Detection of cultural heritage in airborne laser scanning data using Faster R-CNN

Results on Norwegian data

Øivind Due, TRIER, Norwegian Computing Center, Norway[1]
Kristian LØSETH, The Directorate of Cultural Heritage in Norway, Norway

Abstract: A new processing chain for automated archaeological mapping from airborne lidar data is proposed. First, the lidar data was converted to a detailed digital terrain model (DTM), which was then converted to a local relief model (LRM) in which cultural heritage objects may be visible.

Simple faster R-CNN was used as the basis for the detection method. This deep neural network was pre-trained on the ImageNet labelled image database. Additional training was done on LRM images containing known locations of grave mounds, pitfall traps and charcoal kilns.

The classification performance was 87 % consumer's accuracy on a test set not seen during training. At the same time, the producer's accuracy was 75 %. However, all the test set images contained at least one cultural heritage object. In most landscapes, the majority of image patches of the same size may contain no cultural heritage objects visible in the DTM. Thus, the estimated producer's accuracy of 75 % may be too optimistic. On the other hand, the number of false positives appeared to be low on the Øvre Eiker unlabelled test data. In conclusion, it was demonstrated that faster R-CNN is well suited, in terms of consumer's accuracy, for automated detection of cultural heritage objects such as charcoal kilns, grave mounds and pitfall traps in high resolution airborne lidar data. However, one may expect that the method must be improved in terms of producer's accuracy in order to limit the number of false positives when applied on large areas for detailed archaeological mapping.

Keywords: *grave mounds—hunting systems—charcoal kilns—automated detection— lidar—Norway*

CHNT Reference: Øivind Due Trier, Kristian Løseth. 2021. Detection of cultural heritage in airborne laser scanning data using Faster R-CNN. 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

The goal of this research was to develop automated tools for improving the cultural heritage mapping in Norway. The existing cultural heritage mapping in Norway is incomplete. Some selected areas are

[1] Author's address: Øivind Due Trier, Norwegian Computing Center, Gaustadalléen 23 A, P.O. Box 114 Blindern, NO-0314 Oslo, Norway; email: trier@nr.no.

mapped well, while the majority of areas only contain chance discoveries, often with bad positional accuracy.

A note on terminology: This work is multidisciplinary in the overlap between (1) computer vision and pattern recognition, (2) archaeology and cultural heritage, and (3) geographic information systems. These disciplines use the terms ‘object’, ‘artefact’ and ‘feature’ in different ways. In this paper, the term ‘object’ is used to denote something that one wants to detect and map.

Automated methods for detecting some types of cultural heritage objects from airborne laser scanning (ALS) data have previously been developed (Trier, Salberg, and Pilø, 2018). These have contributed to increasing the number of areas that are mapped well. However, the methods have a number of issues that have prevented them from being used systematically on all available ALS datasets. Template matching has been used to map pitfall traps of deer hunting systems, charcoal burning pits of iron extraction sites (Trier and Pilø, 2012) and Iron Age grave mounds (Trier, Zortea and Tonning, 2015). However, the number of false positives was high, and varied between different types of landscape. A deep neural network has been used to detect charcoal kilns (Trier, Salberg and Pilø, 2018), performing better than template matching both in terms of reduced false positive rate and reduced false negative rate, but the method was very slow.

All of Norway will soon be covered by ALS data for the purpose of creating a new national elevation model. The Norwegian Directorate for Cultural Heritage wants to use this opportunity to obtain a more complete and accurate mapping of cultural heritage in the landscape. The focus is on Iron Age grave mounds and deer hunting systems, as these are automatically protected by Norwegian law due to their age. The protection by law also applies for such sites even if they are not yet properly mapped; thus, a complete and accurate mapping is needed to manage the protection by law.

The following challenges were identified:

1. develop an automated processing chain,
2. reduce processing time,
3. reduce the number of false positives and false negatives, and
4. develop detection methods that may be applied on all Norwegian landscapes.

A recent development in deep neural networks for object detection in natural images is the region-proposing convolutional neural network (R-CNN; Girshick et al., 2014), which may also be used for cultural heritage detection in ALS data. Verschoof-van der Vaart and Lambers (2019) use Faster R-CNN (Ren et al., 2017) to detect prehistoric barrows and Celtic fields in ALS data from the Netherlands.

