Objective comparison of relief visualization techniques with deep CNN for archaeology

Alexandre Guyot, Marc Lennon, Laurence Hubert-Moy

Journal of Archaeological Science: Reports 38 (2021) 103027

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Objective comparison of relief visualization techniques with deep CNN for archaeology

Alexandre Guyot [a,b,*], Marc Lennon [b], Laurence Hubert-Moy [a]

[a] Laboratoire LETG - UMR 6554, Université Rennes 2, Place du recteur Henri Le Moal, 35043 Rennes, France
[b] Hytech-Imaging, 115 Rue Claude Chappe, 29280, Plouzané, France

ARTICLE INFO

Keywords:
Archaeological prospection
Remote-sensing
Computer vision
Deep convolutional network
LiDAR
Airborne laser system

ABSTRACT

Archaeology has been profoundly transformed by the advent of airborne laser scanning (ALS) technology (a.k.a airborne LiDAR). High-resolution and high-precision synoptic views of earth's topography are now available, even in densely forested environments, to identify and characterize landform patterns resulting from past human occupation. ALS-based archaeological prospection relies on digital terrain model (DTM) visualization techniques (VTs) that highlight subtle topographical changes perceived and interpreted by archaeologists. An increasing number of VTs have been developed, and they have been evaluated to date mainly based on subjective human perception. This study developed a new approach based on state-of-the-art computer-vision algorithms to benchmark VTs using objective metrics. Thirteen VTs were applied to a ALS-derived DTM, and a deep convolution neural network (deep CNN) was implemented and trained to automatically detect and segment archaeological structures from these images. Visual interpretation of the images showed that the most informative VT was e2MSTP, which combined a multiscale topographic analysis (MSTP) with a morphologically explicit image and a slope-invariant relief detrending technique. The deep CNN approach confirmed these results and provided objective performance metrics. This study indicates that the computer vision approach opens new perspectives in the objective selection of the most suitable VT for archaeological prospection.

1. Introduction

In the past few decades, archaeology has been profoundly transformed by the advent of airborne laser scanning (ALS) technology (a.k.a airborne LiDAR). High-resolution and high-precision synoptic views of earth's topography are now available, even in densely forested environments, to identify and characterize landform patterns resulting from past human occupation. Common ALS-based archaeological prospection relies on digital terrain model (DTM) visualization techniques (VTs) to highlight subtle topographical changes that are visually interpreted by archaeologists (Štular et al., 2012).

Representation of the characteristics archaeological structures (e.g. size, shape, orientation, landscape context, topographic position) varies greatly among VTs (Kokalj and Hesse 2017). However, selecting the most suitable VT for enhancing the perception of archaeological structures remains challenging. Several studies have provided valuable assessment of multiple VTs applied to ALS-derived DTM via visual comparison (Bennett et al., 2012; Devereux et al., 2008; Doneus, 2013; Hesse, 2010; Orengo and Petrie, 2018; Štular et al., 2012; Zakšek et al., 2011). However, this approach is limited due to the subjectivity and bias of visual interpretation (Grammer et al., 2017; Risbøl, 2013), which can influence identification and characterization decisions. To our knowledge, only one study to date has addressed this concern with an objective approach: Mayoral et al. (2017) assessed VTs analytically based on local contrast and zonal statistics. Their approach provided useful information about the ability of VTs to perceive variations in local slope or roughness based on pre-selected topographic conditions. However, it did not address global objective assessment of VTs and did not consider an automatic-detection framework. These limitations and the growing number of available VTs (Kokalj and Somrak, 2019) increase the need to develop new objective assessment tools and methods.

The computer-vision field has also changed profoundly in recent years, especially with the development of deep convolutional neural networks (deep CNNs) to solve complex image-analysis tasks. CNN is a type of artificial neural network whose connections are roughly inspired by biological processes in the visual cortex (Hubel and Wiesel, 1962). Deep CNNs are composed of many connected layers that can learn hierarchical representations of data with multiple levels of abstraction

* Corresponding author.
E-mail address: alexandre.guyot@univ-rennes2.fr (A. Guyot).

https://doi.org/10.1016/j.jasrep.2021.103027
Received 8 February 2021; Received in revised form 26 April 2021; Accepted 27 April 2021
Available online 8 May 2021

(LeCun et al., 2015). While emerging in the 20th century, it is only in the past decade that implementation of deep CNNs resulted in ground-breaking results in image classification (Krizhevsky et al., 2012) and object detection (Girshick. 2015). For details on deep CNNs, see Goodfellow et al. (2016).

The scientific community's increasing interest for such high-performance computer vision capability has resulted in the publication of multiple open source state-of-the art deep CNN software frameworks. Among them, Mask R-CNN, available in different open source Python implementations (Abdulla, 2017; Wu et al., 2019), is designed for object detection and instance segmentation. In deep CNN, object detection predicts the presence and location (surrounded by a bounding-box) of an object in the image, while instance segmentation adds a contour (also called "mask" or "segment") to the detected object.

The first approaches using deep CNN applied to ALS-based archeological prospection were developed to automatically detect charcoal kilns (Trier et al., 2016). They were then evaluated in different contexts and configurations to detect archaeological structures (Gallwey et al., 2019; Kazimi et al., 2018; Trier et al., 2018, 2021; Verschoof-van der Vaart et al., 2020; Verschoof-van der Vaart and Lambers, 2019; Bonhage, 2021). In these studies, a single type of input data was used (either raw elevation data or VT), which had been selected empirically based on intuition or the visual perception provided by the input data. Very few studies evaluated the use of various VTs with deep CNNs in archaeological context. Somrak et al. (2020) applied deep CNN models with different ALS-derived inputs for classifying images of ancient settlements. However, to our knowledge, no studies questioned the potential of computer-vision approaches, especially object detection and segmentation, for the objective assessment of VTs and the relation between computer-based and human-based perception.

