Classifying objects from ALS-derived visualizations of ancient Maya settlements using convolutional neural networks

Maja SOMRAK, Research Centre of the Slovenian Academy of Sciences and Arts, Jožef Stefan International Postgraduate School, Slovenia
Žiga KOKALJ, Research Centre of the Slovenian Academy of Sciences and Arts, Centre of Excellence for Space Sciences and Technologies, Slovenia
Sašo DŽEROSKI, Jožef Stefan Institute, Slovenia

Keywords: *archaeology—deep learning—convolutional neural networks—digital elevation model—airborne laser scanning—lidar*

CHNT Reference: Somrak, Maja; Kokalj, Žiga, and Džeroski, Sašo. 2021. Classifying objects from ALS-derived visualizations of ancient Maya settlements using convolutional neural networks. Börner, Wolfgang; Kral-Börner, Christina, and Rohland, Hendrik (eds.), Monumental Computations: Digital Archaeology of Large Urban and Underground Infrastructures. Proceedings of the 24th International Conference on Cultural Heritage and New Technologies, held in Vienna, Austria, November 2019. Heidelberg: Propylaeum.
doi: 10.11588/propylaeum.747.

Introduction

Archaeologists engaging with remote sensing data rely heavily on manual inspection. This presents a major bottleneck in the data analysis pipeline, preventing them from keeping up with ever increasing data volumes, and creating a pressing need for computational methods that can automate data annotation and analysis.

Within archaeological community, critics of computational methods like to stress the superiority of human vision for identification of archaeological objects from remote sensing data. For the purposes of manual inspection, raw data is often converted to a representation that is most intuitive for visual interpretation. In this regard, the airborne laser scanning data (ALS) data used in archaeological prospection is usually visualized, e.g. the ALS-derived digital elevation model (DEM) represented as Visualization for Archaeological Topography (Kokalj and Somrak, 2019). Meanwhile, in the field of computer vision we see proliferation of deep convolutional neural networks (CNNs), which mimic the human visual system and achieve state-of-the-art performance on optical images. While there have already been applications of CNNs in remote sensing, only a handful of them concern ALS data in archaeological prospection (Verschoof-van der Vaart and Lambers, 2019; Trier et al., 2019).

We therefore propose a CNN-based method to distinguish among various types of anthropogenic objects in ALS visualizations. Development of such a CNN requires incorporating manually annotated data with thousands of samples. While manual annotation of data from our own ALS survey took several man-months, having in mind that thousands of square kilometers of similar data have already been collected in other surveys, further manual work can be minimized if the annotation process is automated.

Data and Methods

220 km² of ALS data have been collected around Chactún, one of the largest Maya urban centers known so far in the central lowlands of the Yucatan peninsula. More than 12,000 anthropogenic objects of three different types (building, platform, and aguada—artificial rainwater reservoirs) were manually annotated over 130 km² (hereinafter referred to as the analyzed area).

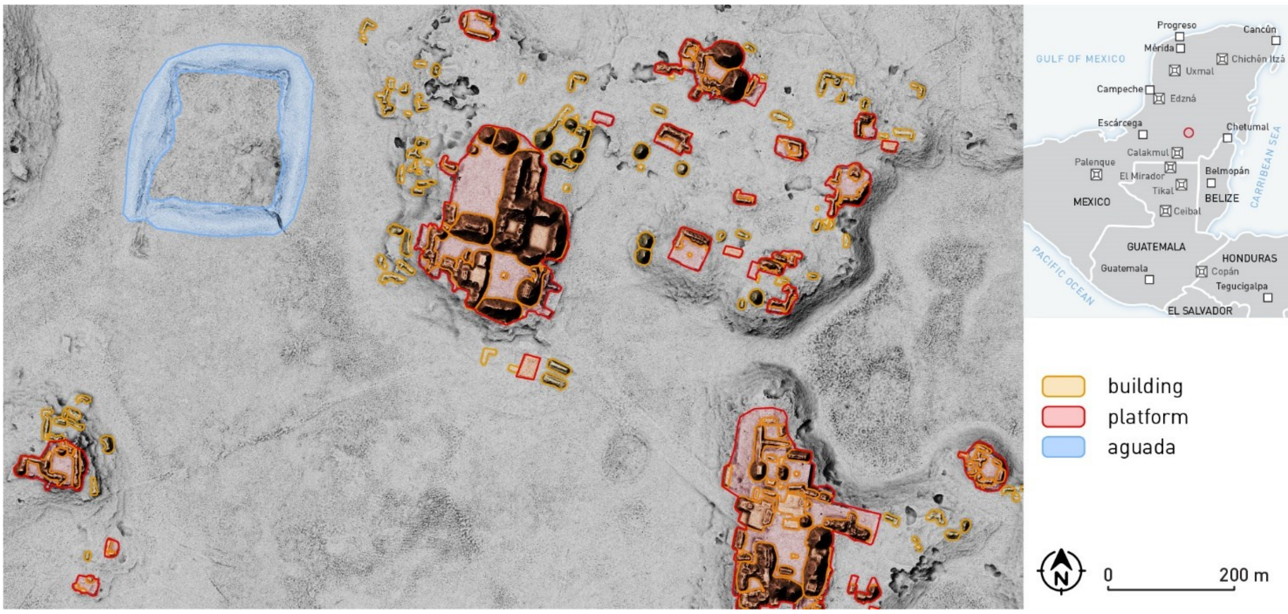

Fig. 1. Chactún with annotated data. (© M. Somrak)

