Note

Automating archaeological object detection

Proof of concept – Arran survey

Note no. SAMBA/08/18

Authors Øivind Due Trier (NR), Anders Ueland Waldeland (NR), Dave Cowley (HES)

Date 18 May 2018

Authors

Øivind Due Trier is senior research scientist at the Section for Earth Observation, Norwegian Computing Center (in Norwegian: Norsk Regnesentral (NR)). He is developing automated processing methods for various remote sensing data, including airborne laser scanning data, airborne hyperspectral data, optical satellite data and synthetic aperture radar satellite data. Applications include cultural heritage mapping, daily monitoring of snow cover, forest vegetation mapping, and oil spill monitoring.

Anders Ueland Waldeland is research scientist at the Section for Earth Observation, Norwegian Computing Center. He is developing image analysis and machine learning for applications within remote sensing, computer vision, and geophysical and marine data. Methodologically, his main focus is on deep learning and gradient based optimization.

Dave Cowley is at the Aerial Survey and Remote Sensing Team, Historic Environment Scotland (HES). He began working on archaeological field survey projects across Scotland in 1989, taking over the management of the Aerial Survey Programme in 2005. Research interests and publications focus on the development of the landscape and approaches to survey, reflected in edited books on aerial archaeology, historic aerial photographs, remote sensing for heritage management and 3D data for archaeology. Co-author of the popular Above Scotland aerial photograph book, with many of his aerial photographs included in Scotland's Landscapes, and published extensively on prehistoric settlement, on aerial survey methodology and on applications of historic aerial imagery. Co-chair of the European Archaeological Council working party on Remote Sensing for Archaeological Heritage Management, and an honorary member of staff in the Department of Archaeology, Glasgow University.

Norsk Regnesentral / Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Historic Environment Scotland / Àrainneachd Eachdraidheil Alba

Historic Environment Scotland is the lead public body set up to investigate, care for and promote Scotland's historic environment. The Aerial Survey and Remote Sensing Team focuses on two main strands of research. The first is improving knowledge in areas where the aerial perspective is especially powerful. For example, many of the known archaeological sites in the Scottish Lowlands have been levelled by agriculture. As a result, most of these are only visible from the air. The second is to actively research emerging technologies that have the potential to aid us in our work programmes. These include: 3D topographic data produced by airborne laser scanning; hyperspectral imaging, which could help us to record sites that are invisible to the naked eye; and more speculative areas such as synthetic aperture radar. Through our aerial survey and remote sensing work, we can better: improve our knowledge base through the discovery of previously unknown sites, understand landscape changes, assess the impact of climate and coastal changes on the historic environment, understand the condition and management of sites, and improve the efficiency of survey methodology.

Title	**Automating archaeological object detection**
Authors	**Øivind Due Trier (NR), Anders Ueland Waldeland (NR), Dave Cowley (HES)**
Quality assurance	Øivind Due Trier
Date	18 May 2018
Year	2018
Publication number	SAMBA/08/18

Abstract

This project seeks to develop heavily automated analysis of digital topographic data to extract archaeological information and to expedite the creation of national-scaled mapping.

Drawing on developments in computer vision this has the potential to fundamentally recast the capacity of archaeological prospection and survey to cover large areas and deal with mass data, breaking a dependency on human resource. Without such developments the potential of the vast amount of archaeological information embedded in large topographic and image-based datasets cannot be realised to inform our knowledge and understanding of Scotland's Historic Environment. A heavily automated computational approach and the increasing availability of large datasets put the creation of systematic national-scaled archaeological mapping of Scotland within reach.

The purpose of this proof of concept project was to run an assessment of existing developments in a Norwegian case study against digital topographic data for Arran, providing outputs that may be assessed for their applicability at a national scale.

Keywords	Deep learning, neural networks
Target group	Cultural heritage managers and researchers, remote sensing researchers
Availability	Open
Project number	220 836 Pilot Arran
Research field	Remote sensing, cultural heritage
Number of pages	34

Table of Content

List of figures

1 Introduction

Archaeological prospection and survey has long relied on human observation, whether in the field or through desk-based work, for the identification of objects of interest. Rates of coverage are inherently limited by the availability of human resource. This means that achieving a systematic national mapping of Scotland's archaeological remains is a distant prospect, unachievable over even many decades. At the same time, extensive high-resolution topographic data is becoming available at such a rate that it has outstripped the capacity of human observer-based approaches to explore it. However, developments in computer vision offer a way forward to efficiently and rapidly explore this data and identify archaeological information. For archaeological survey, computer vision offers the potential for a step-change in rates of coverage, and a mechanism to exploit the vast amount of archaeological information embedded in large topographic and image-based datasets.

The Scottish Government will be acquiring complete coverage of Airborne Laser Scanning (ALS, aka LiDAR) data, and as a 3D digital topographic dataset, this carries enormous potential for archaeological mapping. However, to exploit this potential requires development of analytical methods that can efficiently deal with the mass data. Computational approaches drawing on Convolutional Neural Networks (CNNs) offer a way forward for archaeological prospection, demonstrated in pioneering work by the Norwegian Computing Center (https://www.nr.no/nb/node/849). CNNs require minimal pre-processing, drawing on learning sets to 'look' at data in a manner inspired by the organization of the animal/human visual cortex.

The case addresses the full range of HES Corporate Plan strategic themes, most specifically:

1. Understand: Aspects of Our Place in Time and Scotland's Archaeology Strategy cannot be progressed without better knowledge of where the material remains of past activities survive. This knowledge gap limits HES's ability to tell Scotland's story in a comprehensive way.

2. HES's remote sensing, archaeological survey and digital documentation are well-established, and the national record of historic environment (NHRE) is recognised as a high-quality record. It is, however, a partial record built up piecemeal over more than a century. The proposed approach to mass data analysis provides a mechanism to explore topographic data and imagery to provide a comprehensive, truly national scale of coverage within a relatively short period of time (perhaps a decade).

3. Lead: In carrying out this pilot automated archaeological object detection project, HES would play a lead role in providing comprehensive information about Scotland's historic environment. At a worldwide level it would provide an exemplar of how comprehensive national and regional databases can be built in a resource-effective way that shows the benefits of computational approaches and mass data provision.

4. This proposal represents a significant advance in archaeological prospection, developing from the human observer-based approach of the 20th century to one that will expedite object detection in vast datasets, improving knowledge of the historic environment and increasing the capacity of archaeological survey to cover very large areas.

5. Protect: Systematic survey in most parts of Scotland generates large increases (e.g. up to 10 fold) in the numbers of known monuments. However, only about 10% of the country has been covered in this way, and that means that there are hundreds of thousands of unrecorded monuments preserved in the micro-topography of the landscape. The proposed approach could pave the way to providing this base-level record in a matter of years, and this would provide the means, through knowledge of what survives in the landscape, to better understand, manage and protect our archaeological assets.

The Scottish Government wishes to actively promote the wider, deeper relationships with Nordic and Baltic Countries, including Norway, for innovative, environmentally sustainable solutions to shared problems. This proposal fits within that framework.

Norsk Regnesentral (NR) has previously developed semi-automatic methods for the detection and mapping of cultural heritage remains in airborne laser scanning (ALS) data, including pitfall traps in deer hunting systems (Trier and Pilø, 2012), grave mounds (Trier, Zortea and Tonning, 2015; Trier, Pilø and Johansen, 2015), charcoal burning pits in iron extraction sites (Trier and Pilø, 2015), and charcoal burning platforms (Trier, Pilø and Johansen, 2015; Trier, Salberg and Pilø, 2018). Also, NR has developed semi-automatic mapping of levelled grave mounds in cereal fields from optical very high resolution satellite data (Trier, Larsen and Solberg, 2009). The purpose of cultural heritage mapping is twofold: (1) to increase the understanding of the past, and (2) to reduce the negative impact of modern land use on the cultural remains. The vision is to develop methods that may be used in a national infrastructure for semi-automatic mapping of cultural heritage (Kermit, Hamar and Trier, 2018)

Two different strategies have been used in the above research:

1. Template matching

2. Deep convolutional neural networks

Template matching was successful for the detection of pitfall traps on sand deposits south of the lake Olstappen in Nord-Fron municipality, Oppland County, Norway (Trier and Pilø, 2012). In this landscape, the pitfall traps stood out as unique, man-made structures in the digital terrain model (DTM) obtained from the ALS data. However, the template matching also gave some false positives. Many of these were removed by including additional tests on the shape of the DTM in a local neighbourhood.

