Mag-Net

Improving magnetometer interpretation workflows with semantic segmentation

Julien WOLF, Magnitude Surveys, United Kingdom
Finnegan POPE-CARTER, Magnitude Surveys, United Kingdom
Paul S. JOHNSON, Magnitude Surveys, United Kingdom

Abstract: This paper presents a novel application of convolutional neural networks to the detection of archaeological ring ditches in gradiometer data in the UK. The Unet and ResUnet models applied in this study achieved an overall accuracy of 0.7 mIoU on a test dataset with minimal cost spent in pre-processing the data, demonstrating that machine learning on gradiometer data is commercially viable and can pave the way for a reduction in costs and time spent manually digitising features.

Keywords: *Machine Learning—Semantic Segmentation—Automatic Interpretation—Magnetoetry—Commercial Geophysics*

CHNT Reference: Wolf, J., Pope-Carter, F., and Johnson, P. S. (2022). 'Mag-Net. Improving magnetometer interpretation workflows with semantic segmentation', in Börner, W., Rohland, H., Kral-Börner, C. and Karner, L. (eds.) *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, held online, November 2020*. Heidelberg: Propylaeum.
doi:10.11588/propylaeum.1045.c14483

Introduction

Practitioners of commercial geophysics in Britain have met increasing demands for large-scale surveys with the development of multi-sensor, cart-based, motorised, and modular survey systems to improve efficiency and speed of data-collection. By rising to these physical challenges, however, the concomitant increase in the quantity of data collected poses new logistical challenges in keeping reporting turnover within an acceptable timescale without undermining the quality of interpretations produced. While there has been some success in addressing these challenges by restructuring internal reporting workflows and by improving collaborative multi-interpreter environments (Harris and Pope-Carter, 2019), large datasets encompassing multiple square-kilometres of readings remain challenging for entirely manual interpretation. The response of Magnitude Surveys to this challenge has been, in part, to explore the use of convolutional neural networks (CNN) for the detection of anomalies in geophysical data, producing outputs then intended for secondary archaeological classification performed by human specialists.

The aim of this study is to assess the possibility of reliably detecting archaeological features in gradiometer data using an approach based on semantic segmentation, and the feasibility of this approach for commercial use. This research project is ongoing, but this paper presents some promising early results. Some of the broad questions that this project aims to address are:

How the quantity and quality of the data affects the development of CNN, and where is this affected specifically by the nature of geophysical data?

What types of network architectures and hyper parameter choices are particularly suited to detecting anomalies in magnetometer data?

How the development of a natively archaeological CNN, trained exclusively on archaeological geophysical data, may improve the algorithm's robustness in detecting anomalies of potentially archaeological nature in comparison with a transfer-learned network?

Background

Manual classification of geophysical data is a labour-intensive task that requires specialist training. Research into improving methods of classifying geophysical data have focused on themes as diverse as the statistical characterisation of noise in order to assist with interpretation of anomalies (e.g. Schmidt et al., 2020), to automating classification of anomalies through the use of object-orientated approaches based on multi-scalar data-segmentation of observed values within magnetic datasets (e.g. Pregesbauer et al., 2014). One of the issues with using only observed geophysical values to classify anomalies is that it requires those of interest to be clearly and quantitatively differentiated from the background and any other interference or noise. While this may be possible on certain geologies, and with certain feature-types, these "ideal" circumstances cannot be guaranteed, particularly where survey is being undertaken as part of developer-led site investigations.

The use of CNN addresses some of the issues which remain with the use of other approaches and algorithms, particularly in dealing with large variety of geologies and soil conditions that would otherwise require manual site-by-site pre-processing of the data to improve feature contrast. CNN have seen a considerable amount of progress over the last two decades and there have been some promising applications in archaeological geophysics: Green and Cheetham (2019) applied the ResNet152 and Google's Inception architectures to successfully detect the presence of Irish medieval graves in GPR B-scan data. In a similar vein, Verdonck (2019) leveraged Faster R-CNN to detect the location of hyperbolas within GPR profiles. Lastly, Küçükdemirci and Sarris (2019) adapted Ronneberger et al.'s (2015) Unet architecture to detect anomalies in GPR time slices. In the wider field of remote sensing there have been numerous applications of CNNs, most numerously so on LiDAR data, using a variety of architectures including, amongst others, ResNet (Trier et al., 2018), Faster R-CNN (Verschoof-van der Vaart and Lambers, 2019) and VGG-19 (Somrak et al., 2020). Each of the above address specific challenges and research questions and all report success metrics in the form of F1 scores and validation accuracy, amongst others.

