

Monitoring Looting at Cultural Heritage Sites: Applying Deep Learning on Optical Unmanned Aerial Vehicles Data as a Solution

Social Science Computer Review
2024, Vol. 42(2) 480–495
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08944393231188471
journals.sagepub.com/home/ssc



Mark Altaweel¹, Adel Khelifi², and Mohammad Maher Shana'ah²

Abstract

The looting of cultural heritage sites has been a growing problem and threatens national economies, social identity, destroys research potential, and traumatizes communities. For many countries, the challenge in protecting heritage is that there are often too few resources, particularly paid site guards, while sites can also be in remote locations. Here, we develop a new approach that applies deep learning methods to detect the presence of looting at heritage sites using optical imagery from unmanned aerial vehicles (UAVs). We present results that demonstrate the accuracy, precision, and recall of our approach. Results show that optical UAV data can be an easy way for authorities to monitor heritage sites, demonstrating the utility of deep learning in aiding the protection of heritage sites by automating the detection of any new damage to sites. We discuss the impact and potential for deep learning to be used as a tool for the protection of heritage sites. How the approach could be improved with new data is also discussed. Additionally, the code and data used are provided as part of the outputs.

Keywords

heritage, archaeological sites, deep learning, security, looting, unmanned aerial vehicles, optical imagery

Introduction

The looting and destruction of heritage sites is a growing concern for authorities and communities worldwide (Al-Ansi et al., 2021; Basnet Silwal, 2021). The consequences of such damage can have far-reaching social impact, including damage to local economies that depend on heritage tourism, destruction of sites used for scientific and heritage research, attacks on ethnic or religious

¹Institute of Archaeology, University College London, UK

²Department of Computer Science and Information Technology, Abu Dhabi University, UAE

Corresponding Author:

Adel Khelifi, Abu Dhabi University, Abu Dhabi Campus, P.O Box 59911, Zayed City, Abu Dhabi, UAE.

Email: adel.khelifi@adu.ac.ae

identity affiliated with heritage places, and the erasure of the past that can be important for social cohesion and community building (Kersel and Hill 2020; Brusasco, 2012; Taniguchi 2017). Additionally, looted objects can be sold on the black market, resulting in their permanent loss and even funding of other illegal activities that can harm societies where looting might be prevalent (Clarke & Szydlo, 2017). Given these threats to cultural heritage sites and their broader social and economic implications, there is a critical need for new technologies that can enhance site protection.

Unmanned aerial vehicles (UAVs) have the potential to provide a solution to the threats against cultural heritage by being platforms that police and other authorities can use to monitor for potential damage or impact against heritage sites, particularly in remote locations or areas with limited resources (Silverman and Ruggles 2007). However, a problem remains in that not all personnel can easily determine if looting has occurred, especially if damage is not easily detected, or authorities are not trained on what to look for. Therefore, there is a need to develop technologies that address these limitations, including in monitoring and responding to threats against heritage sites.

In this study, we propose a new deep learning tool that processes optical imagery from unmanned aerial vehicles and demonstrates the ability to identify looting damage to heritage sites. The purpose of this application is not only to identify potential damage to sites but also to monitor live imagery to alert authorities of potential damage, which can then be used to dispatch protection or develop an appropriate response to damaged cultural heritage sites. We argue that such computational tools will be increasingly needed as looting and damage to heritage sites accelerate and pose a real threat to countries' heritage.

We begin our presentation by providing background on heritage site looting and the use of deep learning methods for monitoring threats. Our approach is then presented, along with relevant data used to demonstrate the efficacy of our work. We then present results to show the accuracy, precision, and recall of the tool. A discussion on the implications of such tools for protecting heritage is then given. Future work and direction of research are also discussed.