Data

Three types of cultural heritage objects were used in this study (Fig. 1): grave mounds from the Iron Age (approx. 500 BC–1000 AD), pitfall traps from deer hunting systems (approx. 1–1500 AD) and charcoal kilns (approx. 1550–1900 AD). These three types of cultural heritage were selected since they are numerous in the Norwegian landscape, and thus suitable candidates for automated mapping.

ALS point cloud data was downloaded from http://hoydedata.no. This internet site provides free access to all ALS data in Norway. Vector maps of known locations of grave mounds and pitfall traps were provided as ESRI shape files by the Norwegian Directorate for Cultural Heritage. These data were extracted from the national cultural heritage database, which may also be accessed from https://kulturminnesok.no. This internet site provides search and view functionality. Vector maps of charcoal kiln locations were provided by Oppland County Administration. None of the vector maps are freely available.

The data were split into three parts, named 'training', 'validation', and 'test' (Table 1). The neural network parameters were learned from the training data iteratively by minimising a loss function. The validation data were used to select the best set of neural network parameters. The test data were then used to estimate detection performance on data not seen during training.

Methods

Preprocessing

The ALS data consists of a large number of individual (x, y, z) points, each being labelled with 'ground' or 'other', and also whether it was a first return from a laser pulse. By keeping only the points labelled 'ground', one may create a detailed digital terrain model (DTM) in which it is possible to see, e.g., cultural heritage structures that would otherwise be hidden by tree vegetation (Fig. 2).

The ALS point cloud data were converted to a digital terrain model (DTM) with 0.25 m pixel spacing. The DTM was converted to a simplified local relief model (LRM; Hesse 2010) by subtracting a smoothed version of the DTM. The LRM enhances local elevation differences while suppressing the general landscape topography. Thus, cultural heritage objects including grave mounds, pitfall traps and charcoal kilns may be visible (Fig. 3).

For each cultural heritage object in the vector data, a 150 m×150 m image was extracted from the LRM. In order to mimic practical use of the detection method, where the object's location is not known in advance, the image centre was moved by a random distance in both x and y, with the restriction that the entire object still be within the image. For some 150 m×150 m images, there were more than one cultural heritage object clearly visible. All these were included in the image annotations.

Increasing the size of the training set

A common problem in automatic object detection is to obtain a sufficiently high number of training examples. With too few training examples, the neural network may perform badly on data not seen during training. A common trick to provide eight times as many training examples is to create rotated (0, 90, 180 and 270 degrees) and flipped versions. This was done for the training and validation subsets, but not for the test subset.

Fig. 1. Top: a grave mound in Norway's largest Viking Age grave field at Vang, Oppdal municipality, Trøndelag County. Middle: a pitfall trap, Nord-Fron municipality, Oppland County. Photo: Lars Holger Pilø, Oppland County Administration. Bottom: Charcoal kiln, Lesja municipality, Oppland County.

	number of objects						
object type	training		validation		test		sum
charcoal kiln	773	73 %	190	18 %	95	9 %	1058
grave mound	545	52 %	302	29 %	199	19 %	1046
pitfall trap	613	41 %	565	38 %	303	20 %	1481
sum	1931	54 %	1057	29 %	597	17 %	3585

Table 1. Summary of vector data used for neural network training and evaluation.

