Implementing state-of-the-art deep learning approaches for archaeological object detection in remotely-sensed data: the results of cross-domain collaboration

Olivier, M.; Verschoof, W.B.

Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data: The Results of Cross-Domain Collaboration

journal of computer applications in archaeology

RESEARCH ARTICLE

ubiquity press

MARTIN OLIVIER

WOUTER VERSCHOOF-VAN DER VAART

*Author affiliations can be found in the back matter of this article

ABSTRACT

The ever-increasing amount of remotely-sensed data pertaining to archaeology renders human-based analysis unfeasible, especially considering the expert knowledge required to correctly identify structures and objects in these type of data. Therefore, robust and reliable computer-based object detectors are needed, which can deal with the unique challenges of not only remotely-sensed data, but also of the archaeological detection task.

In this research – across-domain collaboration between archaeology and computer science — the latest developments in object detection and Deep Learning — for both natural and satellite imagery — are used to develop an object detection approach, based on the YOLOv4 framework, and modified to the specific task of detecting archaeology in remotely-sensed LiDAR data from the *Veluwe* (the Netherlands). Experiments show that a general version of the YOLOv4 architecture outperforms current object detection workflows used in archaeology, while the modified version of YOLOv4, geared towards the archaeological task, reaches even higher performance. The research shows the potential and benefit of cross-domain collaboration, where expert knowledge from different research fields is used to create a more reliable detector.

CORRESPONDING AUTHOR:
Martin OLIVIER
ESIEA, FR
martin.olivier1997@gmail.com

KEYWORDS:
Remote Sensing; Deep Learning; YOLO; LiDAR; Cross-domain collaboration

TO CITE THIS ARTICLE:
Olivier, M and Verschoof-van der Vaart, W. 2021. Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data: The Results of Cross-Domain Collaboration. *Journal of Computer Applications in Archaeology*, 4(1): 274–289. DOI: *https://doi.org/10.5334/jcaa.78*

1 INTRODUCTION

Remote sensing has become an essential part of archaeological spatial research, to locate and characterise the surviving physical evidence of past human activity in the landscape (Verhoeven 2017). Consequently, the manual analysis of remotely-sensed data is a common practice in local and regional scale archaeological research and heritage management (Cowley 2012; Opitz and Herman 2018). However, the amount of open access, high-quality remotely-sensed data is continuously growing and exceeds the capacity of general hardware, software, and/or human resources, i.e., it can be considered Geospatial Big Data (McCoy, 2017). Therefore, alternative strategies, which for instance rely on citizen science or computational approaches, are needed to effectively and efficiently analyse these datasets and find and document the overwhelming number of potential archaeological objects therein (Bennett, Cowley & De Laet 2014; Bevan 2015).

In recent years, Computer Vision and more generally Machine Learning — which in turn falls under the broad category of Artificial Intelligence — has made enormous progress thanks to the advent of Deep Learning techniques, which are based upon Artificial and Convolutional Neural Networks (CNNs; Krizhevsky, Sutskever & Hinton 2012; LeCun, Bengio & Hinton 2015). The latter are hierarchically structured algorithms, consisting of multiple layers, which generally comprise a (image) feature extractor and classifier, loosely inspired by the animal visual cortex (Ball, Anderson & Chan 2017). CNNs have been applied in multiple domains, for a variety of tasks ranging from face recognition to medical imaging (Perrault et al. 2019). Comparable to other Machine Learning approaches, a CNN learns to generalise from a large set of examples (generally labelled images) rather than relying on human feature engineering — a crucial and very time-consuming part of classical Computer Vision approaches. A common task in Computer Vision is object detection, where a model has to predict the presence and location of an object, or a class of objects, in an image. Object detection has been successfully implemented on remotely-sensed LiDAR data in archaeology (Bonhage et al. 2021; Trier, Reksten & Løseth 2021; Verschoof-van der Vaart and Lambers 2019; Verschoof-van der Vaart et al. 2020). However, archaeological object detection in remotely-sensed data is not as straightforward as more general object detection tasks, such as finding persons, animals, or household objects in photographs (see Everingham et al. 2010; Lin et al. 2014). Many challenges remain, related to the objects of interest and the data used. These challenges are not restricted to archaeology but are also prevalent in other domains, such as earth observation or environmental sciences (Sumbul et al. 2019; Van Etten 2018).

1.1 CHALLENGES OF OBJECT DETECTION IN REMOTELY-SENSED DATA

Recently, multiple domains have endeavoured to adapt Deep Learning techniques for automated detection in remotely-sensed data, especially in satellite imagery (Ball, Anderson & Chan 2017). For instance, Van Etten (2018) published an extensive review of challenges and possible solutions for using CNNs in satellite-based earth observation. Many of these challenges are not domain specific, but are also prevalent in archaeological automated detection (see Verschoof-van der Vaart in press). The most common challenge is the fact that although the remotely-sensed images are massive in size, the objects of interest are generally very small. This is especially true in comparison to the size of images and objects found in more general purpose datasets, such as Microsoft COCO (Lin et al. 2014) or Pascal VOC (Everingham et al. 2010). These general purpose datasets contain 'natural images' (i.e., photographs of scenes seen in normal, every-day settings) in which objects are generally large and prominent, and occupy a major portion of the image. Traditional object detection methods take advantage of this by downscaling (and pooling) the images when they pass through the CNN to greatly reduce the computational cost. However, this also removes small objects, rendering them impossible to detect. In addition, the objects of interest are often densely clustered but scarcely distributed (e.g., cars in a parking lot or barrows in heathland), generally lack a consistent orientation, are often occluded (by trees, human-made objects, etc.), and in the case of archaeology are also in various states of preservation (Verschoof-van der Vaart in press).

Consequently, most research in other domains centres around dealing with these small objects and with reducing the information loss that occurs during downsampling and pooling, the latter being necessary to deal with the large resolution of the images. Multiple strategies have been developed to address this issue. The most prominent of these is *Feature Fusion* in which information (i.e., feature maps) from earlier layers are added to latter layers in the CNN to increase performance (Qian et al. 2020). Other strategies use *Dilated Convolutions* to increase the focal field of the convolution layers without increasing the kernel size (Ju et al. 2019). On the other hand, Van Etten (2018) uses two modified YOLOv3 detection frameworks (Redmon & Farhadi 2018), where one is responsible for detecting small objects and the other for large objects.

Archaeological automated detection could potentially greatly benefit from incorporating these state-of-the-art developments from other domains. However, as the field of Deep Learning moves at a staggering rate, considerable time investment and up-to-date knowledge is required to evaluate and incorporate these latest developments. Consequently, archaeological automated detection generally 'borrows' methods and techniques

that are thoroughly proven, widely available, and easily implementable, but are not necessarily fully understood or state-of-the-art (Cowley, Verhoeven & Traviglia 2021).

1.2 AIM

Based on the above, the main aim of this paper is to develop a reliable, accurate, and fast workflow geared towards the detection of multiple classes of archaeological objects in remotely-sensed data, by using the latest developments from other domains, and from Deep Learning object detection in general. To accomplish this, the research presented is a joint effort between archaeologists from the Faculty of Archaeology at Leiden University and computer scientists from the French Graduate Engineering School ESIEA.

A state-of-the-art object detection approach, called YOLOv4 (Bochkovskiy, Wang & Liao 2020), has been modified, both in architecture and training regime, for the specific problem of archaeological object detection in LiDAR data. The developed approach is trained and tested on a dataset from the Netherlands (***Figure 1***). In addition, the performance of this optimised model is compared to a general version of the same model, i.e., without any modi-fications, and other object detection approaches used in recent archaeological research. These comparisons allow us to investigate the benefits of using state-of-the-art techniques and modifications. Finally, the potential of cross-domain collaboration between archaeologists and computer scientists can be evaluated.

In the next Section (2), the research area and datasets are introduced, followed by a detailed overview of the object detection approach and modifications used (Section 3). In Section 4 and 5 the results of the experimental evaluation are presented and discussed. The paper ends with an overview of the main insights and future research planned (Section 6).

