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

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, especially 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.

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

CHNT Reference: Kazimi, B., Malek, K., Thiemann, F., and Sester, M. (2022). ‘Effectiveness of DTM Derivatives for Object Detection Using Deep Learning’, in Börner, W., Rohland, H., Kral-Börner, C. and Karner, L. (eds.) *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, held online, November 2020.* Heidelberg: Propylaeum.
doi:10.11588/propylaeum.1045.c14480

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 neighbouring 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 related work in deep learning and its applications in archaeology, contributions of this research as a case study, evaluation of results and finally a summary and outlook for future research directions.

Related Work

Deep learning methods have reached a considerable level of maturity on natural language processing and computer vision tasks for natural images. Recent indicators of this progress in natural language processing are Transformers (Vaswani et al., 2017) and other Transfer-based models, e.g. BERT (Devlin et al., 2019) and GPT-3 (Brown et al., 2020). Their counterparts in computer vision tasks are Image-GPT (Chen et al., 2020), HRNet (Sun et al., 2019; Wang et al., 2019), and EfficientNet (Tan and Le, 2019), among others. Inspired by this progress, researchers in other domains continuously find methods to incorporate deep learning techniques in their own research and projects. Archaeology is also among research fields that could make use of the progress in deep learning techniques. Verschoof-van and Lambers (2019) use deep learning to automatically detect archaeological objects in LiDAR data. The authors train the Faster R-CNN model (Ren et al., 2015) on a dataset of archaeological objects such as barrows, celtic fields, and charcoal kilns in the Netherlands. Trier, Cowley, and Waldeland (2019) use deep residual networks (ResNet) by He et al., (2016) to classify archaeological objects such as roundhouse, shieling hut, and small cairns in LiDAR data from Arran, Scotland. Other example applications of deep learning in archaeological research include the works of Bundzel et al. (2020), Gallwey et al. (2019) and Maxwell et al. (2020a; 2020b), among others.

Case Study

This research explores the use of object detection methods in archaeology using DTMs. Generally, deep learning models could be fed directly with DTM patches and be trained for automated detection of objects and structures of interest. DTMs contain elevation values for points on the terrain and do not have a fixed range of values, rendering them hard to perceive with the naked eye. For analysis and visualization, researchers use DTMs to generate relief rasters such as hillshade, SLRM, SVF, LD, Openness, slope and aspect, among others. The main concept behind this study is to see the impact of such relief rasters on detection of archaeological objects with deep learning. It is hoped that such rasters help deep learning models make better predictions, similar to how they help with visualization and analysis.

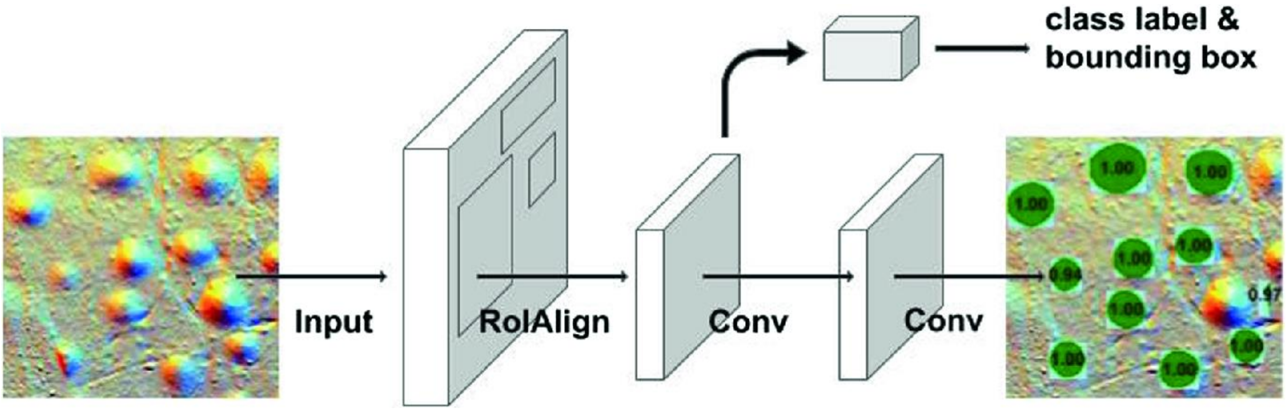

Fig. 1. Mask R-CNN architecture (© Kazimi et al., 2019a)

The well-known deep learning model 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. Mask R-CNN is designed to take input images and predict bounding box locations, segmentation masks and class labels for every object contained in the input. The architecture is illustrated in Figure 1.