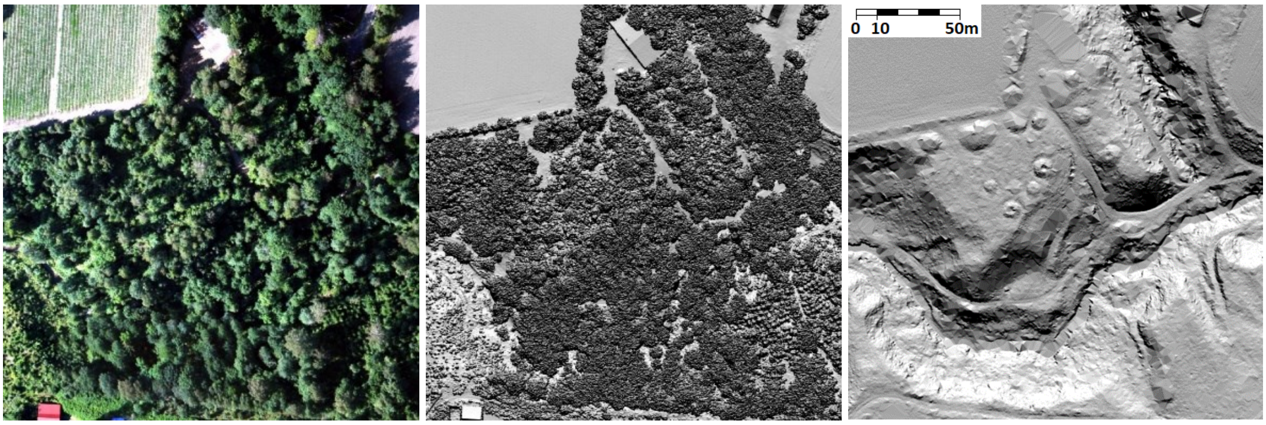

Fig. 2. A forested area in Larvik municipality, Vestfold County. Left: air photo. Middle: digital surface model from airborne lidar data, first returns. Right: digital terrain model from airborne lidar data, by keeping only points labelled 'ground'.

Detection

For detection, the python code library *simple faster R-CNN* was downloaded from https://github.com/chenyuntc/simple-faster-rcnn-pytorch. For each detected object the R-CNN predicted a bounding box, a class label and a score value in the range 0.0–1.0. The score value indicated how likely it was that the detected object was a cultural heritage object, with 1.0 meaning 'very likely' and 0.0 'not likely'. Detected objects with score values lower than 0.7 were discarded.

A few modifications had to be done:

1. The list of class labels was changed to match the class labels used in the image annotations, i.e., 'grave mound', 'pitfall trap' and 'charcoal kiln'. The original code uses class labels for objects that often occur in natural outdoor images, such as 'car', 'bicycle', 'cat', 'dog', etc.
2. The downloaded code crashed if there were no detected objects within an image. We suspect that the code, which is developed for object detection in photographs, was never run on images not containing any objects of interest. However, in the context of cultural heritage detection, the normal situation in many parts of the landscape is the absence of cultural heritage objects. Thus, if-tests had to be added for the case that no objects were detected in an image.

Pre-training of the neural network was done by importing parameters learned from the VGG16 deep neural network (Simonyan and Zisserman, 2015) on the ImageNet dataset of photographs with labelled objects such as cars, dogs, etc. (Russakovsky et al., 2015). Additional training was then done on the LRM images with labelled cultural heritage objects. The neural network was then used to predict the locations and sizes of grave mounds, pitfall traps and charcoal kilns in LRM images of size 600 × 600 pixels (Fig. 4).

