Learning to look at LiDAR: combining CNN-based object detection and GIS for archaeological prospection in remotely-sensed data

Verschoof-van der Vaart, W.B.

4

Applying automated object detection in archaeological practice: a case study from the southern Netherlands

Abstract — Within archaeological prospection, Deep Learning algorithms are developed to detect objects within large remotely-sensed datasets. These approaches are generally tested in an (ideal) experimental setting, but have not been applied in different contexts or 'in the wild', i.e., incorporated in archaeological prospection. This research explores the applicability, knowledge discovery—on both a quantitative and qualitative level—and efficiency gain resulting from employing an automated detection tool called WODAN within (Dutch) archaeological practice. WODAN has been used to detect barrows and Celtic fields in LiDAR data from the Dutch *Midden-Limburg* area, which differs in archaeology, geo-morphology, and land-use from the *Veluwe* in which it was developed. The results show that WODAN was able to detect potential barrows and Celtic fields, including previously unknown examples, and provided information about the structuring of the landscape in the past. Based on the results, combined Human–Computer strategies are argued, in which automated detection has a complementary, rather than a substitute role, to manual analysis. This can offset the inherent biases in manual analysis and deal with the problem that current automated detection methods only detect objects similar to the pre-defined target class(es). The incorporation of automated detection into archaeological prospection, in which the results of automated detection are used to highlight areas of interest and to enhance and add detail to existing archaeological predictive maps seems logical and feasible.

This chapter has been published as:

Verschoof-van der Vaart, W. B. & K. Lambers (2021), "Applying automated object detection in archaeological practice: a case study from the southern Netherlands", *Archaeological Prospection*, DOI: 10.1002/ARP.1833.

"R2-D2, you know better than to trust a strange computer."

C-3PO
(Star Wars: Episode V, The Empire Strikes Back, Lucasfilm Ltd., 1980)

4.1 Introduction

In the Netherlands, archaeological prospection—sometimes called archaeological evaluation in other countries—generally follows a stepped scheme of: 1) a desk-based assessment; followed by 2) a field survey, i.e., borehole surveys and/or field walking; and finally, 3) test trenches (Lauwerier et al., 2017). Only sporadically this strategy is supplemented with geophysical surveys (Rensink, 2019), although attempts are made to further incorporate geophysics in Dutch archaeological practice (Jelsma et al., 2021). While remote sensing normally plays an important role within archaeological prospection, in the Netherlands the focus has always been on using coring and field walking. This limited application of remote sensing techniques is mainly due to the geo(morph)ology of the Netherlands (where archaeological traces are frequently covered by thick layers of subsoil; see Berendsen, 2004) and the complex development and long-term, dynamic land-use (Risbøl, 2013), although practical and financial factors also have played a role (Waldus, 2006). This changed with the release of the *Actueel Hoogtebestand Nederland* or AHN—a LiDAR (Light Detecting And Ranging; Crutchley & Crow, 2018) dataset covering the entire Netherlands—in 2003. Nowadays, consulting the AHN is common practice within desk-based assessments. However, this LiDAR data is generally only superficially analyzed on a site level, while the full potential of this data source for large-scale landscape analyses has been underutilized. This is mainly due to the complications surrounding manually documenting and analyzing the overwhelming amount of potential archaeological objects within these large, continually improving and expanding datasets (Bennett et al., 2014; Bevan, 2015).

The last decade has seen an increase in the development of research strategies that either rely on crowd-sourced and expert-led manual brute force methods or on computational approaches to (semi-)automatically detect archaeological objects in remotely-sensed data (Casana, 2014, 2020). Recent applications of the former mainly involve the use of citizen science for the classification of remotely-sensed data (Forest et al., 2020; Lambers et al., 2019; Stewart et al., 2020). Within the latter a trend towards Deep Learning (Goodfellow et al., 2016; LeCun et al., 2015) can be observed (Fiorucci et al., 2020). This subfield of Machine Learning predominantly utilizes Convolutional Neural Networks (CNNs), hierarchically structured algorithms that generally consist of a (image) feature extractor and classifier, and are loosely inspired by the animal visual cortex (Ball et al., 2017; Guo, 2017). These algorithms learn to generalize from given examples, i.e., a large set of labeled images, rather than relying on a human operator to set parameters or formulate rules. A major advantage of CNNs is the possibility to use transfer-learning (Razavian et al., 2014), where a CNN is pre-trained on a large, generic dataset and subsequently is fine-tuned on a small, specific dataset. In archaeology, transfer-learning has been successfully implemented on different types of remotely-sensed data from Europe (Bonhage et al., 2021; Gallwey et al., 2019; Guyot et al., 2021b; Kazimi et al., 2019; Trier et al., 2019;

Verschoof-van der Vaart & Lambers, 2019; Verschoof-van der Vaart et al., 2020; Verschoof-van der Vaart & Landauer, 2021; Zingman, 2016; Zingman et al., 2016) and further abroad (Bundzel et al., 2020; Caspari & Crespo, 2019; Somrak et al., 2020; Soroush et al., 2020; Trier et al., 2021, 2018). To date these approaches are generally tested in an (ideal) experimental setting, but have not been applied in different contexts or 'in the wild', i.e., incorporated in archaeological prospection, although the latter is the main aim of most initiatives (see Trier et al., 2019). However, research has shown that when these approaches are used beyond an ideal experimental setting, the performance decreases (Chapter 3). Furthermore, one of the main questions that remains is the transferability of these methods (Cowley et al., 2020; Kermit et al., 2018). Therefore, studies 'in the wild' and in different environments are important to investigate the true potential of automated approaches for archaeological practice.

4.1.1 Aim

In this chapter, the application of a Deep Learning tool within archaeological practice will be addressed. Furthermore, the knowledge discovery—on both a quantitative and qualitative level—and efficiency gain resulting from applying an object detection model will be explored. The object detection model WODAN (**W**orkflow for **O**bject **D**etection of **A**rchaeology in the **N**etherlands; Chapters 2 & 3), developed in one area of the Netherlands (the *Veluwe*) will be used to detect two classes of archaeology (barrows and Celtic fields) in the Dutch *Midden-Limburg* area (Fig. 4.1). This area has been chosen because it has different archaeological, geo(morpho)logical, and land-use conditions (Section 4.2). The results of the automated detection (Section 4.3) will be compared to two reference datasets: an inventory of documented archaeological sites and a manual analysis of the LiDAR data, conducted in the framework of this research. The knowledge discovery and efficiency gain will be analyzed (Section 4.3) and discussed (Section 4.4).

4.2 Materials and Methods

4.2.1 Research Areas

The Veluwe area (Table 4.1) comprises the western part of the province of Gelderland in the Netherlands (Fig. 4.1, Red). It consists of ice-pushed ridges formed in the Saale glacial period (ca. 350,000 to 130,000 BP), which were subsequently partially covered with coversand during the Weichselian glacial period (ca. 115,000 to 11,500 BP; Berendsen, 2004). Nowadays, this area, ca. 1100 km^2, is predominantly covered by forest and heath, interspersed with villages and towns of various size, and agricultural fields (for a detailed overview of the area see Lambers et al., 2019). The Midden-Limburg area (ca. 265 km^2, Table 4.1) covers the municipalities Echt-Susteren, Roerdalen, and Roermond in the province of Limburg in the southern part of the Netherlands (Fig. 4.1, Black). The western boundary of the area consists of the Meuse river, while the area is demarcated in the east by the Dutch-German border (Fig. 4.1).

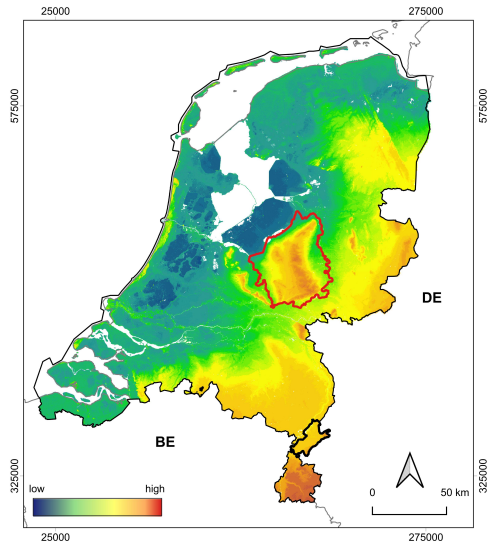

Figure 4.1: The Midden-Limburg (outlined in black) and Veluwe (outlined in red) areas on a height model of the Netherlands (source of the height model: Nationaal Georegister, 2021; coordinates in Amersfoort/RD New, EPSG: 28992).