Background

Looting of Heritage Sites

Key reasons why heritage looting can be destructive to communities where such events occur include damage to aesthetic, historical, and/or spiritual characteristics of given sites (Al-Ansi et al., 2021; Byrne, 2016; Silverman and Ruggles 2007). Furthermore, economic loss from heritage damage or looting could make it difficult for communities to recover and create employment opportunities, particularly those that had greatly depended on heritage for income (Brodie, 2010). Looted items from cultural heritage sites also generate illicit revenue, often for criminal gangs (Kersel and Hill 2019; Brodie & Renfrew, 2005). A relatively recent UNESCO-sponsored estimate has stated that sales of heritage items amounts to over \$50 billion USD, with a large percentage of this likely being illegal sales. In 2020 alone, over 800,000 cultural heritage objects were seized globally by law enforcement, with many of these objects coming from illegally excavated or looted sites (UNESCO 2021). This suggests that heritage site looting has likely increased or continues to pose a significant threat to communities and their social and economic well-being. Among various international treaties, an important one is the UNESCO 1970 Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property treaty, which was intended to help limit the sale of illegally obtained heritage objects and thereby potentially limiting damage to heritage sites where cultural objects come from (UNESCO 2022). While over 140 countries have now ratified this

international treaty, enforcement and attitudes towards enforcing this and other similar treaties have been argued to be weak, including in relative priorities for law enforcement (Runhovde 2022). This potentially necessitates more time efficient tools that demand less resources to aid law enforcement, particularly if the priority in allocating financial and personnel resources in preventing looting or illicit removal of cultural objects from sites is likely to remain low in the future.

Deep Learning and Heritage Monitoring

There is a need to increase efficiency in the monitoring and protection of cultural heritage sites. Machine learning-based implementations are more cost-effective and time-efficient because they eliminate the need for labor-intensive manual feature engineering and extraction from raw data, such as images and videos. Deep learning (DL), also known as deep structured learning, has the capabilities to transform the traditional input pipeline requiring data pre-processing and manual feature extraction. Recent trends within the field have shifted towards end-to-end (E2E) deep learning that promises potentially better performance (Boloor et al., 2020). Supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning techniques can be utilized to create DL models that combine automatically learned input data representations and artificial neural networks (ANNs). The deep learning model is comprised of two components, with the first being feature extraction and second is the learning component (Sarkar and Shenoy 2020).

The superior performance of DL over traditional machine learning techniques has revealed its potential for automating processes across a wide variety of applications. Work has shown that DL can improve performance in image classification, computer vision, machine translation, speech recognition, image captioning, healthcare domain applications, prediction of events, and in many other areas of application (Rajesh et al., 2021; Sarkar 2021). In most cases, DL can work with 2D image data (LeCun et al., 2015; Krizhevsky et al., 2018), 3D point cloud data (Charles et al., 2017), hyper-spectral imagery (Chen et al., 2014), and interpret sequence to sequence data (Tandler et al., 2019). Among deep learning models, variations of convolutional neural network (CNN) models have been proven to work well on image data, while stacked autoencoders are suitable for learning features from input data and encoding them to a compressed vector, from which a decoder learns to generate the original input data. Deep learning models usually require a lot of training data to be able to generalize and classify given images. In many applications where the number of training data are not sufficient, transfer learning can be applied (Tammina 2019). Transfer learning refers to training a model in which the neural network is trained using a large set of training data first and subsequently using the pre-trained model as a feature extractor for a new application, where the new application often has a smaller dataset. Disadvantages of DL have included the large number of training data required as well as long processing times sometimes required to train models.

In heritage-related topics, automated classification has been applied to remotely sensed images, including in topics related to the detection of looting on heritage sites. For instance, research has applied hierarchical categorization and localization as a method to detect looting pits (or holes) on satellite imagery for archaeological sites (Bowen et al., 2017). Other work uses autocorrelation, unsupervised classification, and segmentation on satellite imagery (Lasponara and Masini 2018). A relatively fast change detection technique was used to also automate detection of looting and heritage damage at sites (Cerra et al., 2016). Other machine learning techniques have also been used, including those that deploy multiple methods to derive optimal results (El-Hajj, 2021). Currently, most DL methods have been used on satellite-based data, although within remotely sensed data most of this work is not related to looting detection but general feature detection and extraction (Guyot et al., 2021; UNESCO 2022). The application of DL to optical data from UAVs has been deployed in cultural heritage research, but this has been limited mainly to non-looting

contexts and detection (Altaweel et al., 2022). Here, we assess the utility of DL as a technique for looting monitoring using optical data from UAVs and the applicability of this technique for site protection. What this entails and the contributions we make in applying UAV data are further discussed in the methods section.