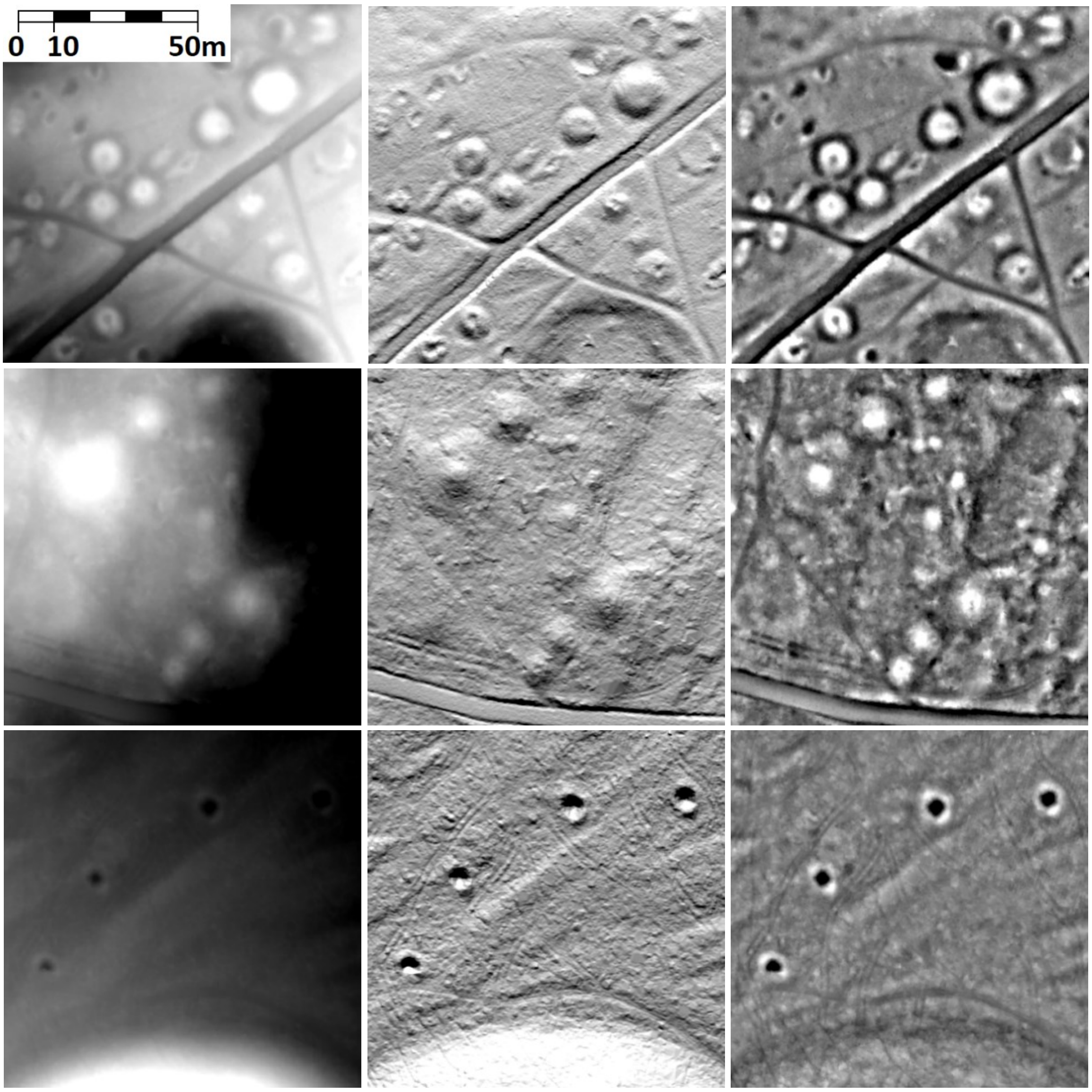

Fig. 3. DTM visualisations. Left column: terrain elevation. Middle column: hillshade. Right column: local relief model. Top row: from Bøkeskogen, Larvik municipality, Vestfold County. Several grave mounds are visible. Middle row: from Omsland, Larvik municipality, Vestfold County. Several grave mounds are visible. Bottom row: from Nord-Fron municipality, Oppland County. A deer hunting system with pitfall traps is visible.

Processing chain

The preprocessing and detection methods were integrated into a python script that may be called from QGIS or started from the Linux command line. The input was a collection of LAS files, and the output was two ESRI shape files for each object type; centre points in one file and object outlines in another file. Each object outline was obtained by converting the predicted bounding box to a circle.

