Semi Supervised Learning for Archaeological Object Detection in Digital Terrain Models

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, Object Detection, Semi Supervised Learning, Convolutional Autoencoder

Abstract

Cultural heritage preservation is crucial for appreciating past human accomplishments, and learning from their actions. A step towards this goal is identification and registration of archaeological monuments. While experts are able to manually inspect archaeological sites and document their findings, different measures have been taken to automate the process while leveraging human expertise, leading to a more efficient documentation of interesting archaeological landscape structures and monuments. One effective method is using LiDAR data or one of its derivatives, such as Digital Terrain Models (DTMs). Archaeologists use these data to manually identify, label, and keep records of interesting monuments and structures. In an attempt to automate the process, Meyer et al. (2019) used LiDAR data and the eCognition tool by Trimble (2014) to detect monuments such as ridge and furrow areas, burial mounds, and Motte-and-Baily castles. To automate the process even further, labelled DTM data can be used by different techniques in artificial intelligence, specifically deep learning, to train a model that learns to distinguish distinct objects. The trained models can then detect similar objects and label them in regions not inspected manually by archaeologists.

Deep learning models can learn to classify, i.e., produce a label or category for a given segment of the DTM. It can also learn to categorize each point in the given DTM to a specified class. The former technique is referred to as classification while the latter is called semantic segmentation. Additionally, another technique called instance segmentation takes an input, and gives as output a bounding box, semantic segmentation mask, and a class label for each instance in the input. While all of the three techniques have proved to be effective, they highly depend on a large volume of labelled data to learn recognizing objects.

In our research, the task is to detect archaeological structures. In this work, we focus on examples of bomb craters, charcoal kilns, and barrows in the Harz mining region of Lower Saxony, Germany, which is home to the UNESCO world heritage site, “Historic Town of Goslar, Mines of Rammelsberg, and the Upper Harz Water Management System”. While a DTM for the whole state of Lower Saxony is available with a resolution of half a meter per pixel, only parts of the region of interest in Harz are labelled with known object categories. There are 38024 unlabelled DTM patches with a dimension of 2016x2016 pixels. The categories for which ground truth labels are available are listed in Table 1.

Category	Examples	Minimum Diameter	Average Diameter	Maximum Diameter
Bomb Craters	157	2.5 meters	6 meters	10 meters
Charcoal Kilns	1044	4 meters	11.5 meters	19 meters
Barrows	431	4 meters	17 meters	32 meters

Table 1. Data Statistics for labelled examples

Previous works by Politz et al. (2018), Trier et al. (2019), and Kazimi et al. (2018) show satisfactory results using the classification technique. Semantic segmentation also produces promising results, as discussed in the works of Kazimi et al. (2019a). Finally, instance segmentation technique applied on DTM data by Kazimi et al. (2019b) show comparable results. While these results are significant, they depend only on the limited amount of available labelled data. In the current research, a semi supervised deep learning technique, deep convolutional autoencoder, is used that makes use of unlabelled DTM data available for the whole state of Lower Saxony. This approach involves two steps. In the first step, referred to as unsupervised pre-training, a model is trained to learn to produce a compressed, lower dimensional representation of a given DTM input, and use this representation to reconstruct the original DTM. In the second step, supervised training, the model is further trained on the labelled data to learn labels for each point in the input, as in semantic

segmentation approach. In this approach, the model generalizes better, and learns to capture the data distribution from the unlabelled data well enough to create a distinct and representation for each input DTM segment (Erhan et al., 2010, Vincent et al., 2008). This is a more general representation than just using the labelled data, and hence, it improves the result of semantic segmentation in the second step of the approach. The architecture of the model is depicted in Fig. 1. This is the well-known DeepLabV3+ architecture by Chen et al. (2015). The preliminary results for the experiments in this research are shown in Fig. 2 and Table 2.