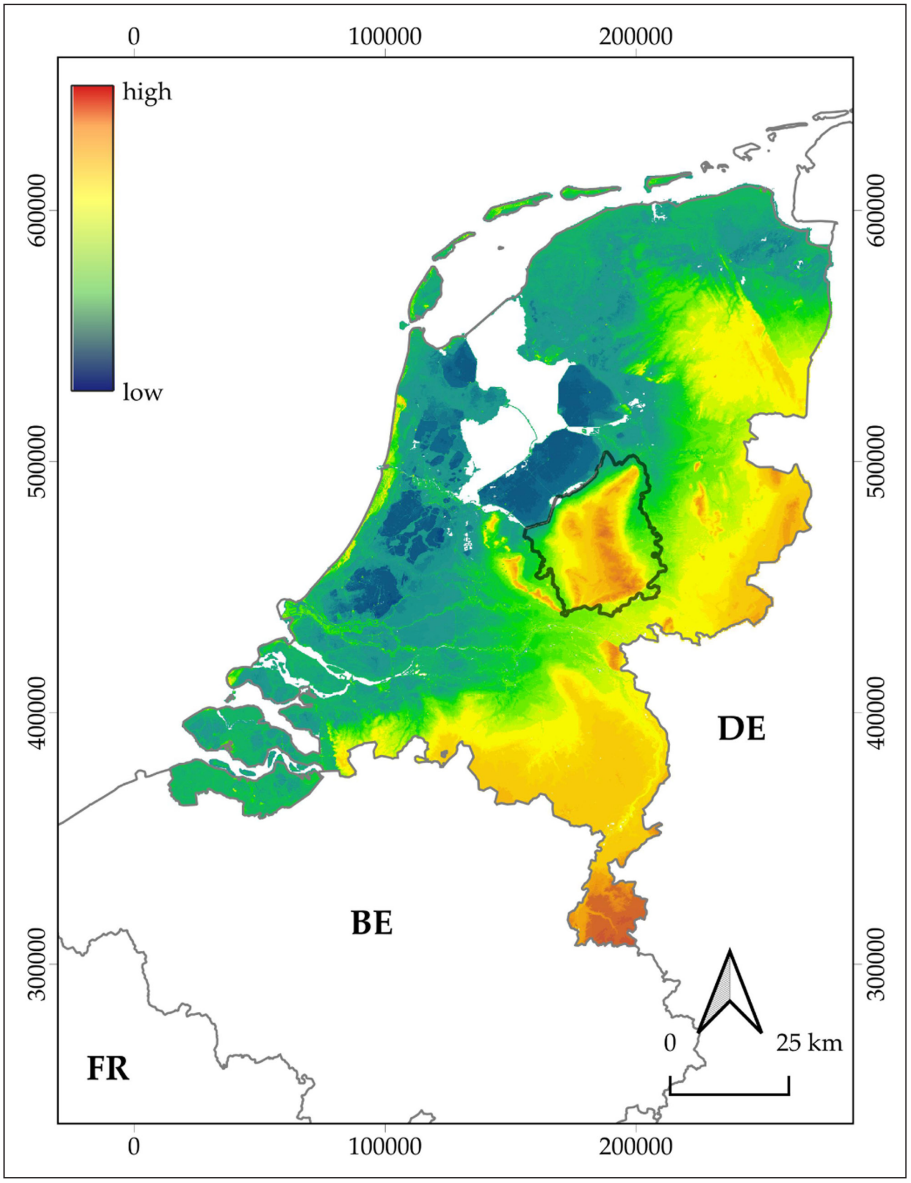

Figure 1 The research area (black outline) on an elevation model of the Netherlands (source of the elevation model: Nationaal Georegister, 2021; coordinates in Amersfoort/RD New, EPSG: 28992; amended from Lambers, Verschoof-van der Vaart & Bourgeois 2019).

2 RESEARCH AREA AND DATASETS

2.1 RESEARCH AREA

In this research, LiDAR data from the *Veluwe*, a region (ca. 2200 km^2) in the central part of the Netherlands is used (***Figure 1***). Nowadays, this region consists predominantly of forest and heathland, interspersed with agricultural fields, built-up areas of various size, and major and minor roads (for a detailed overview of the area see Lambers, Verschoof-van der Vaart & Bourgeois 2019; Verschoof-van der Vaart and Lambers 2019). The Veluwe holds one of the densest concentrations of archaeological objects in the Netherlands, including prehistoric barrows (Bourgeois 2013) and Celtic fields (Arnoldussen 2018), (post) medieval charcoal kilns (Deforce, Groenewoudt & Haneca 2020), hollow roads (Verschoof-van der Vaart & Landauer 2021; Vletter & van Lanen 2018), as well as more recent traces of conflict (van der Schriek & Beex 2017).

Three classes of archaeological objects are of interest for this research: barrows, Celtic fields, and charcoal kilns (***Figure 2***). Barrows are small, round or oval mounds of raised earth (or stone), which are placed over one or several burials. These funerary monuments are fairly common in north-western Europe and date, in the case of the Veluwe, from around 2800 to 1400 cal. BCE (Bourgeois 2013). The diameter of barrows in our datasets is on average about 18 m, although their size varies considerably (Verschoof-van der Vaart in press). The second class of archaeological objects are Celtic fields: agricultural field systems from later prehistory (Late Bronze Age until the Roman Period; ca. 1100 cal. BCE – 200 CE). These are composed of roughly rectangular, embanked plots of fairly regular size (on average 40–50 m) that form extensive, checkerboard-shaped systems (Arnoldussen 2018). In our datasets, these individual plots are considered as a single Celtic field object, not the entire field. This greatly increases the number of examples in the dataset and allows for easier detection, since a single plot is generally rectangular, while the shape of the entire field often is not (see Verschoof-van der Vaart & Lambers 2019). Charcoal kilns are circular platforms or mounds surrounded by a shallow ditch or circle of pits, used for the production of charcoal by heating wood at a specific temperature, while covered with soil (Hirsch et al. 2020). Charcoal kilns found on the Veluwe are on average 14 m in diameter (Verschoof-van der Vaart in press), and are often found in groups or rows along forest paths. These kilns were mainly in use from the late Middle Ages until the second half of the 20th century (1250–1950 CE; Deforce, Groenewoudt & Haneca 2020) and can be found all over Europe as well as in north-eastern America (see Raab et al. 2017).

2.2 DATA

The datasets used in this research (see also Verschoof-van der Vaart et al. 2020) are constructed using LiDAR data, freely available from the online geospatial data repository PDOK (Nationaal Georegister 2021). The data is disseminated as an interpolated Digital Terrain Model (DTM) in GeoTIFF images of 10,000 by 12,500 pixels (5 by 6.25 km). In this research, the DTM has first been processed with the *fill no data* tool in QGIS (QGIS Development Team 2017) to remove pixels without values from the raster image. Subsequently, the DTM was visualised with the Simple Local Relief Model (SLRM; Hesse 2010) visualisation from the *Relief Visualisation Toolbox* (Kokalj & Hesse 2017).

2.2.1 Dataset Generation

One of the challenges of this research is to efficiently and effectively transform (often large) remotely-sensed images into a specific format usable by a Deep Learning approach. Careful choices have to be made in order to have the correct balance of quantity versus quality of examples in the dataset used to train a CNN-based approach, as this is of direct influence to the performance of the resulting model (Kumar & Manash 2019). In order

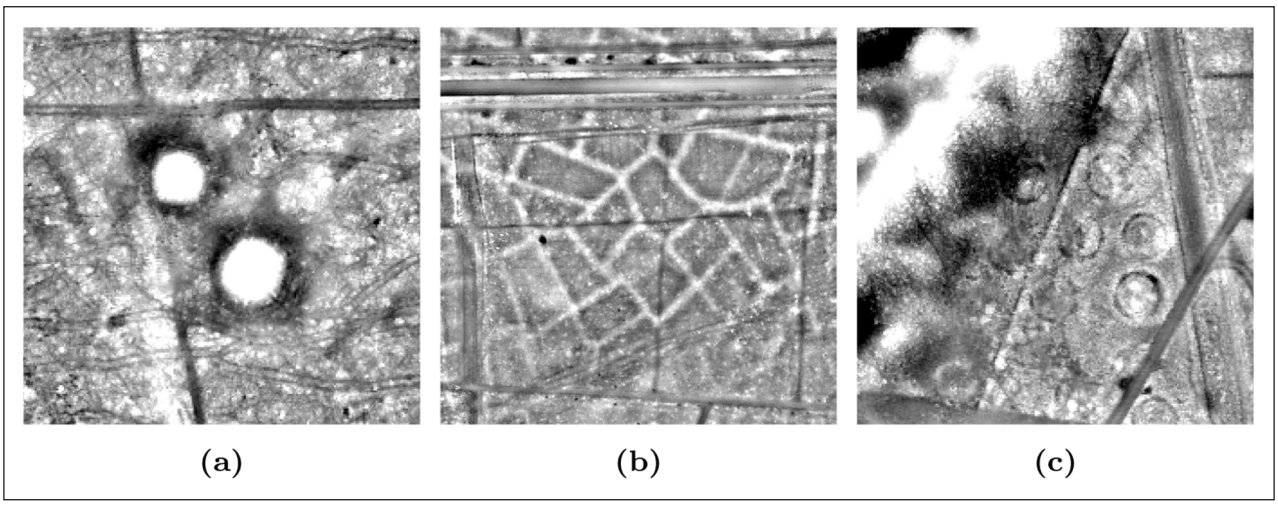

Figure 2 Excerpts of LiDAR data, visualised with Simple Local Relief Model (Hesse 2010), showing: **(a)** barrows; **(b)** Celtic fields; and **(c)** charcoal kilns (source of the elevation model: Nationaal Georegister 2021).

to correctly train the model, a sufficient amount of images that contain one or more examples of a class of object, are needed. Therefore, the intent is to have as many examples and images as possible, although each has to be sufficiently different from the others to prevent overfitting — where the model memorises the examples instead of learning to generalise (Goodfellow, Bengio & Courville 2016).