Table 4.1: Main characteristics of the Veluwe and Midden-Limburg research areas.

Area	Area (km^2)	Images	General terrain	Main land-use
Veluwe	375	1152	Ice-pushed ridges, partly covered with coversand	Forest and heath
Midden-Limburg	265	4405	River terraces, partly covered with coversand, loess, and river deposits	Agricultural fields

The Midden-Limburg area comprises a highly diverse landscape, which results from eolian, fluvial, and tectonic processes (Fig. 4.2). In the north, the area is dissected by a northwest-southeast orientated geological fault line, the *Peelrandbreuk*. Immediately to the south of the research area lies another fault line, the *Feldbissbreuk*. Therefore, the majority of the region is part of the subsiding *Roerdalslenk* or Roer Valley Graben, while the northeastern part lies on the *Peelhorst*, which experiences tectonic uplift (Berendsen, 2004). The subsoil and landscape in the research area mainly formed by repeated deposition and incision of the Meuse river, starting in the Holstein interglacial (ca. 400,000 till 380,000 BP) up till the present (Fig. 4.2). This has resulted in a series of river terraces and escarpments: the higher terrace consists of coarse river deposits from the Middle Pleistocene (ca. 400,000 till 130,000 BP). This terrace mainly follows the Dutch-German border and is only found in the extreme northeastern and eastern part, the *Meinweg* nature reserve, of the research area. A steep escarpment, spanning a height difference of up to 23 m, separates the higher terrace from the middle terrace.

The latter can be divided into a higher part, formed in the Middle and Late Pleistocene (ca. 380,000 till 15,000 BP), and a (few meters) lower part which was formed during the warmer period at the end of the Weichselian (ca. 15,000 till 11,500 BP). The middle terrace consists of Meuse river deposits (coarse sand and gravel). During the Late Pleistocene (ca. 130,000 till 11,500 BP) the higher part of the middle terrace and the higher terrace were subjected to eolian processes and became partly covered with coversand and loess. Finally, the lower terrace contains the active stream valley of the Meuse river and mostly consists of clay (Ellenkamp & Tichelman, 2008). In the Holocene (ca. 11,500 BP till the present) stream valleys were formed by smaller river courses such as the *Roer* or the *Swalm*, which deposited loam and sand. To a lesser extent the area was covered by peat formations (Berendsen, 2004). Today, the research area is predominately covered by agricultural fields, urbanized areas of various size, and to a lesser extent with forest (Fig. 4.2).

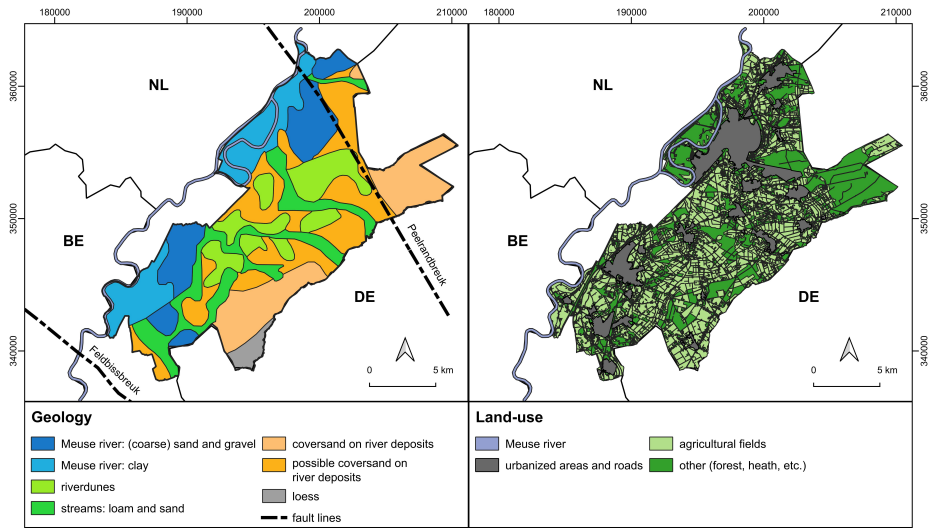

Figure 4.2: Overview of the geology (left) and current land-use (right) of the Midden-Limburg research area; amended from Ellenkamp & Tichelman (2008).

4.2.2 Datasets

LiDAR Data

LiDAR data of both research areas is freely available as an interpolated Digital Terrain Model (DTM) from the online repository PDOK (Nationaal Georegister, 2021). The data has an average ground point density of 6–10 per m^2, a spatial resolution of 50 cm, and a vertical and planimetric accuracy of 5 cm (van der Zon, 2013). In this research the second generation of Dutch LiDAR data (AHN2, released in 2012) is used. The third generation (AHN3), with equal resolution but with a higher accuracy compared to prior generations is currently being made available on a nation-wide level. For the training of our object detection model (see Section 4.2.3), a dataset of 1152 LiDAR images (600 by 600 pixels) was used. This is a selection of images that contain archaeological objects, from various parts of the Veluwe (spread over an area

of ca. 375 km^2). Images without archaeological objects from this area were excluded from the training dataset. The total Midden-Limburg area (ca. 265 km^2) was used to form a test dataset of 4405 LiDAR images (600 by 600 pixels).

Archaeological Inventory

The Midden-Limburg area is rich in archaeology from many time periods. The following overview is limited to the Neolithic period until the start of the Roman period (ca. 4900–12 BC; Louwe Kooijmans et al., 2005), focusing on three types of archaeological sites: settlements, burial sites, and Celtic fields (Fig. 4.3). These types were selected as they are either objects detected by WODAN (barrows, urnfields, and Celtic fields) or are related to detected objects (settlements, other burials sites). Therefore, comparing this overview with the results of the object detection can provide information about the knowledge gain (see Section 4.3.4). The overview was assembled by consulting the two principal Dutch archaeological databases (ArchIS and the AMK; Rijksdienst voor het Cultureel Erfgoed, 2021b), the archaeological predictive maps of the three municipalities (Ellenkamp & Tichelman, 2008; Verhoeven et al., 2010a,b), and recent archaeological grey literature (Arnoldussen, 2013; Arnoldussen et al., 2014; Meurkens & Tol, 2016; Verhart & Janssen, 2010).

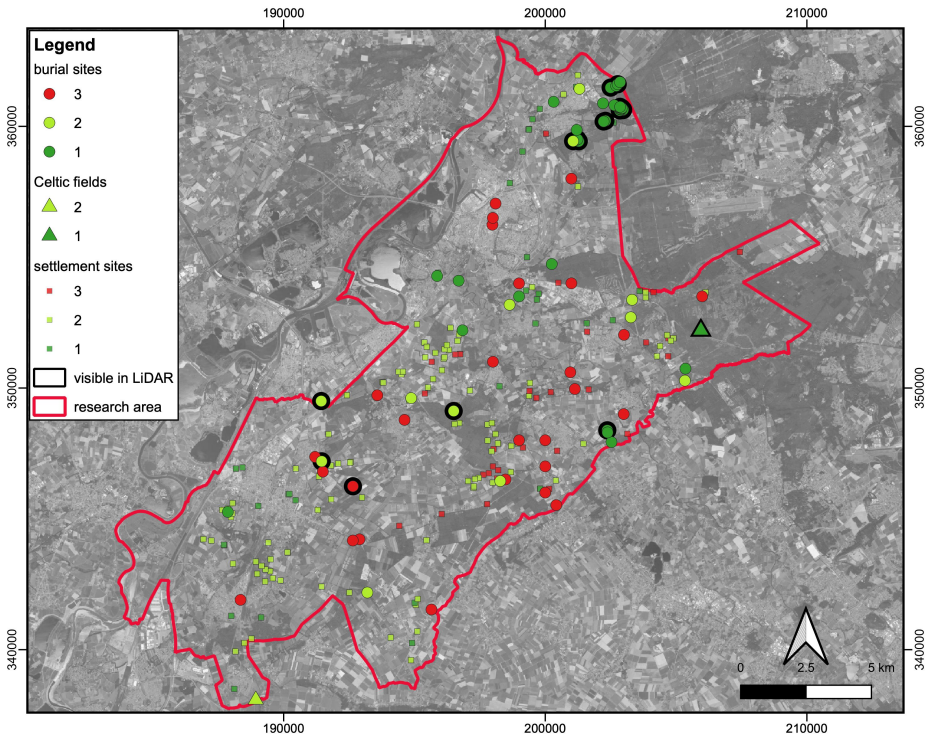

Figure 4.3: The distribution of burial sites (circles), Celtic fields (triangles), and settlement sites (rectangles) in the research area (red outline) on a recent aerial photograph; sites visible in LiDAR data are outlined in black (source of the photograph: Nationaal Georegister, 2021; coordinates in Amersfoort/RD New, EPSG: 28992).

