Received: 9 November 2020 | Revised: 4 January 2021 | Accepted: 7 January 2021

DOI: 10.1002/arp.1806

RESEARCH ARTICLE

WILEY

A modified Mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland

Alexander Bonhage[1] | Mahmoud Eltaher[2] | Thomas Raab[1] | Michael Breuß[2] | Alexandra Raab[3] | Anna Schneider[1]

[1]Brandenburg University of Technology Cottbus-Senftenberg - Chair of Geopedololgy and Landscape Development, Cottbus, Germany

[2]Brandenburg University of Technology Cottbus-Senftenberg - Chair of Applied Mathematics, Cottbus, Germany

[3]Brandenburg University of Technology Cottbus-Senftenberg - Chair of Environmental Economics, Cottbus, Germany

Correspondence
Mahmoud Elthaher, Brandenburg University Technology Cottbus-Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, Germany.
Email: mahmoud.eltaher@b-tu.de

Funding information
German Research Foundation (DFG), Grant/Award Number: RA 931/8-1

Abstract

Due to complicated backgrounds and unclear target orientation, automated object detection is difficult in the field of archaeology. Most of the current convolutional neural network (CNN) object-oriented detection techniques are based on a faster region-based CNN (R-CNN) and other one-stage detectors that often lack adequate processing speeds and detection accuracies. Recently, the two-stage detector Mask R-CNN technique achieved impressive results in object detection and instance segmentation problems and was successfully applied in the analysis of archaeological airborne laser scanning (ALS) data. In this study, we outline a modified Mask R-CNN technique that reliably and efficiently detects relict charcoal hearth (RCH) sites on light detection and ranging (LiDAR) data-based digital elevation models (DEMs). Using image augmentation and image preprocessing steps combined with the deep learning-based adaptive gradient method with a dynamic bound on the learning rate (AdaBound) optimization technique, we could improve the model's accuracy and significantly reduce its training time. We use DEMs based on high-resolution LiDAR data and the visualization for archaeological topography (VAT) technique that give images with a very strong contrast of the terrain and the outline of the sites of interest in the North German Lowland. Therefore, the model can identify RCH sites with an average recall of 83% and an average precision of 87%. Techniques such as the modified Mask R-CNN method outlined here will help to greatly improve our knowledge about archaeological site densities in the realm of historical charcoal production and past human-landscape interactions. This method provides an accurate, time-efficient and bias-free large-scale site mapping option not only for the North German Lowland but potentially for other landscapes as well.

KEYWORDS
anthropogenic geomorphology, automated detection, charcoal hearth, deep learning, Mask R-CNN, object detection

1 | INTRODUCTION

The increasing availability of airborne high-resolution light detection and ranging (LiDAR) data has led to an ever-growing interest in applying remote sensing to the archaeological domain, in which the use of machine learning techniques is now also increasing (e.g., Cowley et al., 2020; Davis, 2018; Opitz & Herrmann, 2018). Considering the trends in the developments of neural networks in recent years, deep learning techniques have initiated a significant change in the field of computer vision. Catalyzed by the use of convolutional neural networks (CNNs), strong advances in many classic visual inspection tasks have been made recently, for example, in object detection, object localization, semantic segmentation and object instance segmentation tasks (He et al., 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014; Zagoruyko et al., 2016). Focussing on neural networks for object detection, progress made in recent years is almost completely linked to the basic CNN-based model and its extensions, namely, the region-based CNN (R-CNN), the fully convoluted network, the Fast R-CNN and the more efficient variant Faster R-CNN (Girshick, 2015; Girshick et al., 2014; Long et al., 2015; Ren et al., 2015). Mask R-CNN expands the Faster R-CNN architecture by adding an algorithmic branch for predicting an object segmentation mask parallel with the existing region proposal stage (He et al., 2017). A detailed description of the Mask R-CNN technique is given elsewhere (e.g., see, Kazimi et al., 2019). Because of its relatively easy trainability and efficiency, Mask R-CNN has seen a surge in popularity for use in object detection (Ahmed et al., 2020; Johnson, 2018; Sorokin, 2018; Yu et al., 2019) and in the analysis of archaeological airborne laser scanning (ALS) data (Gong & Zhang, 2020; Pham & Lefèvre, 2018; Verschoof-van der Vaart & Lambers, 2019).

In recent years, small anthropogenic landforms, that is, so-called relict charcoal hearths (RCHs, sometimes also called charcoal-burning platforms or kilns), which are mainly found in forests and result from historical charcoal production, have attracted the attention of archaeologists and soil scientists in the North German Lowland (Raab et al., 2015). RCHs are part of landscapes, the so-called sociocultural fingerprint (Tarolli et al., 2019), and an important source of anthracological information (Gocel-Chaltè et al., 2020; Smidt et al., 2017), providing insight into historical land use practices (e.g., Tolksdorf et al., 2020; Deforce et al., 2020). Recent studies in soil science focus on RCH site-specific changes of soil chemical properties, such as increases in soil organic matter contents and element stocks (e.g., Donovan et al., 2021), changes in soil physical properties (e.g., Schneider, Hirsch, et al., 2020) and effects on vegetational and faunal growth (e.g., Buras et al., 2020; Gießelmann et al., 2019).

In flat terrain, RCHs are generally circular in shape, with a wide range of diameters (up to 30 m and averaging 12 m), are elevated several decimetres above the earth's surface and are often surrounded by a shallow circular ditch or multiple small pits (Hirsch et al., 2020). These morphological properties are favourable for detecting RCHs on LiDAR-based digital elevation model (DEM) visualizations because the circular elevation (positive feature) and the ditch (negative feature) form a strong visual contrast. Various studies manually mapped and digitized RCH sites, mostly by means of shaded relief visualizations (e.g., Carter, 2019; Deforce et al., 2013; Raab et al., 2019; Risbøl et al., 2013; Schmidt et al., 2016) or other visualization techniques (Hesse, 2010). Since high-resolution LiDAR data have become readily available for an increasing number of countries and RCH sites are found in increasing larger forested areas, fully or semiautomated methods are required to decrease the workload of such labour-intensive manual mappings. Some previous (semi)automated mapping approaches for RCHs and other archaeological objects involve template matching (Schneider et al., 2014), geographic object-based image analysis (GEOBIA) (Witharana et al., 2018) and, more recently, deep learning techniques (e.g., Kazimi et al., 2020; Lambers et al., 2019; Trier et al., 2018; Trier et al., 2021; Verschoof-van der Vaart et al., 2020). Mapping RCH sites shares similarities with mapping burial mounds, which are some of the most frequently studied archaeological sites globally (Davis, 2020). However, mounds may vary in their shape more so than RCHs, having rectangular, triangular and trapezoidal elevation profiles (Davis et al., 2019), while RCH sites are predominantly circular elevations. Nonetheless, impressive mound mapping results using machine learning techniques have been achieved (e.g., Guyot et al., 2018; Caspari & Crespo, 2019).