In this study we developed a new approach that uses state-of-the-art computer-vision algorithms to benchmark VTs using objective metrics. To this end, VTs were first visually interpreted and compared to assess their ability to identify archaeological structures. Then, the same VTs were compared using a deep CNN trained for automatic detection and segmentation of archaeological structures.

First, we tested the assumption that visual representation of data, effective from the perspective of human vision, is also effective from the perspective of deep CNNs. Then, we presented the results and discussed the benefits and limits of an objective comparison of ALS-derived relief VTs using deep CNN for archaeology.

2. Materials and methods

2.1. Study area

The study area (Fig. 1) was located in the Morbihan department (Brittany, France), along the Atlantic coast. The region has a complex and fragmented mosaic of landscapes. The hinterland is composed of woodlands, moorlands and farmland that form a rural environment oriented to agriculture. The coastal area is also diverse, with estuaries and small islands near the intricate Gulf of Morbihan and large open, sandy areas in the Bay of Quiberon that concentrate most of the economic activities of tourism and fisheries.

The area is home to a unique megalithic heritage. Erected between the 5th and 3rd millennia BC, the Neolithic architecture (standing stones and megalithic tombs) represents an exceptional corpus of archaeological sites that are candidates for the UNESCO World Heritage List. Beyond this emblematic heritage, the coast of Morbihan includes a wide variety of archaeological sites that encompass several prehistorical and historical periods.

2.2. Dataset

2.2.1. Digital terrain model

The DTM was generated from a ALS point cloud collected with a fixed-wing plane using an Optech Titan ALS sensor operated over the study area in the leaf-off season in 2016. The specifications of the airborne acquisition were defined to obtain a nominal point density of 14 points/m^2. Ground points were filtered from the raw point cloud

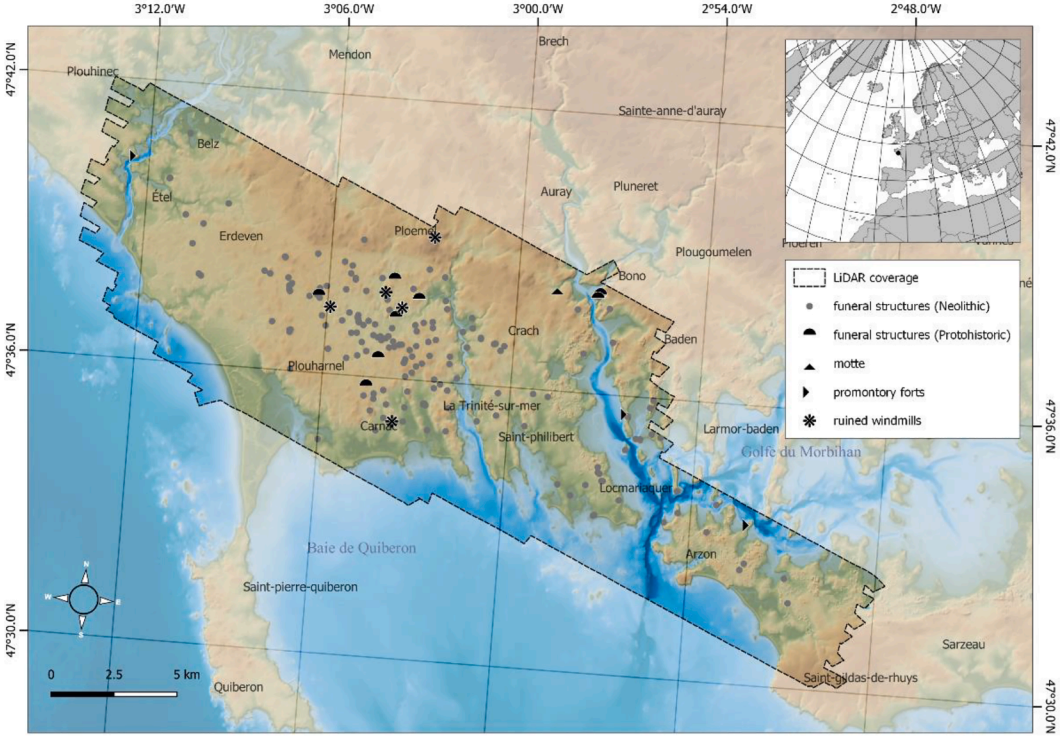

Fig. 1. Airborne Laser Scanning (ALS) coverage of the study area with archaeological reference data used to train and test the Deep convolution neural network models.

using LAStools (Isenburg, 2020) before being interpolated to create a Triangular Irregular Network that was rasterized onto a grid of 50 cm resolution (see Guyot et al., 2018 for processing details).

2.2.2. Archaeological reference data

The reference dataset consisted of 195 georeferenced polygons that represented footprints of known archaeological sites in the study area. The sites were selected from the regional archaeological reference dataset provided by Brittany's Service régional de l'archéologie. Only archaeological structures whose topographic characteristics could be perceived on the ALS-derived DTM were kept (excluding sites related only to small-object deposits and sites considered as aboveground structures with no influence on the bare-earth topography, such as standing stones).

The archaeological sites selected had diverse chronologies, morphologies and landscape contexts. Their state of conservation also varied greatly, from long-known restored monuments to unexcavated little-documented structures. The reference dataset included 195 archaeological structures (Fig. 1): 176 funeral structures attributed to the Neolithic, 10 funeral structures attributed to protohistoric periods, 1 motte, 3 promontory forts and 5 ruined windmills. Note that the great majority of structures are elevated and there are only few depressions.

Given the highly imbalanced dataset (since Neolithic structures dominated), the annotations were intentionally grouped into a single "archaeological structure" class with no further distinction. The reference dataset was converted from a geospatial format to an annotation format (json COCO format) in which each annotation was associated with its corresponding VT tile to be input into the deep CNN architecture. Due to the spatial proximity of some archaeological sites, 150 VT images covered the 195 annotations (a mean of 1.3 annotations per image).