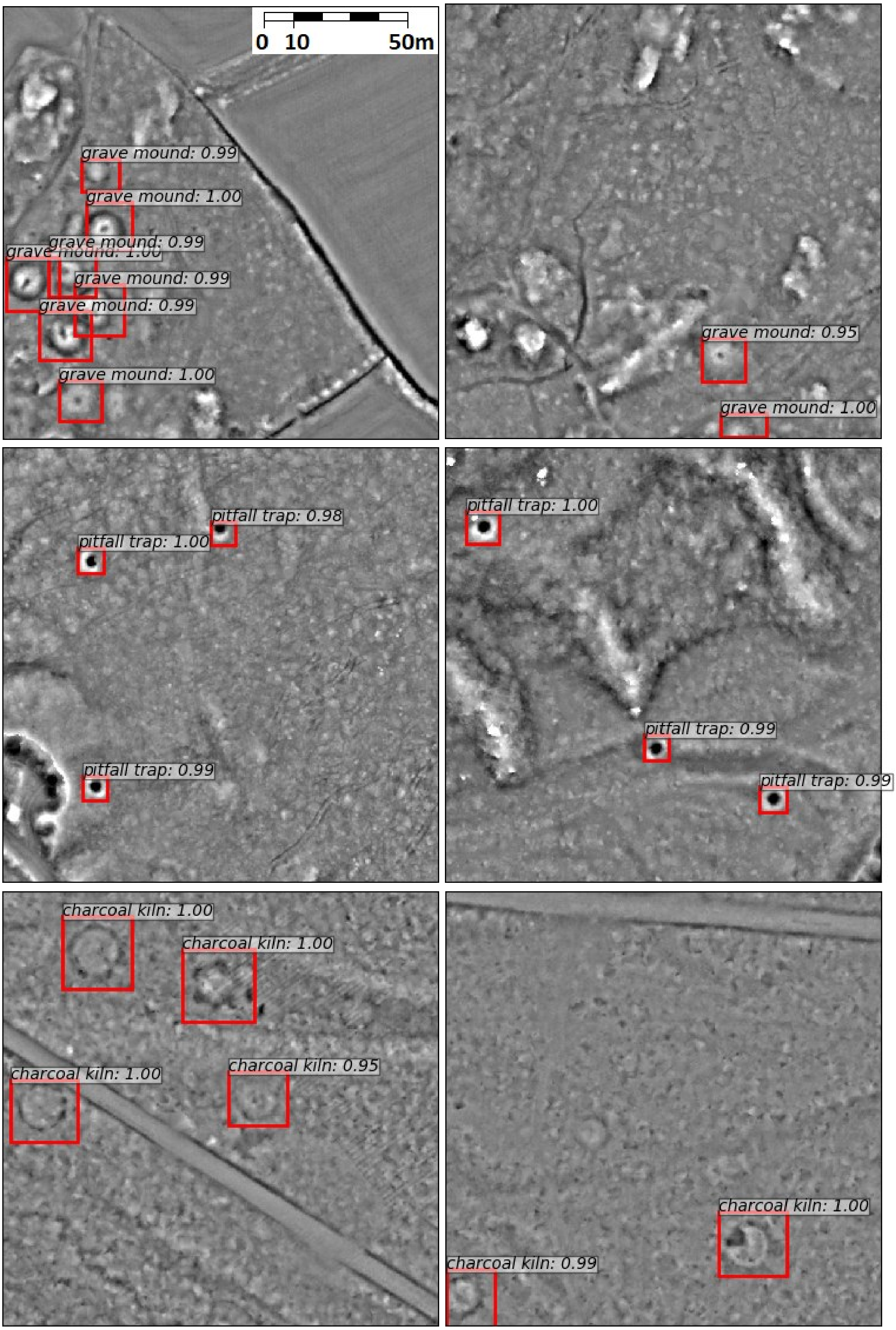

Fig. 4. Automatic predictions, with score values (see text), of cultural heritage objects in local relief model images of 150 m by 150 m areas. Top: Predicted grave mound locations. Middle: Predicted pitfall trap locations. Bottom: Predicted charcoal kiln locations.

Results

By running the method on 737 test images not seen during training, the consumer's accuracy, i.e., how many of the true cultural heritage objects were correctly detected, was 87 %, and for the specific classes, grave mound 84 %, pitfall trap 86 % and charcoal kiln 96 % (Table 2). 13 % of the true cultural heritage objects were missed by the method, while less than 1 % was detected with wrong class. The producer's accuracy, i.e., how many of the predicted cultural heritage objects were true cultural heritage objects and with correct class, was 75 %.

24 % of the objects that the method predicted as being cultural heritage were in fact not. However, the latter figure may be an optimistic estimate of the amount of false positives that the method may provide. All the test images contained at least one cultural heritage object. In operational use, there may be large areas, within an ALS dataset, with no cultural heritage objects visible in the data. Thus, the potential for false positives is much larger. Evaluation of the detection and classification performance in such a setting will be done in the near future.

	predicted class						
true class	charcoal kiln	grave mound	pitfall trap	back-ground	sum	count	rate
charcoal kiln	180	1	0	7	188	180	96 %
grave mound	3	603	0	109	715	603	84 %
pitfall trap	1	6	1073	161	1241	1073	86 %
background	80	252	267	0	599		
consumer's accuracy					2144	**1856**	**87 %**
wrong class					2144	11	0,5 %
false negatives					2144	277	13 %
producer's accuracy	68 %	70 %	80 %		2466	**1856**	**75 %**
false positives					2466	599	24 %
wrong class					2466	11	0,4 %

Table 2. Detection results.

The method was then used on all of Øvre Eiker municipality, an area with few recorded charcoal kilns; thus, no ground truth existed. This is the normal situation for the practical use of the method, in order to discover previously unknown cultural heritage locations. More than 1000 charcoal locations were predicted by the method (Fig. 5). A quick visual inspection (e.g., Fig. 6) confirmed that the large majority, if not all, of the predicted charcoal locations were true. Thus, they were included into the Askeladden database of all cultural heritage locations in Norway (https://kulturminnesok.no/).

Discussion and conclusions

The classification performance was 87 % consumer's accuracy on a test set not seen during training. At the same time, the producer's accuracy was 75 %. However, all the test set images contained at least one cultural heritage object, and each image was only of a 150 m by 150 m area. In most landscapes, the majority of 150 m by 150 m patches may contain no cultural heritage objects visible in the DTM. Thus, the estimated producer's accuracy of 75 % may be too optimistic. On the other hand, the number of false positives appeared to be low on the Øvre Eiker unlabelled test data.