Recently, Mask R-CNN has been used by Kazimi et al. (2019) for the DEM-based identification of archaeological objects such as bomb craters, charcoal hearths, and barrows, constituting a multiobject detection approach. Our study uses a modified Mask R-CNN approach, which we developed independently from the aforementioned study. We propose several modifications and extensions to the standard Mask R-CNN technique to (1) make it adapt easier to new data and reduce overfitting, (2) minimize the training time of the model and (3) improve the model's accuracy by adding image preprocessing and augmentation steps. Therefore, we outline an improved method to detect charcoal hearths that can be easily applied to other objects in LiDAR DEMs.

2 | STUDY AREA

The study area is located in Lower Lusatia in the North German Lowland (Figure 1). It is covered by forests consisting mainly of pine and oak. The southwestern part of the area has been previously mapped based on LiDAR data, which revealed a relatively high density of RCHs of up to 440 RCH sites per square kilometre, but there are also areas featuring much lower densities (Raab et al., 2019). The relief throughout the area is disturbed by former military activities (trenches, bunkers, etc.). RCH sites in the area have been previously mapped and described based on older DEMs with comparably lower quality (1- and 2-m grid sizes) and ground surveys by Bonhage et al. (2020), Schneider, Bonhage, et al. (2020) and Raab et al. (2019). In this study, we mapped RCHs in 20 subareas of 0.17 km^2 each, totalling 3.4 km^2. Ten training and

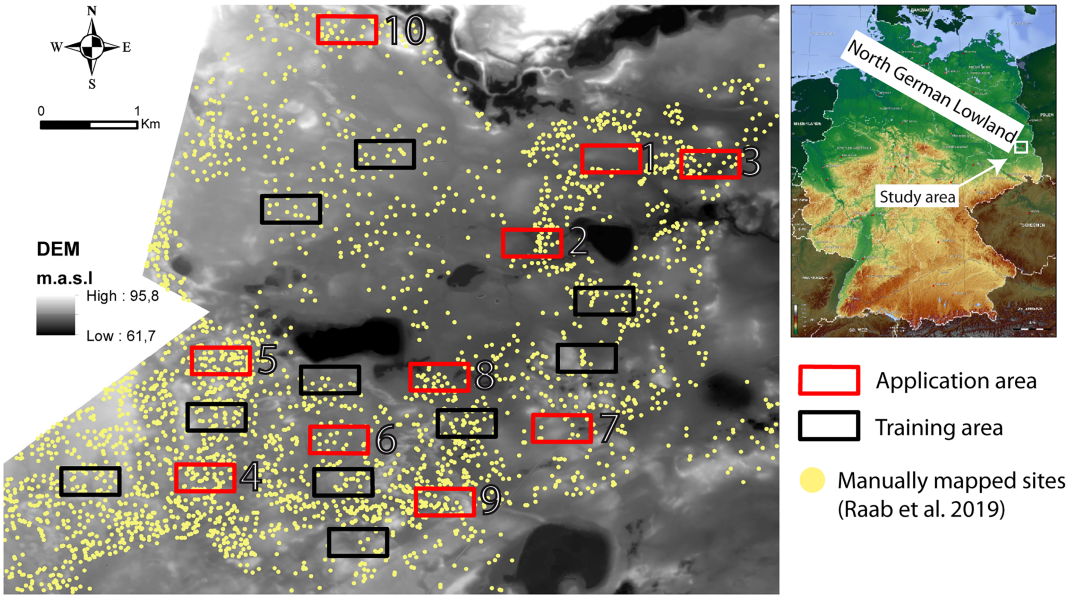

FIGURE 1 Study area overview with the training and validation areas. Light detection and ranging (LiDAR) data courtesy of LEAG. DEM = digital elevation model [Colour figure can be viewed at wileyonlinelibrary.com]

10 validation areas were selected to include a variety of RCH site densities and anthropogenic relief disturbances. The sites in the north are, on average, larger, have a lower spatial density and yield less anthropogenic disturbance on the adjacent relief than the sites to the south. The charcoal produced in the area provided resources for the nearby ironwork in Peitz, which operated from the mid-16th to the mid-19th century. Consequently, site dating using dendrochronology revealed that the RCH ages were between 1654 and 1852 (Raab et al., 2015; Raab et al., 2019).

3 | METHODS

3.1 | LiDAR data DEM visualization and mapping

DEM visualizations were created based on a LiDAR data-derived 0.5-m grid using the Relief Visualization Toolbox, version 2.2.1 (Kokalj & Somrak, 2019). We applied a modified version of the visualization for archaeological topography (VAT) method. It creates a blended image of three DEM visualizations: (1) the sky-view factor, which is a method of diffuse illumination that uses a sky-view factor corresponding to a portion of the visible sky that is limited by the relief (Zakšek et al., 2011); (2) the positive openness, which is a method that highlights the highest and lowest points of features (Doneus, 2013); and (3) regular analytical hillshading. The first two methods highlight features independently from their aspect or the illumination angle of the light source and therefore greatly enhance the visibility of all the relief features (Figure 2). We omitted the slope gradient layer since it did not enhance the visibility of the RCH features and added considerable noise to the images. The images were created and blended according to the default settings of the toolbox and stored as.jpg files (1200 × 600). The RCH sites in all the areas were digitally mapped by an experienced human operator prior to Mask R-CNN analysis. The sites were digitized by drawing a circle around the edge of the RCH platform (positive relief feature), which was saved in a shapefile. We excluded the ditch (negative relief feature), since its outer edge DEM signature is sometimes disturbed and/or barely visible. Bounding boxes for training the model were created with the Python-based LabelImg (Tzuatlin, 2015) application. The bounding boxes were drawn to include the platforms and the ditch DEM signature.