One of the advantages of CNN is the availability of transfer learning: models can be pre-trained on large open datasets such as the ImageNet library, which reduce the quantity of archaeological data required by pre-training a network on larger non-archaeological datasets. This pre-training allows for the application of CNN to the smaller datasets, which are more common in archaeology, and speeds up the process as the network is already partly trained to some degree of accuracy before it sees the archaeological dataset. Pre-training is not without drawbacks however, as the nature of the data in libraries such as ImageNet can prove too divergent from archaeological data, both in terms of data-formatting as well as complexity and distinctness, which can lead to decreased model performance (Trier et al., 2018, p. 227; Verschoof-van der Vaart and Lambers, 2019, p. 38). The authors would therefore expect networks trained solely on archaeological data to be more robust in their predictions. Conversely, the process of training a CNN on the above-mentioned architectures can be time-consuming and resource-intensive, hence this project also aims to investigate whether significant improvements can be made in this manner with minimal input costs in time and effort. While this project has been driven by commercial imperatives, its potential outcomes and applications would be equally applicable within a purely academic context as time- and financial-constraints increase.

Semantic segmentation is a subfield of CNN applications focusing on the pixelwise classification of input images. Segmentation is of specific interest as it allows the identification of the shape of a target feature within an image (e.g. Küçükdemirci and Sarris, 2019), as opposed to classifying images (e.g. Somrak et al., 2020) or detecting the location of a predetermined feature within an image (e.g. Verdonck, 2019). The architectures required for this task involve some additional steps following the classic series of convolutions native to the CNN family, including some form of up-sampling to return the output to the same size as the network input. The network architectures investigated in this paper are based on variations of those introduced by Shelhamer et al. (2016) and Ronneberger et al. (2015) which both employ up-sampling to produce output labels with the same dimensions as the given inputs.

Methodology

Datasets

For this pilot study, a small corpus of geophysically detected ring ditches and larger circular anomalies was chosen as a dataset. These types of anomalies were selected for having distinct, recognisable features, while also being morphologically more complex than simple linear, or point, anomalies. The dataset was comprised of 150 such features located across the UK, all of which were discovered between 2016 and 2020 by Magnitude Surveys (MS) across a representative sample of commercial archaeological investigations. Collection of the data thus varied in terms of scale, survey conditions, and geological backgrounds, amongst other factors. Broad natural anomalies or anthropogenic activities in the background, such as modern manuring practices, can further complicate the data and reduce the likelihood of features of interest being detected by machine algorithms. Data processing has been consistent across all sites and follows MS' standard processing workflow (Pope-Carter et al., 2017).

Only ring ditches that could confidently be identified as such were included in this dataset, accounting for the relatively small number compared to larger image datasets such as the above-mentioned ImageNet.

A centre point was manually located and applied to each anomaly. The data, stored in MS' in-house archive, was located using these points, which comprised gradiometer data minimally processed in accordance with standards established by Historic England for 'raw or minimally processed data' (David et al., 2008, p. 11), and in line with EAC guidelines (Schmidt et al., 2015, p. 16). Unfortunately, the authors were not able to use true values in this study, as the Keras library used to create the models had difficulties processing the ASCII grids produced by MS. Instead, geotiffed greyscales clipped to a range of -1 to 2 nT were used as data inputs. The labels for each feature comprised of the original interpretations generated manually by MS staff during the reporting process, and stored as polygons in a PostGIS database.

None of the original interpretations were modified, nor were the gradients submitted to any further processing. This approach was taken in order to assess whether this methodology would be feasible with as minimal modification of the input datasets as possible. Although the magnetic data might have benefited from additional processing to enhance contrast and emphasise the target features, this was deliberately omitted due to the high variability within the dataset, and to maintain a consistent baseline against which to assess the efficacy of the methodology. Variability in the dataset was caused by the many factors which can affect magnetic data, such as geology and soils, agriculture, and magnetic interference from modern features. Any pre-processing would therefore also have been necessarily undertaken on a site-by-site basis, a time intensive process, which would also have worked against the premise of this feasibility study.