One advantage that UAVs provide over most satellite imagery is that they can be used to more rapidly inform on damage occurring to heritage sites. Data, in fact, could potentially be used almost instantaneously, where automated messaging from detected damage can be used to alert authorities, potentially even in near real-time. While short-range UAVs may not be ideal to cover vast distances for monitoring of potential looting, long-distance UAVs, which can travel tens or hundreds of kilometers, could be utilized to access remote regions. By decreasing the time between the detection of damage to sites and alerting relevant authorities, UAVs and deployed software can play a crucial role in the protection of heritage and limiting damage that affects communities and economies. Applying detection algorithms that can identify threats or occurrences of looting are needed to enable automation without the requirement of having trained staff always present to monitor UAV data.

Methods

Deep Learning Approach

We develop a DL technique that processes optical imagery from UAVs, where still images from the Middle East are used to train a deep learning model for heritage sites in that region. Potentially, different imagery types, such as hyperspectral data, can be used in the approach presented, but we purposely limit this to optical data due to the fact such data represent the most common data UAVs produce and our experience is in the Middle East region.

Prior to beginning, we evaluated different detection algorithms for our application. One popular object detection algorithm is You Only Look Once (YOLO); one critique is this approach may struggle with smaller or unusual shapes within smaller images (Francies et al., 2022). Other popular object detection models include Single Shot MultiBox Detector (SSD) (Liu et al., 2016), RetinaNet (Del Prete R et al., 2021), and Faster regional-convolutional neural network (RCNN) (Girshick, 2015). SSD is known for its speed and accuracy in detecting objects in real-time video streams. However, it may struggle with detecting small objects or objects with complex shapes. RetinaNet is a popular choice for detecting objects at various scales and resolutions, making it suitable for a wide range of applications. However, it may require more training data to achieve optimal performance. Faster RCNN is an extension of RCNN and is known for its speed and accuracy in object detection tasks. However, it may be more computationally intensive than other models and may require longer training times.

We utilize a form of RCNN to detect and extract features using segmentation (Girshick et al., 2014). This algorithm works by creating proposed regions in an image that are evaluated to potentially belong to a given object. The regions are sub-segmented using color, texture, size, and shape using a selective search algorithm, where similar regions are then combined to form identified singular objects (Uijlings et al., 2013). Our full training method is a mask region-based convolutional neural network (Mask RCNN) used to train on data that consist of looted regions within cultural heritage sites. Imagery obtained were used to train and validate results. The approach is comparable to earlier work and is summarized here with the code also provided with this work (Altaweel et al., 2022).

Our developed approach enables image segmentation as well as instance segmentation, where images, in this case 2D optical photographs, are divided into multiple segments. This then divides and delineates objects within imagery. Here, we deploy the approach to perform instance

segmentation, which enables instances of objects to be demarcated to show their boundaries. The instances we identify are looting pits or holes that signify damage or presence of damage on heritage sites likely conducted by looters. Our main algorithm is related to and builds on the Fast RCNN method (He et al., 2017; Girshick, 2015). When potential objects, such as pits that suggest looting or damage by looters on an image, are identified, classification is done using a convolutional layer which generalizes data from input images into feature maps via a series of filters and kernels. Within this series of layers, pooling layers are used to downsample feature maps; layers are then connected using references to enable feature reconstruction. These connections enable and incorporate weights and bias input data that translate between layers. Multiple convolutional and pooling layers are created to enable further propagation between layers that enable forward and backward image reconstruction on imagery data. The underlying neural network model used is ResNet101, which means a convolutional model with 101 layers. The regional aspect of the Mask RCNN used applies bounding boxes across regions of interest (ROI) to then classify single or multiple regions into relevant classes, such as looted areas or holes in our case. The mask part of the algorithm creates an output that determines a spatial layout by refining the bounding box at the pixel level.