3.2 | Image preprocessing and augmentation

We applied several image preprocessing steps to increase the success of object detection. First, we vectorized the maps and bounding boxes since the input data and targets must be tensors of floating point precision values. Then, we normalized the tonal range of the maps by converting the pixels from integer grayscale values ranging from 0 to 255 to floating point values and dividing by 255, resulting in final floating point values ranging from 0 to 1. Without the normalization step, the model can trigger large gradient updates that can prevent the network from converging (Sola & Sevilla, 1997). Furthermore, we applied nonlinear diffusion filtering (Perona & Malik, 1990) on the images to further enhance the RCH features. For this step, we set the processing time to 10 s with a lambda value of 0.5. To increase the number of sites and decrease model overfitting, we first augmented the 10 training area maps by using image transformations. The maps were duplicated randomly and modified by image rotation, translation and horizontal and vertical flipping. Therefore, the number of training maps was increased to 200.

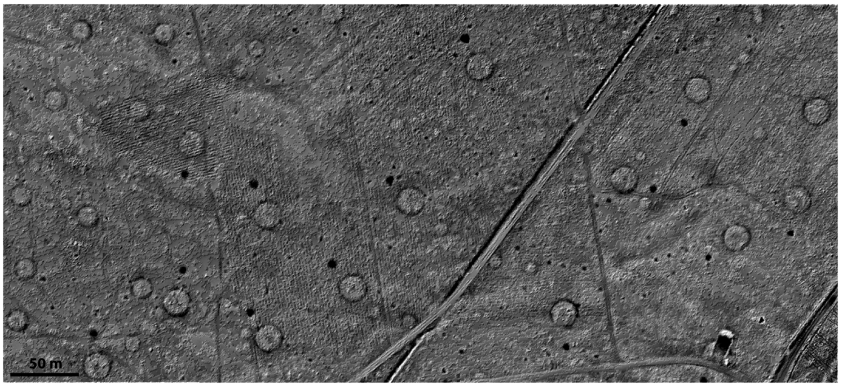

FIGURE 2 Example of modified VAT visualization (blend of sky-view factor, positive openness and hillshade visualizations) for Validation Area 3, showing circular RCH sites of various sizes, disturbances and visibilities. The smaller black dots are from pits resulting from historical military activity, and the larger black dots show pits that are most likely associated with hearth operation. RCH = relict charcoal hearth, VAT = visualization for archaeological topography

3.3 | Mask R-CNN model setup

Mask R-CNN relies on region proposal networks (RPNs) for identifying objects and is therefore a two-stage detector system. The three main steps (Figure 3) of the model are as follows: An RPN is used to decide where to search to reduce the computational requirements of the overall inference process; the region proposals are transformed by means of classification and creating of bounding boxes; and finally, the creation of segmentation masks, in which the algorithm takes positive regions of interest (ROIs) as input and creates 28 × 28 pixel floating-point value masks, which are then scaled up. For the inference process, the algorithm uses a multitask loss function: L = Lcls + Lbbox + Lmask (He et al., 2017). The developed model is based on a present implementation by Matterport, Inc. and released under an MIT licence, which is itself constructed based on the open-source libraries Keras and TensorFlow. In our study, we started by using weights obtained from the Microsoft Common Objects in COntext (MS COCO) dataset (Lin et al., 2014). However, a preliminary test showed unsatisfactory results caused by the specific domain-related problem of our dataset, that is, the occurrence of small objects. To counter this issue, we modified the *anchor scale* to values of 8, 16, 32, 64 and 128. We also modified the *training anchors per image* to values of 64 and 128, allowing for a large number of detected objects. We trained the model for 15 epochs using the stochastic gradient descent (SGD) method with a momentum of 0.9 and starting with a scheduled learning rate to speed up the training process. Furthermore, we compared the training results and training time for the adaptive gradient method with the dynamic bounds of the learning rate (AdaBound) (Luo et al., 2019) and the SGD optimization method. To reduce overfitting, we applied a weight decay since this parameter adds an L2 penalty to the cost, which can effectively lead to smaller model weights. The model runs under the deep learning development framework of TensorFlow and Keras, with an Nvidia 1080 GPU, an Intel® Core™ i7-8700k CPU and 16 GB memory. For model validation purposes, we divided the 200 training maps into a training subset (80% of the images) and a validation subset (20% of the images).

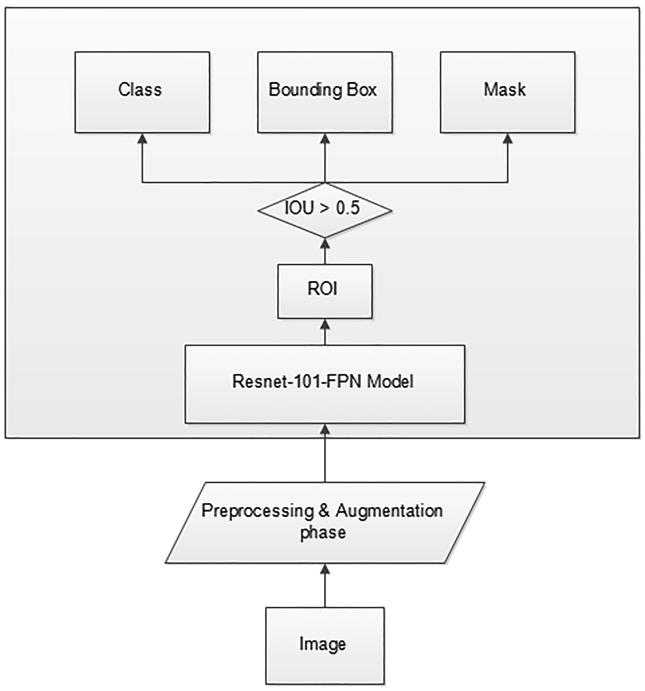

FIGURE 3 Overview of the image processing steps in the Mask region-based convolutional neural network (R-CNN) algorithm. ROI = region of interest