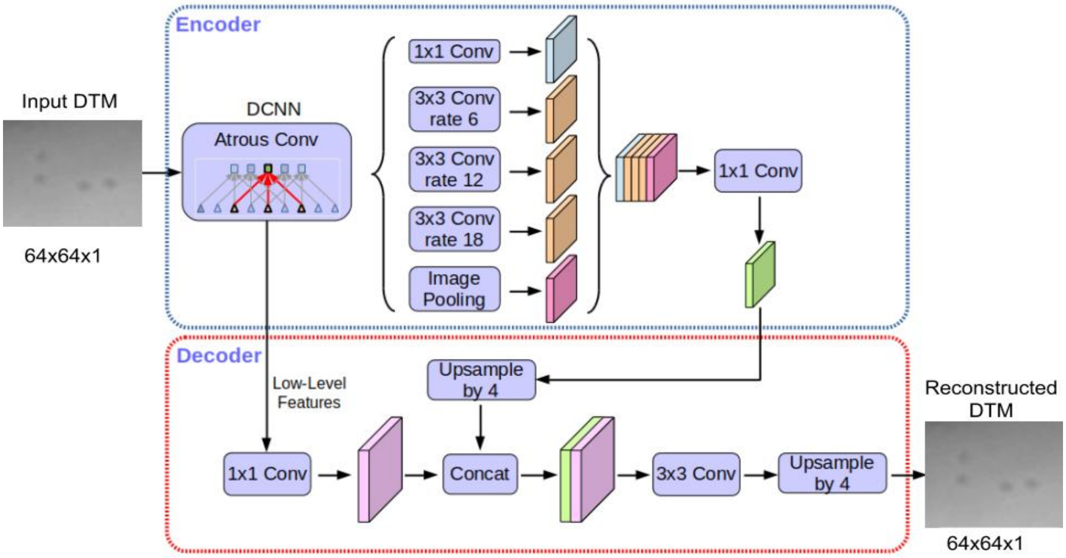

Fig. 1. Architecture of the autoencoder. After unsupervised training, the number of filters in the last convolutional layer is changed to be equal to the number of classes, and the activation function is changed from tanh to softmax. The model is then trained further to perform semantic segmentation and produce labels for each pixel in the input.

Category	Precision DLV3+	Precision Pretrained DLV3+	Recall DLV3+	Recall Pretrained DLV3+	F1-Score DLV3+	F1-Score Pretrained DLV3+	IOU DLV3+	IOU Pretrained DLV3+
Background	99	99	100	99	99	99	99	99
Bomb craters	84	**87**	67	**76**	74	**81**	75	**80**
Charcoal kilns	**94**	80	38	**70**	54	**75**	35	**54**
Barrows	**87**	84	**86**	83	**87**	84	**72**	71

Table 2. Evaluation results. DLV3+ is the original semantic segmentation model trained on 80 percent of the labelled data and evaluated on the remaining 20 percent. Pretrained DLV3+ is the model first used as an autoencoder and trained on the unlabelled data, and then trained as a semantic segmentation model similarly on 80 percent of the labelled data and evaluated on the same 20 percent. In the supervised training step, both models are trained for 100 epochs with the same hyperparameters. The numbers are all in percentage and the best results are in bold. Pretraining with unlabelled data improves model recall and Intersection over Union (IOU), and it is more significant in the case of charcoal kiln examples, which are smaller in area compared to bomb craters and barrows.