Figure 1 demonstrates the steps deployed in the Mask RCNN used; the code provided (see [Supplementary Material](#)) gives further, specific detail on the implementation algorithm. To summarize the main CNN algorithm, connectivity between layers in the neural network and layer transformation is conducted using convolution functions (Figure 2) (Chen & Ho, 2019). The convolution changes pixels into a single value which is then connected with a rectified linear activation function (ReLU), which is a piecewise linear function that outputs positive input or zero for non-positive. The pooling operation layer is used to calculate feature map values, ultimately creating a downsampled map feature. At the end, layers in the CNN are connected with the output

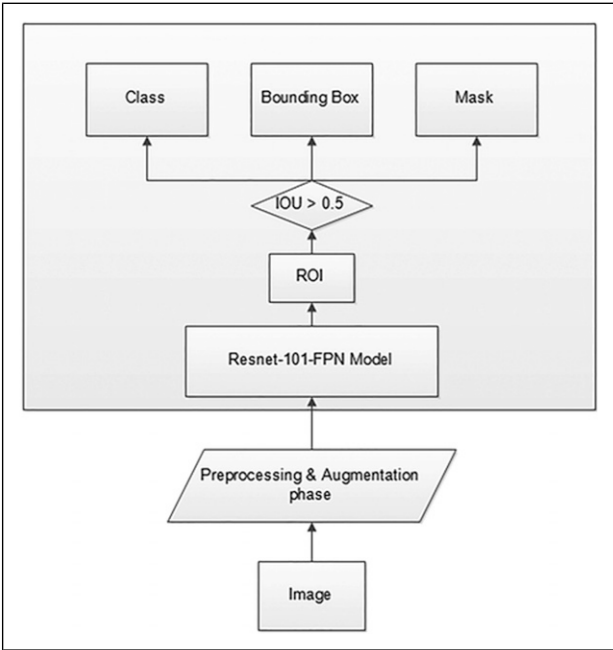


Figure 1. Diagram showing the deployed mask RCNN approach with relevant layers deployed to process and classify data.

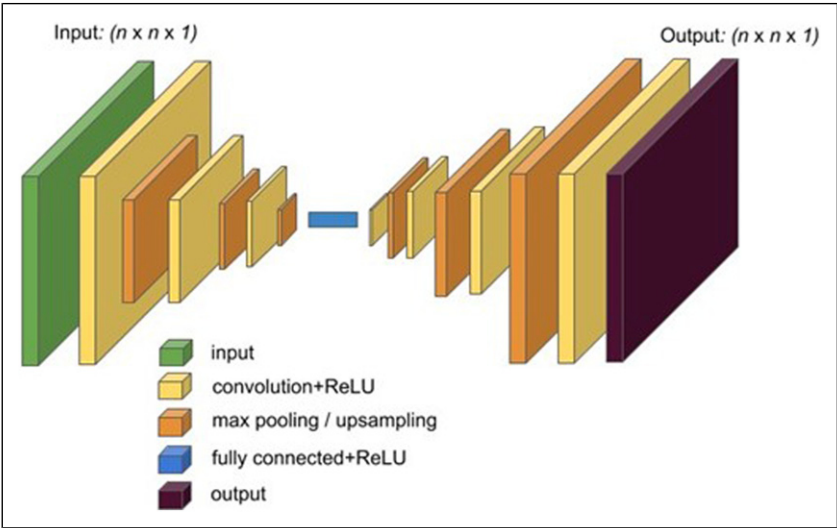


Figure 2. Connectivity between layers and input and output data applying convolution, ReLU, and pooling functions in the CNN.

feature. Going back to Figure 1 and the steps used in our sequence of methods, training images are preprocessed and augmented, which includes preparing and cleaning training images as well as annotating images. Augmentation is a process whereby improvements to images attempt to reduce validation error in training (Shorten and Khoshgoftaar 2019). In our case, this involved creating new images from existing images. Images were flipped left to right, up and down, rotated 45°, rotated 90°, scaled to .5 size, and scaled to 1.5 size. This enabled different sized images and different angles and variations to be used in training data. Once training and validation data were prepared, then a ResNet101 network model was used; in our case, we used the ResNet101-Feature Pyramid Network (FPN). This is a feature extraction model that takes an image of an arbitrary size and produces an output that has proportionally sized feature maps at multiple levels. The result from each network layer is used to reference feature maps for subsequent layers (Lin et al., 2017). The next step is ROI pooling, where the step involves taking input from identified classes and converting given feature maps into fixed dimensions. This helps connect layers and ensure outputs are the same size between layers. The subsequent step involves using intersection over union (IOU), which is used to calculate mean average precision. This is a calculation used to determine the accuracy of object detection in a model relative to ground-truth object annotations. Results over .5 are deemed sufficient or a good prediction, which is used here to evaluate performance of training. After this step, the trained model can be used to identify a relevant class, in this case looting pits/holes with bounding boxes for features designating site damage, and the regional mask, which is an extraction of a regional layout on images. We used a threshold value of 0.5 as the decision point for classifying pixels as looting holes/pits or background data.