Template matching was then used to automatically detect pit and mound structures for semi-automatic mapping of iron extraction sites, grave mounds and charcoal burning platforms. The results were less convincing than for the pitfall traps. There seemed to be two reasons for this:

1. The structures to detect were less distinct than the pitfall traps

2. The structures to detect were more similar to natural terrain features than in the case of the pitfall traps.

The latter explanation did, however, not make sense for the charcoal burning platforms. To a human observer, these stood out as unique. The problem was rather that they had many different appearances in the DTM, and that it was difficult to construct a suitable template or a collection of templates. The strategy that did work to some extent was to apply both heap and pit detection, using small templates for pits and large templates for heaps. Then, many charcoal burning platforms were detected as one of:

1. A central mound with some pits along the circumference

2. A central mound only

3. Some pits in a circular arrangement

However, many charcoal burning platforms were also missed.

Recent advances in computer vision, using deep convolutional neural networks, led us to consider that as an alternative approach. Using a network that had been pre-trained on a million natural images, we discarded the last layer and replaced it with a support vector machine classifier. By training it on 400 examples of charcoal burning platforms and 10,000 random terrain locations, 86% of the true charcoal burning platforms were correctly detected, versus 70% for the template matching approach. The false positive rate was 37%, versus 72% for template matching.

The method was implemented on the Caffe library. The main limitation of our implementation is that it is very slow, in the order of several hours per 1 km by 1 km of DTM data.

Recently, the PyTorch library has emerged as a better alternative than Caffe, offering more flexibility in training and classification.

2 Data

ALS data of all of the Isle of Arran, Scotland were used in this study. Arran is about 30 km long north-south and around 15-20 km wide east-west (Figure 1). The total area is 456 km^2.

The ALS data is organized in 1 km by 1 km tiles of (x,y,z) point data, each point labelled as one of six classes, including ‘ground’, ‘building’ and ‘vegetation’. The average number of ‘ground’ points per square metre was 2.75. However, this varied across the 489 tiles. The per-tile averages varied from 0.43 to 7.44. The main reason for low ‘ground’ point density is dense vegetation, preventing the laser pulses to reach the ground surface. However, the majority of Arran is open land with low vegetation (Figure 2). Another reason low ‘ground’ point density in some areas is the presence of buildings.

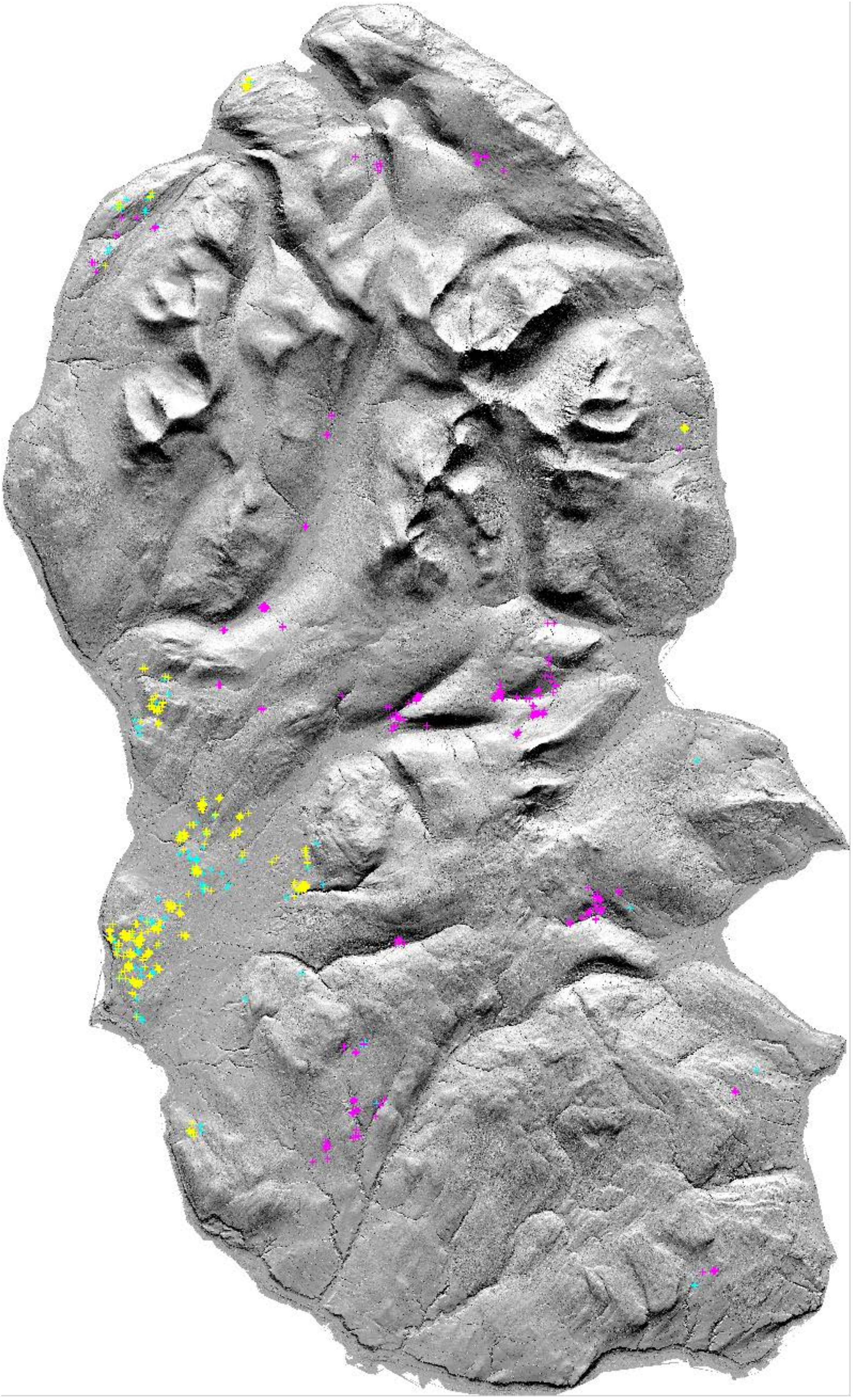


Figure 1. Hillshade relief visualisation of the digital terrain model of Arran, with learning set locations superimposed: cyan=roundhouse, magenta=shieling, yellow=small cairn.

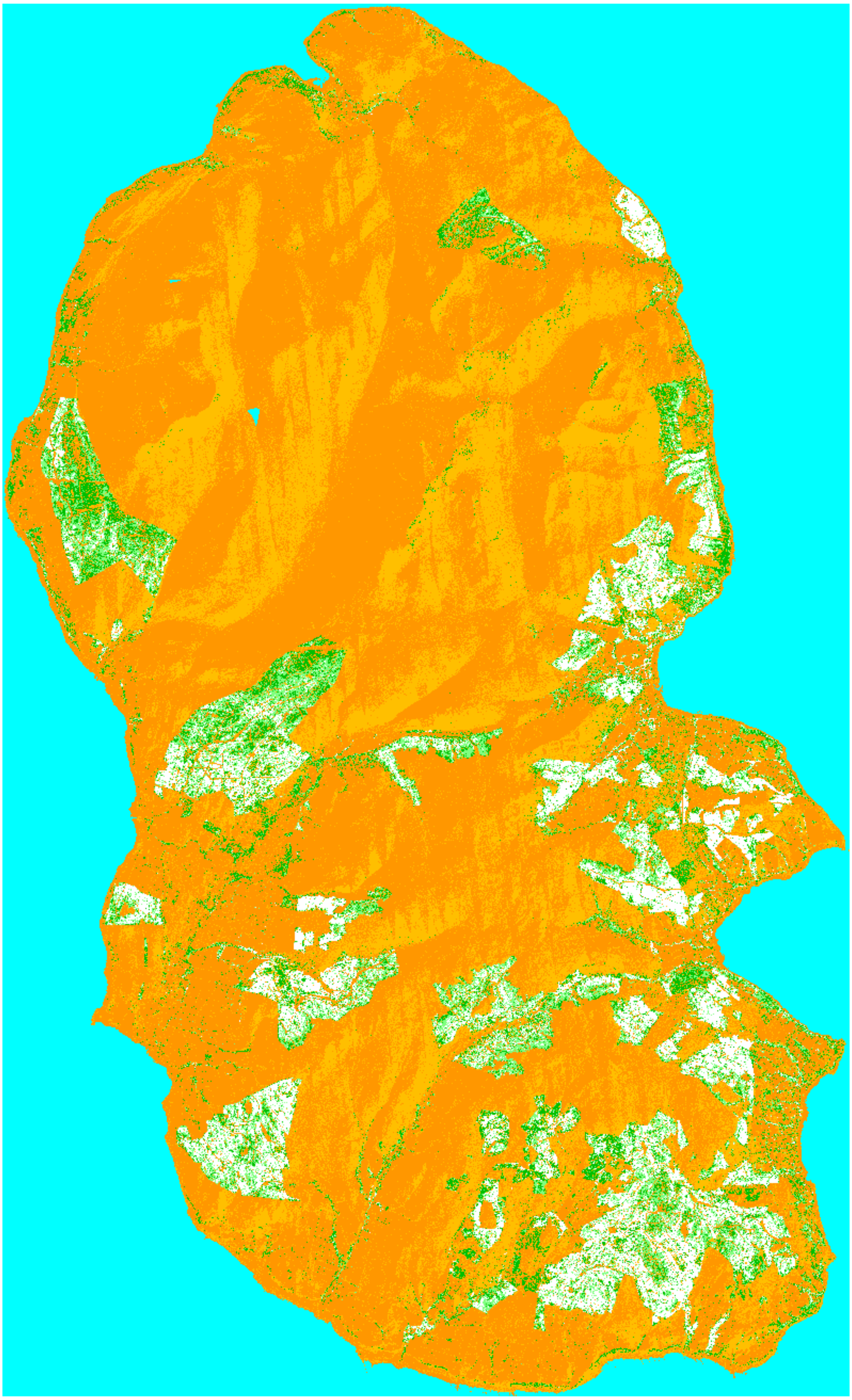


Figure 2. Vegetation height classes. Orange: 0-0.1 m, yellow: 0.1-2.0 m, dark green: 2-7 m, light green: 7-12 m, white: > 12 m.