2.3. Methods

2.3.1. Visualization techniques of the ALS-derived terrain model

We compared 13 VTs: 12 came from archaeological prospection literature, and one (e^2MSTP) was designed during this study. All VTs were applied to the ALS-derived DTM at 50 cm resolution, with specific calculation and visualization parameters (Table 1) using open-source tools such as RVT software (Kokalj, 2020), WhiteboxTools (Lindsay, 2020) and Python blend-mode libraries (Roscheck, 2020).

All VTs (Fig. 2) were normalized to 0–255 using the visualization parameters (Table 1) and converted to 8-bit 3-band images (RGB) to be used as input to the deep CNN. Greyscale VTs were transformed from 8-bit (0–255) single-band to 8-bit (0–255) 3-band images by duplicating the 8-bit single-band images.

The blending techniques were applied using the 3-band normalized and transformed images.

2.3.2. Deep CNN for instance segmentation

The deep CNN was based on a Detectron2 framework (Wu et al., 2019) that implemented a Mask R-CNN architecture for instance segmentation (Fig. 3). Mask R-CNN was chosen for its ability to perform instance segmentation by combining automatic detection and segmentation phases in sequential order. The benefit of instance segmentation for archaeological prospection is that besides automatic detection, predicted segments can be used for morphological or contextual characterization of the terrain anomalies identified.

Detectron2 was configured to use a Resnet-101 + FPN backbone, and training hyperparameters (Table 2) were predefined and remained static for all experiments.

For the transfer-learning strategy, weights of the network were initialized using a model pre-trained with a the Common Object in Context (COCO) dataset (Lin et al., 2015).

A Data augmentation technique was included in the training workflow with randomized flip, crop and rotation transformations

2.3.3. Evaluation metric and cross-validation

Performances of the resulting segmentation were evaluated statistically using the Average Precision (AP) metric (Padilla et al., 2020) for an intersection over union (IoU) threshold of 0.5. This threshold value, commonly used in the literature, was justified in our study by the fuzzy

Table 1
Visualization techniques compared in the study.

	Description	References	Calculation parameters	Visualization parameters
HS	Analytical hillshading	(Yoëli, 1967)	Sun azimuth: 315°; Sun elevation angle: 35°	Linear histogram stretch between 0 and 1
HS_PCA	PCA of multi-analytical hillshading	(Devereux et al., 2008)	Sun azimuths: 16 directions; Sun elevation angle: 35°; Number of principal components: 3	Histogram equalization with 2% cut-off
SLP	Gradient of elevation	(Doneus and Briese, 2006)	No parameters	Linear histogram stretch between 0 and 51°
ON	Negative topographic openness	(Doneus, 2013)	Number of search directions: 16; Search radius: 10 px	linear histogram stretch between 60° and 95°
OP	Positive topographic openness	(Doneus, 2013)	Number of search directions: 16; Search radius: 10 px	linear histogram stretch between 60° and 95°
SVF	Sky view factor	(Kokalj et al., 2011)	Number of search directions: 16; Search radius: 10 px	Linear histogram stretch between 0.64 and 1.00
LD	Local dominance	(Hesse, 2016)	Minimum radius: 10 px; Maximum radius: 20 px	Linear histogram stretch between 0.5 and 1.8
SLRM	Simple local relief model	(Hesse, 2010)	Radius for trend assessment: 20 px	Histogram equalization with 2% cut-off
RRIM	Red relief image map	Based on Chiba et al. (2008)	Source images: openness ((OP-ON) / 2) & SLP (red-toned) Blending: addition with 70% (SLP) and 30% openness	See SLP, OP, ON
MSTP	Multiscale topographic position	(Lindsay et al, 2015; Guyot et al., 2018)	Number of |DEV| calculation: 30; Micro scale (Blue): 3 to 21 px; Meso scale (Green): 23 to 203 px ; Macro scale (Red): 223 to 2023 px	linear histogram stretch between 0 and 3
e^2MSTP	enhanced MSTP	Adapted from Guyot et al. (2021)	Source images: MSTP, RRIM, SLRM. Blending: SLRM blended (screen, 25% opacity) with RRIM blended (softlight, 70% opacity) with MSTP	See MSTP, RRIM, SLRM
VAT	Visualization for Archaeological Topography	(Kokalj and Somrak 2019)	Source images: HS, SLP, OP, SVF Blending: SVF blended (multiply, 25% opacity) with OP blended (overlay, 50% opacity) with SLP blended (luminosity, 70% opacity) with HS.	See HS, SLP, OP, SVF
VAT-HS_channels	3-band, Visualization for Archaeological Topography	(Somrak, Džeroski, and Kokalj 2020)	Source images: SLP, OP, SVF Combined in a 3-band RGB image (Red: SLP, Green:OP, Blue: SVF)	See SLP, OP, SVF