The original dataset consists of geospatial data, i.e., a vector shape file per object class (with the bounding boxes or annotations encoded as X,Y coordinates) and the LiDAR images in GeoTIFF format (see Verschoof-van der Vaart et al. 2020). In order to use the data for the developed approach, it first has to be converted into the format of the Microsoft COCO dataset (Lin et al. 2014). Therefore, the bounding boxes need to be associated with the correct LiDAR image and subsequently the coordinates of the bounding boxes need to be converted from real-world coordinates to image-based (pixel) coordinates. This information can then be used to make input images, i.e., cropped subtiles of the large LiDAR images. Therefore, a script written in *Python* (Van Rossum & Drake 2009) was developed to automate this task (see ***Figure 3***).

In this script, the *gdalinfo* command line utility (GDAL/OGR contributors 2021) is used to retrieve the coordinates of the LiDAR images. Subsequently, the bounding box of every object in the dataset is associated with one of these images, based on their coordinates and it is checked whether the entire bounding box of the object falls inside the image. Using this newly acquired information, a new database is constructed (***Figure 3***), called *CSV of polygons positions*, which contains all necessary information (i.e., the position of the bounding box, the type of object, and the associated image) to construct the final dataset in Microsoft COCO format.

To create cropped LiDAR subtiles, the coordinates of the bounding boxes are first converted to pixel coordinates (between 0 and the length/width of the LiDAR image), using a simple rescaling and translation (Equation 1).

$$f(x) = \frac{x - x_{min}}{x_{max} - x_{min}}\left(x'_{max} - x'_{min}\right) + x'_{min} \quad (1)$$

(where x_{min}, x_{max} are the original range of the value, and x'_{min}, x'_{max} is the desired range)

Subsequently, the LiDAR images are cropped into subtiles of *N* pixels, with *N* being a fixed value, centred around a particular archaeological object. However, having an object in the exact centre of every subtile can introduce a bias in the object detection model. Therefore, a random *jitter* value on the *x* and *y* coordinates, taken uniformly between [100, 100] is added to the coordinates of every object before cropping. This guarantees that the subtile

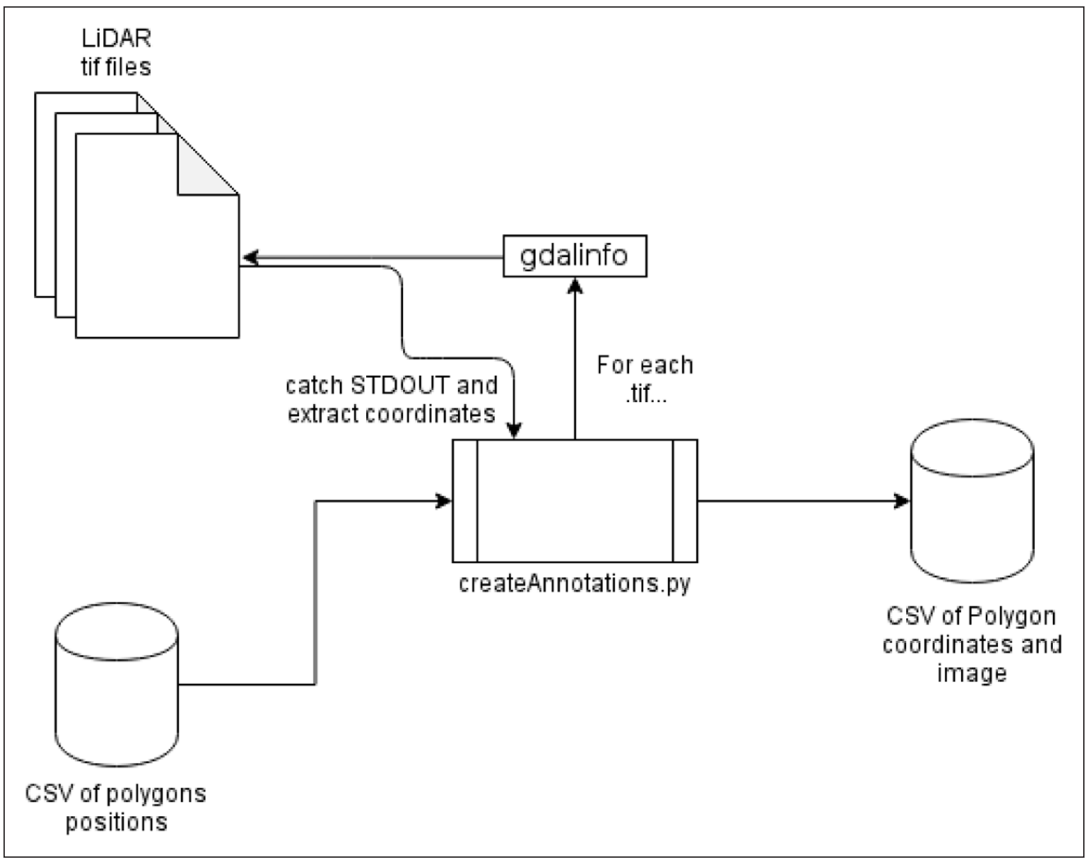

Figure 3 Flow diagram of the dataset creation *Python* script.

is not perfectly centred around an object, reducing bias and creating a more difficult example. After cropping, the script iterates through every entry in the *CSV of polygons positions* database and if an object falls completely within one of the cropped images, the annotation is added to a .txt file associated with that particular image. The resulting dataset, in Microsoft COCO format (Lin et al. 2014), consists of the cropped subtiles, which are all accompanied with .txt files containing the pixel coordinates of the bounding box(es) and class information.

Various elements of the dataset were varied, i.e., different image input sizes, the re-use of objects, and the addition of negative examples, to generate a dataset that resulted in the best performance of our approach. Increasing the input size of the images can reduce the information loss that occurs during the downsampling in the CNN. As archaeological objects are often cluttered together, the cropped subtiles generally contain multiple examples. It is an option to only use images with unique examples, i.e., exclude images with objects that already appeared in another image. However, it might be more beneficial to re-use objects, which already appeared in other images, to increase variability in the dataset. Finally, the use of negative examples, i.e., images with only background and no objects of interest, can teach the CNN the specific texture features of the (non-archaeological) background, which are generally more intricate in remotely-sensed images than in natural images (Gao et al. 2019). This will potentially reduce the number of False Positives.

The final version of the dataset, which produced the best overall performance, consists of images of 1000 × 1000 pixels, incorporates images with re- used objects, and includes circa 20% negatives examples — created by cropping images centred around random coordinates. The total dataset contains around 4500 images, which were randomly split 80/20 in a training and test dataset (***Table 1***). The test dataset is constructed in the same way as the training dataset, in order to more accurately measure the performance of the model.

3 METHODOLOGY

3.1 THE YOLOV4 DETECTION FRAMEWORK

For this research, an object detection framework named YOLOv4 (Bochkovskiy, Wang & Liao 2020) was used. The main concept of YOLO is to approach both the localisation and classification tasks, which together constitute object detection, as a regression problem. Back in 2015, the first version of YOLO (Redmon et al. 2015) marked a breakthrough in object detection as not only the performance of this model was comparable to other methods, but it also had a very impressive inference time, which reached real-time. The YOLO framework is considered a 'one-stage' detector, which means that the object localisation and classification is united in one process. This is contrary to 'two-stage' detectors, e.g., Faster R-CNN (Ren et al. 2017), where this is split in separate processes. In the YOLO framework, input images are downscaled to a fixed resolution and subsequently divided into a grid of $S \times S$ pixels. For each grid cell B bounding boxes (x, y, width, height) with confidence scores are predicted. The confidence prediction represents the Intersection over Union (IoU, see Section 3.2.1) between the predicted bounding box and the ground truth. Each grid cell also predict C conditional class probabilities: P (Class$_i$ Object), which are conditioned on the grid cell containing an object. These parameters (S, B, and C) can be adjusted during training. As the images pass through the CNN only once and the bounding boxes and class predictions are outputted directly as a tensor, without using a separate algorithm such as a Region Proposal Network (Ren et al. 2017), the speed of detections is dramatically increased. Since the release of YOLO, improvements (Redmon & Farhadi 2016; Redmon & Farhadi 2018) have been made, ultimately resulting in the YOLOv4 model (Bochkovskiy, Wang & Liao 2020).

DATASET	IMAGES	NEGATIVE	RATIO
training	3602	691	19.2%
test	931	190	20.4%

Table 1 The datasets used in this research (the columns **Negative** and **Ratio** show the amount and proportion of negative examples respectively).

3.2 VANILLA VERSUS MODIFIED YOLOV4

One of the aims of this research is to investigate the difference between using a general, 'off-the-shelf' object detection model versus a model that uses the latest developments in Deep Learning to be more geared towards the archaeological detection task. Therefore, next to an unmodified 'vanilla' YOLOv4 model (see Bochkovskiy, Wang & Liao 2020), a modified version of YOLOv4 has been developed that uses a variety of augmentations and modifications (***Table 2***). In the following the different modifications and augmentations (bold in ***Table 2***) are discussed in more detail.

PARAMETERS	VANILLA	MODIFIED
backbone CNN	Darknet53	Darknet53
input resolution	416 × 416	**512 × 512**
activation function	Mish	Wish
data augmentation	—	**Cutmix, Mosaic**
regularisation	(DropBlock)	DropBlock
loss function	CIoU	**GIoU**
non-maximum suppression	greedyNMS	**DIoUNMS**

Table 2 Architecture of the vanilla versus modified YOLOv4 model, with the differences between them in bold.