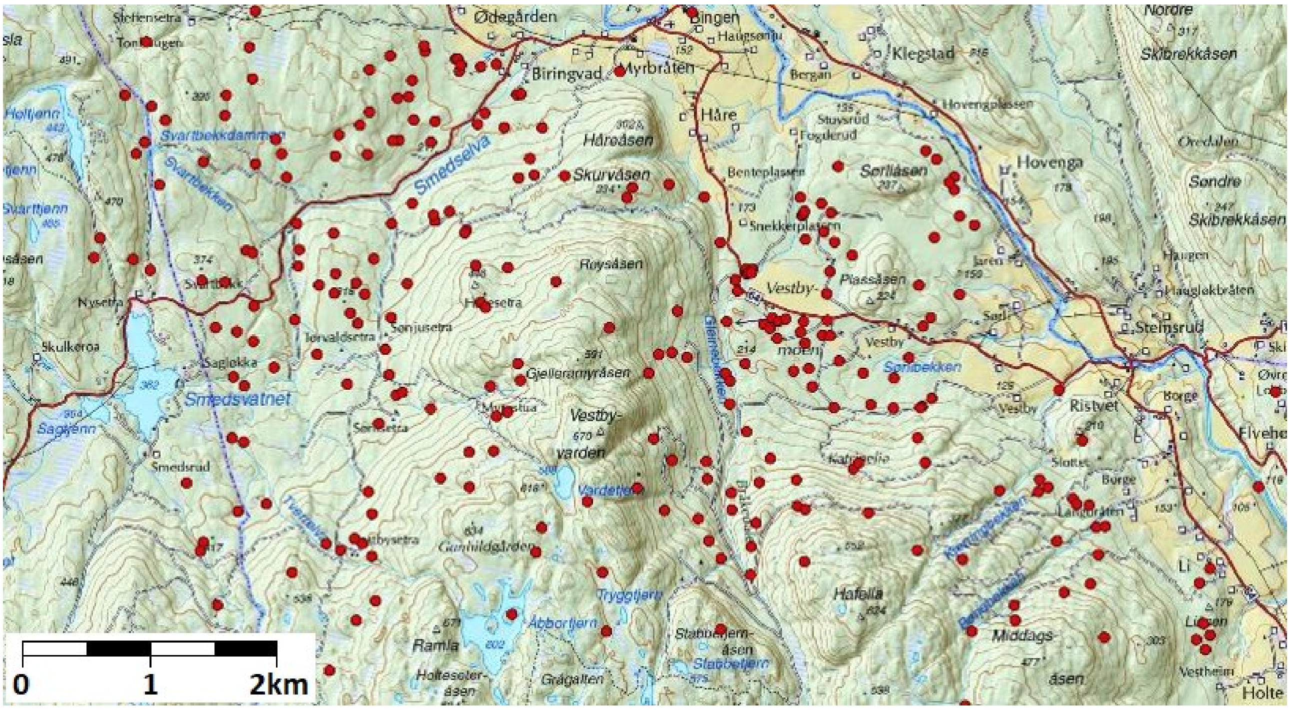

Fig. 5. Some of the predicted charcoal kilns (red circles) in forested areas (pale green) in Øvre Eiker municipality, Buskerud County. The area shown is 10.2 km by 5.6 km.