4 | RESULTS

4.1 | Model validation

The CNN model's efficiency for an object recognition task is given as the mean average precision (mAP). The precision is the ratio of the number of true positives to the total number of positive detections, while the recall gives the ratio of the number of true positives to the total number of ground-truth objects (Henderson & Ferrari, 2016). For the training dataset, the model detects 5019 true positive sites and 447 false positive sites, with a calculated mAP of approximately 81% (Figure 4). For the test dataset, it detects 1239 true positives and 121 false positives, with a mAP of approximately 79%. Using the AdaBound optimization method results in better model efficiency; that is, the model's runtime can be substantially reduced, as seen by the position of the lowest validation loss in Figure 5. For AdaBound,

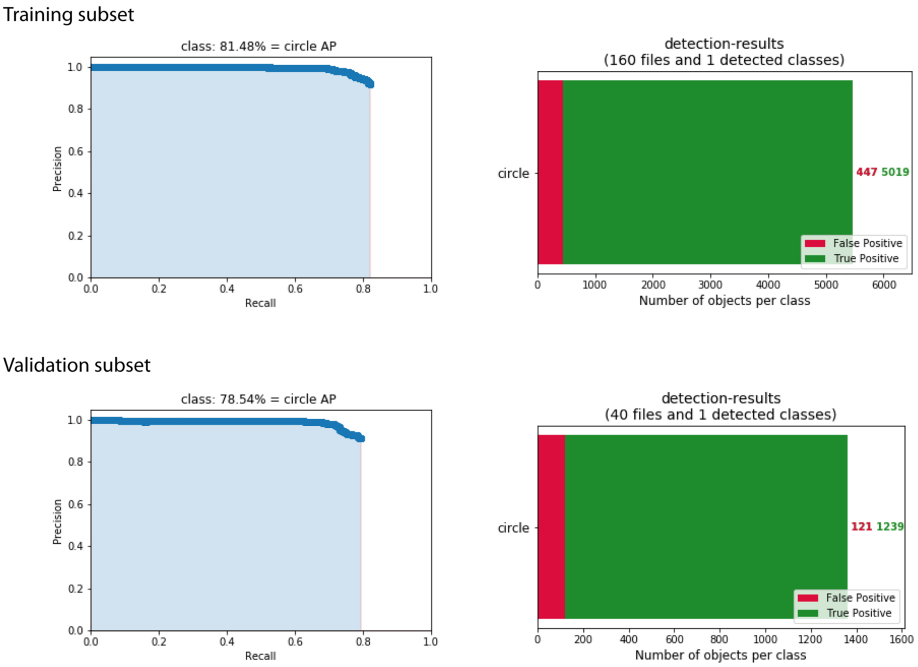

FIGURE 4 Model validation results for the training and validation subsets showing the precision/recall curves and total numbers of true positive and false positive predictions [Colour figure can be viewed at wileyonlinelibrary.com]

28 epochs are required, while for SGD, 36 epochs are required. We propose to stop early after 15 epochs, which constitutes a runtime reduction of approximately 50%. This is not only to reduce the processing time that would otherwise not improve the model loss and accuracy but also to avoid model overfitting. To assess the effect of the image preprocessing steps on the model's performance, we analyzed the training and validation subsets again but without applying any preprocessing steps. The model's mAPs for the training and validation datasets are subsequently reduced to 79% and 74%, respectively. Therefore, the preprocessing steps increased the mAPs of the model by approximately 4% on average.

4.2 | Model application

We applied the trained Mask R-CNN model to 10 application areas (Figure 6) containing 305 manually mapped RCH sites in total. This model detects RCH sites with an average recall of 83% and an average precision of 87% (Table 1); that is, on average, 17% of manually mapped sites are not detected, while 13% of all the detected sites are false positives. Furthermore, the model detects four new sites that were overlooked in manual mapping, raising the total number of known sites in the area by 1.3%. The bounding box output potentially allows the determination of site diameters or RCH areas. However, the extent of the bounding boxes is often larger than the true area of the site, making the determination of areas prone to error, especially for smaller sites. The sites not detected by the model have a smaller average diameter (9.1 m ± 2.8 m) than the detected sites (11.9 m ± 3.4 m). False positive detections are often caused by (partly) circular signatures originating from paths on the forest floor, for example, at intersections of dirt tracks, by slightly puckered terrain and, in one case, by an oval former gun emplacement (Area 4). Furthermore, circular or near-circular artefacts, presumably originating from poor LiDAR data point densities, yielded false positives in 14 cases (Areas 6, 9, and 10). Most of the sites omitted by the algorithm have poor resolution in the DEM. However, for some

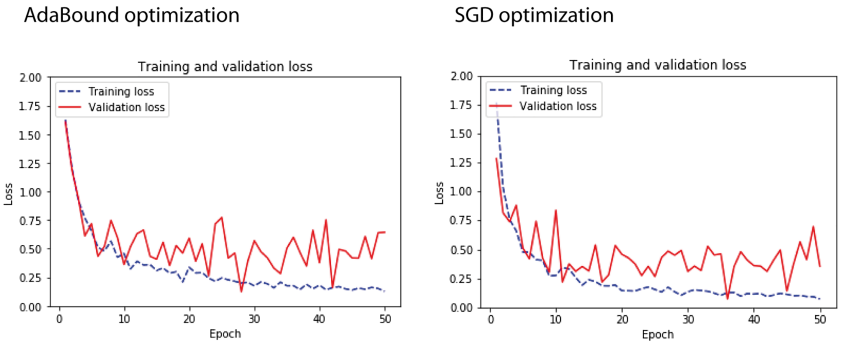

FIGURE 5 Model and validation losses for the model optimization methods: adaptive gradient method with a dynamic bound on the learning rate (AdaBound) and stochastic gradient descent (SGD) [Colour figure can be viewed at wileyonlinelibrary.com]