Every site in the overview has a different confidence level, based on the source (or step in the Dutch archaeological prospection scheme, see Section 4.1) from which the information about the site derives: 3) from indirect sources, such as (historical) literature; 2) minimal destructive (archaeological) research, such as coring and field-walking; or 1) archaeological test-trenches or excavations. As expected the accuracy of the interpretation of these sites varies between these methods: through excavations (1) the site type and date can be specifically determined, while an interpretation based on field-walking (2) or indirect sources (3) is much less certain. Tables 4.2, 4.3, and 4.4 show the known settlements, burial sites, and Celtic fields in the research area divided per time period: Neolithic (4900-2000 BC), Bronze Age (2000-800 BC), Late Bronze Age-Early Iron Age (*Niederrheinische Grabhügelkultur* or NGK; 1100-500 BC), and Iron Age (800-12 BC). Furthermore, the tables show whether the archaeological sites are visible in the LiDAR data (see also Fig. 4.3). In the case of settlement sites, none are discernible, while only 23 of the burial sites and one of the Celtic fields (ca. 0.24 km^2) show up in the LiDAR data. In the results of the automated detection, only these visible archaeological sites will be used. Besides, settlements, burial sites, and Celtic fields are generally hard to discern in the field, but are sometimes visible in other remotely-sensed data, e.g., aerial imagery (see for example Brongers, 1976).

As can be seen in Table 4.2, an abundance of settlement sites (170 in total) are known from the research area. Upon excavation, these areas of habitation generally contain traces of houses or other buildings, additional pits and postholes, and domestic refuse (Louwe Kooijmans et al., 2005). Most of these sites are located within agricultural areas (Fig. 4.3). The table shows an uneven distribution with a higher number of settlements dated in the Neolithic and Iron Age, as compared to the Bronze Age. However, most of these sites have only been roughly dated, as the majority are known from field-walking (ca. 60%), while only about a quarter of these sites have been excavated. On the other hand, most of the burial sites (72 in total) in the research area (Table 4.3) are known from indirect sources (ca. 38%) and excavations (ca. 45%). The majority of these sites concern urnfields (ca. 37%) or barrows (ca. 35%), although isolated burials (ca. 24%) without clear above-ground features, are also known. The high number of urnfields, or barrow cemeteries, is related to the fact that these are one of the most characteristic archaeological phenomena in this region and thus have a rich research history (Theuws & Roymans, 1999). Most burial sites can be dated to the Late Bronze Age or Iron Age. Interestingly, while there is an abundance of settlement- and burial sites from late prehistory, only two (potential) Celtic fields are known from the research area (Fig. 4.3; Table 4.4). One of these concerns the well-investigated Celtic field near the village of Herkenbosch (Arnoldussen, 2013; Verhart & Janssen, 2010). The other site, near the village of Nieuwstadt, is based on anthropogenic soil layers and small pottery fragments in corings. The scarcity of Celtic fields in this region, as compared to other Dutch regions such as the Veluwe, has been attributed to the abundance of natural boundaries—making formal boundaries redundant (van Beek, 2011)—and the predominant geology, subsoil, and hydrology (Spek, 2004), although a lack of research and intensive, degrading agricultural practices also seems to have been of influence (Arnoldussen, 2013).

Table 4.2: Documented settlement sites in the Midden-Limburg research area.

Archaeological period	Confidence level			Number of archaeological objects	Objects visible in LiDAR data
	3	2	1		
Neolithic	13	40	9	62	0
Bronze Age	2	15	8	25	0
NGK	1	4	1	6	0
Iron Age	14	44	19	77	0
total	30	103	37	170	0
percentage	18%	60%	22%	100%	0%

Table 4.3: Documented burial sites in the Midden-Limburg research area.

Archaeological period	Confidence level			Number of archaeological objects	Objects visible in LiDAR data
	3	2	1		
Neolithic				9	4
barrow	0	0	7		
unknown	2	0	0		
Bronze Age				10	7
barrow	0	0	7		
burial	2	0	1		
NGK				23	2
burial	2	0	0		
urnfield	8	4	9		
Iron Age				17	2
barrow	0	0	1		
burial	6	3	1		
urnfield	3	2	1		
Unknown				13	8
barrow	3	2	6		
burial	0	1	0		
unknown	1	0	0		
total	27	12	33	72	23
percentage	38%	17%	45%	100%	32%

Table 4.4: Documented Celtic fields in the Midden-Limburg research area.

Archaeological period	Confidence level			Number of archaeological objects	Objects visible in LiDAR data
	3	2	1		
NGK	0	0	1	1	1
Iron Age	0	1	0	1	0
total	0	1	1	2	1
percentage	0%	50%	50%	100%	50%

Manual Analysis

In the framework of this research, the LiDAR data from the Midden-Limburg area was manually investigated by a researcher with abundant experience in analyzing remotely-sensed data and considerable knowledge of the archaeology of the research area. This analysis enables the comparison between the performance and efficiency of automated detection versus manual analysis. During the analysis the LiDAR data was loaded into *QGIS 3.4 Madeira* (QGIS Development Team, 2017) and visualized with the Simple Local Relief Model visualization (Hesse, 2010) from the *Relief Visualisation Toolbox 2.0* (Kokalj & Hesse, 2017) and all settlement sites, burial sites, and Celtic fields were annotated (Fig. 4.4). The data was evaluated in combination with aerial imagery (25 cm resolution) and geo(morph)logical maps of the research area (source: Nationaal Georegister, 2021).

The manual analysis took 6.75 hours (405 minutes) and resulted in 135 potential barrows. Interestingly, only 16 of the 23 visible barrows on record (see Table 4.3) were recognized as such during the manual analysis. Furthermore, 31 new, demarcated areas of Celtic fields, totaling 3.37 km^2 have been annotated. No settlement sites were annotated during the manual analysis.

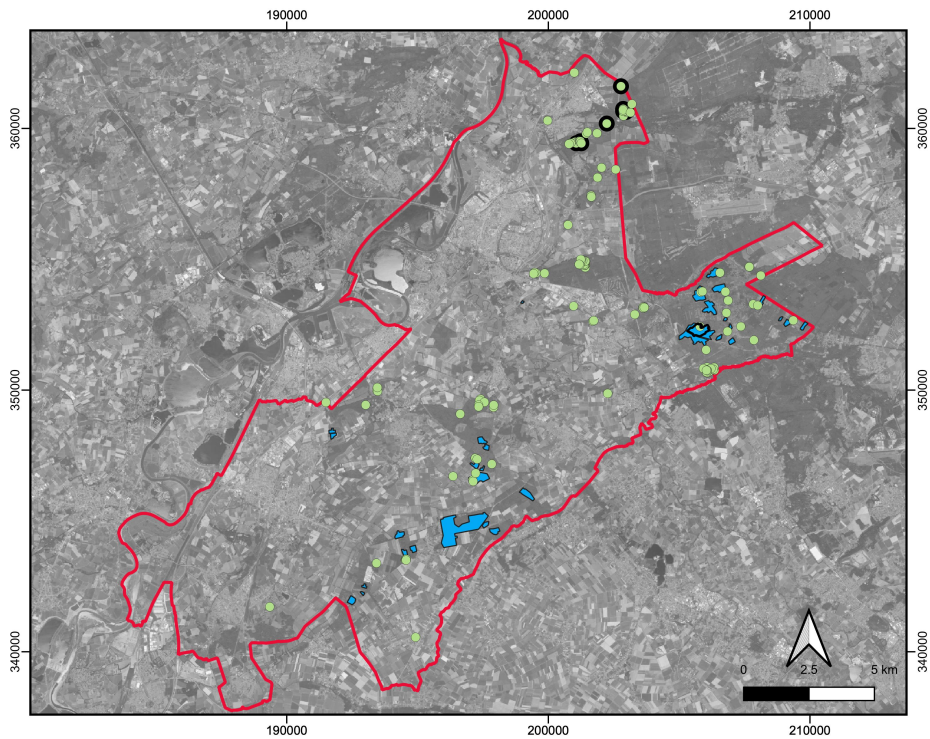

Figure 4.4: The results of the manual analysis with barrows (green) and Celtic fields (blue) in the research area (red outline) on a recent aerial photograph; registered archaeological objects are outlined in black (source of the photograph: Nationaal Georegister, 2021; coordinates in Amersfoort/RD New, EPSG: 28992).

