56.6370	20.7791	0.112	-123.783	%0

Note: RMSE and Fall In CI columns are values for Class2

4 Results

For each point, we collected sentinel-2 RGB images using scale 10 zoom rate. Then, having the surrounding environment around the segmentation predictions was made using relevant models for each image. Such enables semantic information for each investigative pixel. After creating our dataset, we used VAR model for selected index and forecast $h = 12$ steps. The experiment we calculated root mean square error (RMSE), akaike information criterion (AIC) and confidence intervals for each feature using t distribution.

Figure 1 presents the general pipeline of approaches. The segmented image semantic information is added to vector time series models, thus while raw image data seems unchanged significantly, the semantic information allows an additional control mechanism for quality assessment. Cloud probability is often used to remove untruthful images, the same could be done by tracking unchanged situations. The illustrative case in Figure 1 can be seen for index 3 in the Table 1 above. Also one can see in Table 1 that some testing images have high variation in raw data or some data was not overlapped (black/empty image) over specific flight and 0 observations fell in the confidence interval.

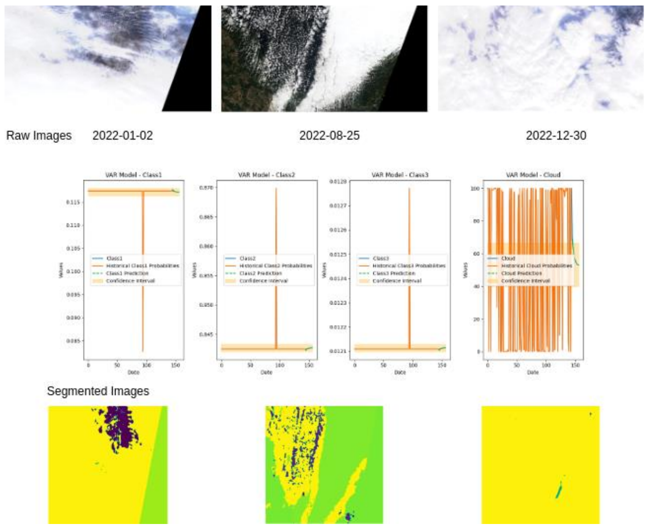

Figure 1. The illustrative example of confidence interval of prediction of the VAR model, which is used to detect the changes in the landscape.

5 Conclusions

In the study we investigate the estimation of non-changing temporal situations in satellite imagery. The publication proposes the addition of additional semantic information usage for tracking changes. The raw and semantic information modelled by vector auto-regressive models. The experiments demonstrated successful usage of the method. The identified change-detection cases are often related to data obscures of not visible areas. The publication is complemented with a reproducible repository of method pipeline in Github[3].

References

[1] Ke, L., Lin, Y., Zeng, Z., Zhang, L., & Meng, L. (2018). Adaptive change detection with significance test. IEEE Access, 6, 27442-27450.

[2] Ridd, M. K., & Liu, J. (1998). A comparison of four algorithms for change detection in an urban environment. Remote Sensing of Environment, 63(2), 95-100.

[3] https://github.com/kursatkomurcu/semantic_segmentation_for-_change_detection_in_satellite_imaging

[3] Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. Genetic Programming and Evolvable Machines, 19(1-2), 305-307.
[4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany.
[5] Peng, D., Zhang, Y., & Guan, H. (2019). End-to-end change detection for high-resolution satellite images using improved unet++. Remote Sensing, 11(11), 1382.
[6] Gong, J., Hu, X., Pang, S., & Li, K. (2019). Patch matching and dense crf-based co-refinement for building change detection from bi-temporal aerial images. Sensors, 19(7).
[7] Daudt, R. C., Le Saux, B., & Boulch, A. (2018). Fully convolutional siamese networks for change detection. Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece.
[8] Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018). Urban change detection for multispectral earth observation using convolutional neural networks. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain.
[9] Zheng, Z., Ma, A., Zhang, L., & Zhong, Y. (2021). Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. Proceedings of the IEEE/CVF International Conference on Computer Vision.
[10] Zhang, C., Wang, L., Cheng, S., & Li, Y. (2022). Swinsunet: Pure transformer network for remote sensing image change detection. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-13.
[11] Liu, J., Xuan, W., Gan, Y., Zhan, Y., Liu, J., & Du, B. (2022). An end-to-end supervised domain adaptation framework for cross-domain change detection. December 2022.
[12] Dabra, A., & Kumar, V. (2023). Evaluating green cover and open spaces in informal settlements of Mumbai using deep learning. Neural Computing and Applications, 1-16.
[13] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. arXiv preprint arXiv:2304.02643.
[14] Wu, Q., & Osco, L. P. (2023). Samgeo: A python package for segmenting geospatial data with the segment anything model (sam). Journal of Open Source Software, 8(89), 5663.
[15] Sentinel-2 overview. (2022).
[16] El-Asmar, H. M., & Hereher, M. E. (2011). Change detection of the coastal zone east of the Nile Delta using remote sensing. Environmental Earth Sciences, 62(4), 769-777.
[17] Tian, S., Zhong, Y., Zheng, Z., Ma, A., Tan, X., & Zhang, L. (2022). Large-scale deep learning based binary and semantic change detection in ultra-high resolution remote sensing imagery: From benchmark datasets to urban application. ISPRS Journal of Photogrammetry and Remote Sensing, 193, 164-186.
[18] Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2019). Multitask learning for large-scale semantic change detection. Computer Vision and Image Understanding, 187, 102783.
[19] Toker, A., Kondmann, L., Weber, M., Eisenberger, M., Camero, A., Hu, J., Hoderlein, A. P., Senaras, C., Davis, T., Marchisio, G., Zhu, X. X., & Leal-Taixe, L. (2022). DynamicEarthNet: Daily multispectral satellite dataset for semantic change segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA.
[20] Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., & Murphy, R. R. (2021). FloodNet: A high-resolution aerial imagery dataset for post-flood scene understanding. IEEE Access, 9, 89644-89654.

[21] Wang, J., Zheng, Z., Ma, A., Lu, X., & Zhong, Y. (2021). LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. October 2021.
[22] Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). Global land use/land cover with Sentinel 2 and deep learning. 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 2021.
[23] Schmitt, M., Hughes, L. H., Qiu, C., & Zhu, X. X. (2019). SENI2MS - A curated dataset of georeferenced multispectral Sentinel-1/2 imagery for deep learning and data fusion. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W7, 153-160.
[24] Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. Econometrica: Journal of the Econometric Society, 1371-1395.