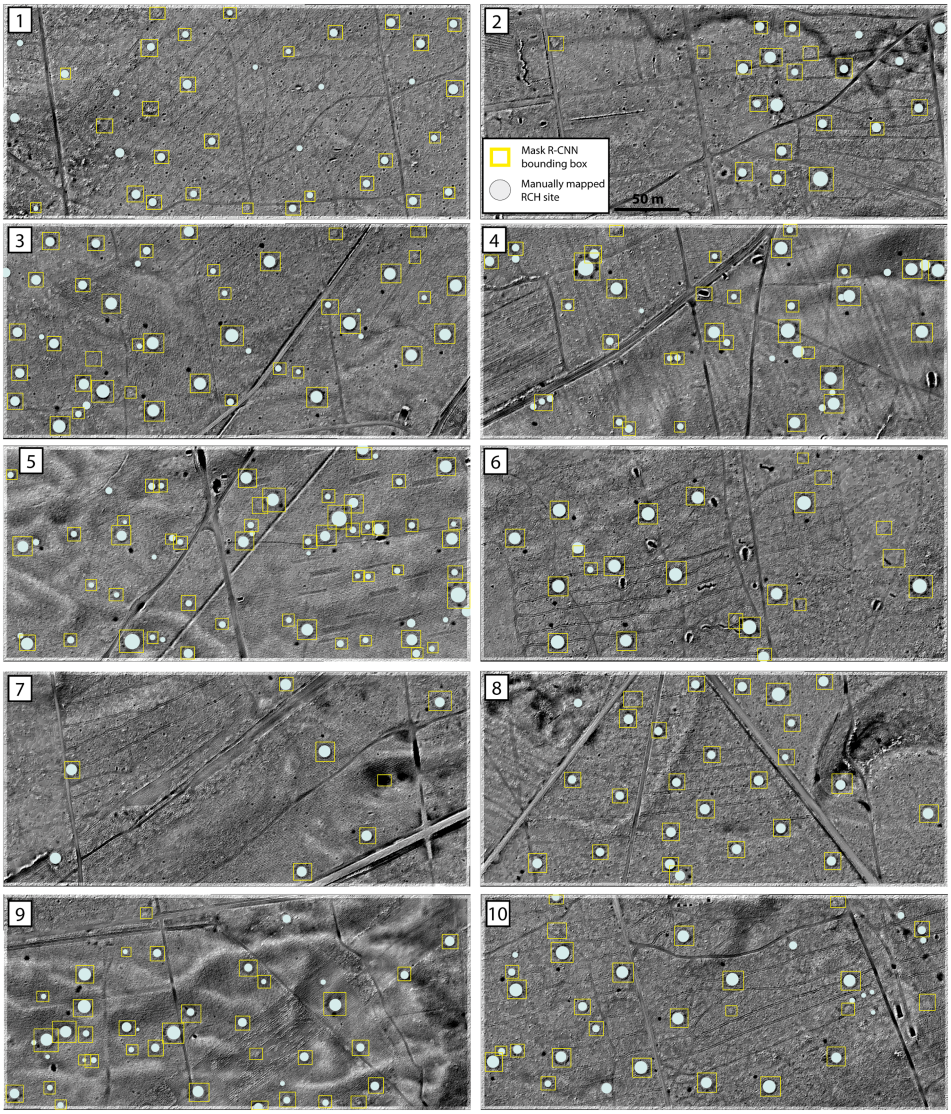

FIGURE 6 Mask region-based convolutional neural network (R-CNN) bounding box output and manually mapped relict charcoal hearth (RCH) sites in the validation areas [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Mask R-CNN mapping results for 10 application areas

Subsection	Sites mapped manually (m)	Sites mapped by the Mask-R-CNN (M)	True positives (TP)	False positives (FP)	New sites	Recall (TP/m) (%)	Precision (TP/M) (%)
1	32	28	24	4		75	86
2	18	17	14	3		78	82
3	41	39	34	5		83	87
4	40	32	30	2		75	94
5	59	49	48	0	1	81	98
6	16	22	16	4	2	100	73
7	7	7	6	1		86	86
8	25	25	24	1		96	96
9	37	34	31	3		84	91
10	30	26	21	4	1	70	81
Average						83	87

Abbreviation: R-CNNN = , region-based convolutional neural network.

sites, it is unclear why they were omitted by the algorithm, as their resolution seems to be on par with similar sites in the vicinity (e.g., in Area 9).

The Mask R-CNN method excels even in complex reliefs; for example, it reliably detects sites in Area 6 that feature an abundance of anthropogenic relief disturbances such as bomb craters and former gun emplacements (Figure 6). Furthermore, it detects sites in close spatial proximity (within ~25 m of each other) or overlapping sites, as seen in Area 5. These spatial microclusters of sites are typical in the larger vicinity of the study area (Raab et al., 2019). Even sites that have been disturbed by forest management activities, sites with imperfect circular ditches and only partially visible sites on the map edges are detected by the algorithm.

5 | DISCUSSION

We have shown that the Mask R-CNN technique can be specifically tailored for the detection of charcoal hearths and that the model's runtime can be reduced by approximately 50% without a loss in accuracy during training. Technical restrictions can arise when dealing with smaller targets, which is a known problem in automated image detection (e.g., Ju et al., 2019). Smaller RCH sites in the area tend to give poorer signatures on DEM maps caused by shallower ditch or elevation profiles. Although we can observe a trend towards smaller sites being omitted by the algorithm, the main risk factor for poor detection seems to be LiDAR data quality and not the size of the sites. A LiDAR point density with lower resolution can be caused by changes in the canopy growth density or the amount of deadwood and undergrowth on the forest floor.

