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

Abstract: The use of deep learning techniques for detection of objects in imagery has spread to many disciplines, including archaeology. Deep learning models are exploited in detection of objects and structures in archaeology using natural and satellite images, as well as aerial and terrestrial laser scanning data. A well-known limitation of such models, specifically deep supervised models, is that they highly depend on large volumes of labelled data. For tasks with a small amount of labelled data and a huge amount of unlabelled data, unsupervised pretraining or transfer learning can be used. In this work, a product of airborne laser scanning data, i.e., Digital Terrain Models (DTM) is used to detect structures such as bomb craters, charcoal kilns, and barrows in the Harz region of Lower Saxony, Germany. Labels for only a small area are available while the majority of the region is unlabelled. Therefore, the large number of unlabelled examples are used to pretrain an auto-encoder model in an unsupervised fashion, and then a supervised training is performed using the labelled data. This combination of unsupervised learning and supervised learning is hereafter referred to as Semi Supervised Learning (SSL). Experiments in this study show that SSL helps gain up to 9 % improvement in performance compared to using supervised training alone.

Keywords: *Archaeology—Object Detection—Semi Supervised Learning—LiDAR—Digital Terrain Model*

CHNT Reference: Kazimi, Bashir; Malek, Katharina; Thiemann, Frank, and Sester, Monika. 2021. Semi Supervised Learning for Archaeological Object Detection in Digital Terrain Models. 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

Cultural heritage preservation is crucial for appreciating past human accomplishments and learning from their actions. A step towards this goal is the identification and registration of archaeological monuments. While archaeologists are able to detect archaeological sites in the field and document them, different measures have been undertaken to improve the process leading to more efficient documentation of interesting archaeological landscape structures and monuments. One effective method is using LiDAR data or one of its derivatives, such as 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 different objects. The trained models can then detect similar objects and label them in regions not inspected manually. In the next step, the proposed structures can be checked directly in the field.

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 this work, a large volume of unlabelled DTM data is leveraged by an unsupervised pretraining technique followed by a supervised training with a small amount of labelled data to detect archaeological objects. The focus of this research is the detection of bomb craters, charcoal kilns, and barrows in the Harz mining region of Lower Saxony, Germany. This region has been shaped by mining for thousands of years, which is represented by a huge amount of archaeological sites (Malek, 2019). It is also home to the UNESCO world heritage site, “Historic Town of Goslar, Mines of Rammelsberg, and the Upper Harz Water Management System” (Bergwerk Rammelsberg Altstadt Goslar Oberharzer Wasserwirtschaft, Goslar: Stadt Goslar, 2017). For the experiments presented in this paper, three types of structures, as previously mentioned, are chosen which are not only typical but also occur in large numbers. The rest of the paper is organized to include the proposed method and related works, experiments and results followed by a conclusion and hints towards future research directions.

Proposed Method

Previous works show successful applications of deep learning techniques using laser scanning data. Politz et al. (2018) use deep classification models to detect road segments and water bodies. Trier et al. (2019) detect archaeological structures such as kilns, mounds, cairns, and cattle feed stance, among others. Kazimi et al. (2018) use deep classifiers to detect streams, lakes, and road segments. Kazimi et al. (2019a) apply semantic segmentation on DTM data to identify man-made landscape structures. Finally, instance segmentation technique is used by Kazimi et al. (2019b) to retrieve pixel-wise labels as well as boundary lines for archaeological objects.

Methods explained above are examples of supervised learning where corresponding labels are present for each example. In our study, DTM data for Lower Saxony is available with labels for only a small portion of the region. Therefore, we make use of the unlabelled data by pretraining an auto-encoder model and then use the pretrained model for semantic segmentation with the labelled data. An auto-encoder model is a model that learns abstract representations of high dimensional inputs, encodes them to a compressed, low dimensional vector which is then used to reconstruct the original input. The advantage of such models is that the compressed encodings can be used for other learning tasks or visualization purposes. Maschi et al. (2011) used stacked convolutional auto-encoders for extracting features from images and using the features for object and digit recognition tasks.