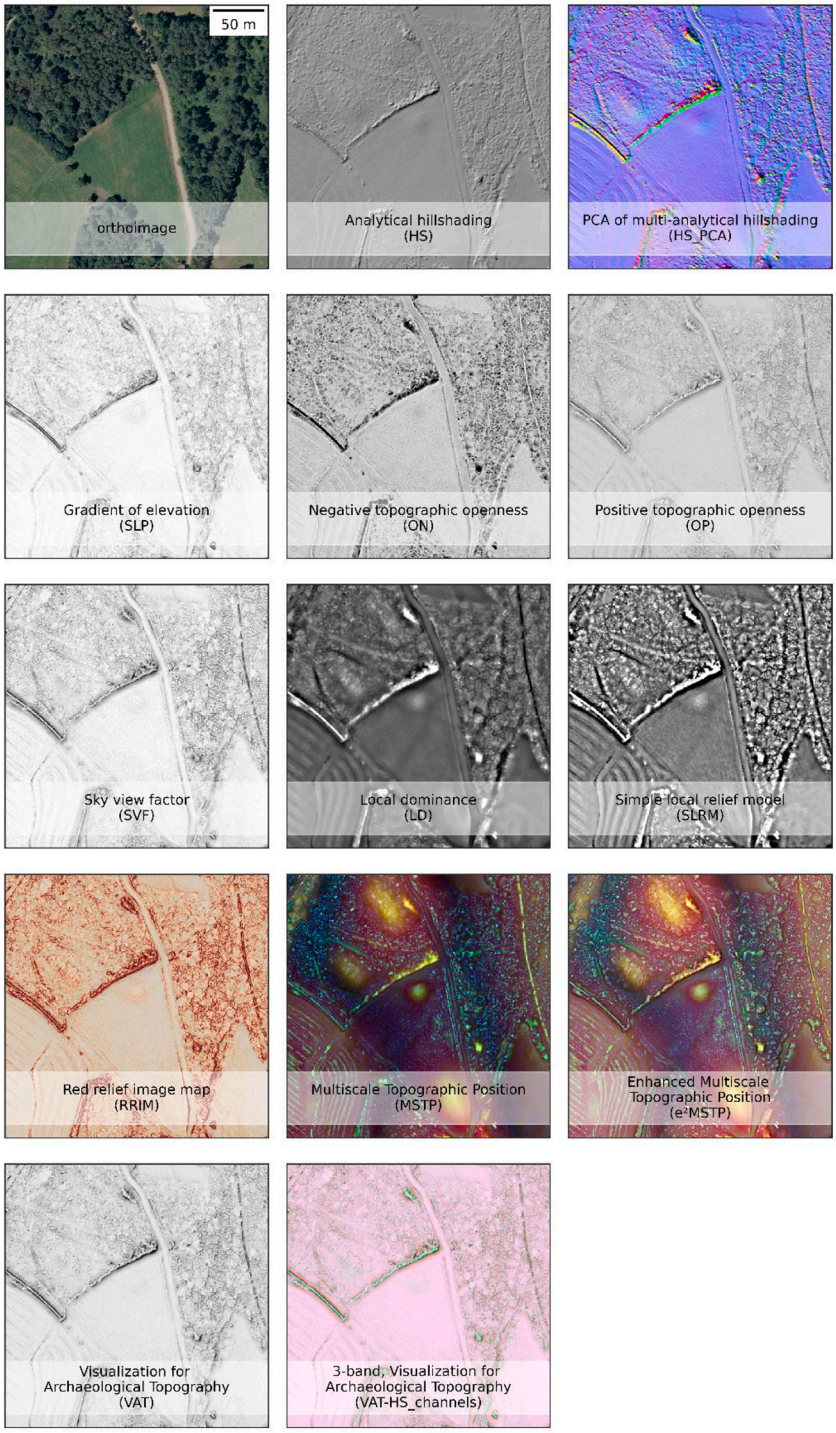

Fig. 2. Visualization techniques applied to the Le Pusso area, Carnac (France). The area has a variety of landforms, including multiple archaeological structures.

nature and spatial uncertainty associated with the archaeological reference dataset (Guyot et al., 2021). The metric, called AP@IoU0.5, refers to the area under the precision-recall curve.

Cross-validation was performed using a K-fold (K = 5) strategy with a 80%/20% train/test split (120/30 images, respectively) to assess the performance stability of each VT (Rodriguez et al., 2010). Since model hyper-parameters were not tuned (only model parameters were fine-tuned), no model selection was performed. Therefore, a split between validation and test sets was not required.

Each VT (Fig. 4) was visually interpreted by a human to empirically assess their performance for visual perception. The same VT configurations (visualization techniques and visualization parameters) were used as input for the deep CNN and as a basis for visual interpretation.

3. Results

3.1. VT assessment through human-based interpretation

Visually, multiscale VTs (MSTP, e^2MSTP) had the best perception performances for archaeological sites with subtle positive topographic variations, regardless of their size or morphology. This result was

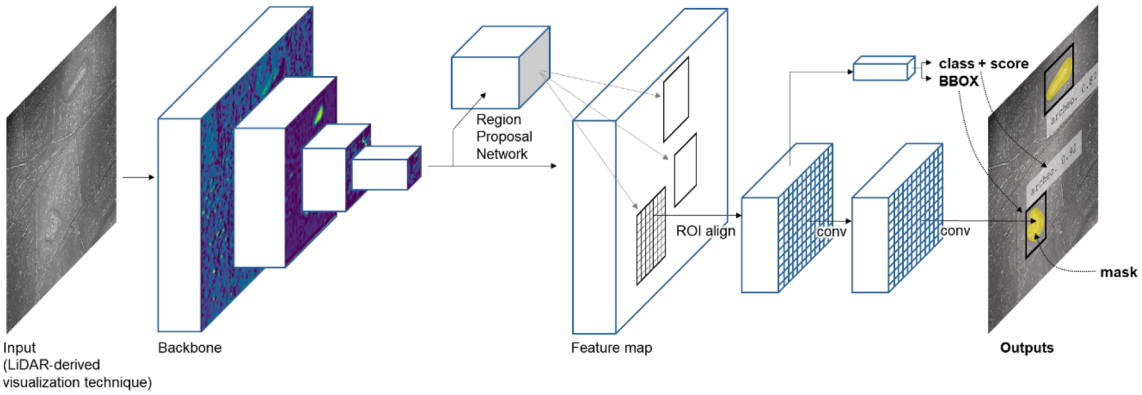

Fig. 3. Architecture of Mask R-CNN for instance segmentation on visualization technique images.

Table 2
Main hyperparameters used to train the deep convolution neural network.