The dataset is created from DTM data acquired from Lower Saxony, Germany and annotations for four structures including bomb craters, charcoal kilns, barrows and mining sinkholes. The DTM has a resolution of half a meter per pixel. Using the Relief Visualization Toolbox (RVT) (Kokalj and Somrak, 2019), relief rasters such as SLRM, SVF, LD, and Openness (positive and negative) are calculated from the DTM. Training examples of 128 × 128 pixels are cropped from the DTM and also the corresponding relief rasters from regions that contain an instance of the aforementioned four classes. The dataset is divided into training, validation and testing sets with splits of 80, 10 and 10 percent. Statistics for the annotated examples are shown in Table 1.

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

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

The model is trained using the original DTM, as well as the aforementioned relief rasters separately. For each raster type, the model is trained for 100 epochs with a batch size of 4 and the Stochastic Gradient Descent (SGD) optimization algorithm (Bottou, 2012). Random rotations and flipping are used as data augmentation techniques to avoid overfitting. Evaluations are performed 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 the DTM and the DTM derivatives or relief rasters. Additionally, examples of predictions by each model are illustrated for qualitative analysis. Table 2 shows best mAP for each input type on the test data using best learned parameters for training and validation data during training.

*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.*

	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

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. Figure 2 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, this research is conducted to see the impact of relief rasters created from DTMs on detection of archaeological objects. The main idea is that if relief rasters help humans in visualization and analysis of objects and structures, they should also help deep learning models perform better. Moreover, the study specially helps find the types of relief rasters more suitable for detecting structures such as bomb craters, charcoal kilns, barrows and mining sinkholes. The mAP scores shown in Table 2 and the illustrations in Figure 1 indicate that LD and SLRM are better derivatives, among those studied in this research. It is also concluded that relief rasters lead to better detection results compared to the original DTM. 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.