From previous field work and visual inspection of the data, several hundred locations of historical structures and some modern structures had been identified (Table 1). E.g., within an 1100 m × 835 m area (Figure 3), 13 roundhouse structures and 14 small cairns have been identified (Figure 4).

Table 1. Archaeological and modern structures in the learning set.

type of structure	count
roundhouse	121
shieling	267
small cairn	384
burial cairn	6
burnt mound	24
cattle feed stance	24
enclosure	11
horse platform	1
possible kiln	5
rectangular building	15

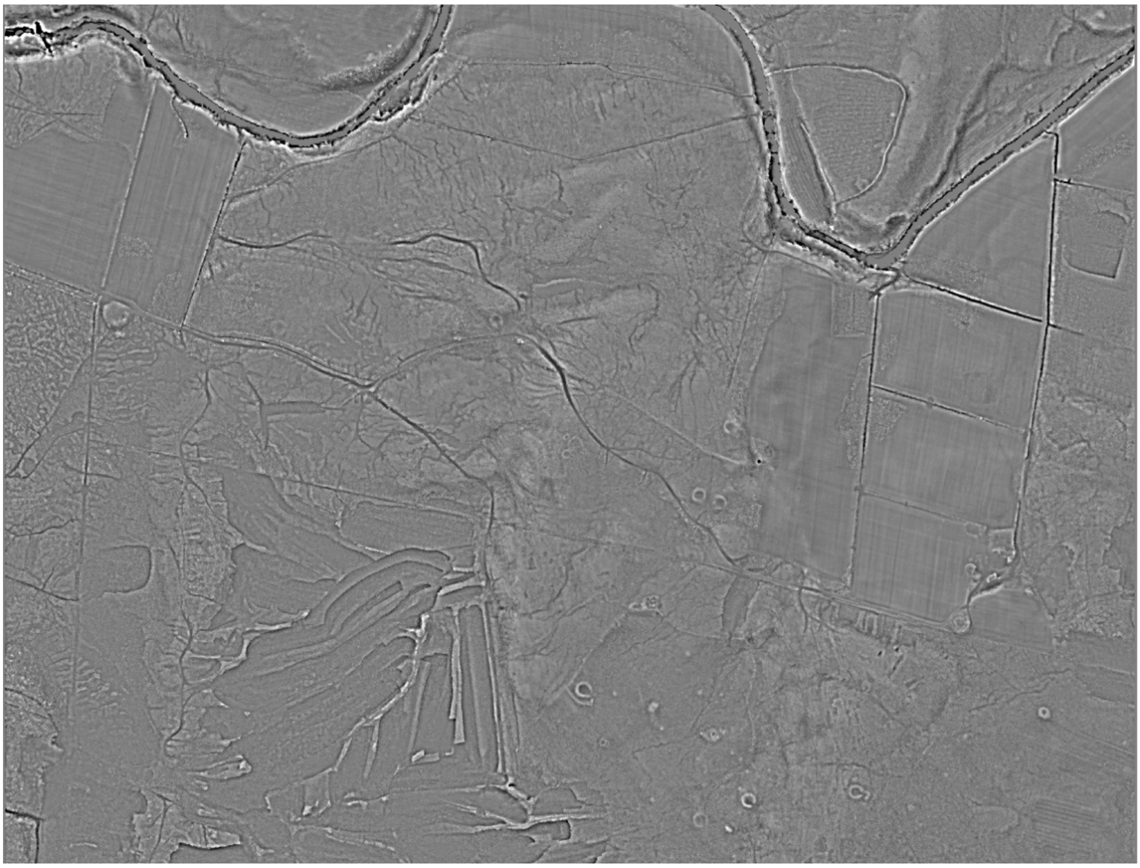


Figure 3. Detail of DTM, local relief visualisation, with visible, circular roundhouse structures. The 1100 m × 835 m image is centred on 190,600 east, 632,450 north (OSGB national grid).

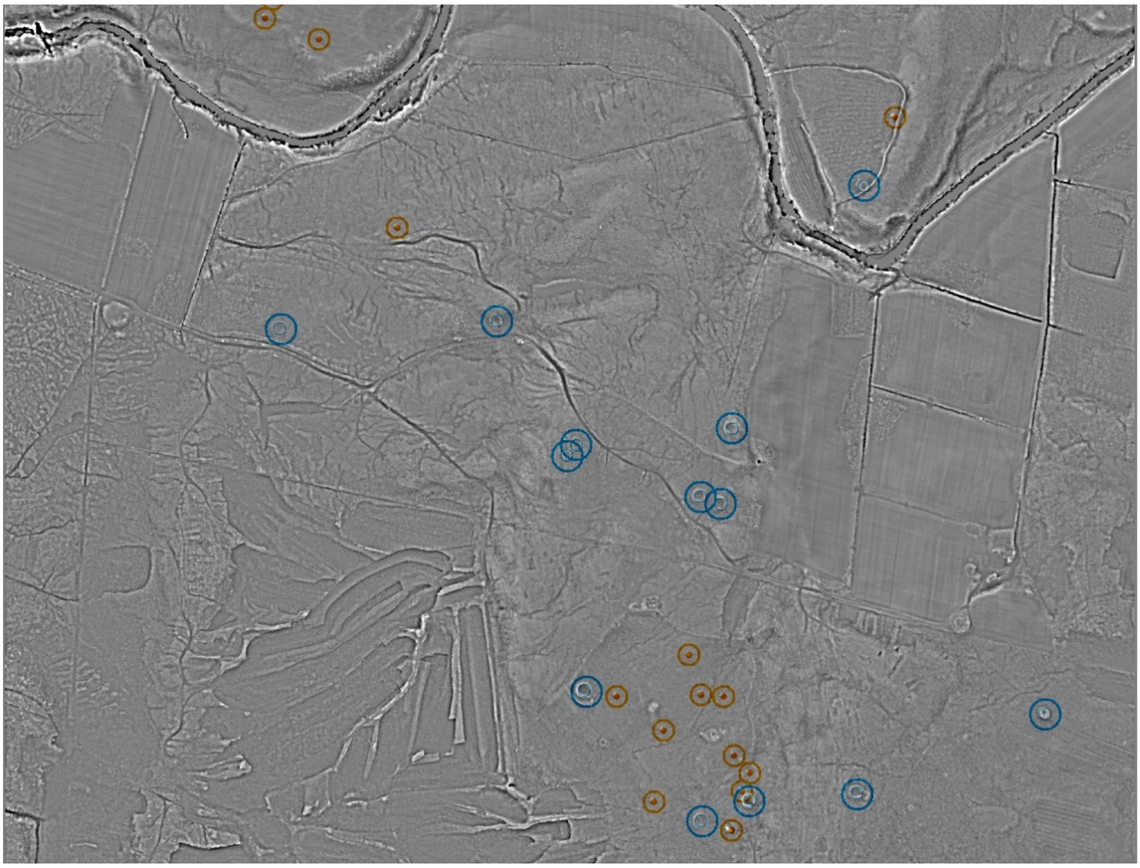


Figure 4. Same as Figure 3, with verified roundhouses (blue) and small cairns (brown).

3 Methods

3.1 Preprocessing of ALS data

In order to allow for detection of cultural heritage structures at tile boundaries, each tile was extended by including data from neighbouring tiles within 50 meters from the tile boundary.

Then, for each extended tile, all ALS points labelled as ‘ground’ were used to create a digital terrain model (DTM) at 0.25 m grid spacing. For this, the ENVI (https://www.harrisgeospatial.com) functions TRIANGULATE and TRIGRID were used.