3.2.1 Modifications and Augmentations

One of the main issues with employing CNNs for archaeological object detection is the loss of information due to downscaling of the images (see Section 1.1), e.g., in the YOLOv4 framework images are downscaled to a fixed input size, normally 416 × 416 pixels. Multiple experiments were done trying to measure the impact on performance caused by varying this input size. Since the main purpose is to limit the information loss, it follows that a larger input size and a smaller downscaling will improve detection performance. However, this also comes with a very severe increase of computational memory cost and longer training and testing times. Empirically, it was determined that an input size of 512 × 512 was the largest possible size without exceeding the available memory (i.e., 'Out Of Memory' errors). Presumably, an even larger input size (e.g., 608 × 608) could even further improve detection performance at the cost of increased memory needed.

Another issue with (archaeological) objects in LiDAR images is the fact that these are often occluded, fragmented, and/or do not appear with the same clarity as examples seen in other (remotely-sensed) imagery. This means that we need to construct and train a network that is robust to changes in context and occlusion. Fortunately, research has focused on developing data augmentation techniques (Goodfellow, Bengio & Courville 2016) that attempt to solve these issues. These augmentation techniques generally 'mix' different training images by covering parts of one another. This makes it harder for the model to correctly detect objects but also improves performance. An example of this type of augmentation is *CutMix* (Yun et al. 2019). Another example, *Mosaic* (Bochkovskiy, Wang & Liao 2020), creates a new image out of four, by creating a sort of 'mosaic', where the four images can take varying portions of the new image. Both augmentation techniques were used in the modified YOLOv4 model.

A further modification, regularisation, was implemented to deal with possible overfitting (and increases complexity of the CNN). Overfitting is a common issue in Deep Learning, especially when small training datasets are used, as is often the case in archaeological automated detection (see Section 1.1). Regularisation usually involves 'dropping' random activations in a CNN and train without those connections, through a technique known as DropOut (Srivastava et al. 2014). A comparable, more suitable method that was used in this research is DropBlock (Ghiasi, Lin & Le 2018). This technique works by first choosing random seed points in an annotation, and dropping a continuous region around those points. This is more effective than removing purely random activations, as spatially close activations contain related information.

During the training of a CNN, the loss function — a function that calculates the penalties of incorrect classifications into a single number (Goodfellow, Bengio & Courville 2016) — is optimised. A low loss function is generally regarded as an indication for a well-trained approach and therefore high performance (Guo et al. 2016). For object detection, the Intersection Over Union (IoU) is often used (Equation 2).

$$IoU = \frac{area(B_{PT}) \wedge area(B_{GT})}{area(B_{PT}) \vee area(B_{GT})} \quad (2)$$

(*With B_{GT}, B_{PT} being the ground truth bounding box and predicted bounding box respectively*)

However, recent research has focused on improving the IoU, because while this metric gives a good indication for bounding box quality, it is not a complete one. For example, predicted bounding boxes that do not overlap with a ground truth bounding box get a score of 0, even if the prediction is very close to the ground truth, i.e., the IoU completely disregards the positional relation. Our modified version of YOLOv4 uses the Generalised IoU (GIoU), introduced by Rezatofighi et al. (2019), which is an improvement over the IoU. The GIoU also takes the distance between the predicted bounding box and the ground truth into account by using the size of a box enclosing the prediction and the truth.

Finally, the last modification concerns the Non Maximum Suppression or NMS (Goodfellow, Bengio & Courville 2016). Generally, an object detection models generates many detection proposals (bounding boxes) that can often be redundant, i.e., a single object is detected multiple times. The NMS serves to reduce the number of redundant detections by filtering them. Traditional NMS (or GreedyNMS) takes the detection with the highest confidence score, compares it with all other detections, and removes all detections whose IoU is over a threshold. However, this NMS again does not take distance into account. Therefore, in our research we used DIoUNMS (Zheng et al. 2019), which works comparably to NMS, but replaces the IoU with the Distance IoU that incorporates the distance between bounding boxes.

3.3 POST-PROCESSING

As archaeology relies heavily on GIS, such as QGIS (QGIS Development Team 2017), to manage, analyse, and visualise archaeological information (Gillings, Hacıigüzeller & Lock 2020), it was deemed important to develop a post- processing step to integrate the detection results into this kind of software, i.e., turn the results of the object detection into geospatial data (see also Verschoof-van der Vaart & Lambers 2019). A simple solution to incorporate the results into GIS is to convert the detections into a CSV file, which can be imported as a shape file in most GIS software. Therefore, the *detectionsToCSV* script was developed (see ***Figure 4***), which allows users to convert the results of the object detection into three CSV databases, one for each class.

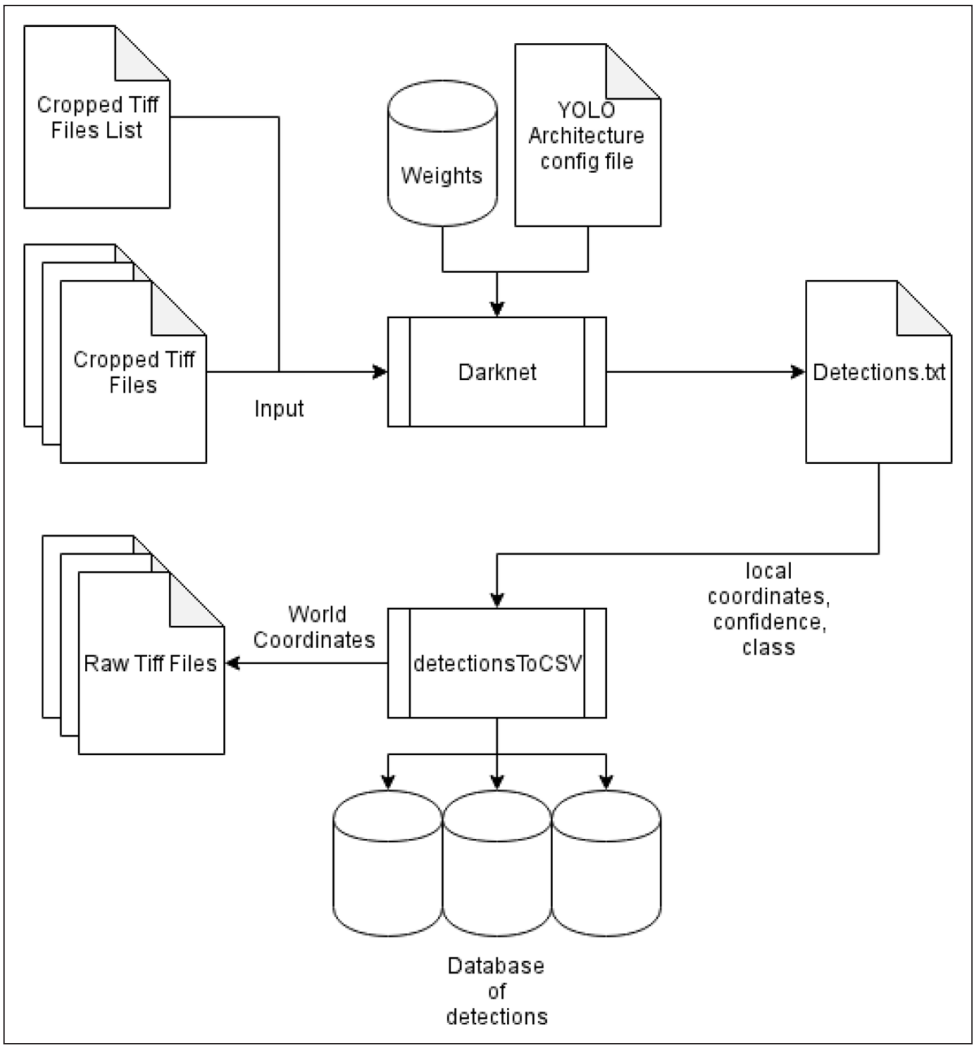

Figure 4 Flow chart showing the testing of the developed workflows: The cropped subtiles of the test dataset are inputted into the Darknet53 CNN, which uses the trained weights and a configuration file to perform inference on the images. The detection results are stored in a .txt file, which is subsequently transformed into three CSV databases.

The detection results per subtile, i.e., a class, confidence score, and location of the bounding box, are saved in a .txt file. The script first connects the subtiles to the LiDAR images from which they originally derive. Then it uses the *gdalinfo* command line utility (GDAL/OGR contributors 2021) to get the coordinates of the main image, and based on that compute the real-world coordinates of the bounding boxes, which are then converted into Well Know Text (WKT) format and written into the corresponding database. Additional post-processing is performed to remove erroneous detections, in particular for Celtic fields, consisting of bounding boxes with a very small width and height but with a high confidence score. These detections are simply discarded by thresholding the ratio between the width and height of all bounding boxes.