4.2.3 WODAN

In this research the object detection model WODAN—the result of a PhD in the Data Science Research Programme at the Faculty of Archaeology, Leiden University (Chapters 2 & 3)—was used to detect archaeological objects in the Midden-Limburg area. The latest version, WODAN2.5 (Chapter 3.8), consists of four parts (Fig. 4.5): 1) A preprocessing part that converts LiDAR data into input images; 2) an object detection part; 3) a post-processing part that turns the results of the prior step into geospatial vectors, directly usable in a GIS; and 4) an additional post-processing step called Location-Based Ranking (LBR; see Section 4.2.3) that incorporates domain knowledge into the workflow, to reduce false positives caused by specific zones within the research area (Chapter 3).

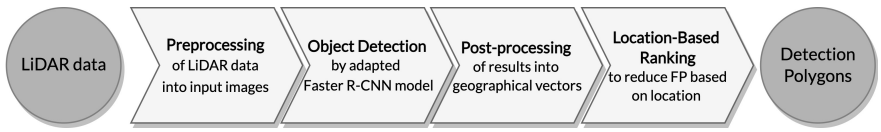

Figure 4.5: Schematic representation of the WODAN2.5 workflow; after Verschoof-van der Vaart et al. (2020).

The object detection part of the WODAN workflow consists of an adapted version of the Faster R-CNN architecture (Ren et al., 2017). This so-called Region-based CNN or R-CNN (Girshick et al., 2014) is able to localize and classify multiple, adjacent or even overlapping objects within a single image—as opposed to general CNNs that give a single classification for the entire input image (Guo et al., 2016). Faster R-CNN consists of two parts: a fully connected convolution Region Proposal Network (RPN) and the Fast R-CNN model (Girshick, 2015). The former generates object proposals, i.e., it selects regions within the image that potentially contain an object of interest. The latter model is used for feature extraction and classification of these candidate regions. Both the RPN and Fast R-CNN are trained simultaneously during the training of Faster R-CNN (for an detailed overview of Faster R-CNN, see Ren et al., 2017).

Location-Based Ranking

In order to use WODAN in large-scale archaeological mapping over different types of complex terrain, Location-Based Ranking (LBR) was developed to reduce false positives caused by 'objects of confusion' with morphology comparable to the archaeological objects of interest (Chapter 3). LBR involves determining, ranking, and mapping of (present-day) landscape characteristics, such as subsoil and current land-use, which have had an impact on the preservation and/or visibility of archaeological objects of interest. These characteristics can be determined based on prior research in the formation of the archaeological landscape and/or by a broad-brush landscape characterization (Cowley, 2011) of the research area. The subsequently assigned ranks, from 3 (low) to 1 (high), correspond to the potential for the occurrence of specific types of archaeological objects within that zone (for a detailed overview of LBR, see Chapter 3).

For the Midden-Limburg area five landscape features were identified: disturbed areas (quarries, etc.), agricultural fields, urbanized or built-up areas, areas with (late) Holocene deposits (stream valleys, driftsand), and modern roads (Fig. 4.6; Table 4.5). The most detrimental are disturbed areas and agricultural fields. Built-up areas, areas with Holocene deposits, and roads have had a less negative impact. While Celtic fields are generally intersected by roads, this has had a limited negative impact on the preservation of the overall objects. The best chance for survival of archaeological objects can be found in the remaining areas (Table 4.5, other), such as forest.

Table 4.5: Different landscape features and their rank in the Location-Based Ranking map for Midden-Limburg research area.

Landscape Features			Rank	
Type	**Area (km^2)**	**Ratio of Research Area (%)**	**Barrow**	**Celtic Fields**
disturbed areas	2.49	0.94	3	3
agricultural fields	122.96	46.38	3	2
built-up areas	34.95	13.18	2	2
holocene deposits	18.77	7.08	2	2
roads	8.73	3.29	2	1
other	77.24	29.13	1	1
total	**265.14**	**100%**		

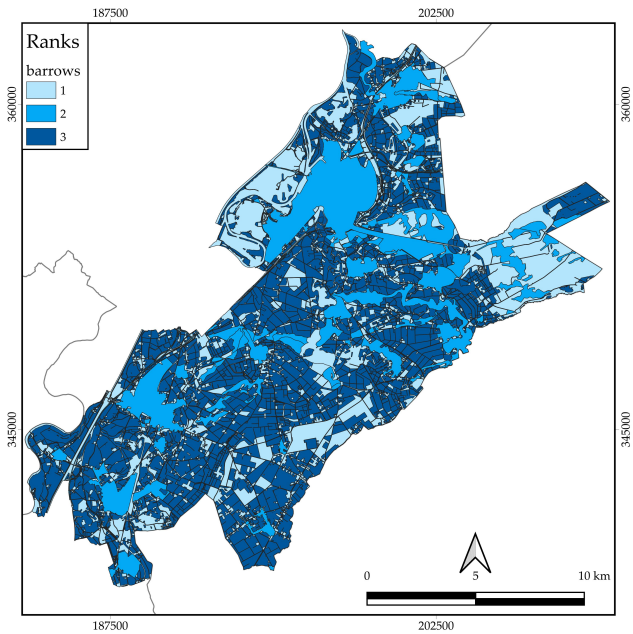

Figure 4.6: Location-Based Ranking map for the Midden-Limburg research area showing the ranks of the zones for barrows in shades of blue (see legend; coordinates in Amersfoort/RD New, EPSG: 28992).

Based on the above a ranked map of the Midden-Limburg area was created based on open-source geo(morph)ological and topographical data from the online spatial data repository PDOK (Nationaal Georegister, 2021). The assigned ranks correspond to the potential for the occurrence of archaeological objects within that zone. Subsequently, all detections from WODAN were compared to this map and assigned to one of the ranks. Detections in high-ranking zones (Rank 1) are more likely to be archaeological objects, while detections in low-ranking zones (Rank 2 or 3) have a much higher likelihood of being false positives. Therefore, LBR can be used to reduce the number of false positives by ignoring detections in low-ranking zones (Chapter 3).

4.3 Results

4.3.1 Implementation Details

To investigate the application of WODAN, the workflow was trained on data from the Veluwe and used to detect barrows and Celtic fields in the Midden-Limburg area, which has different archaeological, geo(morpho)logical, and land-use conditions as compared to the Veluwe (Table 4.1). In recent research a difference in the performance of CNNs was observed when trained and tested on particular LiDAR visualizations (Kazimi et al., 2020b; Somrak et al., 2020, Chapter 5). In prior research with WODAN, data visualized with Simple Local Relief Model (Hesse, 2010) was successfully used, while recent research showed promising results using un-visualized Digital Terrain Model data (Chapter 5). Therefore, two versions of WODAN were used: one model was trained and tested on the un-visualized Digital Terrain Model (WODAN_DTM), and one on data visualized with Simple Local Relief Model (WODAN_LRM). For both versions the LiDAR data was turned into input images following the same pre-processing approach as in Chapter 3.8. In addition, the input images were normalized by subtraction of the central pixel value so that each snippet has pixel (or grayscale) values between 0 and 255 (following Chapter 3.8). For the object detection, the Faster R-CNN architecture (Ren et al., 2017) was used with VGG16 (Simonyan & Zisserman, 2015), pretrained on the ImageNet dataset (Russakovsky et al., 2015), as the backbone network. Faster R-CNN was transfer-learned using Stochastic Gradient Descent with the Adam optimizer (Kingma & Ba, 2015), implemented in *Keras* (Chollet, 2015). Additionally, Focal Loss (Chen et al., 2018; Lin et al., 2020) was implemented in the RPN (see Chapter 3.8). Empirically the learning rate was adjusted to 1×10^{-5} and the number of epochs to 18 (see Goodfellow et al., 2016). In the training process, the sizes of the anchor boxes were adjusted following Chapter 3.3.1. During training, the input images were randomly flipped horizontally and vertically, as well as rotated to augment the data. In the post-processing step the detections were turned into geospatial vectors and subsequently ranked with LBR, based on their location (for a detailed overview, see Chapter 3).