Opitz and Herrmann (2018) discuss a distrust in automated feature extraction systems in archaeology based on technological and social factors, which can result from reported accuracy rates of automated systems below approximately 75%, the novelty of the utilized techniques and, presumably, their opaqueness for a nonexpert in the science of deep learning. They raise the important question as to what level of accuracy would be required from automated systems to be acceptable for a specific purpose. For reasons of heritage protection and decision making as well as archaeological and paleoenvironmental research, the detection of as many objects as possible is preferred. This stipulation is important for not only current but also future studies. In particular, because of increasing land consumption and the possible effects of climate change, the irretrievable loss of many archaeological objects is threatened. In the realm of RCH geoarchaeological research, most studies with a focus on site mapping include rather small areas of usually well under 50 km^2, and only three studies have incorporated mappings on a larger scale so far (see tab. 2 in Rutkiewicz et al., 2019). The efficiency of automated mapping approaches seems scale dependent. The vast majority of recent studies focus on soil science and anthracological-related objectives on the pedon scale and usually analyze a few sites (e.g., see introduction of Buras et al., 2020, for an overview). Schneider, Bonhage, et al. (2020), Rutkiewicz et al. (2019), Johnson, Ouimet, and Raslan (2015) and Raab et al. (2019) increased the scale of research with objectives regarding the assessment of RCH site distributions on the landscape scale. Disregarding the amount of time necessary for creating DEM visualizations, which is the same for both methods, automated mapping approaches, such as the modified Mask R-CNN approach, seem mostly viable for image processing on the scale conducted by these studies, that is, thousands of square kilometres or more. For smaller areas, the compromise between mapping accuracy and runtime seems unreasonable; for example, it took an experienced person only approximately 30 min to map all the sites in the 20 subareas of this study. Regarding the mapping accuracy, automated mapping systems applied in archaeology generally fall behind manual mapping results (see tab. 1 in Trier et al., 2018) with oftentimes very high rates of false positives; although comparisons to RCH studies are difficult to evaluate, LiDAR data quality and the amount and types of archaeological objects vary considerably between these studies. However, as Opitz & Herrmann (2018) discuss, it must be considered that the DEM-based mapping of RCHs and other archaeological objects is inherently incomplete, as recent studies have shown. Comparisons between manual digital RCH mappings and ground surveys show detection rates of 40–60% for study areas in Germany (Schneider, Bonhage, et al., 2020); that is, even in the best-case scenario of humans mapping sites digitally, technical limitations, such as the LiDAR ground point density, vegetational cover disturbances of the relief and biological limitations of the human operator, prevent the identification of the true number of sites in the real world. Schneider, Bonhage, et al. (2020) mapped RCH sites in the entire German Federal State of Brandenburg, of which our study area is also part. Under the assumption of homogeneous LiDAR data quality throughout the federal state and taking into account the average recall and precision values, the trained Mask R-CNN model would have been very efficient in mapping sites, saving working hours and money (not considering the time needed to develop the model). Presumably, their study would have reached the same conclusion of finding a close spatial relationship between RCHs and historical charcoal-consuming industries in homogenous landscape settings. Verschoof-van der Vaart et al. (2020) discuss the feasibility of automated systems and conclude that they should not replace the manual mapping of archaeological sites. Rather, they can help assess site densities over large areas and in complex terrain. Davis (2020) argues that automated mapping techniques in archaeology will help to improve the reproducibility of site mappings by providing a semantically consistent metalanguage, a factor that is oftentimes lacking during manual mappings by different researchers. Although we have no experience transferring our modelling approach to other areas as of yet, we see great potential in achieving consistent and bias-free mapping results for large areas.

Recent studies took the first step to transfer site specific observations of RCH research to a larger scale by determining the increase in the landscape soil carbon stock caused by historical charcoal production as well as by creating a model that allows for the calculation of the RCH volume based on its shape and topographical position (Bonhage et al., 2020; Bonhage, Hirsch, Schneider, et al., 2020).

Completely assessing the effect of the historical charcoal production legacy on current soil landscapes, however, requires knowledge of the large-scale spatial site density and distribution. For this, automated mapping approaches, such as the presented Mask R-CNN technique, are needed to drastically reduce workload and mitigate the effects of human mapping error.

6 | CONCLUSION

The Mask R-CNN technique coupled with contrast-rich LiDAR data-based DEM visualizations gives very promising mapping recall and precision values for flat terrain areas in the North German Lowland. Changing the model's optimization method drastically reduces the model's training time by approximately 50%, allowing the potentially fast inclusion of new site classes. Using image optimization and augmentation techniques in preprocessing further enhances the model's precision. Validations in other regions with known occurrences of RCHs in flat terrain are needed to assess interregional applicability, especially under conditions of regionally dependent changes in the LiDAR data quality. Furthermore, future studies must assess the applicability of the technique when other RCH types are included, such as the commonly described RCH platforms on slopes in the middle mountain ranges of Central Europe and the northern United States.

CONFLICT OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

We want to thank three anonymous reviewers that helped to improve this manuscript. This research was supported by the German Research Foundation (DFG) under grant number RA 931/8-1.

Open access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

The funding source had no involvement in the study besides offering financial support.

REFERENCES

Ahmed, B., Gulliver, T. A., & alZahir, S. (2020). Image splicing detection using mask-rcnn. *Signal, Image and Video Processing*, *14*, 1035–1042. https://doi.org/10.1007/s11760-020-01636-0

Bonhage, A., Hirsch, F., Raab, T., Schneider, A., Raab, A., & Ouimet, W. (2020). Characteristics of small anthropogenic landforms resulting from historical charcoal production in western Connecticut, USA. *Catena*, *195*. https://doi.org/10.1016/j.catena.2020.104896

Bonhage, A., Hirsch, F., Schneider, A., Raab, A., Raab, T., & Donovan, S. (2020). Long term anthropogenic enrichment of soil organic matter stocks in forest soils—Detecting a legacy of historical charcoal production. *Forest Ecology and Management*, *459*. https://doi.org/10.1016/j.foreco.2019.117814

Buras, A., Hirsch, F., Schneider, A., Scharnweber, T., van der Maaten, E., Cruz-García, R., Raab, T., & Wilmking, M. (2020). Reduced aboveground growth and wood density but increased wood chemical concentrations of Scots pine on relict charcoal hearths. *Science of the Total Environment*, *717*, 137189. https://doi.org/10.1016/j.scitotenv.2020.137189

Carter, B. (2019). Identifying landscape modification using open data and tools: The charcoal hearths of the Blue Mountain, Pennsylvania. *Historical Archaeology*, *53*, 432–443. https://doi.org/10.1007/s41636-019-00171-1

Caspari, G., & Crespo, P. (2019). Convolutional neural networks for archaeological site detection – Finding "princely" tombs. *Journal of Archaeological Science*, *110*, 104998. https://doi.org/10.1016/j.jas.2019.104998

Cowley, D., Banaszek, Ł., Geddes, G., Gannon, A., Middleton, M., & Millican, K. (2020). Making liGHT Work of large area survey? Developing approaches to rapid archaeological mapping and the creation of systematic national-scaled heritage data. *Journal of Computer Applications in Archaeology*, *3*(1), 109–121. https://doi.org/10.5334/jcaa.49

Davis, D. S. (2018). Object-based image analysis: A review of developments and future directions of automated feature detection in landscape archaeology. *Archaeological Prospection*, *26*, 155–163. https://doi.org/10.1002/arp.1730