4 RESULTS

4.1 IMPLEMENTATION DETAILS

The two versions of the YOLOv4 detection framework (vanilla and modified; see ***Table 2***) were trained and tested on the developed datasets (see ***Table 1***). Both models use a Darknet53 CNN as backbone network, pretrained on the Microsoft COCO dataset (Lin et al. 2014), and are implemented in the *C* programming language. The models were transfer-learned (Razavian et al. 2014) on the training dataset for 10,000 epochs on a Nvidia GTX 1660 GPU. Training times varied between 20 to 25 hours. Subsequently, the fine-tuned models were used to detect archaeological objects in the test dataset of cropped subtiles of 1000 × 1000 pixels (see ***Figure 4***). To test the speed of our model, the interference or testing time was measured. Testing with the YOLOv4 framework went at a rate of 18 images per second. In comparison the testing time of Faster R-CNN on comparable LiDAR images was about one second per image (Verschoof-van der Vaart & Lambers 2019). This high speed of the former is due to YOLO's design as a fast and accurate object detector. While speed is not the focus of this research, a shorter inference time (without a loss in performance) is beneficial, especially when considering the exceptional size of most remotely-sensed datasets. The results of the object detection were post-processed and evaluated (***Figure 5***).

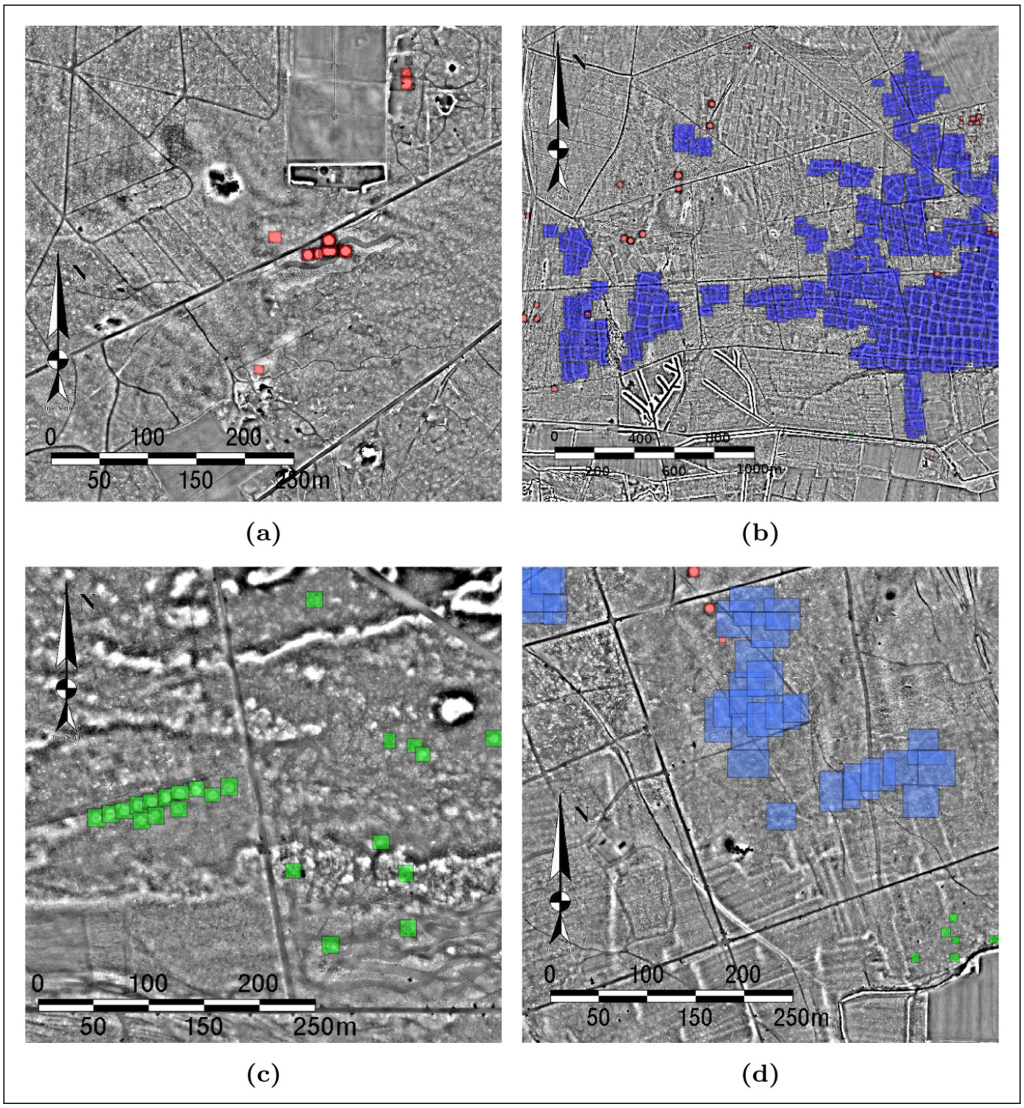

Figure 5 Excerpts of LiDAR data, visualised with Simple Local Relief Model (Hesse 2010), showing successful detections of: **(a)** barrows (red); **(b)** barrows (red) and Celtic fields (blue); **(c)** charcoal kilns (green); **(d)** barrows (red), Celtic fields (blue), and charcoal kilns (green); source of the elevation model: Nationaal Georegister 2021).

4.2 METRICS

To evaluate the performance of the models — and to be able to compare the results to other developed methods — different metrics were calculated, based on the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) in a confusion matrix (Gong 2021). Whether a prediction falls in one of these categories is determined by the amount of overlap between the generated bounding box and the ground truth bounding box, i.e., the. Intersection Over Union (IoU; Equation 2). The threshold for a detection being a TP is normally set to an overlap of 0.5. If the overlap is less, the detection is considered a FP. The (average) IoU can not only be used as a measure for loss during training, but also gives an indication of the quality of the bounding boxes.

The following commonly used metrics for measuring the quality of class predictions of object detection models were calculated (Sammut & Webb 2010): Recall (Equation 3), Precision (Equation 4), and the F1-score (F1; Equation 5). Recall gives a measure of how many relevant objects are selected, while Precision measures how many of the selected items are relevant. The F1-score is the harmonic mean of the Precision and Recall and a measure of the model's performance per class (Sammut & Webb 2010). These measurements are normally restricted between 0 and 1, with higher values indicating a better performance. For readability, the values for all metrics are presented in percentages.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{5}$$

Furthermore, the mAP@50 metric, a popular metric for object detection, was also calculated to measure the quality of class prediction. To compute the mAP, first the

Average Precision (AP) is calculated: a curve of all the predictions done by the model is made, sorted by the predicted confidence level. by moving on this curve, the Recall value will increase, as it is the proportion of True Positives over all possible positives. Precision will have a 'zig-zag' pattern, where it will go up with True Positives but down with False Positives. The AP is defined as the area under this curve. The mAP@50 is simply the mean AP over all classes, and for predictions whose IoU with the ground truth is over 0.5.

4.3 GENERAL PERFORMANCE

Table 3 shows the results, averaged on all object classes, on the testing dataset of the vanilla and modified YOLOv4 models (see ***Table 5*** for a breakdown of the metrics per class). As can be seen, both models perform fairly well (***Figure 5***) with an average F1-score of 0.69 and 0.76 respectively. Both models obtain a high mAP@50 score between 0.75–0.87, and a good average IoU between 0.45–0.58.

The results show that the modified version of YOLOv4 significantly outperforms the vanilla version of YOLOv4 (***Table 3***). A more detailed analysis shows that of the different modifications implemented, the highest performance boost is gained by using the CutMix and Mosaic data augmentations. However, the use of these augmentation techniques does have a negative impact on the training time of the model, which increases up to 25%. To a lesser extent, increasing the input size also resulted in an improvement of performance, albeit with a lower IoU in comparison to models with the same additional modifications but lower input size. The improved performance highlights the benefits of modifying a model towards a specific task. However, it should be noted that developing and finding the optimal modifications is a time-consuming task, which is mostly done by trial and error as there is no predefined 'best way' of adapting a model to a specific task. However, some hyper-parameter tuning could be done using a systematic method, like the Grid Search technique (Liashchynskyi & Liashchynskyi 2019), but this was out of the scope of this research.

Even though performance is reasonably high, some errors do still occurs (see for instance ***Figure 5***, D). Interestingly, while most classes are predicted accurately (see ***Table 4***), there seems to be confusion between barrows and Celtic fields, i.e., barrows are often (circa 10%) confused with Celtic fields. It might be due to a class imbalance present in the dataset, i.e., there are more Celtic Field examples than there are of any other class. Another possible explanation could be the similarities in appearance in LiDAR data between barrows and small fragments of Celtic fields, i.e., both are relatively strong positive elevations. Also the LiDAR data, especially in this format, might not retain enough information to allow the model to correctly distinguish between the two classes. On the other hand, this inter-class confusion does not occur in other research on the detection of both these classes, using the same LiDAR data but a different Deep Learning architecture (see Verschoof-van der Vaart et al. 2020). Finally, the confusion might be related to the training regime used in this research.