4.3.2 General Results

After training both versions of WODAN were tested on the entire Midden-Limburg area (see Fig. 4.7). On average it took only 50 minutes to run the model on the test dataset (on a single GPU per version of WODAN), post-process the results into geospatial vectors, and to implement LBR. Table 4.6 and 4.7 show a comparison of the results of the automated detection in the Midden-Limburg area and the two reference datasets (the archaeological inventory and the manual analysis, see Section 4.2.2). Contrary to other research the performance of WODAN is not evaluated through metrics such as F1-score or Accuracy as we lack a (field) validated baseline to determine performance. Furthermore, the scope of this research is not to evaluate performance, but to investigate the application, efficiency, and knowledge discovery of using automated detection. Therefore, the ratio of overlap between the results of the automated detection and the two reference datasets is given. For instance, WODAN_DTM detected 40.9% of the barrows on record and visible in the LiDAR data, and 30.7% (35) of the barrows annotated during the manual analysis. These results indicate that WODAN is able to detect barrows and Celtic fields in the Midden-Limburg area, when trained on data from a different area in the Netherlands, i.e., the Veluwe. However, the performance of WODAN still has room for improvement, especially when compared to the overlap between the results and the reference datasets (Table 4.6 & 4.7). This level of performance can partly be explained by the fact that WODAN was tested on the entire Midden-Limburg area, which generally causes a decrease in performance (see Chapter 3). Furthermore, other research in which the transferability of comparable methods was tested also observed a decrease in performance when object detection model was tested on an unrelated area (Chapter 6). The different versions of WODAN, using either DTM or SLRM data, seem to have a different performance. Using DTM data, instead of data visualized with SLRM, improves the detection of barrows (compare Table 4.6 and 4.8). However, for Celtic fields the use of DTM data seems detrimental to the performance (compare Table 4.7 and 4.8). This is contrary to the results of prior research on the detection of hollow roads (Chapter 5), but in line with research on the effectiveness of different visualizations (Kazimi et al., 2020b). These results indicate that the performance of CNNs, using different visualizations, is not only related to the visualization itself but also to the type of archaeological object that is detected (see also Chapter 7).

Table 4.6: The results of the automated detection of barrows in the Midden-Limburg area.

Method	Total	Rank			Overlap	
		1	2	3	Archaeological inventory	Manual analysis
arch. inventory	23	21	1	1	100%	–
manual analysis	135	114	20	1	69.6%	100%
WODAN_LRM	696	323	325	41	30.4%	28.9%
WODAN_DTM	780	267	374	139	43.5%	30.7%

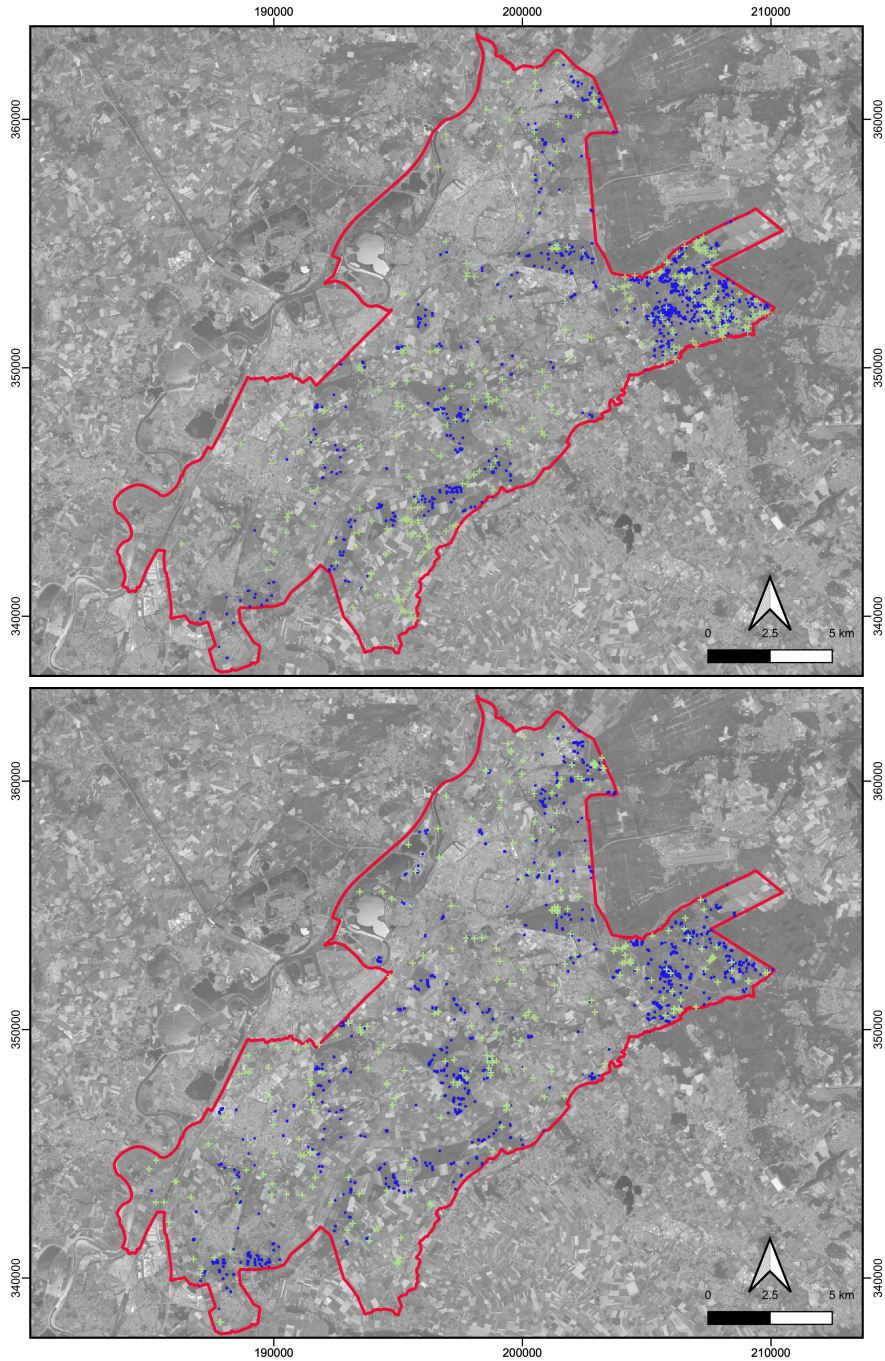

Figure 4.7: The results of the automated detection (top: WODAN_LRM; bottom: WODAN_DTM) showing all Rank 1 detections of barrows in green and Celtic fields in blue in the research area (red outline) on a recent aerial photograph (source of the photograph: Nationaal Georegister, 2021; coordinates in Amersfoort/RD New, EPSG: 28992).

Table 4.7: The results of the automated detection of Celtic fields (in m^2) in the Midden-Limburg area.

Method	Total	Rank			Overlap	
		1	2	3	Archaeo-logical inventory	Manual analysis
arch. inventory	235881	235881	0	0	100%	—
manual analysis	3372470	3277094	95376	0	100%	100%
WODAN_LRM	4290724	2780730	1503210	6784	66.2%	24.4%
WODAN_DTM	12638926	3325880	9210820	102226	58.4%	13.6%

4.3.3 Efficiency Gain

In this research the duration of the manual analysis (405 minutes) and the two versions of WODAN (on average 50 minutes) was recorded. For the latter the time includes testing, post-processing, and implementing LBR. It does not include the preprocessing—as for both the automated detection and the manual analysis the LiDAR data needed to be pre-processed, e.g., visualized—or the training time of WODAN, as the latter could be compared to the training of an operator to analyze LiDAR data. As shown, WODAN (on average 50 minutes) is ca. eight times faster than the manual analysis (405 minutes). Moreover, during the actual running of WODAN on the test dataset the operator does not need to be actively involved, which makes the automated detection even more time efficient. This shows the major potential of automated object detection as a tool to assist in the rapid mapping of archaeological objects over extensive areas (see Soroush et al., 2020). It can reduce the time invested in actually mapping objects, so that the operator's time can be reallocated to analysis, validation, and interpretation of the results.

4.3.4 Knowledge Discovery

In the following the knowledge discovery, i.e., the extraction of implicit, previously unknown, and potentially useful information (McCoy, 2017), resulting from using automated detection in the Midden-Limburg area is presented. Knowledge discovery can be either of a quantitative or a qualitative nature (Huggett, 2020a). The former concerns the locating of hitherto undocumented archaeological objects. The latter concerns a better *understanding* of the patterns and relations between the uncovered archaeological objects and between these objects and the surrounding landscape, through the interpretation of the gained data (Cowley, 2011).

