Effectiveness of DTM Derivatives for Object Detection Using Deep Learning

Bashir KAZIMI, Institute of Cartography and Geoinformatics, Leibniz University Hannover, Hannover, Germany
Katharina MALEK, Lower Saxony State Service for Cultural Heritage, Mining Archaeology, Goslar, Germany
Frank THIEMANN, Institute of Cartography and Geoinformatics, Leibniz University Hannover, Hannover, Germany
Monika SESTER, Institute of Cartography and Geoinformatics, Leibniz University Hannover, Hannover, Germany

Keywords: *Archaeology — Digital Terrain Models — Deep Learning — Object Detection*

Abstract

Deep learning models have achieved significant performances in identification and localization of objects in image data. Researchers in the remote sensing community have adopted such methods for object recognition in remote sensing data, specially raster products of Airborne Laser Scanning (ALS) data such as Digital Terrain Models (DTM). Small patches of larger DTMs, where pixels represent elevations, are cropped to train deep learning models. However, due to the variation in elevation values for the same object in two different regions, deep learning models either fail to converge or take a long time to train. To alleviate the problem, a local preprocessing step such as normalization to a fixed range or local patch standardization is necessary. Another solution is to first calculate other raster products where the pixel values are calculated based on the surrounding pixels within a certain range. Examples of such rasters are Simple Local Relief Models (SLRM), Local Dominance (LD), Sky View Factor (SVF), and Openness (positive and negative). In this research, the effect of using the aforementioned DTM derivatives are studied for detection of historical mining structures in the Harz Region in Lower Saxony. The well-known Mask R-CNN model is trained to produce bounding boxes, labels, and segmentation maps for each object in a given input raster.

Introduction

Deep neural networks achieve tremendous results in object detection. They work well with color images, but researchers have shown that elevation data obtained from airborne laser scanning can also be used by deep neural networks for detection of objects and relevant structures. One of the main products of airborne laser scanning data is Digital Terrain Model (DTM), which represents the ground surface as a rectangular grid where each pixel is assigned an elevation value. The elevation values are exploited to extract useful properties to infer the types, and shapes of different objects and structures in the terrain. Deep learning has been used to detect structures related to historical mining, and archaeology, among others (Kazimi, Thiemann, and Sester, 2019a, 2019b; Kazimi et al., 2018; Politz, Kazimi, and Sester, 2018). The pixel values in natural images range from 0 to 255, and for the deep learning models to converge, it is important to scale the images to have a smaller range, usually $[0,1]$ or $[-1,1]$. Scaling a large DTM globally in one of these ranges cause the values to have very small variations relative to their neighboring pixels and thus, it makes it hard for the model to learn. A common method is to divide the DTM into smaller grids and scale the values locally (Kazimi, Thiemann, and Sester 2019b). Other researchers used a derivative of the DTM called Simple Local Relief Model (SLRM) to train deep learning algorithms for detection of archaeological objects (Trier, Cowley, and Waldeland, 2019; Verschoof-van and Lambers, 2019). SLRM normalization removes the effect of absolute height differences and helps the model learn better. Other derivatives of DTM include Local Dominance (LD), Sky View Factor (SVF), and Openness (positive and negative), among others, each of which help visualize objects and structures in the terrain in a different manner (Kokalj and Hesse, 2017). The goal of this research is to find out which of these DTM derivatives help in automating detection of objects and structures. This article is organized to include a case study, results of the experiments and finally a summary and outlook for future research directions.

Case Study

In this study, the well-known architecture called Mask R-CNN by He et al. (2017) is trained to detect terrain structures related to historical mining in the Harz Region in Lower Saxony. The DTM data has a resolution of half a meter per pixel. Data statistics are are shown in Table 1.

Categories	Training (80 %)	Validation (10 %)	Test (10 %)	Total (100 %)
Bomb craters	909	113	113	1135
Charcoal kilns	836	104	104	1044
Barrows	1058	132	132	1322
Mining holes	2132	267	267	2666

Table 1. Data statistics. Four categories, examples of which are split to 80, 10 and 10 percent for training, validation and testing, respectively.