Furthermore, False Positives do occur, generally caused by 'objects of confusion', i.e., anthropogenic or natural landscape elements with a comparable morphology to the archaeological (Casana 2020). In the Veluwe area, one of the main problems lies with roundabouts being confused for with barrows (see ***Figure 6***), an issue already noted in Verschoof-van der Vaart & Lambers (2019). This problem might be related to the training of YOLOv4, although more likely is the fact that the LIDAR images simply does not hold enough information to discriminate between those two type of objects. This ties in with the fact that we are not training YOLOv4 to detect archaeological objects, but simply objects. The model is incapable of inferring the archaeological origin of an object based purely on appearance, as is exemplified with the roundabouts being misclassified as barrows. Creating a model capable of such a feat is much harder than simply

MODEL	AV. IOU	RECALL	PRECISION	F1	MAP@50
vanilla YOLOv4	0.45	0.84	0.58	0.69	0.75
modified YOLOv4	0.57	0.93	0.64	0.76	0.86

Table 3 The results, averaged on all classes, of the testing of the vanilla and modified YOLOv4 models on the test dataset.

		PREDICTIONS		
		BARROW	CELTIC FIELD	CHARCOAL KILN
Truth	barrow	0.88	0.10	0.02
	Celtic Field	0.03	0.97	0.0
	charcoal kiln	0.05	0.00	0.95

Table 4 Confusion Matrix of the results of the modified YOLOv4 model on the test dataset. A perfect model should obtain an identity matrix.

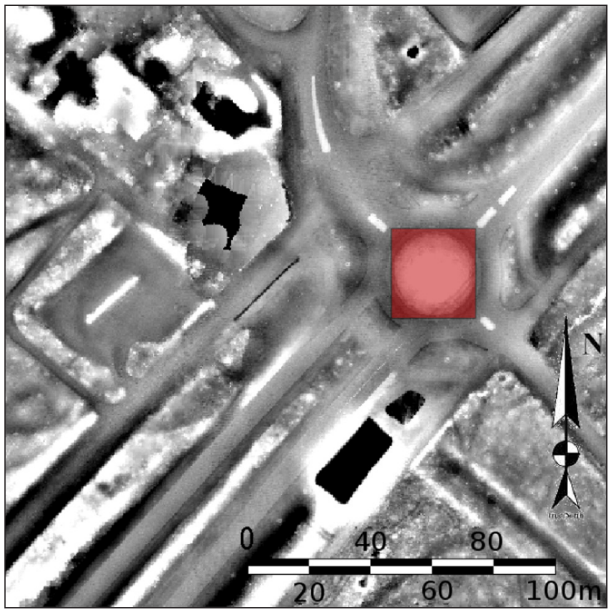

Figure 6 Excerpt of LiDAR data, visualised with Simple Local Relief Model (Hesse 2010), showing the problem of objects of confusion causing False Positives: a roundabout is classified as a barrow due to the similarity of the two in LiDAR data (source of the elevation model: Nationaal Georegister 2021).

training an object detector. A more complex system is needed to understand the archaeological properties of objects, or at least extracting more general and global information than can be done using local convolutions. This could be done by adding domain knowledge to the post-processing of the detection results which eliminates False Positives, such as the Location-Based Ranking approach (Verschoof-van der Vaart et al. 2020).

5 DISCUSSION

5.1 PERFORMANCE COMPARISON

The results of the experiments conducted with the different YOLOv4 models show that these perform well, with a top performance (average F1-score) of 0.76 by the modified YOLOv4 model (see ***Table 3***) and F1-scores of 0.75–0.82 for barrows, 0.72–0.76 for Celtic fields, and 0.59–0.84 for charcoal kilns (***Table 5***). The best performance in terms of F1-score is gained on the charcoal kiln class, which are very characteristic and therefore seem easily recognisable. Interestingly, other researchers have found this 'high speciality' (or complexity) of charcoal kilns as detrimental to detection (Trier, Salberg & Pilø 2018). The performance on the barrow and Celtic field class is lower. This is partly caused by the confusion between the two classes (see ***Table 4***). The main issue with Celtic fields also seems to be the low Precision, due to many False Positives. This could be resolved by using additional post-processing (see Verschoof-van der Vaart et al. 2020).

It is possible to compare this performance to other methods developed for the detection of archaeological objects in LiDAR data (see Bonhage et al. 2021; Trier, Reksten & Løseth 2021; Verschoof-van der Vaart et al. 2020). An important side-note to make is that a direct quantitative comparison often cannot be made, as these methods generally use different training and testing data, and/or detect different classes. For instance, Bonhage et al. (2021) apply a Mask R-CNN (He et al. 2018) on LiDAR data (with different properties) from eastern Germany to detect charcoal kilns, but no other classes. On the other hand, Trier, Reksten & Løseth (2021), use Norwegian LiDAR data (with different properties) to train and test a Faster R-CNN model to detect barrows, charcoal kilns, and hunting traps, but not Celtic fields. In the research of Verschoof-van der Vaart et al. (2020) the same LiDAR dataset and classes were used as in this research. However, the division of the data into a training, validation, and test dataset was different, and therefore the density of archaeological objects in the datasets, i.e., a different ratio of positive and negative examples (see also Soroush et al. 2020), varied with this research. The same is the case for the test dataset used by Trier, Reksten & Løseth (2021, Table 7), in which all images contained at least one archaeological object of interest, i.e., the test dataset contained only positive examples. As noted by the researchers, this results in a bias towards less False Positives. Such differences in the density of objects in the test dataset can be of significant influence on the performance of a Deep Learning approach (Verschoof-van der Vaart et al. 2020).

Even though, in ***Table 5*** a 'rough' comparison between the results (F1-score per class) of our developed YOLOv4 models and these other object detection methods is presented. As shown, our models obtain better results than most of these methods, even when detecting multiple classes. A significant improvement, by circa 25 points on the F1-score, can be observed over the WODAN workflows (Verschoof-van der Vaart et al. 2020) — which offers the closest comparison to our models. On a metric level, the largest improvement lies in the Recall, especially for barrows and charcoal kilns, although generally the Precision is also higher. Especially, in the latter class, charcoal kilns, a significant improvement in performance

METHOD	CELTIC FIELDS			BARROWS			CHARCOAL KILNS		
	R	P	F1	R	P	F1	R	P	F1
YOLOv4 V	0.75	0.75	0.75	0.92	0.65	0.76	0.44	0.90	0.59
YOLOv4 M	0.79	0.86	0.82	0.99	0.57	0.72	0.78	0.92	0.84
WODAN1.0	0.53	0.90	0.15	0.43	0.21	0.28	–	–	–
WODAN2.0	0.45	0.57	0.50	0.40	0.52	0.46	0.35	0.12	0.18
Trier	0.84	0.70	0.76	–	–	–	0.96	0.68	0.80
Bonhage	–	–	–	–	–	–	0.83	0.87	0.85

Table 5 Performance (R: Recall, P: Precision, F1: F1-scores) per archaeological class. Higher is better.

over WODAN can be observed. The low performance of WODAN on charcoal kilns is attributed to the low number of examples in the training dataset (Verschoof-van der Vaart et al. 2020). The results show that YOLOv4 is better in dealing with a small number of examples and is better at correctly detecting and classifying this type of small object.

In comparison to the other research (Bonhage et al. 2021; Trier, Reksten & Løseth 2021), it is shown that the performance on barrows is comparable or higher (0.76 versus 0.75–0.82 F1-score), while the performance on charcoal kilns is comparable (circa 0.85 F1-score). However, we again urge a certain caution in comparing these results as the testing methodology of Trier, Reksten & Løseth (2021) might not accurately represent the performance of the model 'in the wild' (see Verschoof-van der Vaart et al. 2020). Indeed on when tested on another dataset, where negative examples were included, Trier's model obtained a 14% Accuracy on barrows, with 86% False Positives. This highlights the need for a robust and unbiased testing methodology when evaluating the performance of those models (see Verschoof-van der Vaart in press). This comparison does show the benefit of employing state-of-the-art approaches and the implementation of additional modifications geared towards a specific task.

5.2 CROSS-DOMAIN COLLABORATION

The field of Deep Learning moves at an exceptional pace, and original, potentially ground-breaking research is published on a daily basis. This necessitates a considerable time investment in the monitoring of the state-of-the-art, and an ability to properly evaluate recent — often not thoroughly evaluated — research papers and techniques. Consequently, archaeology cannot keep up with these developments. Therefore, in archaeological automated detection, architectures and methods are 'borrowed' that are thoroughly proven, widely available, and easily implementable, but are not necessarily fully understood or state-of-the-art (Cowley, Verhoeven & Traviglia 2021).

This research has shown the potential of cross-domain collaboration, and more specifically, joint research between archaeologists and computer scientists. Within such a collaboration archaeologists can define the exact problems and goals of a project, ensure the implementation of the results in digital archaeological practice (e.g., GIS), and can provide domain knowledge that is generally out of reach of computer scientists (Bennett, Cowley & De Laet 2014). This can inform and help the latter to build more accurate detection methods, which are well-adapted to the specific task. In turn, computer scientists can use their up-to-date knowledge of data science and software engineering to streamline the process of model creation and training, and pinpoint and explain more accurately the issues that can arise during this process (see Trier, Cowley & Waldeland 2019). Especially in Deep Learning, a specialist is needed to correctly create a dataset, select the best CNN architecture, evaluation methods and metrics, and training regime. All those moving parts are very domain dependant, and while there are archaeologists who are familiar with Deep Learning, collaboration between experts from different fields can greatly improve the results of the developed method (see also Gattiglia 2015). The type of collaboration as conducted in this research seems to be especially fruitful, and is able to produce high-quality, high performance models. Since both parties are able to approach the problem with their own tools and knowledge, they are able to see and solve issues in a way that the other can not. In a more general sense, as archaeology generates more and more data it also necessitates proper data analysis techniques (Huggett 2020; McCoy 2017). With the increased collection, aggregation, and processing of data, the occurrence of errors, inconsistencies, and bias also increases (Gattiglia 2015), which can in turn produce wrong or misleading results. Therefore, a collaboration with computer or data scientists might be of extra benefit to properly catch these errors and biases.