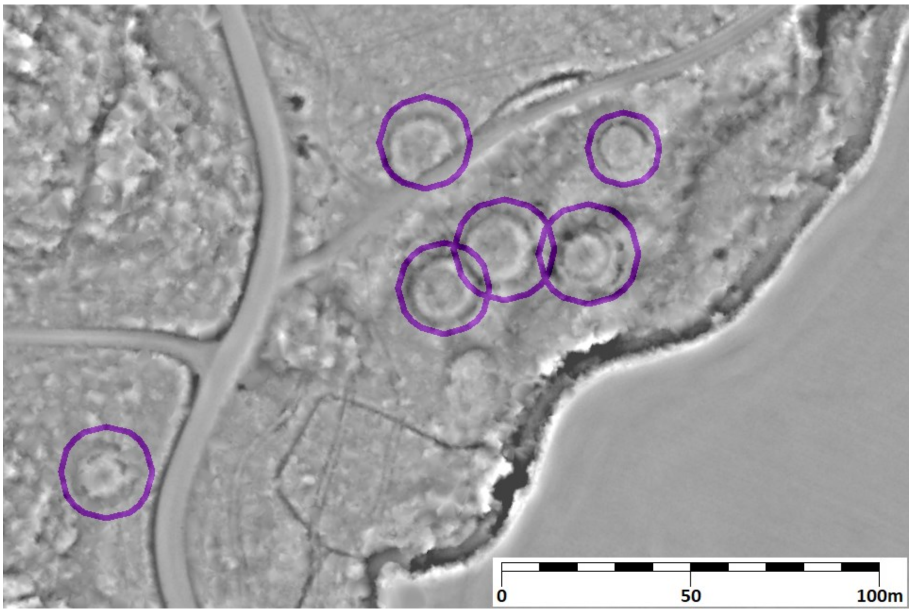

Fig. 6. Visual inspection of six predicted charcoal kilns (purple circles) in Øvre Eiker municipality. Local relief visualization of DTM. The area shown is 240 m by 160 m.

The method has been used on a number of ALS datasets covering a variety of landscape types, including forest, mountain, rural, agricultural and coastal areas. Although a detailed quantification of detection performance has not yet been performed, some trends were observed through practical use of the method for detailed archaeological mapping. The method performed better on charcoal kilns than on the other object types. In the inland, the method performed well on pitfall traps. This included many areas that are lacking detailed cultural heritage mapping. An unexpected bonus was that charcoal pits / tar pits were detected, albeit as pitfall traps. For grave mounds, the method was less successful. Confusion between natural knolls and grave mounds was the main problem. Still, the method may be useful by giving an overview of locations in the landscape with structures resembling grave mounds. These could then be checked visually by an experienced archaeologist, who could spot which locations need to be checked by field visits.

There are some recent projects that involve citizen volunteers to help identify which automatically detected structures are true archaeological remains. In the Chilterns in England (Morrison and Peveler, 2019; Peveler and Morrison, 2019; https://chilternsbeacons.org/wp/), citizens use an internet portal to view different lidar DTM visualisations of an area to identify and map archaeology. In the Veluwe area in the centre of the Netherlands (Lambers et al., 2019; https://www.zooniverse.org/projects/evakap/heritage-quest), an internet portal is also used. Participants are asked to mark every potential barrow, charcoal kiln and Celtic field within a 300 m by 300 m subimage. Each individual image is checked by at least eight different users. These two projects may provide ideas on how to involve citizen volunteers for visual verification of automatic detection results in Norway.

The proposed method was based on transfer learning, but in a setting that may not be optimal. The deep neural network was pre-trained on natural scene images, followed by training on DTM visualisations with labelled cultural heritage remains. As the two types of image are quite different, there is a potential for improvement by pre-training the deep neural network on a large image set that is more similar to the DTM visualisations that we used.

An issue that was observed at terrain discontinuities, e.g., a cliff, was that the local relief model visualisation might hide archaeological objects that were close to the terrain discontinuity. A possible solution could be to use another ALS visualisation, e.g. openness (Doneus, 2013).

Landauer and Hesse (2019) obtain very low false positive rates on a set of 29 000 labelled, possible charcoal kilns, with 95 % detection rate. The labels ranged from 0 = ‘certainly not’ to 4 = ‘definitely yes’. For each 40 m by 40 m image of a possible charcoal kiln, the final label is the average of the labels provided by several human users. Of the 30 false positives (i.e., images labelled with 0, but detected as charcoal kiln by the deep neural network), 15 were in fact charcoal kilns and thus wrongly labelled 0.

In order to estimate detection results for operational use, the plan is to run automatic detection on entire LAS files and not only on small image portions which contain at least one cultural heritage object. The ground truth data must be valid for all the lidar data included in the test set. The predicted cultural heritage objects will be compared with the ground truth data. One may expect to see more false positives, and thus, lower producer’s accuracy than we reported in the results section. However, one may still expect the method to be able to detect roughly the same percentage of the true cultural heritage objects.

One possible workaround to reduce the number of false positives may be to add confusion classes. This may be done by running the detection method on training and validation areas. Then, false positives may be labelled with new class names, e.g., ‘natural mound’, ‘natural pit’ and ‘natural platform’. In addition, it may be necessary to check if any false positives are actually true positives. With the confusion class objects added to the training and validation sets, training will be re-run.