Socher et al. (2011) used recursive auto-encoders for predicting sentiment distributions in textual data. Zhou and Paffenroth (2017) used auto-encoders for anomaly detection in data. In this study, we train an auto-encoder for feature extraction in DTM data. We then use the extracted features for semantic segmentation and producing pixel-level class labels for a given DTM input. The model used for auto-encoding and semantic segmentation in this research is the well-known architecture called Deeplab v3+ proposed by Chen et al. (2018). It is illustrated in Fig. 1.

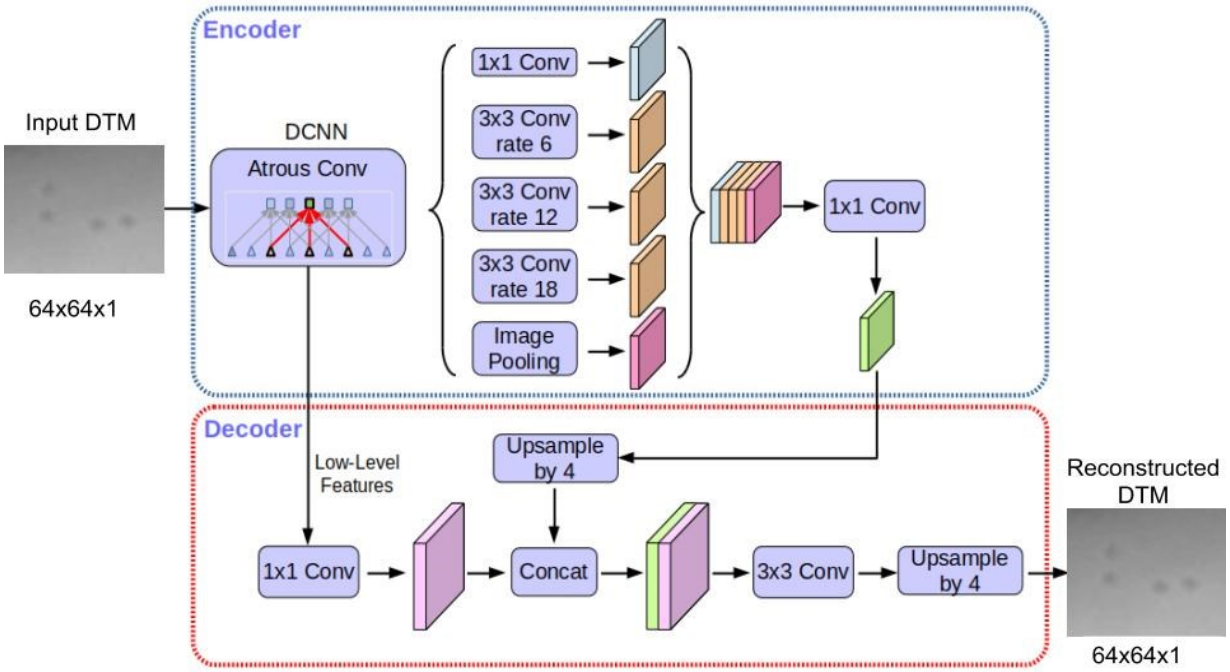

Fig. 1. The architecture of the model. It is first used as an auto-encoder with the unlabelled data. Then, the number of filters in the last convolutional layer is changed to match the number of object categories, and it is trained further with labelled data in order to perform semantic segmentation and produce labels for each pixel in the input DTM.

Deeplab v3+ is originally used for semantic segmentation in natural images, and it follows an encoder-decoder approach. The encoder extracts low-dimensional compressed features from images using combinations of multiple deep convolutional layers and spatial pyramid pooling (He et al., 2015). The features are then fed to the decoder which learns pixel-wise label maps for the given input by a few layers of bilinear upsampling and convolution, in addition to making use of low-level features extracted from the earlier layers of the feature extraction network.