During the course of this study, the authors created two datasets based on the same input data. In a first run at the start of 2020, the tests were conducted on the polygon source labels including both the ring ditches and any surrounding archaeology, whether the latter could be associated with the ditch or not. This approach presented the least expenditure of effort in the preparation of the data. Following initial results with unstable training and low accuracy predictions, however, a second set of experiments were run on the same dataset, but this time only the polygons describing the actual ring ditches were incorporated. In both cases, the original magnetic gradients were used without any modification.

Pre-Processing

The label polygons were rasterised into 244 × 244 pixel PNGs centred around the manually generated feature centre-points. These labels were classified into three categories – ‘no feature’ where there were no anomalies present, ‘feature’ where there was an anomaly present, and ‘out-of-bounds’ where the gradient included an area of nodata values. The latter was included as some early test runs indicated a tendency to categorise nodata areas as archaeological features. Similarly, the gradients were clipped to the same size – for most gradients included in this survey, this was equivalent to a radius of ca. 25 m at 0.125 m/px, although for a few older datasets collected at 0.25m/px this equated to double the size. The resulting data and label pairs were then randomly sampled into training, validation, and prediction sets at a ratio of 60 %, 30 % and 10 % respectively, or 90 training, 45 validation, and 15 testing image pairs.

Machine Learning

The authors decided on two target architectures for this pilot study: Unet by Ronnenberger et al. (2015) named after their characteristic symmetrical convolutional and up-sampling paths, which was designed specifically for semantic segmentation tasks. This network was discussed by the authors as being particularly advantageous for small datasets (Ronneberger et al., 2015, p. 235), which seemed perfect for an archaeological application and previously has been successfully applied to GPR time-slices in a very similar problem set by Küçükdemirci and Sarris (2019). The second network, ResUNet was developed by Diakogiannis et al. (2020) and expands on the Unet architecture by adding residual skips at each layer to counteract the issue of vanishing gradients which can occur with particularly deep networks. The authors hoped this would allow for developing deeper and more complex networks.

As neither of these networks were published in Python, the authors’ programming language of choice, both these networks were recreated by the authors using Python’s Keras library based on details given in their respective publications. A series of experiments were conducted through parameter exploration in order to tune the hyperparameters for the neural network. The parameters in question are described in Table 1 below.

Table 1. List of parameters explored in this study and their descriptions.

Parameter	Description
Batch Size	Number of image + label pairs processed in a batch, before weights are adjusted within an epoch
Number of Filters	Number of filters added per layer of the network; filters increased by a factor of two per layer, e.g. if the first layer had 8, the second would have 16, then 32 filters and so on
Network Depth	Number of layers within the network
Learning Rate	Rate at which parameters are tuned in order to minimise loss
Loss Function	Function to calculate the loss, which is minimised during training to improve network performance

Training was performed on two NVIDIA RTX 2060S with 8 GB of memory each. Each training iteration lasted 100 epochs, unless no convergence or over-fitting was detected. During each processing run, input parameters as well as the state of machine learning software, represented by git hashes, were stored alongside the outputs in order to enable replication of the results at a later date should this prove necessary.

Metrics

Conventionally, accuracy in Keras is given as a percentage of pixels correctly predicted as belonging to a class, where 1.0 indicates a perfect match. The authors found that the models had a tendency to over-predict pixels as background due an imbalance between background and archaeology in the data, which a simple percentage metric cannot convey. Instead, the authors prefer to use mean Intersection over Union (mIoU), a metric commonly used in semantic segmentation (cf. Ronneberger et al., 2015, Shelhamer et al., 2016), which ranges from 0, no match, to 1.0, perfect match. This metric calculates the number of correctly predicted pixels, the intersection between prediction (P) and ground truth (T), divided by the difference between the sum of all pixels in the prediction and ground truth and the intersection. For multiple classes, the metric is calculated for each category and then averaged.

$$IoU = \frac{|T| \cap |P|}{|T| + |P| - |T| \cap |P|}$$

Results

The initial experiments using the first dataset based on a simple rasterised clip of MS' interpretation achieved only moderate results of 0.494 mIoU. This prompted the authors to refine the dataset as described above, and to increase the number of inputs via augmentation. This second round of experiments achieved much better results (Table 2) with the highest recorded accuracy rising to 0.703, achieved by a ResUnet model, while the Unet architecture achieved an mIoU of 0.676. Both these results were achieved under the same parameters detailed in Table 3.