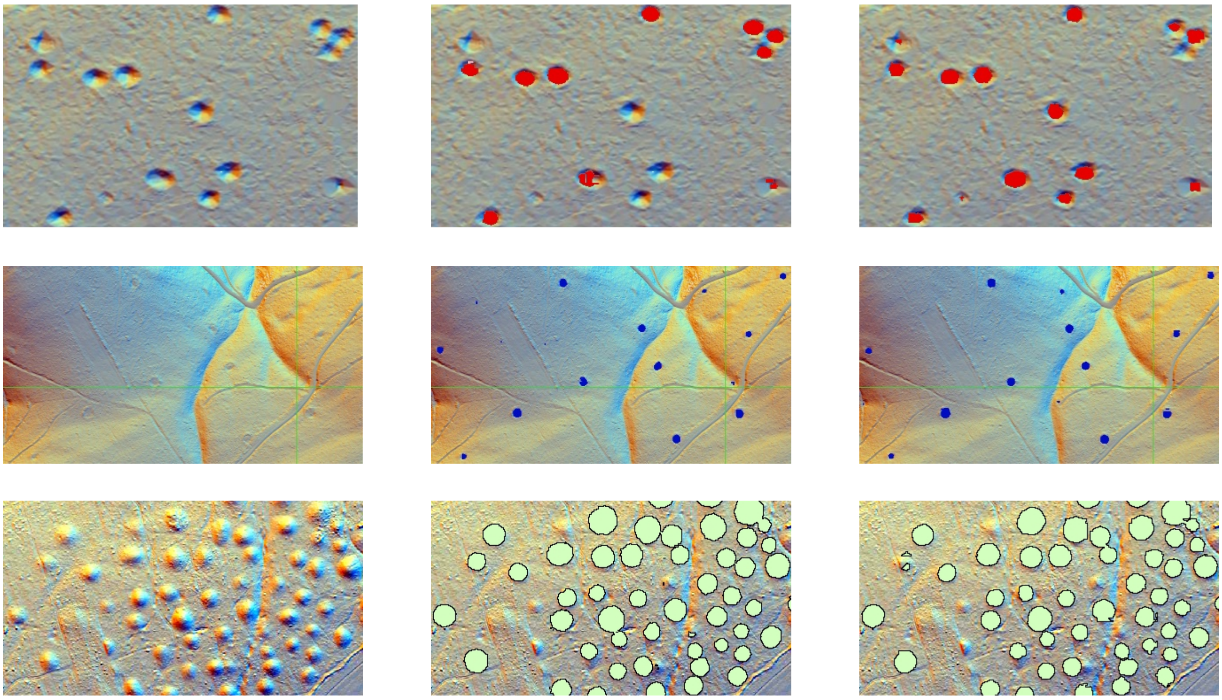

Fig. 2. Qualitative results. Rows represent examples of bomb craters, charcoal kilns and barrows, respectively. Columns indicate ground truth, prediction by DLV3+, and predictions by pretrained DLV3+, respectively.

References

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018). Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 801-818).

eCognition Developer, T. (2014). 9.0 User Guide. *Trimble Germany GmbH: Munich, Germany*.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P. and Bengio, S., (2010). Why Does Unsupervised Pre-training Help Deep Learning? *Journal of Machine Learning Research*, *11*(Feb), pp.625-660.

Kazimi, B., Thiemann, F. and Sester M. (2019a). Semantic Segmentation in Airborne Laser Scanning Data. *To appear at ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2019.*

Kazimi, B., Thiemann, F. and Sester M. (2019b). Object Instance Segmentation in Digital Terrain Models. *To appear at Proceedings of the 18th International Conference on Computer Analysis of Images and Pattern, CAIP 2018.*

Kazimi, B., Thiemann, F., Malek, K. and Sester M. (2018). Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data. *Proceedings of the 2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage co-located with 10th International Conference on Geographical Information Science (GIScience 2018), Alberto Belussi, Roland Billen, Pierre Hallot, Sara Migliorini (eds.), CEUR Workshop Proceedings,* (pp. 21-35). Available at http://ceur-ws.org/Vol-2230/paper_03.pdf

Politz, F., Kazimi, B. and Sester, M. (2018). Classification of Laser Scanning Data Using Deep Learning. In *PFGK18-Photogrammetrie-Fernerkundung-Geoinformatik-Kartographie, 37. Jahrestagung in München 2018* (pp. 597-610). Available at https://www.ikg.uni-hannover.de/fileadmin/ikg/Forschung/publications/49_PFGK18_P07_Politz_et_al.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).

Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.A. (2008). Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th international conference on Machine learning (pp. 1096-1103). ACM.*