The encoder-decoder-like architecture of Deeplab v3+ makes it suitable for us to use it in both stages of our approach, namely unsupervised pretraining with unlabelled data, and supervised semantic segmentation with limited labelled data. In the pretraining stage, the model is trained to learn a compressed representation of given DTM examples using the encoder and reconstruct the original DTM from the features using the decoder. Hence, the number of input channels and output channels in the network is the same (i.e., 1), and the model is optimized to minimize the mean squared error function. In the second stage, the learned encoder parameters are fixed, but the decoder is trained to use the extracted features by the encoder and generate segmentation maps for the given DTM. Therefore, the number of output channels depends on the number of categories or labels in the task at hand. In our study, we investigate 3 types of structures: bomb craters, charcoal kilns, and barrows, but we have an additional category for background pixels, leading to a total of 4 categories.

To evaluate the approach, in addition to running experiments on the unlabelled and labelled data with the proposed method, we also conducted experiments using the pure semantic segmentation approach trained only on the labelled data with randomly initialized parameters. The results of the experiments are given in the following sections in detail.

Experiments

Experiments are performed using the proposed workflow with DTM data for Lower Saxony. Details of the dataset and experiment set-ups for the SSL and pure supervised step are given in the following sections.

Dataset

The data used in this study is DTM data with a resolution of half a meter per pixel acquired from Lower Saxony, Germany. The DTM has a resolution of half a meter per pixel and covers an area of approximately 47000 square kilometres. Areas in the Harz mountains are labelled to indicate structures such as bomb craters, charcoal kilns and barrows. Statistics for labelled examples are shown in Table 1.

Category	Examples	Minimum Diameter	Average Diameter	Maximum Diameter
Bomb Craters	1135	2.5 meters	6 meters	10 meters
Charcoal Kilns	1044	4 meters	11.5 meters	19 meters
Barrows	1322	4 meters	17 meters	32 meters

Table. 1. Statistics for labelled examples.

For training the auto-encoder in the first step, random patches of size 64 × 64 pixels are extracted from the unlabelled regions. In the second step, i.e., in supervised training with labelled data, patches of 64 × 64 pixels are extracted such that they contain either a whole object (one of bomb craters, charcoal kilns, or barrows) or at least a quarter of the object. The data is divided into 80 %, 10 %, and 10 % split for training, validation, and testing.

Training Set-up

In this experiment, Python and Keras (Chollet, 2015) library are used for training and evaluation. Both models, auto-encoding and semantic segmentation are trained for 100 epochs each with input sizes of 64 × 64 pixels, Adam optimization with default arguments, and batch sizes of 32. The objective and metric function for the auto-encoder is mean squared error. The objective function for semantic segmentation model is categorical cross entropy, and the metric function is Intersection over Union (IoU).

Evaluation and Results

The SSL and Pure Supervised Learning (PSL) methods are both compared using the standard evaluation metrics for semantic segmentation, namely Intersection over Union (IoU). IoU, also referred to as Jaccard index, is defined as the percentage of overlap between true labels and those predicted by the deep learning models. It is a ratio of the number of common pixels between the true label map

and the predicted label map as shown in Equation 1. The values range from 0 (bad) to 100 (good) showing the percentage of overlap.

$$IoU = \frac{True\ Labels \cap Predicted\ Labels}{True\ Labels \cup Predicted\ Labels} * 100 \quad (Eq. 1)$$

The semantic segmentation model in the PSL, and the second stage of the SSL are both trained, validated, and tested on the exact same set of examples. During training, the parameters leading to the best IoU values for the validation data are saved to disk and used after training for evaluating the prediction results on the test data.

IoU results for SSL and PSL evaluated on the test set are shown in Table 2. Since the majority of the pixels in the given DTM patches are background pixels, the mean IoU could be quite deceptive. A model could produce a background label for all the pixels in an input and could still get a mean IoU of above 50 percent. Therefore, in addition to the mean IoU, we list individual IoU results for each class for better verification. It is clear from the IoU results that SSL improves detection performance in general, and more specifically for small structures like charcoal kilns.