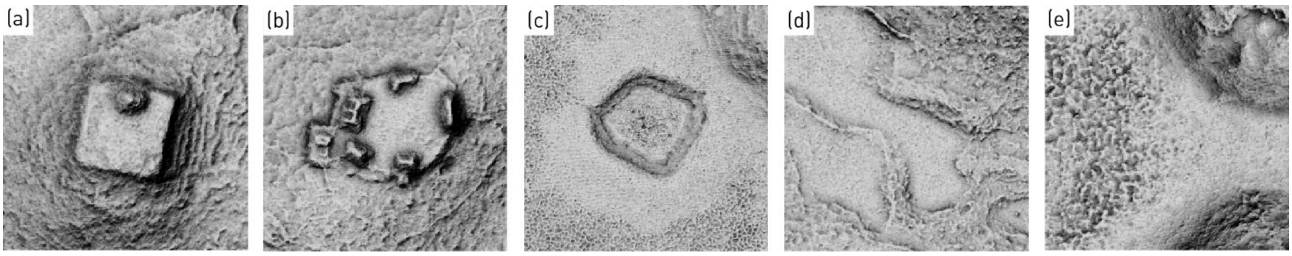

Fig. 2. Examples of generated data samples for different classes: a) platform, b) platform, c) aguada, d) terrain, e) terrain. Platforms on images a) and b) contain one and six buildings, respectively. (© M. Somrak)

Method development and implementation is split into several stages, of which the first is currently being developed and is presented in this paper. Advancements at every stage will build upon the previous to produce results, which are gradually closer to the final desired outcome—a model, which could potentially replace manual annotating. Method development includes: 1) Image classification *(classification of ALS visualizations of individual objects);* 2) Object recognition (*recognition and localization of different objects on ALS visualization of a larger area);* 3) Semantic segmentation *(location and exact boundaries of objects);* 4) Instance segmentation *(segments joined into individual instances, counting instances*); and 5) Repetition of previous implementation steps using elevation data (DEM) instead of ALS visualizations.

The analyzed area was split 80/20, to account for training and validation sets. Samples were generated for each class of anthropogenic objects; square bounding boxes were enlarged with a 30 pixels buffer. For negative samples, i.e. samples of natural terrain, it was assumed that areas of 128 × 128 pixels that do not contain any annotated objects can be considered. The total number of terrain samples equals the combined number of all anthropogenic samples (Table 1).

Class			Number of samples	Percentage of annotated samples from total area of 220 km^2	Class recall (at 4 classes)	Class precision (at 4 classes)
anthropogenic objects	structure	building	9303	~ 60 %	96 %	93 %
		platform	2110	~ 60 %	63 %	78 %
	aguada		173	100 %	79 %	70 %
other / natural	terrain		11,586	/	98 %	98 %

Table 1. Types and number of different objects annotated in our ALS survey data.

The VGG16 model (Simonyan and Zisserman, 2014) was used for transfer learning for the deep CNN model. VGG16 is a network with 16 convolutional layers and very small (3 × 3) convolutional filters. It has already been successfully applied in a few other remote sensing studies, where it obtained state-of-the-art results for some specific datasets (Wang et al. 2017). The VGG16 model used in this study was implemented in Python, using Keras and TensorFlow libraries, with modifications. Our dataset images were rescaled for input size of 128 × 128 pixels just before they were fed into the network. Three variations of the model were made to distinguish between two, three, or four different classes. Output of each of these model variations is one of the class values:

a) *terrain* or *structure*—where class *structure* is introduced as a joined class of buildings and platforms
b) *terrain*, *building* or *platform*
c) *terrain*, *building*, *platform,* or *aguada.*

Results

Model performance was tested for class recall, class precision and overall model accuracy. Accuracy was lower for two and three different classes and higher at four classes (Table 1), achieving overall accuracy of 94 % at four classes.

Discussion

The neural network's low performance on the aguada class is likely a result of its low sample count, therefore data augmentation should be considered to overcome the issue. All of the samples also have to be checked for accuracy of annotation. The overlap between the majority of building and platform objects resulted in the two classes' datasets containing very similar images. Probably as a result, the neural network did not so successfully distinguish between the two classes. Misclassified terrain samples were usually classified as structures (or buildings/platforms). After manual inspection of misclassified samples, it was evident that terrain samples did indeed contain anthropogenic objects that have not been annotated. This was a known potential issue from the beginning, as some objects—such as terraces, tracks, walls, quarries and other less common anthropogenic features were known to be present but were not annotated.

Conclusion

It was shown that CNNs can be successfully used to classify ALS-derived visualizations of different anthropogenic objects against natural terrain. Main misclassification issues have been identified and further work on data annotation, data augmentation, neural network fine-tuning and design modifications should further improve the classification accuracy.

Acknowledgements

This research was partly funded by the Slovenian Research Agency core funding No. P2-0406, and by research project No. J6-9395.

References

Kokalj, Ž. and Somrak, M. (2019). Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping. *Remote Sensing*, 11(7), p. 747.

Simonyan, K. and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*. Available at: http://arxiv.org/abs/1409.1556 (Accessed: June 29, 2019).

Trier, Ø.D., Cowley, D.C. and Waldeland, A.U. (2019). Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2), pp. 165–175.

Verschoof-van der Vaart, W.B. and Lambers, K. (2019). Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1), pp. 31–40.

Wang, J. et al. (2017). Transferring Pre-Trained Deep CNNs for Remote Scene Classification with General Features Learned from Linear PCA Network. *Remote Sensing*, 9(3), p. 225.