Hyperparameters	Value
Learning rate (LR)	BaseLR = 0.002; 100 warmup iterations then 0.1xBaseLR, 0.01xBaseLR, 0.001xBaseLR at 500, 1000 and 1500 iterations respectively
Total iterations	2000
Batch size	2
Epochs*	33 (for 120 images)
Anchors size	16, 32, 64, 128, 256, 512 px
Data augmentation	flip, crop and rotation transformations (50% probability)

*Epochs = total iterations * batch size/total number of images.

especially evident for a levelled tumulus (example 1, Fig. 4) whose remaining trace spreads over an area 80 m long and 50 m wide. The multiscale VTs were also, by design, the only ones that provided information about the wider topographical context (example 2, Fig. 4), by highlighting the dominant position of the Neolithic funeral structures in the landscape. The combination of multiscale representation and local morphological information (e^2MSTP) allowed for better interpretation of structured terrain; for example, terrain with pits, narrow ditches or embankments (example 2, Fig. 4). Detrending techniques (LD and SLRM) were highly informative for small scale-terrain variations; however, their limits were apparent for the following geoarchaeological configurations: (1) small archaeological structures (smaller than the radius of analysis) within highly textured terrain (usually in undergrowth vegetation) and, (2) large archaeological structures (larger than the radius of analysis) considered to be the natural trend of the terrain and thus not enhanced. Terrain openness and its variants (OP, ON, SVF, RRIM) appeared to be effective for small archaeological structures, especially for small mounds (<10 m radius) with a central pit; however, the lack of overall contextual information (such as multiscale topographic position) reduced the ability to interpret these structures. However, these VTs were not visually informative for continuous and subtle variations that occur at a larger scale with the presence of leveled tumulus. This limitation was also observed for standard VTs, such as slope (SLP), hillshadings (HS and, to a lesser extent, HS_PCA) as well also for the combined visualization of archaeological topography (VAT and its 3-band variation VAT-HS_channels). VAT was designed to improve the recognition of small topographical features (Kokalj and Somrak 2019). While it was informative for the visual description of small structures or local morphological characteristics, the results we obtained showed that it was not adapted for the visualization of larger subtle topographic variations or for specific topographic positions.

3.2. VT assessment through computer-based analysis

The AP@IoU0.5 performance results by VT (Fig. 4) showed a minimum average of 24% (analytical hillshading to a maximum average of 65% (e^2MSTP), with per-fold individual performance extremes ranging from 19% (analytical hillshading, fold 4) to 76% (e^2MSTP, fold 5). Monochromatic VTs, such as HS (24%), OP (26%), SVF (28%) and ON (33%), had lower performances than other VTs and were thus considered less informative by the deep CNN model. VAT (the combined monochromatic VT) showed an average performance of 28%, which was higher than the performance of its components, except for SLP (38%). The VAT-HS_channels (a colored VT) showed a better average performance of 39%, confirming the value of multi-band information. However, monochromatic VTs based on terrain detrending such as SLRM (42%) and LD (48%) had higher performances than VAT_HS_channels and other colored VTs such as HS_PCA (34%), based on virtual illumination, or RRIM (41%), based on morphological representation. However, monochromatic VTs based on terrain detrending such as SLRM (42%) and LD (48%) had higher performances than VAT_HS_channels and other colored VTs such as HS_PCA (34%), based on virtual illumination, or RRIM (41%), based on morphological representation. Only the multiscale approaches, MSTP (58%) and e^2MSTP (65%), obtained performances that exceeded 50%. The e^2MSTP, based on a combination of multiscale information with morphological (RRIM) and local detrending (SLRM) representations, was an improvement over the standard MSTP version (+7%).

The statistical performance of deep CNN obtained using different VTs enabled the VTs to be ranked by the mean AP@IoU.5 value (Fig. 4). This metric-based ranking was similar to the subjective human-based assessment presented in 3.1, thus confirming the initial assumption that visual representation of data, effective from the perspective of human vision, is also effective from the perspective of deep CNNs.

4. Discussion

4.1. Towards an objective creation of suitable VTs

The best-performing VT was the e^2MSTP generated using multiscale topographic analysis (MSTP) combined (via a blending technique) with a morphologically explicit image (RRIM) and a slope-invariant relief-detrending technique (SLRM). This VT was created based on empirical knowledge and iterative selection of parameters that enhance the perception of ALS-derived terrain data for archaeological prospection. The computer-vision approach shows the suitability of such VTs for identifying archaeological structures on ALS-derived terrain models in the study area. In particular, it highlights the utility of using a multiscale approach that provides contextual topographic position information and is more robust for varying structure size (Guyot et al., 2018). It also confirms the advantage of combining complementary VTs to address identified limitations of single VTs (e.g. blending morphological and detrended information with multiscale information). Nevertheless, evaluating the complementarity of VTs and selecting the optimal