Davis, D. S. (2020). Defining what we study: The contribution of machine automation in archaeological research. *Digital Applications in Archaeology and Cultural Heritage 18*, e00152. https://doi.org/10.1016/j.daach.2020.e00152

Davis, D. S., Sanger, M. C., & Lipo, C. P. (2019). Automated mound detection using LiDAR and object-based image analysis in Beaufort County, South Carolina. *Southeast. Archaeol*, *38*, 23–37. https://doi.org/10.1080/0734578X.2018.1482186

Deforce, K., Boeren, I., Adriaenssens, S., Bastiaens, J., Keersmaeker, L. D., Haneca, K., Tys, D., & Vandekerkhove, K. (2013). Selective woodland exploitation for charcoal production. A detailed analysis of charcoal kiln remains (ca. 1300–1900 AD) from Zoersel (northern Belgium). *Journal of Archaeological Science*, *40*, 681–689. https://doi.org/10.1016/j.jas.2012.07.009

Deforce, K., Groenewoudt, B., & Haneca, K. (2020). 2500 years of charcoal production in the Low Countries: The chronology and typology of charcoal kilns and their relation with early iron production. *Quaternary International*. In press. https://doi.org/10.1016/j.quaint.2020.10.020

Doneus, M. (2013). Openness as visualization for interpretative mapping of airborne LiDAR derived digital terrain models. *Remote Sensing*, *5*, 6427–6442. https://doi.org/10.3390/rs5126427

Donovan, S., Ignatiadis, M., Ouimet, W., Dethier, D., & Hren, M. (2021). Gradients of geochemical change in relic charcoal hearths soils, Northwestern Connecticut, USA. *Catena*, *197*. https://doi.org/10.1016/j.catena.2020.104991

Gießelmann, U. C., Borchard, N., Traunspurger, W., & Witte, K. (2019). Long-term effects of charcoal on nematodes and other soil meso- and microfaunal groups at historical kiln-sites—A pilot study. *European Journal of Soil Biology*, *93*. https://doi.org/10.1016/j.ejsobi.2019.103095

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2014.81

Gocel-Chaltè, D., Guerold, F., Knapp, H., & Robin, V. (2020). Anthracological analyses of charcoal production sites at a high spatial

resolution: The role of topography in the historical distribution of tree taxa in northern Vosges mountains, France. *Vegetation History and Archaeobotany*. https://doi.org/10.1007/s00334-020-00769-z

Gong, Z., & Zhang, H. (2020). Research on gpr image recognition based on deep learning. *Matec Web of Conferences*, *309*(03027). https://doi.org/10.1051/matecconf/202030903027

Guyot, A., Hubert-Moy, L., & Lorho, T. (2018). Detecting Neolithic burial mounds from LiDAR-derived elevation data using a multi-scale approach and machine learning techniques. *Remote sensing*, *10*, 225. https://doi.org/10.3390/rs10020225

He, K., Gkioxari, G., Dollár, P., & Girshick R. B. (2017). Mask r-cnn. arXiv: 1703.06870v3.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2016.90

Henderson, P., & Ferrari, V. (2016). End-to-end training of object class detectors for mean average precision. In S. H. Lai, V. Lepetit, K. Nishino, & Y. Sato (Eds.), *Computer Vision – ACCV 2016. ACCV 2016. Lecture Notes in Computer Science*, 10115, 198-213. https://doi.org/10.1007/978-3-319-54193-8_13

Hesse, R. (2010). LiDAR-derived local relief models—A new tool for archaeological prospection. *Archaeological Prospection*, *17*, 67–72. https://doi.org/10.1002/arp.374

Hirsch, F., Schneider, A., Bonhage,, A., Raab, A., Drohan, P., & Raab, T. (2020). An initiative for a morphologic-genetic catalog of relict charcoal hearths from Central Europe. *Geoarchaeology*, in print. https://doi.org/10.1002/gea.21799

Johnson, J. W. (2018). Adapting mask r-cnn for automated nucleus segmentation. ArXiv abs/1805.00500.

Johnson, K., Ouimet, W. B., & Raslan, Z. (2015). Geospatial and LiDAR-based analysis of 18th to early 20th century timber harvesting and charcoal production in southern New England. *Geological Society of America Abstracts with Programs At: Bretton Woods, NH 47*(3), 65.

Ju, M., Luo, J., Zhang, P., He, M., & Luo, H. (2019). A simple and efficient network for small target detection. *IEEE Access*, *7*, 85771–85781. https://doi.org/10.1109/ACCESS.2019.2924960

Kazimi, B., Thiemann, F., & Sester, M. (2019). Object instance segmentation in digital terrain models. In M. Vento, & M. Percannella (Eds.), *Computer analysis of images and patterns. CAIP 2019. Lecture Notes in Computer Science* (pp. 488–495, 11679). https://doi.org/10.1007/978-3-030-29891-3_43

Kazimi, B., Thiemann, F., & Sester, M. (2020). Detection of terrain structures in airborne laser scanning data using deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V2-2020*. https://doi.org/10.5194/isprs-annals-V-2-2020-493-2020

Kokalj, Ž., & Somrak, M. (2019). Why not a single image? Combining visualisations to facilitate fieldwork and on-screen mapping. *Remote Sensing*, *11*(7), 747. https://doi.org/10.3390/rs11070747

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *60*(6), 84–90. https://doi.org/10.1145/3065386

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

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *Lecture Notes in Computer Science*, *8693*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2015.7298965

Luo, L., Xiong, Y., Liu, Y., & Sun, X. (2019). Adaptive gradient methods with dynamic bound of learning rate. *Proceedings of the 7th International Conference on Learning Representations*. arXiv: 1902.09843v1

Opitz, R., & Herrmann, J. (2018). Recent trends and long-standing problems in archaeological remote sensing. *Journal of Computer Applications in Archaeology*, *1*(1), 19–41. https://doi.org/10.5334/jcaa.11

Perona, P., & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, *12*(7), 629–639. https://doi.org/10.1109/34.56205

Pham, M.-T., & Lefèvre, S. (2018). Buried object detection from b-scan ground penetrating radar data using faster r-cnn. *GARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia 6804–6807. https://doi.org/10.1109/IGARSS.2018.8517683