Table 2. Results in mIoU achieved by both Unet and ResUnet architectures after training for 100 epochs on either dataset under the same parameters (see Table 3)

Architecture	Dataset 1 (150 Ring Ditches)	Dataset 2 (150 Ring Ditches + 300 Augmentations)
Unet	0.494	0.676
ResUnet	0.484	0.703

Table 3. Hyperparameters used to achieve the results presented in Table 2.

Parameter	Description
Batch Size	2
Number of Filters	16, 32, 64, 128, 254
Network Depth	5
Learning Rate	0.0001
Loss Function	Sparse categorical cross-entropy

One large drawback of the above experimental setup, which became apparent during training, was that the dimensions of the input data resulted in a significant number of parameters (13M+), effectively requiring more RAM than was available on the two graphics cards which had been allocated for training. This lack of capacity caused restrictions on the number of filters and the depth of any network, not to mention ruling out any more complex architecture. By setting a very low batch size, at 2 images per batch, – which in Keras controls how much data is loaded into RAM at any given time – the authors were able to push the remaining parameters, and those described in Table 3 performed the most consistently.

Another, albeit minor, issue encountered, was that only the pre-defined loss functions would produce good results. The authors experimented with variants of loss functions based on mIoU as well as weighted loss functions, however all either performed poorly or did not converge at all. The authors believe this however, to be an implementation issue which can be fixed in the future. That being said, sparse categorical cross entropy which is implemented in Keras seemed to perform well enough under the conditions of the test. Binary cross entropy unfortunately was not an option due to the multi-categorical nature of the labels.

From a visual perspective, the models seemed to perform particularly well in high contrast scenarios, such as in Figure 1, where both Unet (Fig. 1, bottom left) and ResUnet (Fig. 1, bottom right) very closely matched the target label (Fig. 1, top right). The authors noticed that in some earlier experiments the networks had a tendency to fixate on strongly positive features (intense black) over morphological indicators.