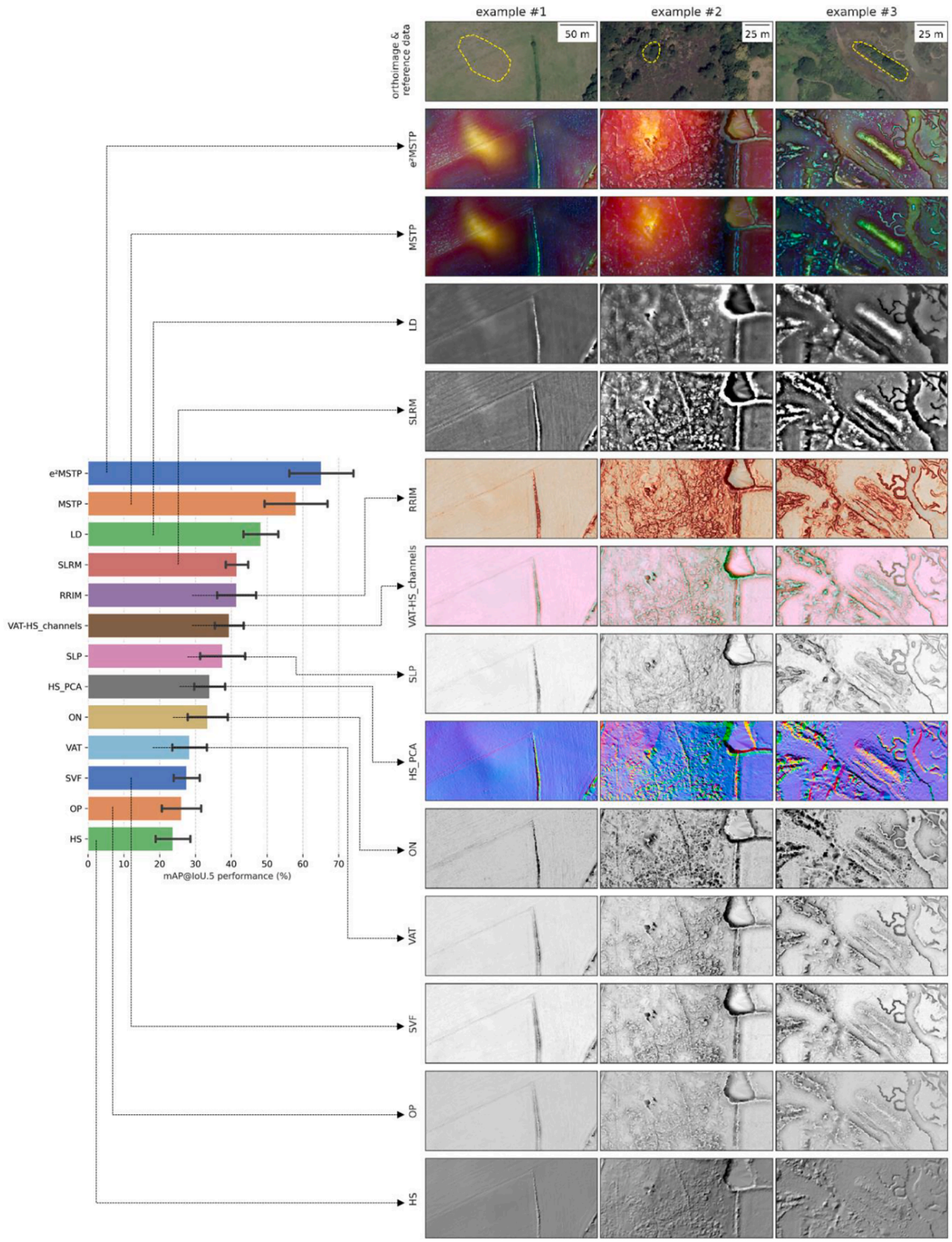

Fig. 4. Performances of detection/segmentation using deep convolution neural network (CNN) for different visualization techniques (VTs). Visual examples of VTs with (right) the reference data and (left) mean deep CNN model performances (mAP@IoU.5) of each VT. Error bars indicate 1 standard deviation. Example 1 is a leveled Neolithic tumulus in an agricultural field. Example 2 is a megalithic complex of 3 dolmens under dense vegetation. Example 3 is a Neolithic tumulus of elongated shape in a marshland area.

blending strategies to emphasis this complementarity remains a challenge. This was illustrated by the results we obtained using VAT and VAT-HS_channels. As expected, VAT showed lower performance than its 3-band variant VAT-HS_channels that is in accordance with the results of Somrak et al. (2020) who compared these two VTs for image classification task. However, VAT-HS_channels, which was produced without blending but by simple stacking of SLP, OP and SVF, showed lower performance than SLP used as a single VT. This could be due to the visual correlation between SLP, OP and SVF. A simple stacking of correlated VTs does not necessarily generate a better performing combined image.

While evaluating all possible VT combinations was out of the scope of this study, the proposed approach could open new perspectives in the

objective selection of the most suitable VT or blending parameters as the remote-sensing archaeological or geomorphological community develops new ALS-derived terrain model visualizations.

4.2. Generalization of the benchmarking approach

The deep CNN was trained on a limited typology of archaeological remains (mostly funeral structures from the Neolithic) and within a limited geographical area. A similar approach applied to different contexts would not necessarily provide the same performances from a deep CNN perspective. It would require new training and evaluation, which may not result in the same ranking of VTs, especially if structures or landforms differ from those in our study area. However, the approach is expected to maintain the relation between the degree of visual perception from VTs and the ability to perform segmentation automatically using Deep CNN.

This deep-CNN-based benchmarking approach has not yet been developed for diverse geographical and archaeological environments. While the available coverage of ALS data is rapidly growing (due to a decrease in the cost of acquisition and an increase in the number of open-access ALS projects supported by public funding), the availability of archaeological reference datasets remains a key issue. The collection, publication and maintenance of labelled archaeological data is not straightforward. Available inventories (e.g. the "Carte Archéologique Nationale" in France) are gradually addressing this issue, but the limits of large-scale archaeological references persist: many sites remain to be discovered, and for many of them, the nature and spatial extent of the archaeological structures could, paradoxically, be confirmed only by destructive archaeological excavation. The "ground truth", which serves as an essential base for all supervised remote-sensing classification or detection approaches, would remain wishful thinking in the archaeological prospection domain. Thus, New paradigms based on fuzzy or partial reference datasets need to be developed.

4.3. Computer vision as a support for human interpretation

This study is based on the initial assumption that the deep CNN-based and human-based processes involved in image interpretation share some similarities (Brachmann et al., 2017; Geirhos et al., 2018; Kim et al., 2019; Zhang et al., 2018). The results confirmed this assumption by showing comparable VT rankings between the computer-based and human-based interpretation. In both cases, the detection performance is related to the ability of a VT to enrich the original data representation with interpretable information. However, the image data is not the only information that influence human-based interpretation.

An expert-based interpretation would include perceptions of the information included in the image, but also external information not available in the data itself (geoarchaeological knowledge of the area or skills based on experience). The computer-vison approach applied to an image cannot encompass the exhaustive aspects that influence archaeological interpretation. Therefore, the proposed approach does not aim at replacing expert-knowledge or imposing a single VT for archaeological interpretation of ALS-derived terrain model, but rather aim at proposing a significant support tool for archaeological analysis. With a high capacity of data processing, a consistent response against similar data and an interpretation bias (even if not entirely absent) that can be measured, the deep CNN approach provides a complementary tool for the identification and characterization of archaeological structures from ALS-derived relief model.