Using the Relief Visualization Toolbox (RVT) (Kokalj and Somrak, 2019), raster derivatives such SLRM, SVF, LD, and Openness (positive and negative) are calculated from the DTM. The model is trained using the original DTM, as well as its aforementioned derivatives separately, and the results are compared. Input data are patches of 128*x*128 pixels. The model is trained for 100 epochs with a batch size of 4 and the Stochastic Gradient Descent (SGD) optimization algorithm (Bottou, 2012). It is evaluated using the Mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 50 % (Everingham et al., 2010). The experiments are conducted using Python programming language, and Keras deep learning library (Chollet, 2015). Results of the experiments are shown in the next section.

Results

In this section, quantitative evaluation results are reported using the mAP value obtained on the test data by each model trained on different DTM derivative. Additionally, examples of predictions by each model are illustrated for qualitative analysis. Table 2 shows best mAP for each data type on the test data using best learned parameters for training and validation data during training.

	LD	SLRM	DTM	SVF	Openness (+)	Openness (-)
Training weights	61.7	**62.8**	59.6	54.1	50.8	50.7
Validation weights	**58.8**	56.0	54.5	45.0	40.2	42.8

*Table 2. Evaluation Results: Models trained on the DTM and its derivatives have been evaluated on the test set using the best parameters obtained during training based on training and validation data. Values show mAP at IoU threshold of 50 % where **bold** indicates best.*

As observed in Table 2, SLRM and LD helps achieve higher mAP scores while SVF and Openness (both positive and negative) scores are lower than that of original DTM. Table 3 shows examples of DTM for each category and the corresponding results and visualizations for each derivative examined in this research. The mAP scores are reflected in the illustrated examples. Even though the true positive rate for all the data types are similar, there are fewer false positives in the case of SLRM and LD than the others.

Summary

In summary, the effectiveness of different DTM derivatives for the purpose of object detection using deep learning has been investigated in this study. The mAP scores shown in Table 2 and the illustrations in Fig. 1 indicate that LD and SLRM are better derivatives, among those studied in this research. Further experiments are required to study the effect of different DTM derivatives for detection of other object categories and those in different regions, different data sources, and data resolutions.