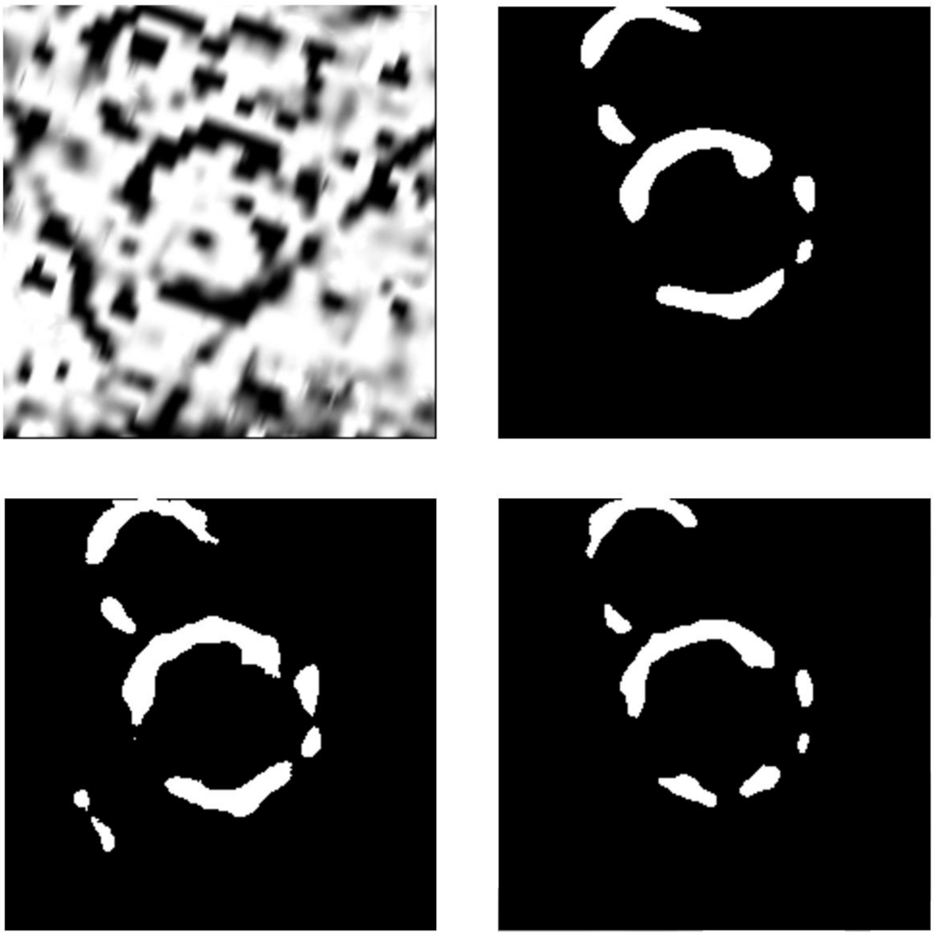

Fig. 1. A previously unseen gradient (top left) with positive readings in black and negative readings in white ranging from -1 to 2 nT. The predictions by the Unet (bottom left) and ResUnet (bottom right) models after 100 epochs of training (cf. Table 2 for parameters) both matched quite closely to the ground truth label (top right) despite the noisy background. (© Magnitude Surveys Ltd 2022)

Where the target feature was quite ephemeral in nature or obscured by surrounding anomalies, unsurprisingly both models struggled. Figure 2 highlights one such example; in this case, the Unet performed much better than the ResUnet, the latter of which picked up some of the positive linear anomalies cutting through this ring ditch.

Discussion

Overall, the authors believe that the results so far have been promising, and demonstrate the potential in the application of semantic segmentation within commercial geophysics. This was achieved with very low effort using existing datasets and no additional processing of the magnetometer data, as well as being based solely on magnetometer data without any prior pre-training on larger image datasets such as ImageNet. There are however a number of areas that will likely result in improvement in the overall performance of the model.

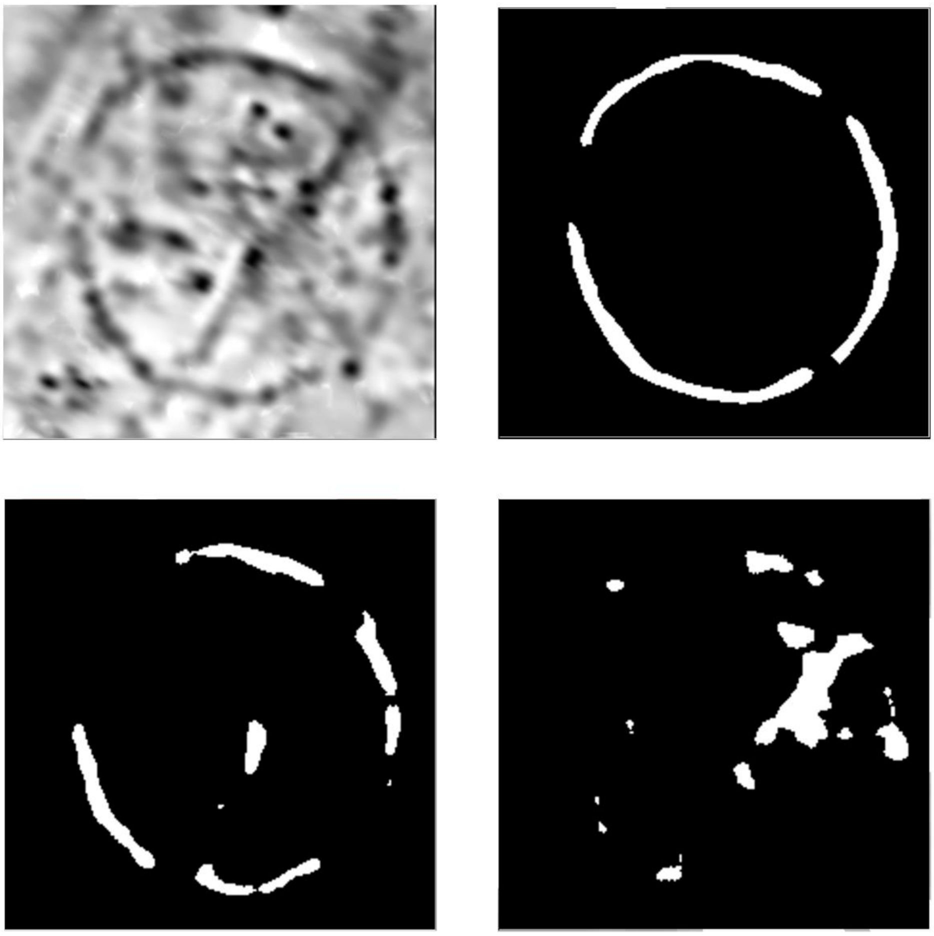

Fig. 2. A previously unseen gradient (top left) and ground truth labels (top right) with predictions by Unet (bottom left) and ResUnet (bottom right) models after 100 epochs of training (cf. Table 2 for parameters). Both models struggled on more ephemeral targets such as this one, with the Unet in this case performing much better at detecting the ring ditch. The ResUnet seems to have been confused by the agricultural feature cutting through the ring ditch.(© Magnitude Surveys Ltd 2022)