Quantitative Knowledge Gain

Obviously, employing automated detection leads to new information on the location of previously unknown archaeological objects, especially if new data sources, e.g., LiDAR, and/or unexplored regions within the research area, for instance forest, are analyzed (see Kenzler & Lambers, 2015). For this analysis, all Rank 1 detections (see Table 4.6 and 4.7) were manually investigated and compared to the reference datasets.

In Table 4.8 the amount of new potential archaeological objects found through the use of automated detection in the Midden-Limburg area is shown. Instead of square meters, Table 4.8 shows the number of demarcated Celtic field areas. This unit is used because in general WODAN locates a number of individual plots within a demarcated Celtic field, but does not detect the entire area. In that sense, using the coverage gives a skewed image of the performance of WODAN. For instance, WODAN_LRM (Table 4.7) only had 24.4% overlap between the detected Celtic field areas and the manually annotated areas. However, it did detect a number of plots within 26 of the 32 manually mapped Celtic fields. Thereby, WODAN adequately indicates the location of Celtic fields in the research area, although the full extent and coverage of these areas is not well presented. Furthermore, Table 4.8 shows (in brackets) how many of potential archaeological objects were not found during the manual analysis. These objects—that are neither in the archaeological inventory nor annotated during the manual analysis—are of special interest, as these show the added benefit of simultaneously using manual and computational methods (see Section 4.4).

Table 4.8: Quantitative knowledge gain of using automated detection on the Midden-Limburg area, showing the number of new potential archaeological objects. In brackets are the number of objects not annotated during the manual analysis.

Model	Potential barrows	Potential Celtic Field areas
WODAN_LRM	38 (8)	36 (11)
WODAN_DTM	35 (7)	32 (8)

Prior to this research 72 burial sites were known in the Midden-Limburg area, of which 63 sites were either barrows or urnfields (see Table 4.3). Using automated detection resulted in 35–38 new potential barrows (either individual mounds or mounds within an urnfield), of which 7–8 were not mapped during the manual analysis (Table 4.8); an increase of more than 50% of the known burial mounds in the area. In general, the detected burial mounds appear clearly in the LiDAR data (Fig. 4.8, a–b). However, some of the detected mounds proved almost indiscernible in the visualized LiDAR maps, and were only recognized in vertical profiles of the LiDAR data (see Fig. 4.8, c).

Using automated detection resulted in a large increase of the number and extent of potential Celtic fields in the research area (Fig. 4.8, d–f). Prior to this research only two Celtic field areas were known of which only one, near Herckenbosch (Arnoldussen, 2013; Verhart & Janssen, 2010), was well-documented (see Table 4.4). Based on the automated detection 32–36 new potential Celtic field areas were detected, of which 8–11 were not mapped during the manual analysis. About half of these potential Celtic fields are located in the vicinity of the known Celtic field at Herckenbosch (Fig. 4.7), which could indicate that this was once a single, much larger system, spanning over three square kilometers. Another concentration of potential Celtic fields has been found between the villages of Maria Hoop, Montfort, and Posterholt.

Most of the potential archaeological objects are found on the higher middle terrace in the eastern and southern part of the area (Fig. 4.7) although some objects can also be found in the central part of the research area: some of the potential barrows seem to be located along the edge of stream valleys, while some of the potential Celtic fields extend onto the riverdunes in the center of the area. However, the question remains to what extent this distribution is the result from the (current) land-use in the area (see Section 4.3.4). For instance, within the part of the research area covered by loess, only one potential barrow was found. Presumably, the absence of archaeological objects within this part is due to intensive agriculture.

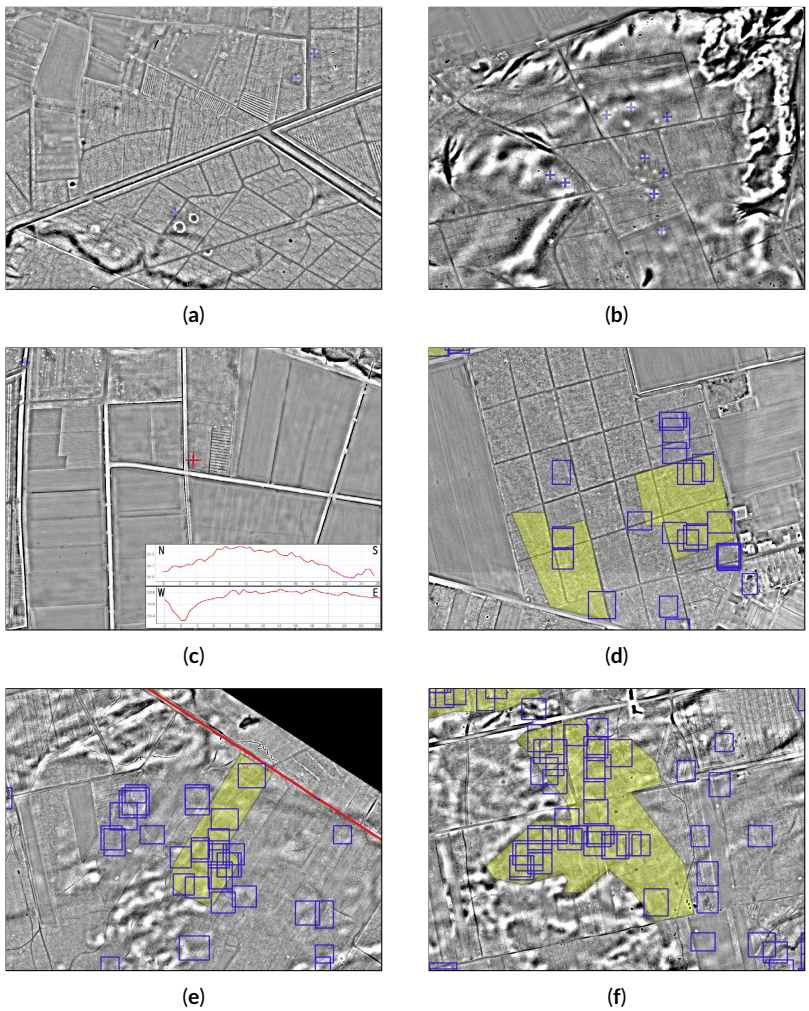

Figure 4.8: Excerpts of LiDAR data, visualized with Simple Local Relief Model (Hesse, 2010), showing barrows (**a–c**) and Celtic fields (**d–f**) detections by WODAN (blue) and Celtic fields annotated during the manual analysis (green; source of the height model: Nationaal Georegister, 2021).

Qualitative Knowledge Gain

As shown, the use of automated detection results in the discovery of previously unknown archaeological objects and contributes to a more complete view (of the distribution) of archaeological objects in the landscape. This data can be used to investigate patterns between these archaeological objects and/or the landscape. It also offers opportunities to investigate the structuring of landscape in the past, especially when archaeological objects such as Celtic fields or hollow roads are mapped. Moreover, it offers insight into the current archaeological research practice and possible biases that result from certain methods and/or interpretations. In the following this is highlighted by two examples.

An example of a more complete view of the patterns and relations between archaeological objects and the surrounding landscape, based on the results of automated detection, is the barrow cluster near the town of Swalmen in the northeastern part of the Midden-Limburg area (Fig. 4.9). Prior to this research 22 burial sites (20 barrows and 2 urnfields), dating between the Late Neolithic and Iron Age, were known from this region. These sites were clustered into several distinct groups. The automated detection and manual analysis yielded 8 and 6 potential barrows in this region respectively. When the region was subsequently manually reexamined an additional 7 potential barrows were discovered, missed during both the manual analysis and the automated detection. The resulting distribution of barrows appears not to be random, but concentrated in a narrow, southwest-northeast orientated zone (Fig. 4.9). Similar formations of barrows, called barrow alignments or barrow lines, are known in the Netherlands, especially from the Veluwe (Bourgeois, 2013). They originate from the Late Neolithic A (2800-2500 cal. BC), and in later periods barrows are added upon it albeit in a more dispersed manner. The different alignments on the Veluwe are comparable in terms of length (a minimum distance of 1–1.5 km) and exhibit a similar placement of barrows, at a fairly regular interval and along a single axis (Bourgeois, 2013). The Swalmen barrow cluster displays several of these characteristics, such as the length (minimum of 2.2 km and maximum of 3.9 km) and placement along a general axis (ca. 240°). However, a regular interval between barrows cannot be observed. Also, it is uncertain which of these barrows date from the Late Neolithic A and form the origin of the potential alignment, as many of the sites are unvalidated, undated and/or of debatable date (see Lanting & Waals, 1974). Therefore, it remains unclear whether the Swalmen barrow cluster concerns a true barrow alignment. Even though, several notions about the cluster might point to a general concern with movement and a predetermined placement of the barrows in the landscape (Løvschal, 2013). For instance, when a simple viewshed analysis (Gillings & Wheatley, 2020) is calculated, using the *Visibility Analysis* plugin (observer height: 1.6 m, target heigth 0.3 m) in QGIS, a person standing on top of the most southwestern potential barrow, situated on the lower terrace of the Meuse river (see Section 4.2.1), would be able to see the group of barrows on the middle terrace and on the group on the high terrace, while the rest of the area is obscured. Furthermore, the line of barrows crosses the *Swalm* stream (valley) at the point where, up until the 1930s, one of the only crossing points of this stream was situated. A similar observation was made for the barrow alignment between Niersen and Epe on the Veluwe, which also led to one and possibly two crossing points (Bourgeois, 2013).