The DTM contains elevation values in meters above sea level. Since we are interested in local elevation differences and not absolute terrain height values, a smoothed version of the DTM was subtracted from the DTM, thus producing a local terrain model (simplified version of local relief model (Hesse, 2010)). The smoothing was done for each pixel by taking the mean value within a 30 × 30 pixels sliding window, i.e., 7.5 m × 7.5 m. The resulting image then contained local elevation deviations from the general, smoothed terrain surface. The values were truncated to the range -2 m to +2 m.

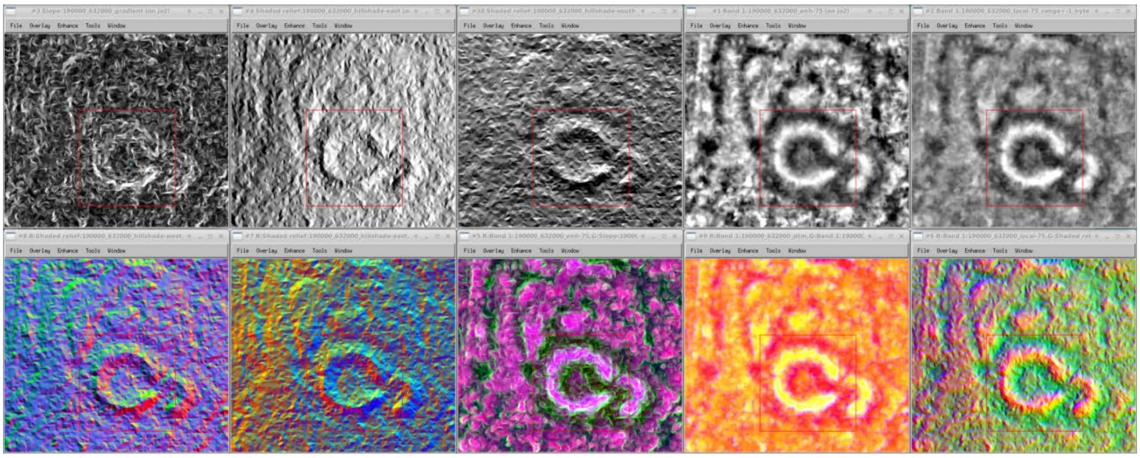


Figure 5. Some alternative DTM visualisations. Top row, from left: (a) gradient, i.e. local steepness, with steep locations in white and flat areas in black; (b) hillshade relief, with illumination from the east; (c) hillshade relief, with illumination from the north; (d) contrast enhancement, i.e., normalized to a fixed mean value and standard deviation within a sliding window; (e) local terrain model (see text). Bottom row: combinations of three visualisations, from left: (f) hillshade-east, hillshade-north, gradient, (g) hillshade-west, hillshade-north and hillshade-east; (h) local, gradient, enhanced; (i) DTM, local, gradient, (j) local, hillshade-east, hillshade-north.

There are other DTM visualisations (Figure 5) that could have been used instead of, or combined with, the local terrain model. Combinations of three bands are possible with the neural network, since it is designed for natural colour images with red, green and blue channels. We have chosen to simply use the local terrain visualization, which is then replicated for the three channels, thus mimicking a grey scale photograph.

3.2 Neural network design

The ResNet18 implementation in pyTorch was used as a starting point. pyTorch includes a tutorial with a ready-to-use implementation which allows the user to take a pretrained network as a starting point, and to refine it with the user's own data. The neural network is pretrained on the ImageNet database of 1.2 million images of natural scenes, each image labelled with one or several labels denoting image content. The total number of unique labels is about 1000.

The ResNet is designed to work on images of size 224 × 224 pixels. We modified the ResNet to accept smaller input images than 224. The input layer acts as a 7 × 7 array of image feature detectors, each of size 32 × 32 pixels. By using 2 × 2 or 3 × 3 feature detectors, input image sizes would be 64 × 64 or 96 × 96 pixels, respectively.

3.3 Training of neural network

It is recommended to run the training phase on a computer with a graphics processing unit (GPU), otherwise training may be very slow.

3.3.1 Extraction of training images

The training of the neural network was done on image extracts of size 101 × 101 pixels centered on known locations of roundhouses, shielings and small cairns. Samples of the 'background' terrain, i.e., excluding any 'foreground' locations, were also extracted, also of size 101 × 101 pixels. To avoid the foreground structures, buffer zones around these locations were made and collected in a mask layer. The locations used for buffer zoning included roundhouses, shielings and small cairns, and also burial cairns, burnt mounds, cattle feed stances, enclosures, horse platforms, possible kilns, and rectangular buildings. A random number generator was used to select x and y coordinates of background locations. Background locations within the buffer zone mask were discarded. The background locations were extracted from the 1 km × 1 km tiles that contained known locations of roundhouse, shieling and small cairn.

The image extracts were divided in two groups: 'training' and 'validation'. The purpose was that the neural network will learn its internal parameters from the training data, and evaluate its detection performance on the validation data, to prevent overfitting on the training data. Overfitting means that the classifier recognizes the training data but may perform badly on data that it hasn't encountered during training.

The number of background examples was large for three reasons:

1. 'background' is the most frequent situation in the landscape

2. there are many different natural terrain structures that we didn't want the neural network to classify as one of the archaeological object types

3. a large variety of background examples may reduce the number of false positives. A false positive is a location that the neural network predicts as being one of the cultural heritage types, but is in fact not.

However, if the number of background examples was too large, the training phase took too long time to complete, thus precluding us from experimenting with various parameter settings, run several training phases and compare the results.

3.3.2 Image augmentations

The neural network contains a large number of parameters that need to be trained. A technique to increase the number of learning examples is to let the neural network perform changes to the image examples in the 'training' set, thus producing a larger, augmented training set. The images in the 'validation' set were, however, kept unchanged. The following image changes were allowed on the training data:

1. Horizontal flip (yes/no)
2. Rotation by 0, 90, 180 or 270 degrees
3. Random scaling 0.95 – 1.00 while keeping the aspect ratio
4. Random scaling 0.95 – 1.00 without keeping thee aspect ratio
5. Random translation by 0, 1 or 2 pixels either to the north or to the south.
6. Random translation by 0, 1 or 2 pixels either to the west or to the east

3.3.3 Image cropping

The training images were centre cropped to size 96 x 96 pixels (for roundhouse detection) or 64 × 64 pixels (for shieling and small cairn detection).

3.3.4 Training and validation iterations

The deep convolutional neural network has many parameters that must be estimated from the training data. The training is done by randomly changing parameters and evaluating the network by classification performance on the validation data. One full iteration is called an epoch. After each epoch, the neural network internal parameters and classification performance are stored.

3.3.5 Storing the result of training

At the end of the training phase, the neural network internal parameters, which were obtained from the epoch that produced the best validation result, were saved to a file. This file could then later be read into the ResNet to restore the exact state of the classifier.

3.4 Running the neural network on entire tiles

The ResNet may be run on different images sizes. It uses a sliding window with 32 pixels (8 metres) stride, thus producing a classification result for each 32 × 32 pixels image block.

The maximum input image size was smaller than 4096 ×4096 pixels on the particular computer jocuda at NR. The maximum image size may be smaller or larger on other computers.

In practice, with 1.1 km × 1.1 km extended tiles, they need to be divided into overlapping sub-tiles. Each sub-tile is fed into the neural network classifier. The lower resolution classification results are expanded to the original image resolution. The overlaps are removed and the classification results are merged into a raster file of the same size as the input extended tile. Finally, the overlaps between extended tiles are removed from the classification results, thus producing non-overlapping classification results.

The classification results are currently, for each input tile, one raster image for each archaeological object type, with values between 0.0 and 1.0.

3.5 Converting classification results

With three archaeological object types: roundhouse, shieling and small cairn, the three classification results may be combined into a single RGB image. The following colour coding was used:

1. while = background
2. cyan = roundhouse
3. magenta = shieling
4. yellow = small cairn.

4 Results

By running the automatic method on the local relief visualisations, predicted locations of roundhouses, shielings and small cairns were obtained (Figure 6).