The performance boost achieved by moving from the first to the second dataset highlighted the importance of selecting good training data, as well as demonstrating the usefulness of data augmentation. Only zoom and pan augmentations were used in this case, however it might be possible to use rotation and flipping, although here the question presents itself whether these are relevant to magnetic data. Orientation is important for the human interpretation of some archaeological features as well as some magnetic features, although it remains to be seen whether this is the case for automatic identification.

While ring ditches are very distinct features found in archaeological geophysics, the obvious next step will be to test different and more numerous features. Linear agricultural features such as ridge and furrow, or historic field boundaries are much more prevalent across the UK and would provide a very large potential dataset with greater morphological diversity. These are particularly interesting targets for automatic identification, as their manual digitisation can be particularly labour intensive.

In addition to increasing the size of the geophysical training data, picking a larger dataset to pre-train the network and/or using transfer learning on a dataset more akin to geophysical data could also provide measurable improvements to model performance. Pre-training on ImageNet was avoided in this study given the concerns raised above, however there are other datasets more comparable to magnetometer data that could potentially be used for pre-training. Gallwey et al. (2010) explored lunar LiDAR data as a potential such pre-training dataset with promising initial results on the Arran dataset. It would be interesting to investigate whether a similar approach would lead to a performance gain here.

The results have also defined some clear limitations in the above experimental setup. The size of the inputs unfortunately limited the parameter choices available in this pilot study. Even with two graphics cards, the hard limit in terms of model complexity was set by the memory of a single graphics unit, at least for the Keras library used in this study. Reducing the size of the images may require some rescaling and thus either loss in area covered or a loss of detail. For the ring ditches in this study, a reduction in size was not possible, as some of the larger examples required a diameter of up to fifty metres. In this case, a loss of detail by scaling down the images to a lower resolution would be required. For linear features such as ridge and furrow, this is not an issue as the images would rarely be able cover the entire feature to begin with.

By reducing the size of the inputs and freeing up more memory, or as the available computing power increases, it will also be possible to increase the batch size. During training, the authors found that having a small batch size with a highly variable training dataset such as this one can lead to very erratic development. Given a bigger batch size, training might be slower, however the authors expect that convergence would be steadier. Moreover, to speed up training again, it would be possible to increase the learning rate.

Conclusion

This paper has shown that automatic identification of archaeological features, specifically ring ditches, found in magnetometer data is possible and commercially feasible with the input of very little time and effort. The experiments presented in this paper achieved a maximum mIoU of 0.703 using a ResUnet neural network. These results were achieved by training the neural network on magnetometer data and pre-existing labels from MS' archives, without using pre-trained weights.

In the discussion above the authors have highlighted several areas of improvements which may yield higher accuracy. Future work will focus on improving the performance of the model, as well as the implementation of these models into MS' magnetometer processing workflow pipeline. The authors' aim is to implement a robust system whereby data are automatically streamed onto in-house servers where basic data-processing and the extraction of features can be undertaken in a matter of minutes, providing specialists more time to spend on qualitative tasks such as interpretation and analysis.

One major issue in archaeology, however, remains the option to quantitatively compare the performance of different approaches against a common benchmark. To that extent, Kramer (2021) has recently published the Arran benchmark dataset based on data collected from the Isle of Arran, Scotland, during the Rapid Archaeology Mapping Programme (RAMP) (Banaszek et al., 2018). This first of its kind benchmark within archaeology may enable a fairer comparison of archaeological CNN models and further accelerate prototyping of new methodologies for archaeological purposes.

References

Banaszek, Ł., Cowley, D.C., and Middleton, M. (2018). 'Towards National Archaeological Mapping. Assessing Source Data and Methodology – A Case Study from Scotland.' Geosciences, 8, 272. doi:10.3390/geosciences8080272.

David, A., Linford, N., Linford, P., and Martin, L. (2008). 'Geophysical survey in archaeological field evaluation: research and professional services guidelines' (2nd edition). Historic England.

Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C. (2020). 'ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data.' ISPRS Journal of Photogrammetry and Remote Sensing, 162, pp. 94–114. doi:10.1016/j.isprsjprs.2020.01.013.