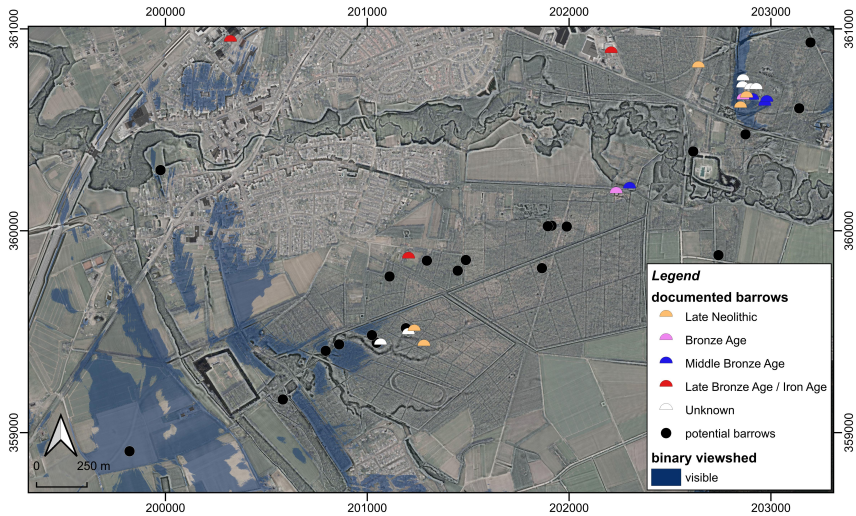

Figure 4.9: Excerpt of LiDAR data, visualized with Simple Local Relief Model (Hesse, 2010), blended with a recent aerial photograph, showing the recorded barrows (hemispheres) and potential new barrows (circles) near Swalmen and a binary viewshed (blue) from the most southwestern potential barrow (observer height: 1.6 m, target height 0.3 m). Note the Roman road running southwest-northeast through the right of the center of the image (source of the height model and aerial photograph: Nationaal Georegister, 2021).

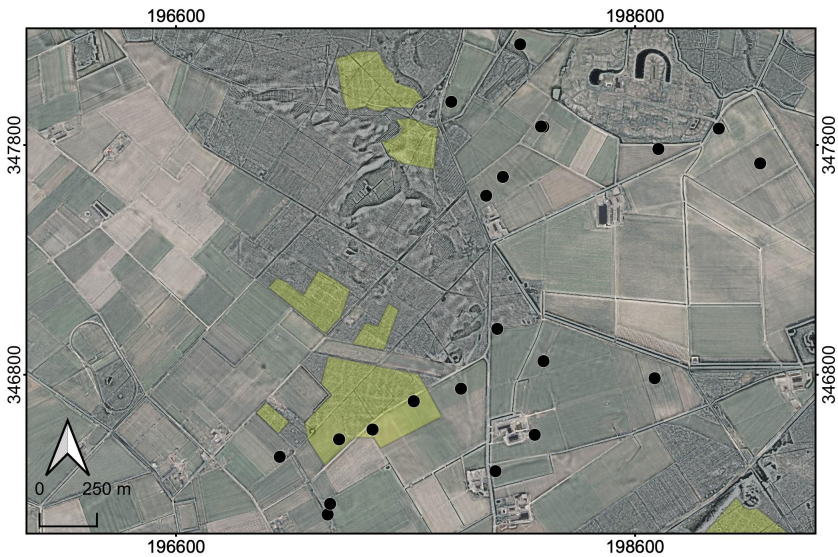

Figure 4.10: Excerpt of LiDAR data, visualized with Simple Local Relief Model (Hesse, 2010), blended with a recent aerial photograph, showing the location of potential Celtic field areas (green) and settlement sites (black; source of the height model and aerial photograph: Nationaal Georegister, 2021).

Certainly one of the major outcomes of this research are the many new potential Celtic fields discovered in the Midden-Limburg area. It shows that, at least in the way areas were parcellated, a comparable agricultural use-strategy was employed in the sand and loess covered Midden-Limburg area as in the sandy regions of the central and northern Netherlands (Arnoldussen, 2018). This significantly changes our view of the Midden-Limburg landscape in later prehistory, as it was assumed that none or a few Celtic fields were present in the area due to the different geology, subsoil, and hydrology (Spek, 2004) and the occurrence of many natural boundaries (van Beek, 2011). If present at all, the Celtic fields were assumed to have been destroyed by (sub)modern agriculture (Arnoldussen, 2013). Although the latter has certainly been the case, as can be seen by the sharp transitions in the presence of Celtic fields in forest and adjacent agricultural fields, it appears that the suggested paucity of research is the main contributing factor (Arnoldussen, 2013). The discovery of the many potential Celtic fields emphasizes a deficiency in the current, local archaeological research practice, which has a primary reliance on field-walking, resulting in an uneven distribution in both site type and site location (see Section 4.2.2), with a clear overrepresentation of settlement sites in modern agricultural areas (see Fig. 4.3). This research, which instead relies on remotely-sensed data, shows an abundance of underrepresented archaeological sites, e.g., Celtic fields, in landscape types, such as forest, that have had little research attention. However, no settlement sites were discovered in the LiDAR data, which shows that a reliance on a single type of archaeological method is detrimental for the knowledge about the archaeology in a particular region. Furthermore, the location of the Celtic fields in relation to modern agricultural fields, necessitates the reconsideration of earlier archaeological interpretations done on the basis of field-walking (Fig. 4.10). For instance, pottery fragments found in the ploughed topsoil of agricultural fields in the vicinity of the detected Celtic fields should not *a priori* be interpreted as settlements, but alternatively as part of debris found in the banks of Celtic field, which were in later periods leveled by agricultural practices. This could especially be the case if only a small amount and/or (strongly) fragmented pieces of pottery are found. Even though, habitation within a Celtic field, such as is attested at other locations in the Netherlands (Arnoldussen & de Vries, 2014), cannot be excluded due to the long period of use of Celtic fields (Arnoldussen, 2018).

4.4 Discussion

In this research the transferability and usability of a Deep Learning object detection tool within archaeological practice was explored by using the WODAN workflow (Chapters 2 & 3), developed in one area of the Netherlands (the Veluwe), on an unrelated area (Midden-Limburg; Fig. 4.1) with a different archaeological record and research history, topography, geo(morpho)logy, and land-use. The results show that WODAN is able to detect barrows and Celtic fields in LiDAR data from the Midden-Limburg area, showing that the workflow can adequately generalize to the new situation.

However, the performance of WODAN still has room for improvement, especially when compared to the overlap between the results and the reference datasets (Table 4.6 & 4.7). This difference between a human and computer in the ability to detect archaeological objects in remotely-sensed data, as shown by the difference between the manual analysis and the results of WODAN, has been noted in earlier research (Chapter 3) and can partly be explained by the fact that a human interpreter, aside from relying on a certain amount of prior experience, archaeological-, and geological knowledge, can observe the vicinity of potential objects, and has the opportunity to consult additional data sources, e.g., aerial imagery (Chapter 3.6.2). Contrary, WODAN has access to a single data source, i.e., LiDAR data. It is therefore hardly surprising that it struggles with certain landscape elements, such as sand dunes and modern field boundaries, that have a comparable appearance in the LiDAR data as barrows and Celtic fields respectively.