Method	Mean IoU	IoU bomb craters	IoU charcoal kilns	IoU Barrows
PSL	67.8	**85.8**	15.5	71.4
SSL	**76.8**	85.2	**48.0**	**75.1**

Table 2. IoU on test set.

Qualitative evaluation results for both methods are illustrated in Fig. 2. Label maps predicted by SSL are smoother and more accurate while those by the PSL are sparse and less accurate, as seen in Fig. 2, especially in the second row. SSL predictions are more compact and a higher number of instances are captured, while predictions by the PSL are generally sparse and the number of undetected examples is higher.

Conclusions

In this research, the effect of semi-supervised learning is studied on the detection of archaeological objects in airborne laser scanning data. The method used is auto-encoder pretraining, followed by supervised fine-tuning. The architecture experimented with is the well-known Deeplab v3+ architecture which, due to its encoder-decoder property, is suitable to be used in both stages of the experiments, namely pretraining and fine-tuning. To evaluate the effect of such an approach in detecting patterns in DTM data, a parallel experiment is conducted using solely the supervised approach, and the results are compared. As observed in the IoU values in Table 2, and qualitative results in Fig. 2, semi-supervised learning improves IoU up to 9 percent. The results are especially significant for small structures like charcoal kilns.

Even though the IoU results prove that leveraging unlabelled data and applying semi-supervised learning techniques help to get better predictions, there is more room for improvement. First of all, the labelled examples are not perfectly created. Knowing the location and average diameter of the known structures, for simplicity, a distance buffer has been used to create circular polygons in ArcGIS marking instances of mentioned categories. The buffering method introduces a lot of false labels, i.e., many background pixels are labelled as one of the classes and for bigger instances of

the structures, some pixels are falsely labelled as background. Additionally, the completeness of the ground truth labels plays a role in the performance of the neural networks. Some instances in the training region that are not easily visible to the human eye have not been labelled, which could cause confusions to the model during training. Correcting the labelling problems will contribute to an increase in prediction accuracy of the models.