Gallwey, J., Matthew, E., Matthew, T., and Coggan, J. (2019). 'Bringing Lunar LiDAR Back Down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning'. Remote Sensing 11(17). doi:10.3390/rs11171994.

Green, A. and Cheetham P. (2019). 'Rise of the Machines: Improving the identification of possible graves in GPR data with interactive survey guidance and machine learning.', in Bonsall, J. (ed.) New Global Perspectives on Archaeological Prospection. 13th International Conference on Archaeological Prospection. Sligo: Archaeopress Publishing Ltd., pp. 300–304.

Harris, C. and Pope-Carter, F. (2019). 'Up-skilling and Up-scaling: the realities of adapting to the challenges of the current environment in British commercial geophysics.', in Bonsall, J. (ed.) New Global Perspectives on Archaeological Prospection. 13th International Conference on Archaeological Prospection. Sligo: Archaeopress Publishing Ltd., pp. 359–360.

Kramer, I. (2021). 'Ickramer/Arran: (v1.0)'. Zenodo. doi:10.5281/zenodo.5138109.

Küçükdemirci, M. and Sarris, A. (2019). 'Automated segmentation of archaeo-geophysical images by convolutional neural networks.', in Bonsall, J. (ed.) New Global Perspectives on Archaeological Prospection. 13th International Conference on Archaeological Prospection. Sligo: Archaeopress Publishing Ltd., pp. 296–299.

Pope-Carter, F., Harris, C., Attwood, G., and Eyre T. (2017). 'Live-streaming for the real-time monitoring of geophysical surveys.', in Jennings, B., Gaffney, C., Sparrow, T., and Gaffney, S. (eds.) AP2017. 12th Archaeological Conference of Archaeological Prospection. Bradford: Archaeopress Publishing Ltd., pp. 198–196.

Pregesbauer, M., Trinks, I. and Neubauer, W. (2014). 'An Object oriented approach to automatic classification of archaeological features in magnetic prospection data.' Near Surface Geophysics. 12, pp. 651–656.

Ronneberger, O., Fischer, P., and Brox T. (2015). 'U-Net: Convolutional Networks for Biomedical Image Segmentation.', in Navab, N., Hornegger, J., Wells, W., and Frangi, A. (eds.). Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Springer, Cham. pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Schmidt, A., Dabas, M., and Sarris, A. (2020). 'Dreaming of Perfect Data: Characterizing Noise in Archaeo-Geophysical Measurements.' Geosciences 10, p. 328. doi:10.3390/geosciences10100382.

Schmidt, A., Linford, P., Linford, N., David, A., Gaffney, C., Sarris, A., and Fassbinder, J. (2015). 'Guidelines for the use of geophysics in archaeology: questions to ask and points to consider.' EAC Guidelines 2. European Archaeological Council: Belgium.

Shelhamer, E., Long, J., and Darrell, T. (2017). 'Fully Convolutional Networks for Semantic Segmentation.' IEEE Transactions on Pattern Analysis and Machine Intelligence. 39, 4 (April 2017), pp. 640–651. doi:10.1109/TPAMI.2016.2572683.

Somrak, M., Džeroski, S., and Kikalj, Ž. (2020). 'Learning to Classify Structures in ALS-Derived Visualisations of Ancient Maya Settlements with CNN.' Remote Sensing. 12, p. 2215. doi:10.3390/rs12142215.

Trier, O.D., Salberg, A.-B., and Pilo, L.H. (2018). 'Semi-automatic mapping of charcoal kilns from airborne laser scanning data using deep learning.', in Matusmoto, M. and Uleberg, E. (eds.) Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology. Oceans of Data – Computer Applications and Quantitative Methods in Archaeology. Oslo: Archaeopress Publishing Ltd., pp. 219–231. Available at: https://www.researchgate.net/publication/326771506_Semiautomatic_mapping_of_charcoal_kilns_from_airborne_laser_scanning_data_using_deep_learning (Accessed: 25 July 2019).

Verdonck, L. (2019). 'Automated detection and analysis of diffraction hyperbolas in ground penetrating radar data.' In Bonsall, J. (ed.). New Global Perspectives on Archaeological Prospection. 13th International Conference on Archaeological Prospection. Sligo: Archaeopress Publishing Ltd., pp. 305–308.

Verschoof-van der Vaart, W.B. 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), pp. 31–40. doi:10.5334/jcaa.32.