Raab, A., Bonhage, A., Schneider, A., Raab, T., Rösler, H., Heußner, K.-U., & Hirsch, F. (2019). Spatial distribution of relict charcoal hearths in the former royal forest district Tauer (SE Brandenburg, Germany). *Quarternary International*, *511*, 153–165. https://doi.org/10.1016/j.quaint.2017.07.022

Raab, A., Takla, M., Raab, T., Nicolay, A., Schneider, A., Rösler, H., Heußner, K.-U., & Bönisch, E. (2015). Pre-industrial charcoal production in Lower Lusatia (Brandenburg, Germany): Detection and evaluation of a large charcoal-burning field by combining archaeological studies, GIS-based analyses of shaded relief maps an dendrochronological age determination. *Quatenary International*, *367*, 111–122. https://doi.org/10.1016/j.quaint.2014.09.041

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Risbøl, O., Bollandsås, O. M., Nesbakken, A., Ørka, H. O., Næsset, E., & Gobakken, T. (2013). Interpreting cultural remains on airborne laser scanning generated digital terrain models: Effects of size and shape on detection success rates. *Journal of Archaeological Science*, *40*, 4688–4700. https://doi.org/10.1016/j.jas.2013.07.002

Rutkiewicz, P., Malik, I., Wistuba, M., & Osika, A. (2019). High concentration of charcoal hearth remains as legacy of historical ferrous metallurgy in southern Poland. *Quarternary International*, *512*, 133–143. https://doi.org/10.1016/j.quaint.2019.04.015

Schmidt, M., Mölder, A., Schönfelder, E., Engel, F., & Fortmann-Valtnik, W. (2016). Charcoal kiln sites, associated landscape attributes and historic forest conditions: DTM-based investigations in Hesse (Germany). *Forest Ecosystems*, *3*(8). https://doi.org/10.1186/s40663-016-0067-6

Schneider, A., Bonhage, A., Raab, A., Hirsch, F., & Raab, T. (2020). Large-scale mapping of anthropogenic relief features—Legacies of past forest use in two historical charcoal production areas in Germany. *Geoarchaeology*, *35*(4), 545–561. https://doi.org/10.1002/gea.21782

Schneider, A., Hirsch, F., Bonhage, A., Raab, A., & Raab, T. (2020). The soil moisture regime of charcoal-enriched land use legacy sites. *Geoderma*, *366*. https://doi.org/10.1016/j.geoderma.2020.114241

Schneider, A., Takla, M., Nicolay, A., Raab, A., & Raab, T. (2014). A template-matching approach combining morphometric variables for automated mapping of charcoal kiln sites. *Archaeological Prospection*, 22(1), 45–62. https://doi.org/10.1002/arp.1497

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556v6

Smidt, E., Tinter, J., Klemm, S., & Scholz, Z. (2017). FT-IR spectral and thermal characterization of ancient charcoals—A tool to support archaeological and historical data interpretation. *Quaternary International*, *457*, 43–49. https://doi.org/10.1016/j.quaint.2016.11.031

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on nuclear science*, *44*(3), 1464–1468. https://doi.org/10.1109/23.589532

Sorokin, A. (2018). Lesion analysis and diagnosis with mask-rcnn. arXiv: 1807.05979v2

Tarolli, P., Cao, W., Sofia, G., Evans, D., & Ellis, E. C. (2019). From features to fingerprints: A general diagnostic framework for anthropogenic geomorphology. *Progress in Physical Geography*, *43*(1), 95–128. https://doi.org/10.1177/0309133318825284

Tolksdorf, J. F., Kaiser, K., Petr, L., Herbig, C., Kočár, P., Heinrich, S., Wilke, F. D. H., Theuerkauf, M., Fülling, A., Schubert, M., Schröder, F., Křivánek, R., Schulz, L., Schulz, A., & Hemker, C. (2020). Past human impact in a mountain forest: Geoarchaeology of a medival glass production and charcoal hearth site in the Erzgebirge, Germany. *Regional Environmental Change*, *20*, 71. https://doi.org/10.1007/s10113-020-01638-1

Trier, Ø. D., Cowley, D. C., & Waldeland, A. U. (2018). 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), 165–175. https://doi.org/10.1002/arp.1731

Trier, Ø. D., Reksten, J. H., & Løseth, K. (2021). Automated mapping of cultural heritage in Norway from airborne lidar data using faster R-CNN. *International Journal of Applied Earth Observations and Geoinformation*, *95*. https://doi.org/10.1016/j.jag.2020.102241

Tzutalin. (2015). LabelImg. Git code. https://github.com/tzutalin/labelImg

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), 31–40. https://doi.org/10.5334/jcaa.32

Verschoof-van der Vaart, W. B., Lambers, K., Kowalczyk, W., & Bourgeois, Q. P. J. (2020). Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from The Netherlands. *ISPRS International Journal of Geo-Information*, *9*(5), 293. https://doi.org/10.3390/ijgi9050293

Witharana, C., Ouimet, W. B., & Johnson, K. M. (2018, GIScience & Remote Sensing). Using LiDAR and GEOBIA for automated extraction of 18th-late 19th century relict charcoal hearths in southern New England. *55*(2), 183–204. https://doi.org/10.1080/15481603.2018.1431356

Yu, Y., Zhang, Y., Yang, K., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural evironment based on mask-rcnn. *Computers and Electronics in Agriculture*, *163*, 10484. https://doi.org/10.1016/j.compag.2019.06.001

Zagoruyko, S., Lerer, A., Lin, T.-Y., Pinheiro, P. O., Gross, S., Chintala, S., & Dollár, P. (2016). A multipath network for object detection. *Proceedings of the British Machine Vision Conference (BMVC)*, 15.1–15.12. https://dx.doi.org/10.5244/C.30.15

Zakšek, K., Oštir, K., & Kokalj, Ž. (2011). Sky-view factor as a relief visualization technique. *Remote Sensing*, *3*(2), 398–415. https://doi.org/10.3390/rs3020398

How to cite this article: Bonhage A, Eltaher M, Raab T, Breuß M, Raab A, Schneider A. A modified Mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland. *Archaeological Prospection*. 2021;28:177–186. https://doi.org/10.1002/arp.1806