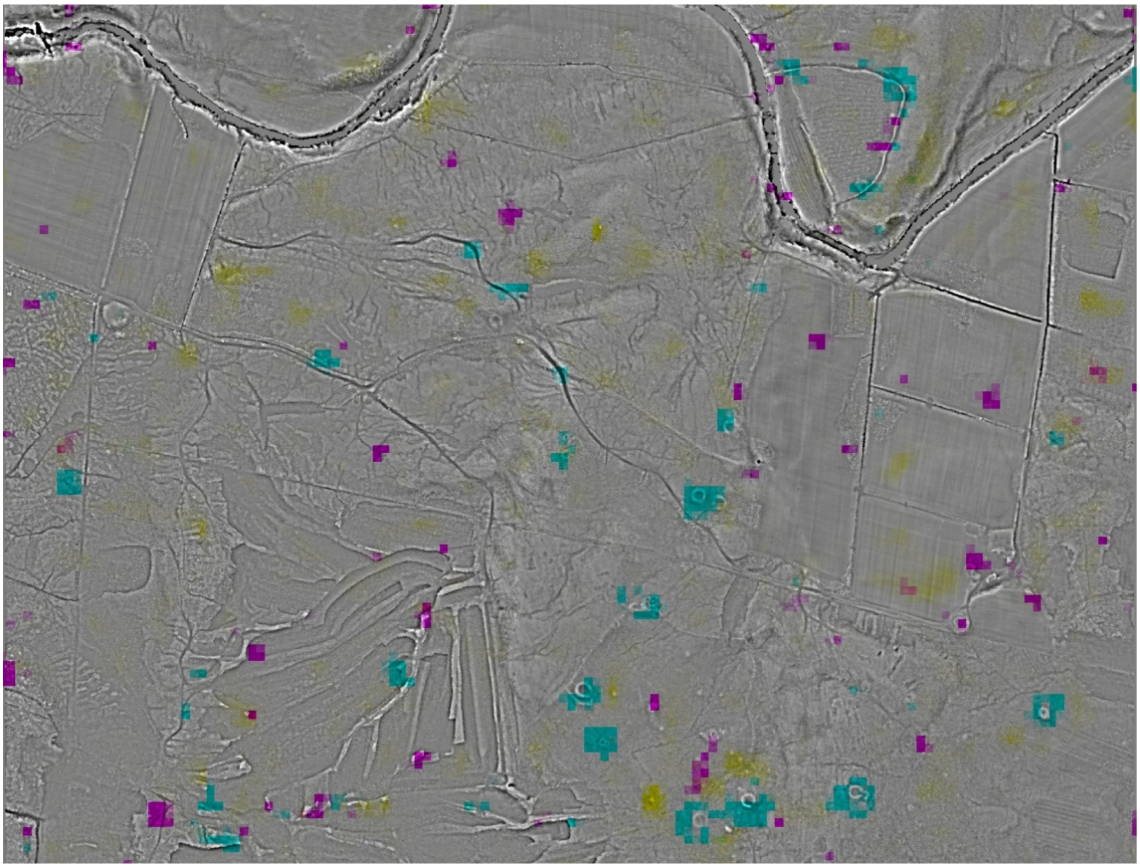


Figure 6. Same as Figure 3, with automatic detections of roundhouses (cyan), shielings (magenta) and small cairns (yellow).

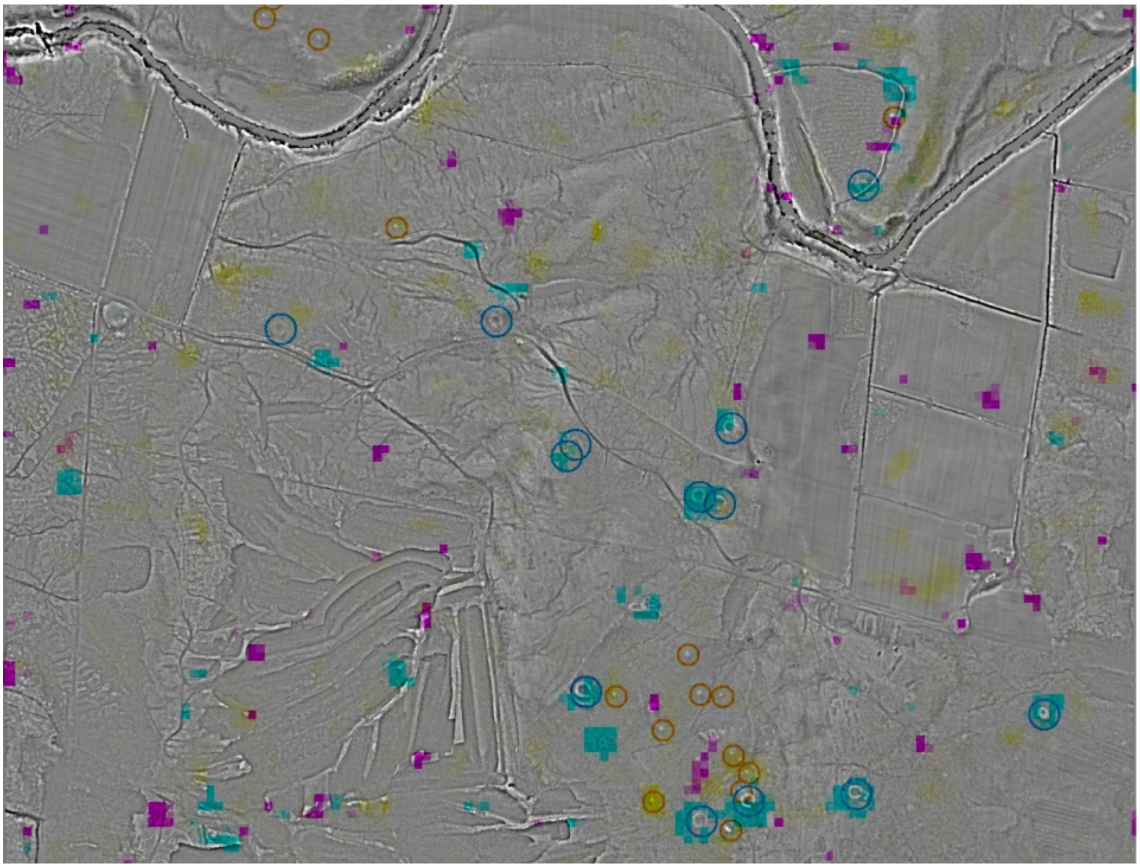


Figure 7. Same as Figure 6, with verified roundhouses (blue) and small cairns (brown).

By comparing the automatic predictions with verified locations of the same types of archaeological structure (Figure 5), it appears that the roundhouse predictions are quite meaningful, in the sense that many true roundhouse locations are identified, and that the number of false positives is not too large. On the other hand, some false negatives also occur, i.e., some true locations of roundhouses are missed by the automatic predictions.

Some of the small cairn predictions match verified locations (Figure 7), but many verified locations are also missed.

For an area with no verified shieling locations (Figure 7) a quite large number of false shieling predictions occur. For another area, containing several shielings, only a few were correctly identified (Figure 8). Also, the number of false predictions was high.

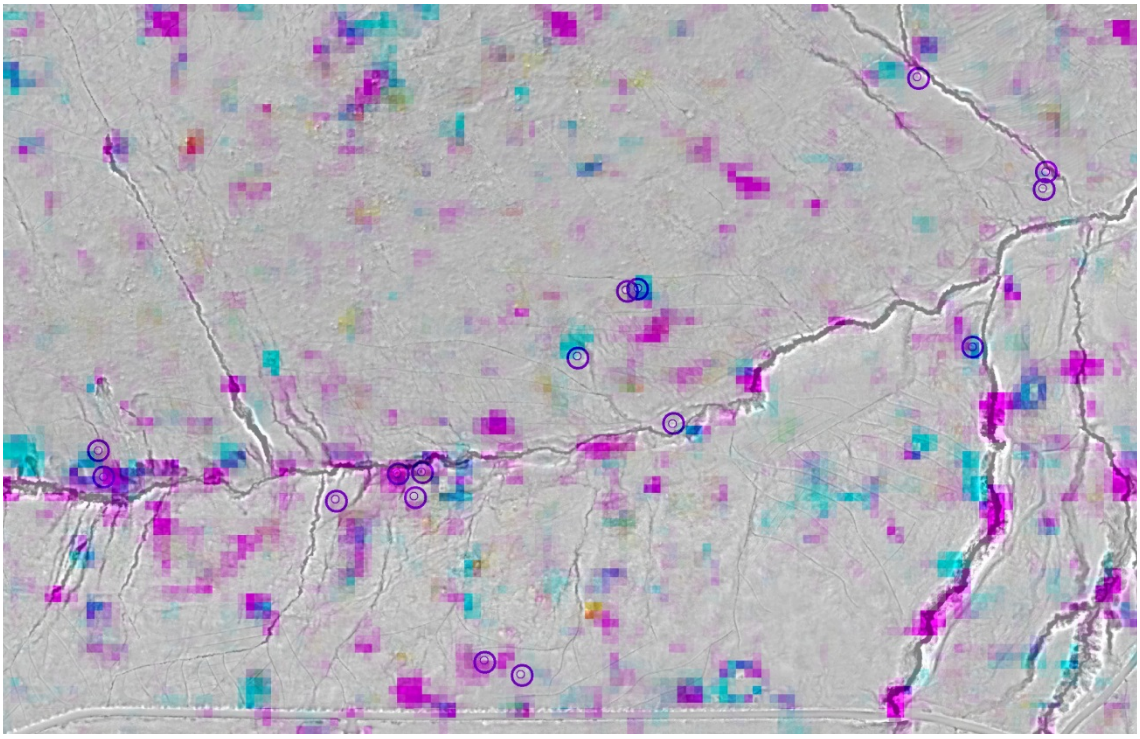


Figure 8. Detection results (coloured overlay) for an area with verified shielings (purple circles). The 1100 m × 705 m image is centred on 198,500 east, 636,355 north (OSGB national grid).