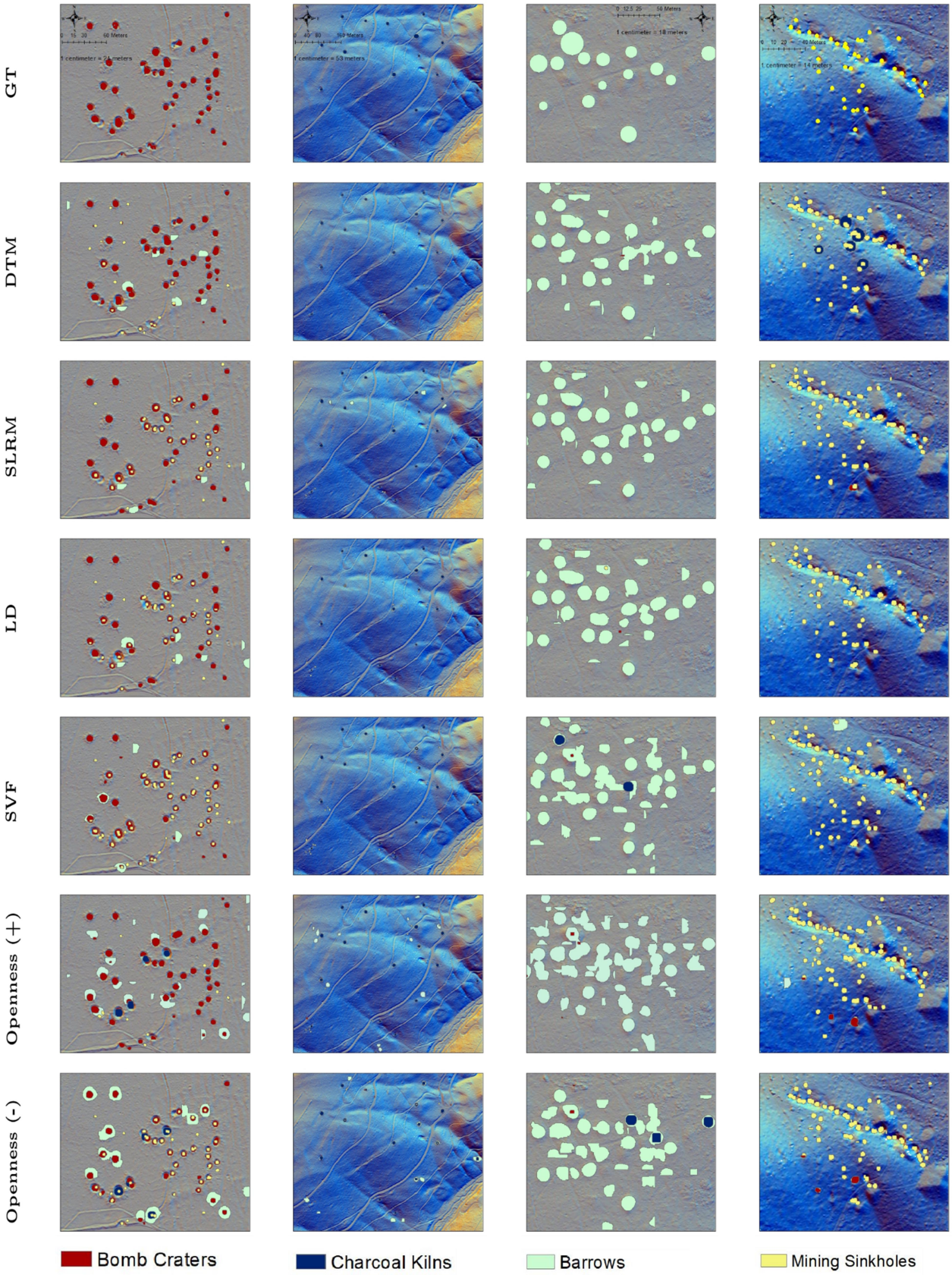

Fig 1. Detection results on four test regions containing the relevant structures. Each column illustrates the hill-shade relief of the region with ground truth (GT) labels and detection results by models trained on the DTM and its derivatives.

References

Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade* (pp. 421-436). Springer, Berlin, Heidelberg.

Chollet, F. (2015). keras.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, *88*(2), 303-338.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Kazimi, B., Thiemann, F., & Sester, M. (2019a). Object instance segmentation in digital terrain models. In *International Conference on Computer Analysis of Images and Patterns* (pp. 488-495). Springer, Cham.

Kazimi, B., Thiemann, F., & Sester, M. (2019b). Semantic Segmentation of Manmade Landscape Structures in Digital Terrain Models. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4.

Kazimi, B., Thiemann, F., Malek, K., Sester, M., & Khoshelham, K. (2018). Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data. In *2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage (COARCH 2018)* (Vol. 2230, pp. 21-35).

Kokalj, Ž., & Hesse, R. (2017). Airborne laser scanning raster data visualization: a guide to good practice (Vol. 14). Založba ZRC.

Kokalj, Ž., & Somrak, M. (2019). Why not a single image? Combining visualizations to facilitate fieldwork and on-screen mapping. *Remote Sensing*, *11*(7), 747.

Politz, F., Kazimi, B., & Sester, M. (2018). Classification of Laser Scanning Data Using Deep Learning. 38th Scientific Technical Annual Meeting of the German Society for Photogrammetry, Remote Sensing and Geoinformation 27. https://pdfs.semanticscholar.org/698c/924265e469d58eb6ffd7e561c2d2b 4814a06.pdf.

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., & 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).