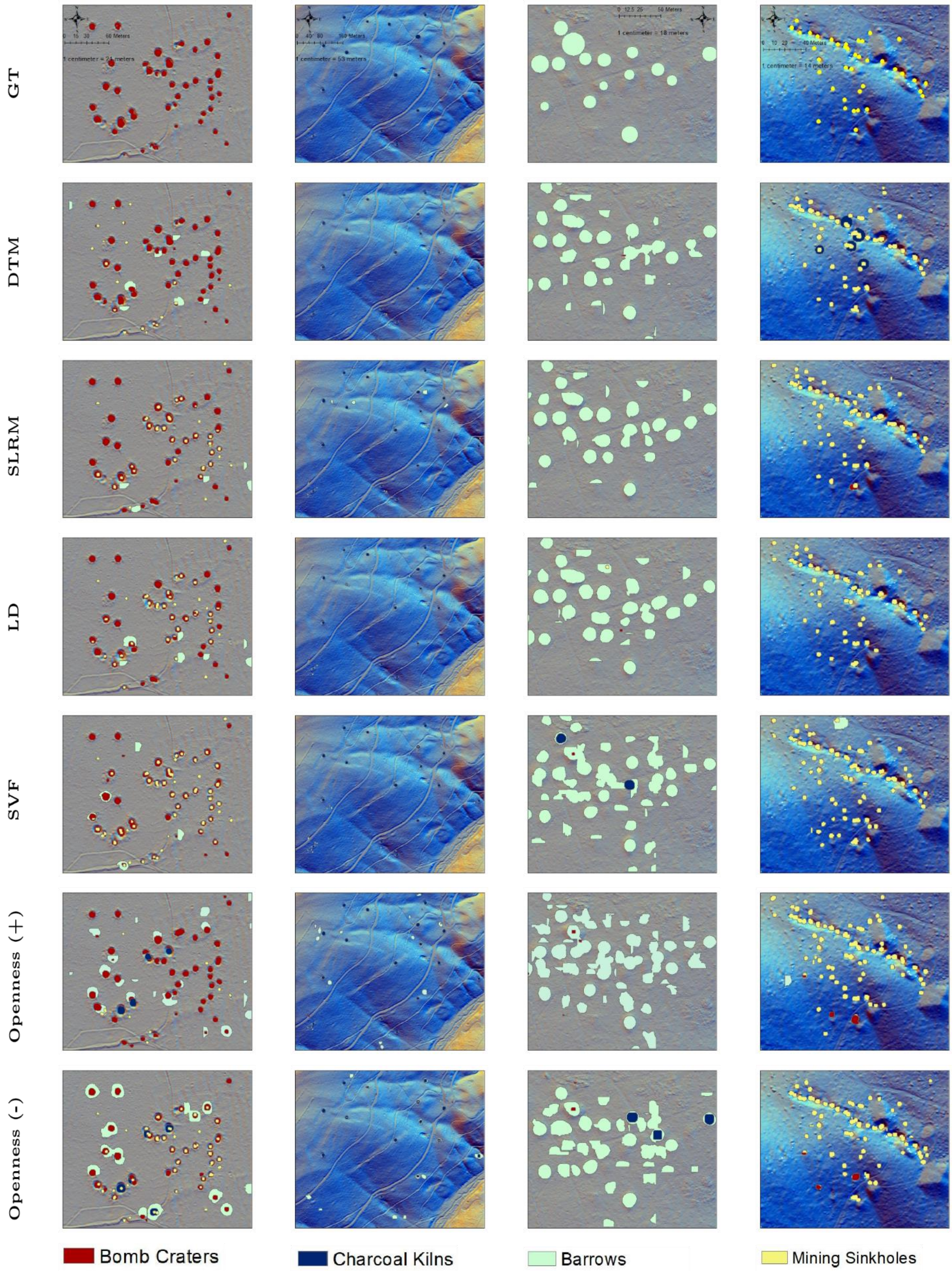

Fig 2. Detection results on four test regions containing the relevant structures. Each column illustrates the hill-shade relief of the region with ground truth (GT) labels and detection results by models trained on the DTM and its derivatives.

Funding

This research project is funded by the Ministry of Science in Lower Saxony.

Conflict of Interests Disclosure

All authors declare that they have no conflicts of interest.

Author Contributions

Conceptualization: Bashir Kazimi and Monika Sester
Data curation: Katharina Malek, Frank Thiemann, Bashir Kazimi
Formal Analysis: Bashir Kazimi
Funding acquisition: Monika Sester, Katharina Malek
Investigation: Bashir Kazimi, Katharina Malek
Methodology: Bashir Kazimi
Project Administration: Monika Sester, Katharina Malek, Frank Thiemann
Resources: Monika Sester, Frank Thiemann, Katharina Malek
Software: Bashir Kazimi
Supervision: Monika Sester
Validation: Bashir Kazimi
Visualization: Bashir Kazimi
Writing – original draft: Bashir Kazimi
Writing – review & editing: Monika Sester, Katharina Malek, Frank Thiemann

References

Bottou, L. (2012). 'Stochastic gradient descent tricks', in Montavon, G., Orr, G.B., Müller, K.R. (eds.) *Neural networks: Tricks of the trade.* Berlin, Heidelberg: Springer, pp. 421–436.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askel, A. et al. (2020). 'Language Models are Few-Shot Learners', *arXiv:2005.14165.*

Bundzel, M., Jaščur, M., Kováč, M., Lieskovský, T., Sinčák, P., and Tkáčik T. (2020). 'Semantic Segmentation of Airborne Lidar Data in Maya Archaeology', *Remote Sensing*, 12(22), p. 3685. [Online]. doi:10.3390/rs12223685.