Before arriving at the present version of the classifier, some parameters were varied to see the effect on the classification results. Here are some of the main findings:

1. 0.1 m pixel size seems to give too much attention to unimportant detail in the ALS data.

2. 0.5 m pixel size seems to make the archaeological structures too small in the learning set

3. 0.2 m and 0.25 m pixel sizes seem to be appropriate for the archaeological structures in this study.

4. 224 × 224 pixels is the image size of the original ResNet code. However this resulted in larger regions of each class in the detection results, thus even less precise location prediction. Smaller images also include less background terrain, but it is unclear if the amount of background terrain within each subimage is an issue.

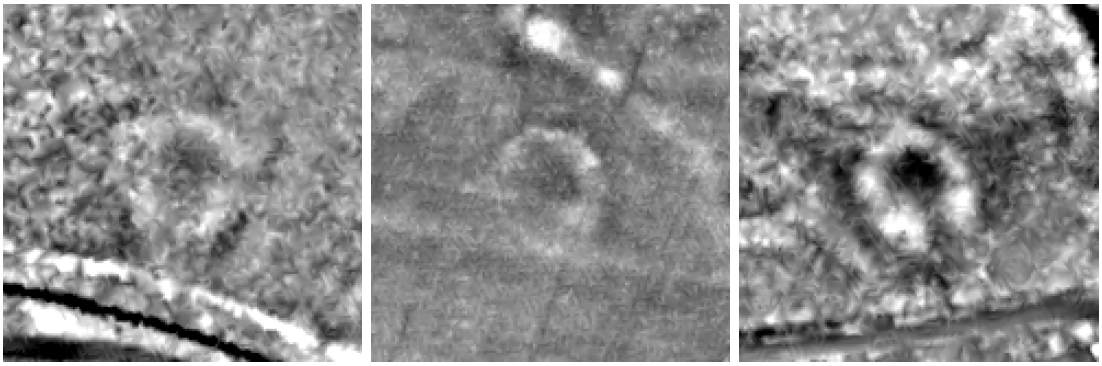
5. Smaller subimage size makes the training phase faster, thus allowing for more experiments.

6. Thus, a number of parameters were indeed varied; however, none of the observed combinations resulted in a clearly superior result.

7. The budget of the pilot study did not allow for a comprehensive, systematic, detailed documentation of all combinations and the resulting classification accuracy. However, we have other suggestions for improvement that may be more important to investigate.

4.1 Detailed inspection of some individual monuments

By inspecting the DTM visualization at roundhouse locations that the automatic method missed (Figure 9), one may gain some insight into what kinds of mistake the classifier makes. In two of the cases (Figure 9a-b), the roundhouse structures appear weak in the DTM visualization. However, in another case (Figure 9c), the structure is not as weak. In all three cases, linear structures appeared close to the roundhouse.



(a) (b) (c)

Figure 9. Roundhouse locations that the automatic method missed. The roundhouse locations are at OSGB coordinates: (a) 190,217 east, 632,640 north; (b) 190,300 east, 633,013 north; and (c) 190,428 east, 632,648 north. The image portions are 80 m × 80 m.

On the other hand, the majority of roundhouse locations in Figure 7 were correctly identified by the automatic method (Figure 10). In two of these cases (Figure 10a, i), there were linear structures close to the roundhouses. Thus, the presence of linear structures does not preclude correct identification of roundhouses. Two of the correctly identified roundhouse structures (Figure 10h, i) are overlapping other structures. They also have varying widths of the circular arcs. However, it is difficult to state precisely why the roundhouses in Figure 10h-i were detected and the one in Figure 9c was not.

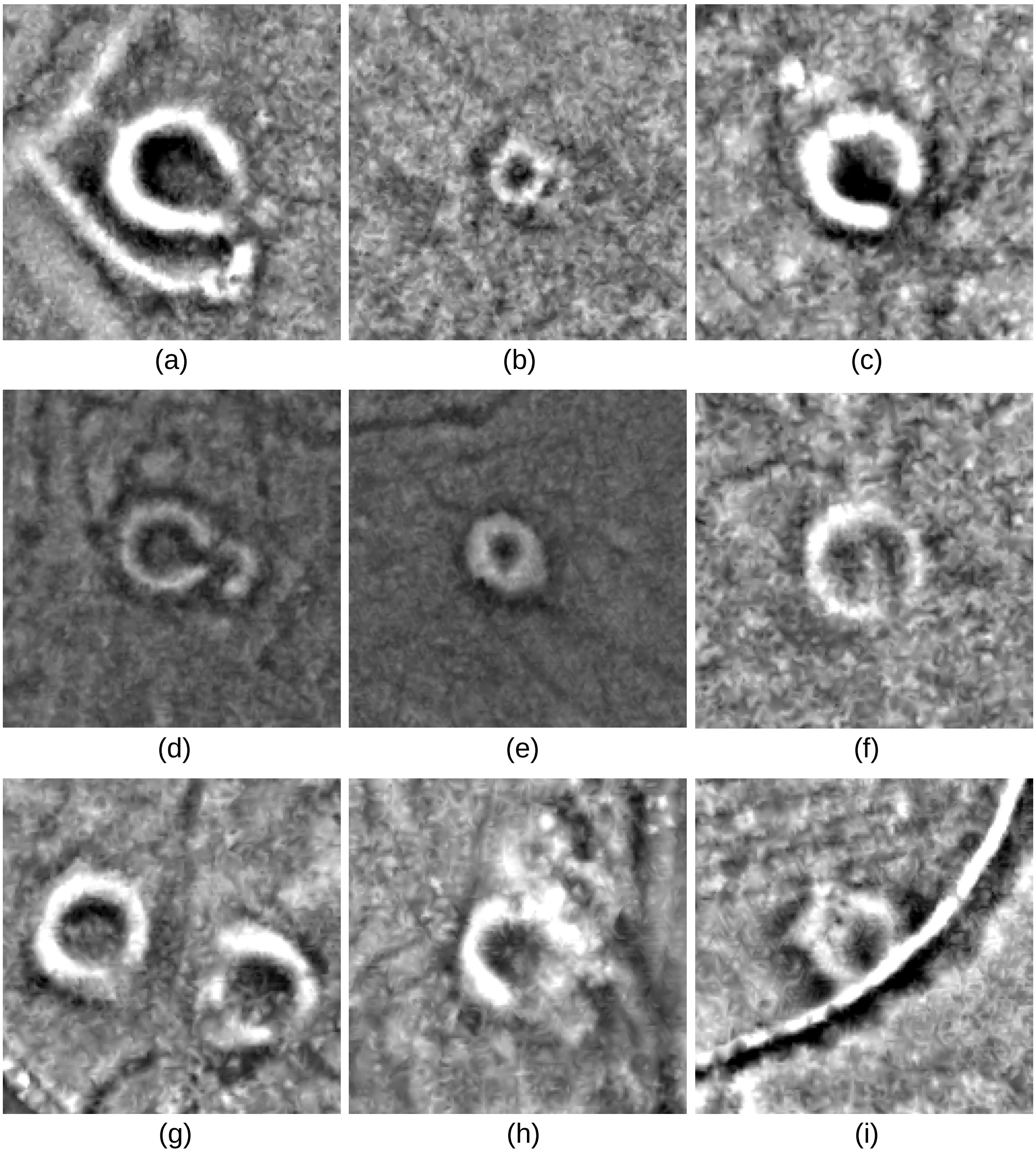


Figure 10. Roundhouse locations that the automatic method correctly identified. The locations are at OSGB coordinates: (a) 190,513 east, 632,288 north; (b) 190,532 east, 632,237 north; (c) 190,674 east, 632,180 north; (d) 190,778 east, 632,188 north; (e) 190,961 east, 632,266 north; (f) 190,620 east, 632,206 north; (g) 190,625 east, 632,477 north and 190,646 east, 632,470 north; (h) 190,656 east, 632,543 north; (i) 190,785 east, 632,778 north. The image portions are 80 m × 80 m.

Some of the roundhouse locations in Figure 10 were also visited in the field (Figure 11 –Figure 16).

4.2 Field visit

A field visit to selected locations on Arran was conducted 12-16 March 2018.