5. Conclusion

In this study, we demonstrated the potential of the deep CNN approach as a tool to objectively assess the utility of ALS-derived VTs in the context of archaeological prospection. We used a state-of-the-art open-source instance-segmentation framework to compare the performances of automatic detection and segmentation of deep CNN models with 11 different VTs used as input data. The results allowed to rank VTs by their performance from an automatic detection and segmentation point of view. This computer-based ranking was compared to a subjective human-based interpretation. Ranking outcomes were comparable and thus confirmed the assumption that the deep CNN perception was similar to the subjective perception of human-based visual interpretation. Based on this confirmation, we showed that deep CNN computer vision approach could be used to objectively assess the ability of VTs ability to enhance the perception of archaeological structures. Although the study was conducted in a limited geo-archaeological context, the approach is expected to be reproducible on different areas and different types of structures or landforms, especially because it is based on relative evaluation of the selected VTs. Moreover, by relying on a non-subjective benchmarking method, the approach developed could help design new or hybrid VTs that could be used to improve the human-based interpretation, or as inputs to the CNN for further automatic extraction tasks on large datasets.

Funding information

This research was co-funded by the Région Bretagne (project "Patrimoine – Mégalithes de Bretagne") and DRAC Bretagne, Service régional de l'archéologie and Hytech-Imaging.

Declaration of Competing Interest

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

Acknowledgments

The authors thank members of DRAC Bretagne, Service régional de l'archéologie for providing access to the raw airborne LiDAR data, archaeological documentation and the reference dataset. OSUR/OSUNA and GeoFIT are also acknowledged for the airborne LiDAR acquisition. The Nantes Rennes LiDAR platform was funded by the Region Bretagne and the Region Pays de la Loire with European Regional Development Fund (ERDF). The authors also thank the two anonymous reviewers for the constructive comments and suggestions, which helped to substantially improve this manuscript.

References

Abdulla, Waleed. 2017. "Matterport/Mask_RCNN: Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow." 2017. https://github.com/matterport/Mask_RCNN.

Bennett, Rebecca, Kate Welham, RossHill, a., Andrew, Ford., 2012. A Comparison of Visualization Techniques for Models Created from Airborne Laser Scanned Data: A Comparison of Visualization Techniques for ALS Data. Archaeol. Prospect. 191, 41–48. https://doi.org/10.1002/arp.1414.

Bonhage, Alexander, 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. Archaeol. Prospect. https://doi.org/10.1002/arp.1806.

Brachmann, Anselm, Barth, Erhardt, Redies, Christoph, 2017. Using CNN features to better understand what makes visual artworks special. Front. Psychol. 8 https://doi.org/10.3389/fpsyg.2017.00830.

Chiba, Tatsuro, Shin-ichi Kaneta, and Yusuke Suzuki. 2008. "Red Relief Image Map: New Visualization Method for Three Dimensional Data." In Proceedings of the XXIth ISPRS Congress, edited by Jun CHEN, Jie JIANG, and Wolfgang KAINZ, 37:1071–76. Beijing: ISPRS Archives. http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/11_ThS-6/08.pdf.

Devereux, B.J., Amable, G.S., Crow, P., 2008. Visualisation of LiDAR terrain models for archaeological feature detection. Antiquity 82 (316), 470–479. https://doi.org/10.1017/S0003598X00096952.

Doneus, Michael, 2013. Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. Remote Sensing 5 (12), 6427–6442. https://doi.org/10.3390/rs5126427.

Doneus, Michael, Briese, Christian, 2006. Full-Waveform airborne laser scanning as a tool for archaeological reconnaissance. BAR Int. Series 1568, 99.

Gallwey, Eyre, Tonkins, and Coggan. 2019. "Bringing Lunar LiDAR Back Down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning." Remote Sensing 11 (17): 1994. 10.3390/rs11171994.

Geirhos, Robert, David H. J. Janssen, Heiko H. Schütt, Jonas Rauber, Matthias Bethge, and Felix A. Wichmann. 2018. "Comparing Deep Neural Networks against Humans: Object Recognition When the Signal Gets Weaker." ArXiv:1706.06969 [Cs, q-Bio, Stat], December. http://arxiv.org/abs/1706.06969.

Girshick, Ross. 2015. "Fast R-Cnn." In Proceedings of the IEEE International Conference on Computer Vision, 1440–48. Santiago, Chile: IEEE. 10.1109/ICCV.2015.169.

Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron, 2016. Deep Learning. MIT Press.

Grammer, Benedikt, Draganits, Erich, Gretscher, Martin, Muss, Ulrike, 2017. LiDAR-guided archaeological survey of a mediterranean landscape: lessons from the ancient Greek Polis of Kolophon (Ionia, Western Anatolia). Archaeol. Prospect. 24 (4), 311–333. https://doi.org/10.1002/arp.1572.

Guyot, Alexandre, Hubert-Moy, Laurence, Lorho, Thierry, 2018. Detecting neolithic burial mounds from LiDAR-derived elevation data using a multi-scale approach and machine learning techniques. Remote Sensing 10 (2), 225. https://doi.org/10.3390/rs10020225.

Guyot, Alexandre, Lennon, Marc, Lorho, Thierry, Hubert-Moy, Laurence, 2021. Combined detection and segmentation of archeological structures from LiDAR data using a deep learning approach. J. Comput. Appl. Archaeol. 4 (1), 1–19. https://doi.org/10.5334/jcaa.64.

Hesse, Ralf, 2010. LiDAR-derived local relief models - a new tool for archaeological prospection. Archaeol. Prospect. 17 (2), 67–72. https://doi.org/10.1002/arp.374.