Another potential for possible improvement is in replacing Faster R-CNN with another deep neural network. He et al. (2017) extend Faster R-CNN into Mask R-CNN by providing, for each detected object, an object mask in addition to the bounding box provided by Faster R-CNN. This may be beneficial if, unlike in the present study, the objects to detect deviate substantially from compact, near-circular structures. Lin et al. (2020) proposes the RetinaNet to address the imbalance of foreground versus background, a common problem in many cultural heritage detection tasks, i.e., that the absence of objects is much more frequent than the presence of objects. Code for Mask R-CNN and RetinaNet is included in Detectron (Girshick et al., 2018).

In conclusion, it has been demonstrated that faster R-CNN is well suited, in terms of consumer's accuracy, for automated detection of cultural heritage objects such as charcoal kilns, grave mounds and pitfall traps in high resolution airborne lidar data. However, one may expect that the method must be improved in terms of producer's accuracy in order to limit the number of false positives when applied on large areas for detailed archaeological mapping.

Acknowledgements

This research was funded by the Directorate for Cultural Heritage in Norway.

References

Doneus, M. (2013). Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. *Remote Sensing* 5(12), pp. 6427-6442. doi: 10.3390/rs5126427.

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, USA, 23–28 June 2014, pp. 580–587. doi: 10.1109/CVPR.2014.81.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P.and He, K., 2018. *Detectron*. https://github.com/facebookresearch/detectron.

He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017). Mask R-CNN. *The IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 22–29 Oct. 2017, pp. 2961–2969. doi: 10.1109/ICCV.2017.322.

Hesse, R. (2010). Lidar-derived local relief models – a new tool for archaeological prospection. *Archaeological Prospection* 17, pp. 67–72. doi: 10.1002/arp.374.

Lambers, K., Verschoof van der Vaart and W. B., Bourgeois, Q.P.J. (2019). Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. *Remote Sensing* 11(7), 794. doi: 10.3390/rs11070794.

Landauer, J. and Hesse, R. (2019). Machine learning for large area archaeological feature detection. Applying transfer learning to airborne lidar data. In: *24th Conference on Cultural Heritage and New Technologies*, 4–6 November 2019, Vienna, Austria. https://www.chnt.at/wp-content/uploads/Machine-learning-for-large-area-archaeological-feature-detection.pdf.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P., 2020. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(2), pp. 318–327. doi: 10.1109/TPAMI.2018.2858826.

Morrison, W. and Peveler, E. (2019). Beacons of the past. Visualising LiDAR on a large scale. In: *24th Conference on Cultural Heritage and New Technologies*, 4–6 November 2019, Vienna, Austria. https://www.chnt.at/wp-content/uploads/Beacons-of-the-Past.pdf.

Peveler, E. and Morrison, W. (2019). Beacons of the past: citizen science and community engagement in the Chilterns. In: *24th Conference on Cultural Heritage and New Technologies*, 4–6 November 2019, Vienna, Austria. https://www.chnt.at/wp-content/uploads/Beacons-of-the-Past-1.pdf.

Ren, S., He, K., Girshick, R. and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6), pp. 1137–1149. doi: 10.1109/TPAMI.2016.2577031.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal on Computer Vision* 115, pp. 211–252. doi: 10.1007/s11263-015-0816-y.

Simonyan, K.and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In: Bengio, Y. and LeCun, Y. (eds). *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, 7–9 May 2015.

Trier, Ø.D. and Pilø, L.H. (2012). Automatic detection of pit structures in airborne laser scanning data. *Archaeological Prospection* 19 (2), pp. 103–121. doi: 10.1002/arp.1421.

Trier, Ø.D., Salberg, A.-B. and Pilø, L.H. (2018). Semi-automatic mapping of charcoal kilns from airborne laser scanning data using deep learning. In: Matsumoto, M and Uleberg, E. (eds). *CAA2016: Oceans of Data. Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology*. Oslo, Norway, 30 March-3 April 2016. Oxford: Archaeopress, pp. 219–231. Available at: http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={6A565CFE-F617-4333-9818-4C13E78B7C1B}.

Trier, Ø.D., Zortea, M., and Tonning, C. (2015). Automatic detection of mound structures in airborne laser scanning data. *Journal of Archaeological Science: Reports* 2 (1), pp. 69–79. doi: 10.1016/j.jasrep.2015.01.005.

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. doi: 10.5334/jcaa.32.