4.2.1 Roundhouse locations

Six roundhouse locations were photographed (Figure 11 – Figure 16). These roundhouses had all been detected by the automatic method.

Figure 11. Roundhouse at 190,513 east, 632,288 north, see Figure 10a.

Figure 12. Roundhouse at 190,532 east, 632,237 north, see Figure 10b.

Figure 13. Roundhouse at 190,674 east, 632,180 north, see Figure 10c.

Figure 14. Roundhouse at 190,778 east, 632,188 north, see Figure 10d.

Figure 15. Roundhouse at 190,961 east, 632,266 north, see Figure 10e.

Figure 16. Roundhouse at 190,627 east, 632 161 north, see Figure 10f.

4.2.2 Other monuments

For one small cairn location (Figure 17), there was a weak indication of small cairn in the automatic detection result. For a mound location (Figure 18), there was a strong indication of shieling. For a chambered cairn location (Figure 19), there was an indication of roundhouse.

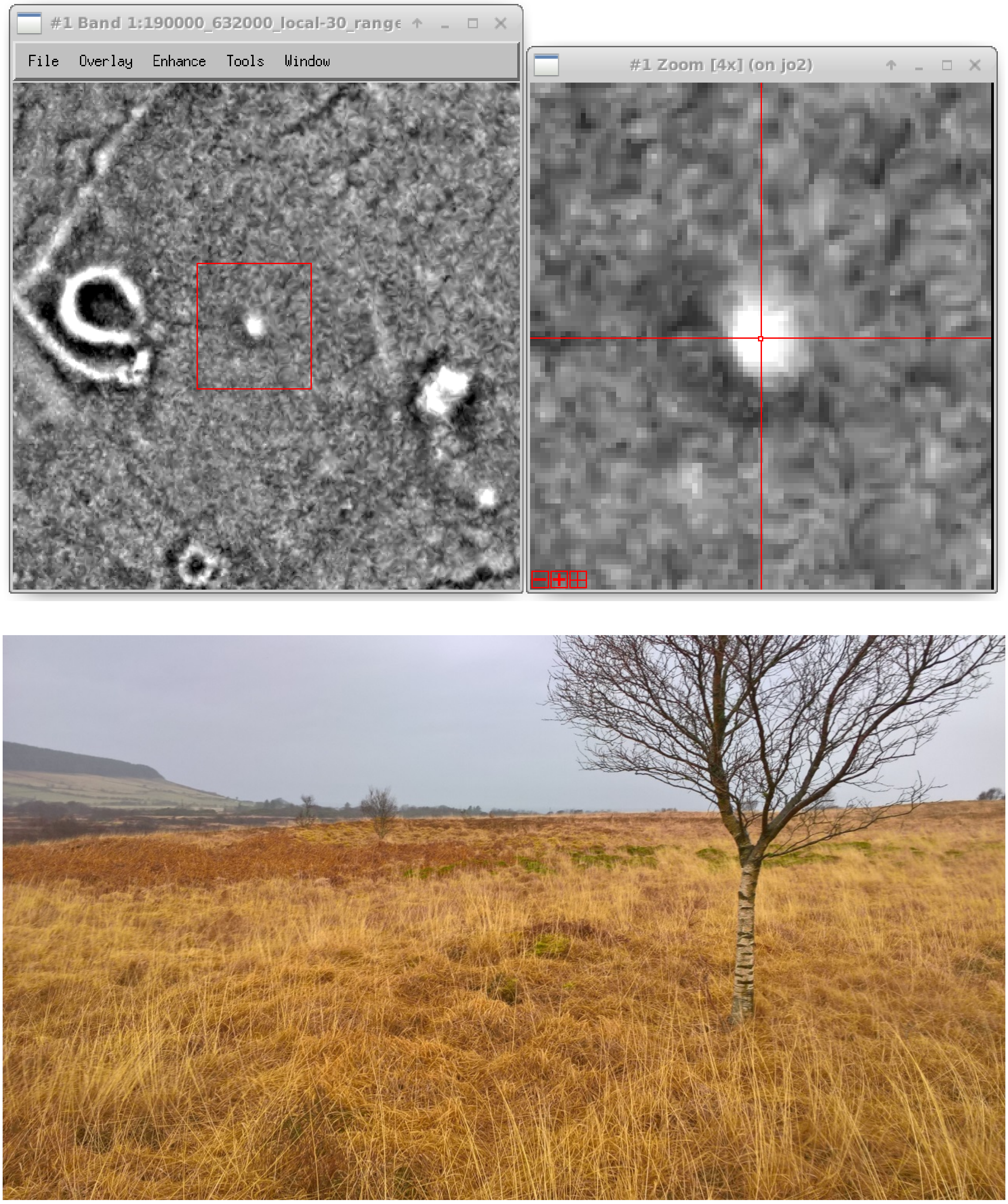


Figure 17. Small cairn at 190,543 east, 632,284 north; with roundhouse in the background at 190,513 east, 632,288 north (Figure 11).

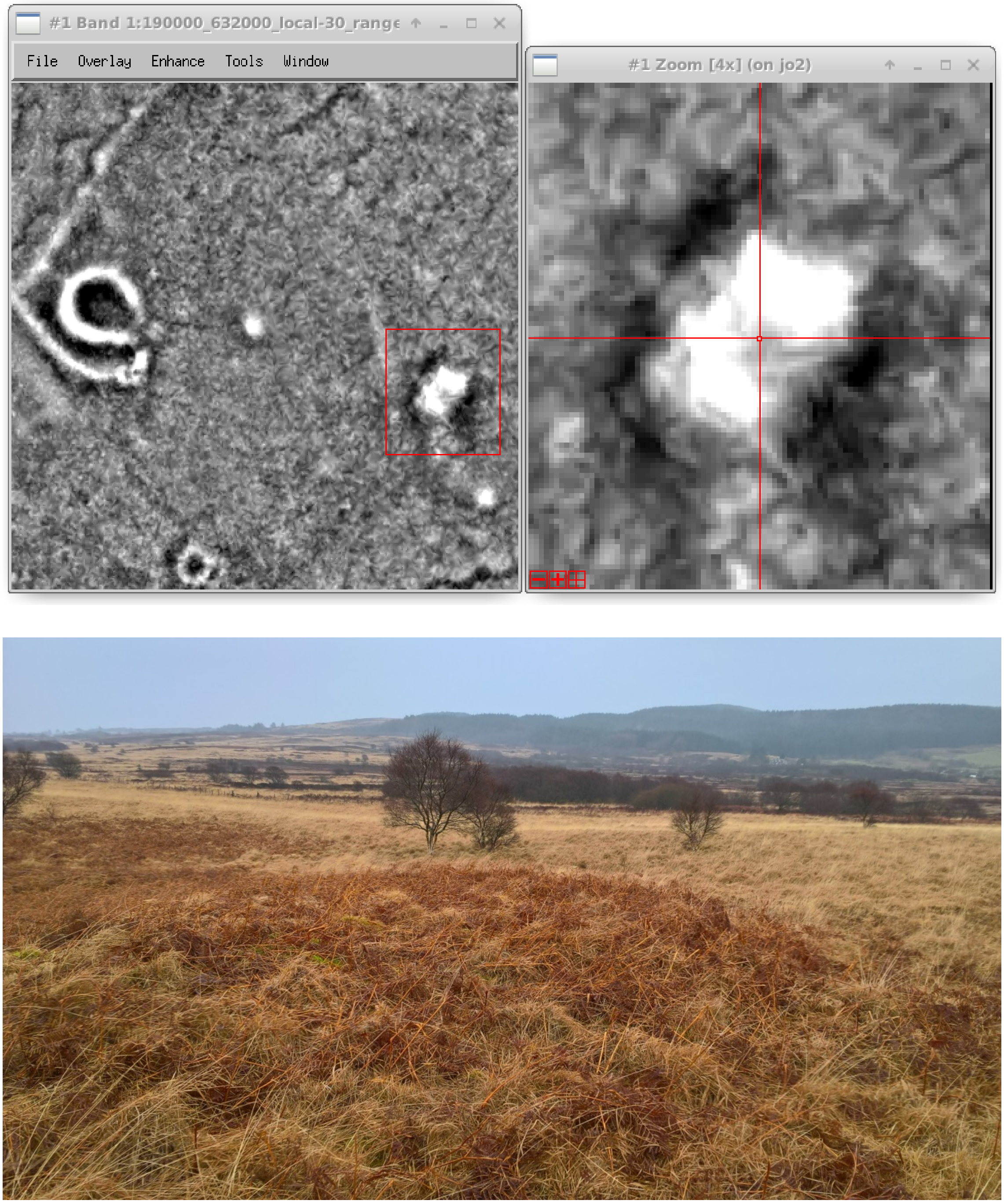


Figure 18. Mound at 190,581 east, 632 271 north.