Hesse, Ralf. 2016. "Visualisierung Hochauflösender Digitaler Geländemodelle Mit LiVT." Pdf. 10.17171/3-34-7.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160 (1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837.

Isenburg, Martin. 2020. Efficient LiDAR Processing Software (version 170322). Rapidlasso. http://lastools.org.

Kazimi, Bashir, Frank Thiemann, Katharina Malek, Monika Sester, and Kourosh Khoshelham. 2018. "Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data." In Proceedings of the 2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage, 15. Melbourne, Australia: CEUR Workshop Proceedings. 10.4230/LIPIcs.COARCH.2018.

Kim, J., Nguyen, A., Lee, S., 2019. Deep CNN-based blind image quality predictor. IEEE Trans. Neural Networks Learn. Syst. 30 (1), 11–24. https://doi.org/10.1109/TNNLS.2018.2829819.

Kokalj, Žiga. 2020. Relief Visualization Toolbox (RVT). https://iaps.zrc-sazu.si/en/rvt#v.

Kokalj, Žiga, Hesse, Ralf, 2017. Airborne Laser Scanning Raster Data Visualization a Guide to Good Practice. Založba ZRC, Ljubljana http://zalozba.zrc-sazu.si/p/P14.

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

Kokalj, Žiga, Zakšek, Klemen, Oštir, Krištof, 2011. Application of Sky-View Factor for the Visualisation of Historic Landscape Features in Lidar-Derived Relief Models. Antiquity 85 (327), 263–273. https://doi.org/10.1017/S0003598X00067594.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In Advances in Neural Information Processing Systems 25, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2015. "Microsoft COCO: Common Objects in Context." ArXiv:1405.0312 [Cs], February. http://arxiv.org/abs/1405.0312.

Lindsay, J.B., Cockburn, J.M.H., Russell, H.A.J., 2015. An integral image approach to performing multi-scale topographic position analysis. Geomorphology 245 (September), 51–61. https://doi.org/10.1016/j.geomorph.2015.05.025.

Lindsay, John. 2020. WhiteboxTools (version 1.4.0). https://jblindsay.github.io/ghrg/WhiteboxTools/index.html.

Mayoral, Alfredo, Toumazet, Jean-Pierre, Simon, François-Xavier, Vautier, Franck, Peiry, Jean-Luc, 2017. The highest gradient model: A new method for analytical assessment of the efficiency of LiDAR-derived visualization techniques for landform detection and mapping. Remote Sens. 9 (2), 120. https://doi.org/10.3390/rs9020120.

Orengo, Hector A., Petrie, Cameron A., 2018. Multi-scale relief model (MSRM): A new algorithm for the visualization of subtle topographic change of variable size in digital elevation models: MSRM: An algorithm for the multi-scale analysis of topographic change. Earth Surf. Proc. Land. 43 (6), 1361–1369. https://doi.org/10.1002/esp.4317.

Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242, 10.1109/IWSSIP48289.2020.9145130.

Risbøl, Ole. 2013. "Cultivating the 'Wilderness'-How Lidar Can Improve Archaeological Landscape Understanding." Interpreting Archaeological Topography: 3D Data, Visualisation and Observation; Opitz, RS, Cowley, DC, Eds, 51–62.

Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity analysis of K-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32 (3), 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Roscheck, Florian. 2020. Blend-Modes: Image Processing Blend Modes (version 2.1.0). Python. https://github.com/flrs/blend_modes.

Somrak, Maja, Džeroski, Sašo, Kokalj, Žiga, 2020. Learning to classify structures in ALS-derived visualizations of ancient maya settlements with CNN. Remote Sens. 12 (14), 2215. https://doi.org/10.3390/rs12142215.

Štular, Benjamin, Kokalj, Žiga, Oštir, Krištof, Nuninger, Laure, 2012. Visualization of lidar-derived relief models for detection of archaeological features. J. Archaeol. Sci. 39 (11), 3354–3360. https://doi.org/10.1016/j.jas.2012.05.029.

Trier, Øivind Due, Cowley, David C., Waldeland, Anders Ueland, 2018. Using deep neural networks on airborne laser scanning data: results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. Archaeol. Prospect. 26 (2), 165–175. https://doi.org/10.1002/arp.1731.

Trier, Øivind Due, Reksten, Jarle Hamar, Løseth, Kristian, 2021. Automated mapping of cultural heritage in Norway from airborne Lidar data using faster R-CNN. Int. J. Appl. Earth Obs. Geoinf. 95 (March), 102241 https://doi.org/10.1016/j.jag.2020.102241.

Trier, Øivind Due, A.-B. Salberg, and L. Holger Pilø. 2016. "Semi-Automatic Mapping of Charcoal Kilns from Airborne Laser Scanning Data Using Deep Learning." In CAA2016 : Oceans of Data, 221–32. Oslo, Norway: Archeopress.

Wu, Yuxin, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. "Detectron2." 2019. https://github.com/facebookresearch/detectron2.

Verschoof-van der Vaart, Wouter B., Lambers, Karsten, 2019. Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR Data from the Netherlands. J. Comput Appl. Archaeol. 2 (1), 31–40. https://doi.org/10.5334/jcaa.32.

Verschoof-van der Vaart, Wouter B., Lambers, Karsten, Kowalczyk, Wojtek, Bourgeois, Quentin P.J., 2020. Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from The Netherlands. ISPRS Int. J. Geo-Inf. 9 (5), 293. https://doi.org/10.3390/ijgi9050293.

Yoëli, P., 1967. The mechanisation of analytical hill shading. Cartograph. J. 4 (2), 82–88. https://doi.org/10.1179/caj.1967.4.2.82.

Zakšek, Klemen, Oštir, Kristof, Kokalj, Žiga, 2011. Sky-view factor as a relief visualization technique. Remote Sens. 3 (12), 398–415. https://doi.org/10.3390/rs3020398.

Zhang, Richard, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." ArXiv:1801.03924 [Cs], April. http://arxiv.org/abs/1801.03924.