Based on this research it seems that archaeology and computer science/engineering synergies very well, which will hopefully result in many promising new approaches for archaeological automated detection.

6 CONCLUSIONS

In this paper, an automated detection workflow, based on the state-of-the-art YOLOv4 detection framework, was presented. The results of the experimental evaluation show that this model is able to accurately detect and categorise archaeological objects in challenging data, such as remotely-sensed LiDAR data. While multiple issues remain to be solved, such as hard to differentiate objects (e.g., roundabouts and barrows), the results are promising. The *Python* scripts to generate the datasets (see Section 2.2.1) and configuration files for the modified YOLOv4 model (see Section 3) are available on the Github repository: *https://github.com/epsln/YOLOv4LiDAR*. The archaeological data used in this study are available from the authors upon reasonable request.

It is also shown that modifying an existing model to a specific task, using the latest developments in Deep Learning object detection, can increase performance over an 'vanilla' off-the-shelf method. Such 'tinkering' is especially fruitful in a cross-domain collaboration, with one party being able to know what kind of architectures, modifications, and hyper-parameters could lead to better performance, while the other's knowledge about the data and the characteristics of objects of interest can lead to new insights on how to better train and construct the method.

6.1 FURTHER IMPROVEMENTS

One promising path for future research is the use of multiple types of data to train and test automated approaches. For example, one could use LiDAR and aerial photography data to train a more robust model. Combining such data might require modifications to the detection model and the use of modified file formats, e.g., images containing four instead of three channels: one for the LiDAR grayscale values, and three for the RGB values of the photographs. Alternatively, multiple models or more elaborate feature fusion could be used.

Another potential avenue for research is the use of semantic segmentation models that would be able to segment objects on a pixel level (see for instance Bundzel et al. 2020; Kazimi, Thiemann & Sester 2019). While bounding boxes are adequate for the localisation of objects, segmentation might offer additional information concerning their size, coverage, etc. The use of bounding boxes is also problematic when objects cover extensive areas and/or have irregular shapes, which makes it problematic to 'catch' these in (rectangular) bounding boxes (Verschoof-van der Vaart in press). Although some potential solutions have been proposed, e.g., rotatable bounding boxes (Liu, Pan & Lei 2017), more precise segmentation could be a useful and interesting study path. Of course, a comprehensive review of the state-of-the-art in segmentation is necessary, and modifications to these methods might be required to achieve acceptable performance on the particular task of archaeological detection in remotely-sensed data.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Martin Olivier
ESIEA, FR

Wouter Verschoof-van der Vaart *orcid.org/0000-0002-1053-3009*
Leiden University, NL

REFERENCES

Arnoldussen, S. 2018. The fields that outlived the Celts: The use-histories of later prehistoric field systems (Celtic Fields or Raatakkers) in The Netherlands. *Proceedings of the Prehistoric Society*, 84: 303–327. DOI: *https://doi.org/10.1017/ppr.2018.5*

Ball, JE, Anderson, DT and **Chan, CS.** 2017. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4): 042609. DOI: *https://doi.org/10.1117/1.JRS.11.042609*

Bennett, R, Cowley, D and **De Laet, V.** 2014. The data explosion: Tackling the taboo of automatic feature recognition in airborne survey data. *Antiquity*, 88(341): 896–905. DOI: *https://doi.org/10.1017/S0003598X00050766*

Bevan, A. 2015. The data deluge. *Antiquity*, 89(348): 1473–1484. DOI: *https://doi.org/10.15184/aqy.2015.102*

Bochkovskiy, A, Wang, C-Y and **Liao, H-YM.** 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection*, 28 April 2020. Available at *https://arxiv.org/abs/2004.10934* [Last accessed 2 November 2021].

Bonhage, A, Raab, A, Eltaher, M, Raab, T, Breuß, M and **Schneider, A.** 2021. A modified Mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland. *Archaeological Prospection*, 1–10. DOI: *https://doi.org/10.1002/arp.1806*

Bourgeois, QPJ. 2013. *Monuments on the Horizon. The Formation of the Barrow Landscape throughout the 3rd and 2nd Millennium BC*. Leiden: Sidestone Press.

Bundzel, M, Jăščur, M, Kováč, M, Lieskovskỳ, T, Sinčák, P and **Tkáčik, T.** 2020. Semantic Segmentation of Airborne LiDAR Data in Maya Archaeology. *Remote Sensing*, 12(22): 3685. DOI: *https://doi.org/10.3390/rs12223685*

Casana, J. 2020. Global-Scale Archaeological Prospection using CORONA Satellite Imagery: Automated, Crowd-Sourced, and Expert-led Approaches. *Journal of Field Archaeology*, 45: 89–100. DOI: *https://doi.org/10.1080/00934690.2020.1713285*

Cowley, D. 2012. In with the new, out with the old? Auto-extraction for remote sensing archaeology. *Proceedings of SPIE*, 8532: 853206. DOI: *https://doi.org/10.1117/12.981758*

Cowley, D, Verhoeven, GJ and **Traviglia, A.** 2021. Editorial for Special Issue: Archaeological Remote Sensing in the 21st Century: (Re)Defining Practice and Theory. *Remote Sensing*, 13(8): 1431. DOI: *https://doi.org/10.3390/rs13081431*

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

Everingham, M, Van Gool, L, Williams, CKI, Winn, J and **Zisserman, A.** 2010. The PASCAL Visual Object Classes (VOC) Challenge. *Internattional Journal of Computer Vision*, 88: 303–338. DOI: *https://doi.org/10.1007/s11263-009-0275-4*

Gao, L, He, Y, Sun, X, Jia, X and **Zhang, B.** 2019. Incorporating negative sample training for ship detection based on deep learning. *Sensors*, 19(3): 684. DOI: *https://doi.org/10.3390/s19030684*

Gattiglia, G. 2015. Think big about data: Archaeology and the Big Data challenge. *Archëaologische Informationen*, 38(1): 113–124. DOI: *https://doi.org/10.11588/ai.2015.1.26155*

GDAL/OGR contributors. 2021. *GDAL/OGR Geospatial Data Abstraction software Library*. Available at *https://gdal.org*. [Last accessed 2 November 2021].

Ghiasi, G, Lin, TY and **Le, QV.** 2018. Dropblock: A regularization method for convolutional networks. *Advances in Neural Information Processing Systems*, 12: 10727–10737.

Gillings, M, Hacıigüzeller, P and **Lock, G.** 2020. Archaeology and spatial analysis. In: Gillings, M, Hacıigüzeller, P and Lock, G (eds.), *Archaeological Spatial Analysis: A Methodological Guide*, 1–16. Oxon/New York: Routledge. DOI: *https://doi.org/10.4324/9781351243858-1*

Gong, M. 2021. A Novel Performance Measure for Machine Learning Classification. *International Journal of Managing Information Technology*, 13(1): 1–19. DOI: *https://doi.org/10.5121/ijmit.2021.13101*

Goodfellow, I, Bengio, Y and **Courville, A.** 2016. *Deep Learning*. Cambridge, MA: The MIT Press.

Guo, Y, Liu, Y, Oerlemans, A, Lao, S, Wu, S and **Lew, MS.** 2016. Deep learning for visual understanding: A review. *Neurocomputing*, 187: 27–48. DOI: *https://doi.org/10.1016/j.neucom.2015.09.116*

He, K, Gkioxari, G, Dollár, P and **Girshick, R.** 2018. *Mask R-CNN*, 13 April 2017. Available at *http://arxiv.org/abs/1703.06870* [Last accessed 2 November 2021].