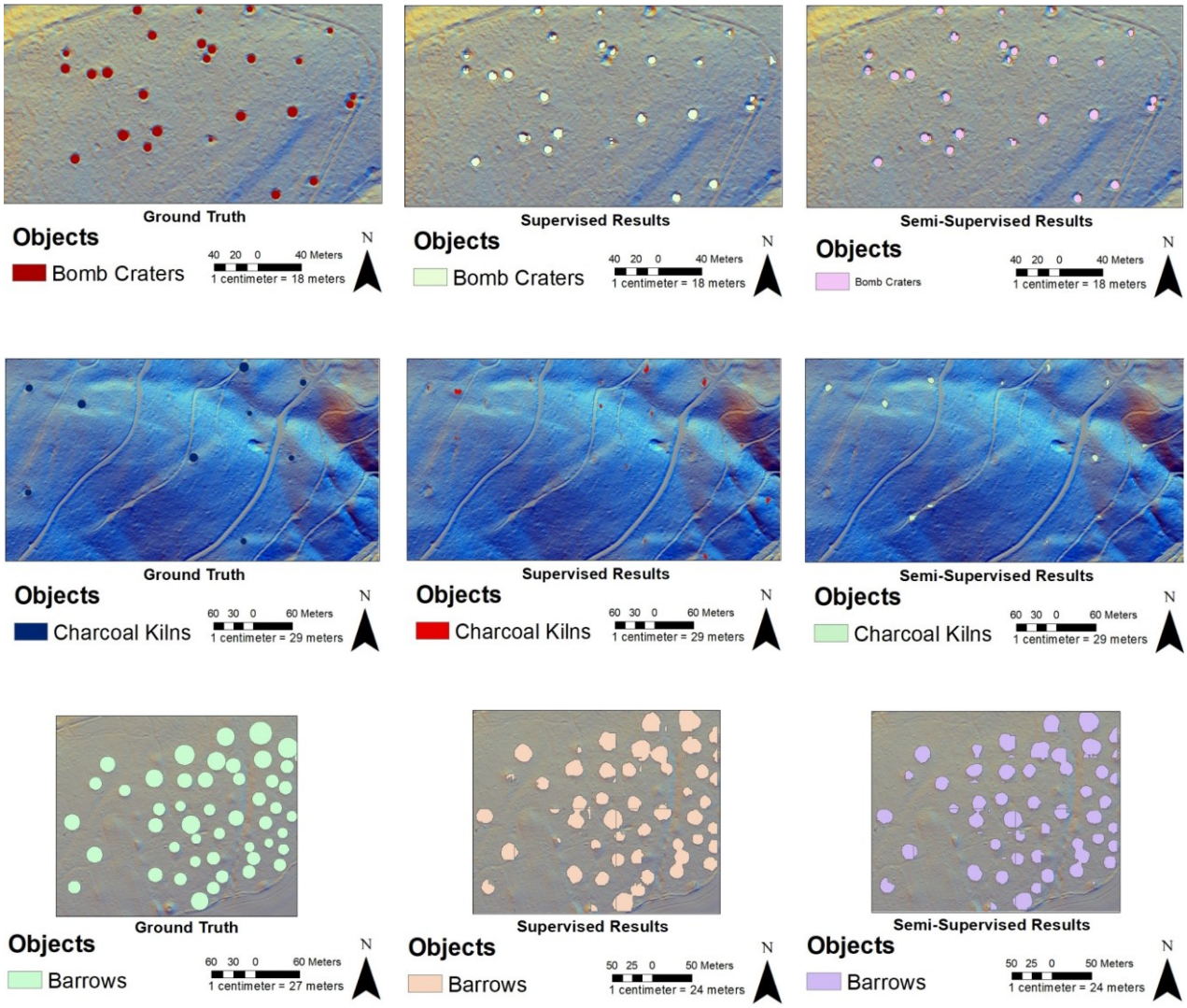

Fig. 2. Qualitative results. Rows represent examples of bomb craters, charcoal kilns and barrows, respectively. Columns indicate ground truth, prediction by PSL, and predictions by SSL, respectively.

Moreover, deep learning models usually work quite well with natural images since the range of values are fixed, i.e., pixels contain values in the range of 0 to 255. In DTM data, however, there is not a fixed range of values, and usual normalization techniques such as scaling input patches to a fixed range (e.g. 0 to 1 or -1 to 1) do not work well for generalizability of the trained model on unseen regions. Use of other raster data such as RGB hillshade, sky view factor, or local relief models may help improve performance since they contain values within a fixed range.

While the verification is taking place in the field, other future research directions include detection of more object classes and investigating SSL effects on tasks with more labelled examples. SSL could

also be used in combination with instance segmentation models such as Mask RCNN (He et al., 2017) where the predicted label maps are more fine-grained than those of semantic segmentation models.

References

Bergwerk Rammelsberg Altstadt Goslar Oberharzer Wasserwirtschaft. Goslar: Stadt Goslar, 2017.

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.

Chollet, F. (2015). Keras, Github. *https://github.com/fchollet/keras*

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.

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

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), pp.1904–1916.

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, and Sara Migliorini (eds.), CEUR Workshop Proceedings, (pp. 21–35). Available at http://ceur-ws.org/Vol-2230/paper_03.pdf

Malek, K. (2019). Der Harz – Ein Mittelgebirge aus bergbauarchäologischer Sicht. *Archäologie in Niedersachsen, 22*, (pp.43–50).

Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In International conference on artificial neural networks (pp. 52–59). Springer, Berlin, Heidelberg.

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

Socher, R., Pennington, J., Huang, E.H., Ng, A.Y., and Manning, C.D. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In Proceedings of the conference on empirical methods in natural language processing (pp. 151–161). Association for Computational Linguistics.

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.*

Zhou, C. and Paffenroth, R.C., (2017). August. Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 665–674.