Chen, M., Radford, A., Child, R., Wu, J., Jun, H., Dhariwal, P., Luan, D., and Sutskever, I. (2020). 'Generative Pretraining From Pixels', in *Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research 119,* pp. 1691–1703. [Online]. Available at: https://proceedings.mlr.press/v119/chen20s.html.

Chollet, F. (2015). 'Keras. Deep Learning for Humans'. Available at: https://github.com/keras-team/keras.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). 'Pre-training of Deep Bidirectional Transformers for Language Understanding', in Burstein, J., Doran, and C., Solorio, T. (eds.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* Vol. 1. Minneapolis: Association for Computational Linguistics, pp. 4171–4186.

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

Gallwey, J., Eyre, M., Tonkins, M., and Coggan, J. (2019). 'Bringing Lunar LiDAR Back Down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning', *Remote Sensing,* 11(17), p. 1994. [Online]. doi:10.3390/rs11171994.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). 'Mask r-cnn', in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988. [Online]. doi:10.1109/ICCV.2017.322.

He K., Zhang, X., Ren, S., and Sun, J. (2016). 'Deep residual learning for image recognition', in *Proceedings of the IEEE international conference on computer vision*, pp. 770–778. [Online]. doi:10.1109/CVPR.2016.90.

Kazimi, B., Thiemann, F., and Sester, M. (2019a). 'Object instance segmentation in digital terrain models', in Vento, M., Percannella, G. (eds.) *Computer Analysis of Images and Patterns. CAIP 2019. Lecture Notes in Computer Science, vol. 11679. Cham:Springer.* doi:10.1007/978-3-030-29891-3_43.

Kazimi, B., Thiemann, F., and Sester, M. (2019b). 'Semantic Segmentation of Manmade Landscape Structures in Digital Terrain Models', *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, IV-2/W7, pp. 87–94. [Online]. doi:10.5194/isprs-annals-IV-2-W7-87-2019.

Kazimi, B., Thiemann, F., Malek, K., Sester, M., and 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)*, *CEUR Workshop Proceedings,* Vol. 2230, pp. 21–35. [Online]. Available at: http://ceur-ws.org/Vol-2230/paper_03.pdf.

Kokalj, Ž. and Hesse, R. (2017) Airborne laser scanning raster data visualization: a guide to good practice. Ljubljana: Založba ZRC (Prostor, kraj, čas, 14). doi:10.3986/9789612549848.

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. [Online]. doi:10.3390/rs11070747.

Maxwell, A.E., M.S. Bester, L.A. Guillen, C.A. Ramezan, D.J. Carpinello, Y. Fan, F.M. Hartley, S.M. Maynard, and J.L. Pyron. (2020). 'Semantic segmentation deep learning for extracting surface mine extents from historic topographic maps', *Remote Sensing*, 12(24), pp. 1–25. doi:10.3390/rs12244145.

Maxwell, A.E., P. Pourmohammadi, and J. Poyner. (2020). 'Mapping the topographic features of mining-related valley fills using mask R-CNN deep learning and digital elevation data', *Remote Sensing*, 12(3), pp. 1–23. [Online]. doi:10.3390/rs12030547.

Politz, F., Kazimi, B., and 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*. [Online]. Available at: https://pdfs.semanticscholar.org/698c/924265e469d58eb6ffd7e561c2d2b 4814a06.pdf.

Ren, S., He, K., Girschick, R.B., and Sun, J. (2015). 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.) *Advances in Neural Information Processing Systems 28*. [Online]. Available at: https://papers.nips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

Tan, M., and Le, Q., (2019). 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Network', in Chaudhuri, K., Salakhutdinov, R., *Proceedings of the 36th International Conference on Machine Learning,* PMLR 97, pp. 6105–6114. [Online]. Available at: http://proceedings.mlr.press/v97/tan19a/tan19a.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.

Vaswni, A., Shazeer, N., Parmer, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). 'Attention Is All You Need', in Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.) *Advances in Neural Information Processing Systems 30,* pp. 5998–6008*.* [Online]. Available at: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

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. [Online]. doi:10.5334/jcaa.32.