Hesse, R. 2010. LiDAR-derived Local Relief Models – a new tool for archaeological prospection. *Archaeological Prospection*, 17(2): 67–72. DOI: *https://doi.org/10.1002/arp.374*

Hirsch, F, Schneider, A, Bonhage, A, Raab, A, Drohan, PJ and **Raab, T.** 2020. An initiative for a morphologic-genetic catalog of relict charcoal hearths from Central Europe. *Geoarchaeology*, 35(6): 1974–1983. DOI: *https://doi.org/10.1002/gea.21799*

Huggett, J. 2020. Is Big Digital Data Different? Towards a New Archaeological Paradigm. *Journal of Field Archaeology*, 45: S17. DOI: *https://doi.org/10.1080/00934690.2020.1713281*

Ju, M, Luo, J, Zhang, P, He, M and **Luo, H.** 2019. A Simple and Efficient Network for Small Target Detection. *IEEE Access*, 7: 85771–85781. DOI: *https://doi.org/10.1109/ACCESS.2019.2924960*

Kazimi, B, Thiemann, F and **Sester, M.** 2019. Semantic Segmentation of Manmade Landscape Structures in Digital Terrain Models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2): 87–94. DOI: *https://doi.org/10.5194/isprs-annals-IV-2-W7-87-2019*

Kokalj, Ž and **Hesse, R.** 2017. *Airborne Laser Scanning Raster Data Visualization: A Guide to Good Practice*. Ljubljana: Založba ZRC. DOI: *https://doi.org/10.3986/9789612549848*

Krizhevsky, A, Sutskever, I and **Hinton, GE.** 2012. ImageNet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, 25: 1106–1114. DOI: *https://doi.org/10.1016/j.protcy.2014.09.007*

Kumar, R and **Manash, EB.** 2019. Deep learning: A branch of machine learning. *Journal of Physics: Conference Series*, 1228: 12045. DOI: *https://doi.org/10.1088/1742-6596/1228/1/012045*

Lambers, K, Verschoof-van der Vaart, WB and **Bourgeois, QPJ.** 2019. Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. *Remote Sensing*, 11(7): 794. DOI: *https://doi.org/10.3390/rs11070794*

LeCun, Y, Bengio, Y and **Hinton, G.** 2015. Deep learning. *Nature*, 521: 436–444. DOI: *https://doi.org/10.1038/nature14539*

Liashchynskyi, P and **Liashchynskyi, P.** 2019. *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*, 12 December 2019. Available at *http://arxiv.org/abs/1912.06059* [Last accessed 2 November 2021].

Lin, T-Y, Maire, M, Belongie, S, Hays, J, Perona, P, Ramanan, D, Dollar, P and **Zitnick, C.** 2014. Microsoft COCO: common objects in context. *Lecture Notes in Computer Science*, 8693: 740–755. DOI: *https://doi.org/10.1007/978-3-319-10602-1_48*

Liu, L, Pan, Z and **Lei, B.** 2017. *Learning a Rotation Invariant Detector with Rotatable Bounding Box*, 26 November 2017. Available at *http://arxiv.org/abs/1711.09405* [Last accessed 2 November 2021].

McCoy, MD. 2017. Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science*, 84: 74–94. DOI: *https://doi.org/10.1016/j.jas.2017.06.003*

Nationaal Georegister. 2021. *Publieke Dienstverlening Op de Kaart (PDOK)*. Available at *https://www.pdok.nl/* [Last accessed 2 November 2021].

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

Perrault, R, Shoham, Y, Brynjolfsson, E, Clark, J, Etchemendy, J, Grosz, B, Lyons, T, Manyika, J, Mishra, S and **Niebles, JC.** 2019. *The AI Index 2019 Annual Report*. Standford, CA: Human-Centered AI Institute, Stanford University.

QGIS Development Team. 2017. *QGIS Geographic Information System*. Available at *http://qgis.org* [Last accessed 2 November 2021].

Qian, X, Lin, S, Cheng, G, Yao, X, Ren, H and **Wang, W.** 2020. Object Detection in Remote Sensing Images Based on Improved Bounding Box Re- gression and Multi-Level Features Fusion. *Remote Sensing*, 12(1): 143. DOI: *https://doi.org/10.3390/rs12010143*

Raab, T, Hirsch, F, Ouimet, W, Johnson, KM, Dethier, D and **Raab, A.** 2017. Architecture of relict charcoal hearths in northwestern Connecticut, USA. *Geoarchaeology*, 32(4): 502–510. DOI: *https://doi.org/10.1002/gea.21614*

Razavian, AS, Azizpour, H, Sullivan, J and **Carlsson, S.** 2014. CNN features off-the-shelf: An astounding baseline for recognition. *IEEE CVPRWorkshop*, 806–813. DOI: *https://doi.org/10.1109/CVPRW.2014.131*

Redmon, J, Divvala, SK, Girshick, RB and **Farhadi, A.** 2015. *You Only Look Once: Unified, Real-Time Object Detection*, 9 May 2016. Available at *http://arxiv.org/abs/1506.02640* [Last accessed 2 November 2021]. DOI: *https://doi.org/10.1109/CVPR.2016.91*

Redmon, J and **Farhadi, A.** 2016. *YOLO9000: Better, Faster, Stronger*, 25 December 2016. Available at *http://arxiv.org/*

abs/1612.08242 [Last accessed 2 November 2021]. DOI: *https://doi.org/10.1109/CVPR.2017.690*

Redmon, J and **Farad, A.** 2018. *YOLOv3: An Incremental Improvement*, 8 April 2018. Available at *http://arxiv.org/abs/1804.02767* [Last accessed 2 November 2021].

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

Rezatofighi, SH, Tsoi, N, Gwak, J, Sadeghian, A, Reid, ID and **Savarese, S.** 2019. *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*, 15 April 2019. Available at *http://arxiv.org/abs/1902.09630* [Last accessed 2 November 2021]. DOI: *https://doi.org/10.1109/CVPR.2019.00075*

Sammut, C and **Webb, GI.** 2010. *Encyclopaedia of Machine Learning*. Boston, MA: Springer. DOI: *https://doi.org/10.1007/978-0-387-30164-8*

Soroush, M, Mehrtash, A, Khazraee, E and **Ur, JA.** 2020. Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. *Remote Sensing*, 12(3): 500. DOI: *https://doi.org/10.3390/rs12030500*

Srivastava, N, Hinton, G, Krizhevsky, A and **Salakhutdinov, R.** 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *JMLR*, 15(1): 1929–1958.

Sumbul, G, Charfuelan, M, Demir, B and **Markl, V.** 2019. Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. *IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904. Yokohama: IEEE. DOI: *https://doi.org/10.1109/IGARSS.2019.8900532*

Trier, ØD, Cowley, D and **Waldeland, AU.** 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2): 165–175. DOI: *https://doi.org/10.1002/arp.1731*

Trier, ØD, Reksten, JH and **Løseth, K.** 2021. Automated mapping of cultural heritage in Norway from airborne lidar data using faster R-CNN. *International Journal of Applied Earth Observation and Geoinformation*, 95: 102241. DOI: *https://doi.org/10.1016/j.jag.2020.102241*

Trier, ØD, Salberg, A-B and **Pilø, LH.** 2018. Semi automatic mapping of charcoal kilns from airborne laser scanning data using deep learning. In: Matsumoto, M and Uleberg, E (eds.). *CAA 2016: Oceans of Data. Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology*, 219–231. Oxford: Archaeopress.

Van der Schriek, M and **Beex, W.** 2017. The application of LiDAR-based DEMs on WWII conflict sites in the Netherlands. *Journal of Conflict Archaeology*, 12(2): 94–114. DOI: *https://doi.org/10.1080/15740773.2017.1440960*

Van Etten, A. 2018. *You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery*, 24 May 2018. Available at *https://arxiv.org/pdf/1805.09512.pdf* [Last accessed 2 November 2021].

Van Rossum, G and **Drake, F.** 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Verhoeven, GJ. 2017. Are we there yet? A review and assessment of archaeological passive airborne optical imaging approaches in the light of landscape archaeology. *Geosciences*, 7(3): 86. DOI: *https://doi.org/10.3390/geosciences7030086*

Verschoof-van der Vaart, WB. In press. *Learning to Look at LiDAR. Combining CNN-based object detection and GIS for archaeological prospection in remotely-sensed data*. Unpublished thesis (PhD), Leiden University.

Verschoof-van der Vaart, WB and **Lambers, K.** 2019. Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1): 31–40. DOI: *https://doi.org/10.5334/jcaa.32*

Verschoof-van der Vaart, WB, Lambers, K, Kowalczyk, W and **Bourgeois, QPJ.** 2020. Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from The Netherlands. *ISPRS International Journal of Geo-Information*, 9(5): 293. DOI: *https://doi.org/10.3390/ijgi9050293*

Verschoof-van der Vaart, WB and **Landauer, J.** 2021. Using CarcassonNet to automatically detect and trace hollow roads in LiDAR data from the Netherlands. *Journal of Cultural Heritage*, 47: 143–154. DOI: *https://doi.org/10.1016/j.culher.2020.10.009*

Vletter, WF and **van Lanen, RJ.** 2018. Finding vanished routes: Applying a multi-modelling approach on lost route and path networks in the Veluwe Region, The Netherlands. *Rural Landscapes: Society, Environment, History*, 5(2): 1–19. DOI: *https://doi.org/10.16993/rl.35*

Yun, S, Han, D, Oh, SJ, Chun, S, Choe, J and **Yoo, Y.** 2019. *CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features*, 28 May 2019. Available at *http://arxiv.org/abs/1905.04899* [Last accessed 2 November 2021]. DOI: *https://doi.org/10.1109/ICCV.2019.00612*

Zheng, Z, Wang, P, Liu, W, Li, J, Ye, R and **Ren, D.** 2019. *Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression*, 19 November 2019. Available at *http://arxiv.org/abs/1911.08287* [Last accessed 2 November 2021].

TO CITE THIS ARTICLE:
Olivier, M and Verschoof-van der Vaart, W. 2021. Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data: The Results of Cross-Domain Collaboration. *Journal of Computer Applications in Archaeology*, 4(1): 274–289. DOI: *https://doi.org/10.5334/jcaa.78*

Submitted: 05 July 2021 Accepted: 25 October 2021 Published: 08 December 2021