4.4.1 Performance and Completeness

Based on the difference between human and computer performance, the question arises what level of performance is acceptable from an object detection tool (see Cowley et al., 2020; Chapter 3.6.3). While this is obviously dependent on the (envisioned) users (see Opitz & Herrmann, 2018, Chapter 3.6.3), the intended task of the tool and the incorporation of the tool (and its results) within the wider archaeological research framework are also of importance (see for instance Banaszek et al., 2018; Cowley et al., 2020; Lambers et al., 2019). Related to this is the question of the level of completeness—how accurately the results reflect the extent of the archaeological objects—that is required. As shown, WODAN is able to detect the majority of demarcated areas of Celtic fields within the Midden-Limburg area, but generally does not detect the full extent, i.e., all the plots within the Celtic field. The same occurs with urnfields, where only several of all barrows within the group are detected. This exclusion of nearby objects might have a technical cause, related to the functioning of the RPN within the Faster R-CNN architecture, e.g., the stride or non-maximum suppression (see Ren et al., 2017). Either way, the exclusion of these objects means a lower level of completeness, which results in low(er) performance when calculating metrics based on the number of objects or the square meters of coverage, even though the method did point to the location of archaeological objects. If automated detection is used independently as the sole source of information, high levels of performance (and completeness) are required. Especially when landscape patterns, such as field systems or roads, are detected. On the other hand, if the method is used to help target limited manual inspection of the data or is used in conjunction with a manual analysis, to enhance the results, a certain degree of exclusion can be afforded.

4.4.2 Combined Human–Computer Strategies for Large-Scale Mapping

This research has shown the potential of automated detection in a complementary, rather than a substitute role, to manual analysis. For instance, WODAN detected additional archaeological objects that were missed during the manual analysis (see Table 4.8). Furthermore, when the northeastern part of the area was reexamined, based on the results of the automated detection, even more potential archaeological objects were discovered, overlooked during both the manual analysis and the automated detection (see for instance Section 4.3.4). This shows that combining these two strategies for large-scale mapping has an added benefit. Furthermore, the feasibility of combined Human–Computer strategies becomes even more apparent when the variability in human detection accuracy (Risbøl et al., 2013) is considered. This variability can lead to a multiplicity of interpretations between different interpreters of the same data (Quintus et al., 2017). Inevitably, manual analysis is biased towards the expectations, experience, knowledge, and observational abilities of the interpreter(s), with the risk of missing or dismissing objects (Halliday, 2013). Contrary, an automated approach, which detects all objects with specific criteria, can offset the aforementioned bias (Bennett et al., 2014), although the pre-defined criteria of the automated detection also come with an inherent bias (see below). Considering the fast run-time of WODAN, it is even possible and efficient to run multiple automated detection models, that can detect different archaeological classes, or multiple versions of the same model (see Chapter 3) during the same time as the manual analysis. Besides, a certain degree of involvement from an archaeological expert is and should remain necessary, in the least to interpret the automated detection results, as the goal of these methods is not to entirely replace the archaeological expert or 'automate archaeology' (Traviglia et al., 2016).

Another benefit of these combined Human–Computer strategies is that it deals with one of the caveats of current automated detection methods: these tools can only detect objects similar to the pre-defined target class(es) while other objects are ignored (Lambers et al., 2019). While automated detection can extend our knowledge on known archaeological classes, it is unable to find potential new and/or unique types of archaeology. This process might unintentionally reinforce the dominance of the objects of interest in the archaeological record, by vastly multiplying the number of examples, while further marginalizing archaeological objects that are not detected (Nuninger et al., 2020b). Especially when automated detection is the sole source of information. For instance, the LiDAR data of the Midden-Limburg area is littered with archaeology from many periods, including modern traces of conflict, e.g, gun emplacements, tank traps, and trenches. These objects belong to a major defense line, the *Maas-Rur-Steilhang-Elmpter-Wald-Stellung* that was constructed in the last part of the Second World War to halt the allied advance into Germany (van der Schriek & Beex, 2017). These traces are not detected by the current model, at least not intentionally, and remain undocumented. The detection of all archaeological objects of interest within an certain area would require either many different models, a model detecting a multitude of classes, or a model that detects a more general class of 'archaeological anomalies'. Recently, attempts at developing the latter have been made (Guyot et al., 2021b).

However, questions remain how applicable these models are in complex terrain, e.g., the Veluwe, where many objects of confusion are present, and how useful such a catch-all category would be in terms of both archaeological research and heritage management (see Trier et al., 2019).

4.4.3 Incorporating Automated Detection into Archaeological Practice

The incorporation of automated detection in the first step of the archaeological prospection scheme, i.e., the desk-based assessment, seems logical. The results of automated detection can be regarded as showing highlighted areas of interest—containing potential archaeological objects that require (field) verification. When used in this way, the results are very comparable, in type and value, to archaeological predictive maps. These maps, used commonly in Dutch archaeological practice (Lauwerier et al., 2017), are based on a quantitative analysis and prior knowledge of the archaeological record, and give a change—low, middle-high, or high—on the occurrence of archaeology within certain, often geomorphological zones (Rijksdienst voor het Cultureel Erfgoed, 2021a). The results of automated detection could be used to add further detail to these maps. When used in this way, the level of competence and especially completeness of an automated detection tool does not have to be extremely high (also see Opitz & Cowley, 2013) as the results are merely one of multiple consulted data sources, that form the basis for subsequent fieldwork.

This research has shown that the employment of automated detection can lead to both quantitative and qualitative knowledge gain of the archaeological record of a certain region. Undeniably, the ability to rapidly map (multiple classes of) archaeological objects in large remote sensing datasets can radically transform archaeological practice, and has broadly positive implications for both research and cultural heritage management (Gattiglia, 2015; Opitz & Herrmann, 2018). Although we should not lose sight of the problems surrounding this shift to a data-intensive approach to science (Huggett, 2020b), from a research standpoint it offers opportunities for spatial analysis and landscape archaeology (Gillings et al., 2020). Through the efficient detecting and mapping of the presence and location of archaeological objects, it facilitates the investigating of trends within the distribution and interrelationship of these objects in the landscape, the emerging of large-scale patterns between different types of objects, and the structuring of the landscape in the past. Understanding this spatial relation between archaeological objects and their surroundings, i.e., the landscape, lies at the core of landscape archaeology (Verhoeven, 2017). The possibility to effectively investigate 'Big' datasets not only means that a phenomenon can be investigated on a wider scale, but also that all available data can be used, instead of a sample, which will let us see details we never could when we were limited to smaller quantities (Gattiglia, 2015). Such a knowledge base is fundamental to effective archaeological research and cultural heritage management (Cowley & Sigurdardóttir, 2011). The obvious benefit for the latter is the possibility to rapidly evaluate a certain region. It can also highlight biases in the existing archaeological record (see Risbøl, 2013), by adding information about underrepresented areas, leading to more appropriate conservation and heritage policy.

4.5 Conclusion

This paper presents the results, efficiency gain, and knowledge discovery of employing a Deep Learning automated detection tool within archaeological practice. The WODAN workflow (Chapters 2 & 3) that has been developed in one area of the Netherlands (the Veluwe), was used to detect two classes of archaeology (barrows and Celtic fields) in the Dutch Midden-Limburg area, which differs in archaeological record and research history, topography, geo(morpho)logy, and land-use from the Veluwe. The results of the automated detection were compared to an inventory of documented archaeological sites and a manual analysis, conducted in the framework of this research, of the same area. The results show that WODAN is able to detect barrows and Celtic fields in LiDAR data from the Midden-Limburg area, and can therefore generalize to this new situation, while being about eighth times faster than the manual analysis. Furthermore, using WODAN led to both a quantitative and qualitative archaeological knowledge gain, by mapping previously unknown potential archaeological objects and by contributing to a more complete view (of the distribution) of archaeological objects in the landscape. The latter can be used to investigate patterns between these archaeological objects and/or the landscape and the structuring of the landscape in the past. Moreover, it offered insight into potential biases within the current archaeological research practice. Future research will focus on improving the performance of WODAN, for instance by combining Citizen Science and automated detection (Lambers et al., 2019).

This research has shown the potential of combining Human–Computer mapping strategies—with automated detection in a complementary, rather than a substitute role, to manual analysis—for the efficient and effective analysis of remotely-sensed datasets of large-scale, complex landscapes. Within current archaeological practice the implementation of automated detection within desk-based assessments seems logical. The results of these methods offer opportunities to enhance and refine existing archaeological predictive maps, which are commonly used in Dutch archaeological practice.