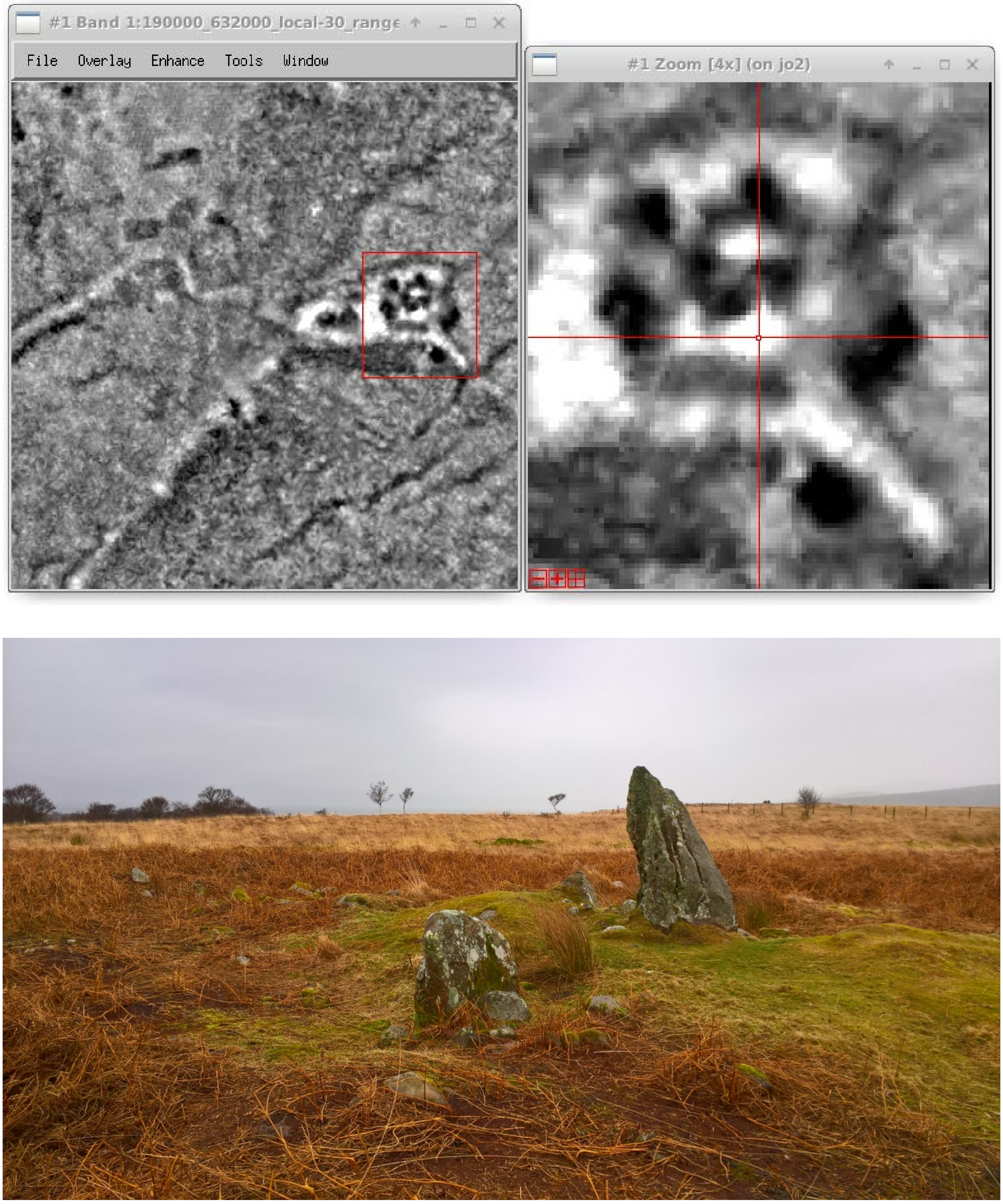


Figure 19. Chambered mound at 190,581, 632,370 north.

5 Discussion and conclusions

The preliminary results of automatic roundhouse detection had some mistakes. However, the method found many true roundhouse locations, and the number of false positives was not overwhelming. Thus, it would be interesting to apply the current version of automatic roundhouse detection on all of Arran. The main purpose would be to evaluate if the current version is perceived as an aid to archaeologists in detecting and mapping currently unknown roundhouse locations. Another purpose is to identify structures that are often mistaken as roundhouses, and to add these false positives to the learning set in one or more confusion classes.

The preliminary results of automatic detection of shielings and small cairns were less convincing. Indeed, shieling and cairns were included since we knew they would be more difficult to detect than roundhouses. Again, adding false positives into additional confusion classes could be a way forward. This may even be done in multiple iterations.

The terrain contains glacial deposits. Some of these are of similar shape as the small cairns, but are usually higher than the man-made cairns. These glacial deposits are an obvious candidate for a confusion class.

The field visit included a mound and a chambered cairn, which were predicted as being shieling and roundhouse, respectively. These examples illustrate what may happen when some classes are not present in the training data. A structure may have an appearance that is closer to one of the classes in the training data than to the general background terrain. If one wishes to fix this confusion, then one needs to include many examples of, in our case, mounds and chambered cairns. Another alternative is to accept that rare classes may be confused with classes that are more numerous.

The latter approach may be used when the true purpose of automatic classification is to identify previously unknown archaeological structures. The various classes of archaeological structure may be viewed as a means to design the classifier to better discriminate between archaeological structures on one hand, and natural terrain features and modern man-made structures on the other hand. Thus, confusion between classes of archaeological structure may be regarded as a minor mistake, compared to false positives and false negatives.

5.1 How to move forward

In order to improve the automatic classification method, we suggest the following:

1. Run automatic roundhouse detection on all of Arran. The strong indications of roundhouse could be labelled as one of:

    a. True roundhouse

    b. Other archaeological structure

    c. Cannot tell if it is archaeological or not

   d. Something else (modern structures or natural terrain features)

2. Identify glacial deposit heaps in the learning set. The 1 km tiles containing small cairns should be inspected first.

3. Add confusion classes to the learning set. This may be done iteratively:

   a. Train deep neural network classifier on learning set, containing known locations of archaeological structures, background terrain, and, optionally, locations of other structures

   b. Apply classifier on entire tiles

   c. Identify false positives, i.e., locations that the classifier predicts as archaeological structures but are in fact not.

   d. Add locations of false positives as 'other structures' to learning set. There may be several named subclasses of 'other structures', e.g., 'glacial deposit'.

   e. Repeat steps a-d until the number of newly encountered false positives is below a threshold, or the total number of 'other structures' in the learning set is above another threshold, or the maximum number of iterations has been reached.

4. More training augmentations. At the moment, only rotations that are multiples of 90 degrees are allowed. Our ResNet implementation allows rotation by any angle. However, the image extracts may need to be larger than 101 × 101 pixels to avoid missing values in the corners of the rotated images.

5. Convert detection result from raster heat map to vector locations. At the moment, the results, in the form of coloured raster layers, are only suitable for visual interpretation as a backdrop for the DTM visualisation.

6. Improve location accuracy of detection results

7. Develop prototype software into a user friendly tool

References

Hesse, R 2010. Lidar-derived local relief models – a new tool for archaeological prospection. *Archaeological Prospection* 17, 67-72.

Kermit, MA, Hamar, JB & Trier, ØD 2018. Towards a national infrastructure for semi-automatic mapping of cultural heritage in Norway. In: *CAA 2016, Proceedings of the 44st Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, Oslo, Norway, 29 March – 2 April 2016, in press.

Trier, ØD, Salberg, A-B & Pilø, LH 2018. Semi-automatic mapping of charcoal kilns from airborne laser scanning data using deep learning. In: *CAA 2016, Proceedings of the 44st Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, Oslo, Norway, 29 March – 2 April 2016, in press.

Trier, ØD & Pilø, LH 2015a. Archaeological mapping of large forested areas, using semi-automatic detection and visual interpretation of high-resolution lidar data. In: *CAA2014. 21st Century Archaeology. Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, Paris, France, 22-25 April 2014, Archaeopress, Oxford, 81-86.

Trier, ØD, Pilø, LH & Johansen, HM 2015b. Semi-automatic mapping of cultural heritage from airborne laser scanning data. *Sémata* 27, 159-186.

Trier, ØD, Zortea, M & Tonning, C 2015c. Automatic detection of mound structures in airborne laser scanning data. *Journal of Archaeological Science: Reports*, 2(1), 69-79.

Trier, ØD & Pilø, LH 2012. Automatic detection of pit structures in airborne laser scanning data. *Archaeological Prospection* 19(2), 103-121.

Trier, ØD, Larsen, SØ & Solberg, R 2009. Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection* 16(1), 1-15.