Table 1 lists key input values used in training and developing the DL model deployed in our case. The final product of the DL algorithm is a model file that can then be used to classify new imagery for the presence of looting on a given site. In this case, we identify relatively new or fresh looting, with an excavated pit being by far the most common pit encountered on heritage sites in the study area. In Table 1, the learning rate is how much the model is updated and changed in response to the estimated error as model weights are updated during training. The Train ROIs Per Image refers to the region of interests being incorporated to the classifier. The learning momentum refers to the value

Table 1. Relevant Input Hyperparameters for the Deep Learning Model Produced.

Input Parameter	Value
Learning rate	.002
Train ROIs per image	700
Learning momentum	.95
Weight decay	.0001
Detection min. confidence	.7
Steps per epoch	150
Num. classes	2
Mask pool size	15
Pool size	8
Validation steps	50
RPN anchor scales	16, 32, 64, 128, 256
RPN anchor ratios	.5, 1, 1.5
IMG. min. dim	512

that helps to improve training speed and accuracy by providing a coefficient for weight updates and helping to prevent sensitive movement in model training. Weight decay is used as regularization to reduce model complexity and overfitting. The detection minimum confidence gives us a score (usually scored between 0 and 1) whereby the detected object is a correct classification, with 1 being a more certain classification. The steps per epoch parameter refers to the number of training samples to run relative to the batch size; the number of classes refers to the classes to identify in imagery (i.e., looting holes and non-pit damage). Masking pool size refers to the sequence-processing layers that are reduced when processing the data. The pool size refers to the pooling operation or filter size (e.g., 2×2 or 8×8) applied to a feature map for given scenes. The validation step refers to the number of backpropagation applied to validation data to track validation accuracy. The anchor scales refers to one of the inputs used to generate candidate boxes for anchor points; this is used to identified given objects (e.g., looting holes). In addition to the scales, the ratios help derive these candidate boxes. Additionally, the image minimum dimension refers to the minimum dimension possible for images used in our examples (e.g., 512×512 resolution).

Data Collection Process and Training Data

Training data were obtained from available imagery from our own data, archival data (APAAME, 2022), and images made available to us. These data allow us to annotate given imagery so that evidence for looting is indicated on images and objects classified as looting are used to train (Figure 3).

The UAV data made available to us and used in this study were collected using a DJI Phantom 4 Pro drone at an altitude of 100 meters above ground level (AGL). The drone was equipped with a 20-megapixel camera capable of capturing high-resolution images with a pixel size of 5472×3648 . The camera was set to capture images in JPEG format with a compression quality of 100%. The study area where we obtained imagery covers a region of approximately 25 square kilometers in the vicinity of archaeological or heritage sites that have experienced looting damage in the form of looting pits or holes. The drone was flown in a grid pattern over the study area, capturing images of the ground at an angle perpendicular to the surface. The images were taken during daylight hours with clear weather conditions to ensure high-quality images. The number of looting pits/holes ranged between 10 and 200 in each image, with over 8000 looting holes in total.

In total, there are 95 images used in training and model validation, with 80% of the samples (76 images) used for training a model and 20% (19) used for validation. This worked out to be over 1500 looting holes for validation and over 6000 for training. This is also what we based our tests on for validation. All images were made to be the same dimensions and underwent quality enhancement, including orienting and adjusting images so that looting damage was clear. Objects classified as evidence for looting are annotated using a tool called LabelMe that enabled classes to be defined as polygons (LabelMe 2022) and used in training/validation. Additionally, as models are created by the end result of the training in the Mask RCNN algorithm, we used a pretrained model derived from the Microsoft COCO dataset. This assists in creating a final trained model using an annotated set that consists of over 330,000 images where pre-defined and pre-trained weights are generated (Microsoft COCO Dataset 2022). The COCO dataset is also used to benchmark our algorithm to compare performance with object detection on deployed images. The class identified is a looting hole/pit, as the images we have from all of our sampled images relate to archaeological or heritage sites that have looting damage in the form of looting pits or holes. This was, therefore, used for training, validating, and testing the final model result.

Data Collection Process

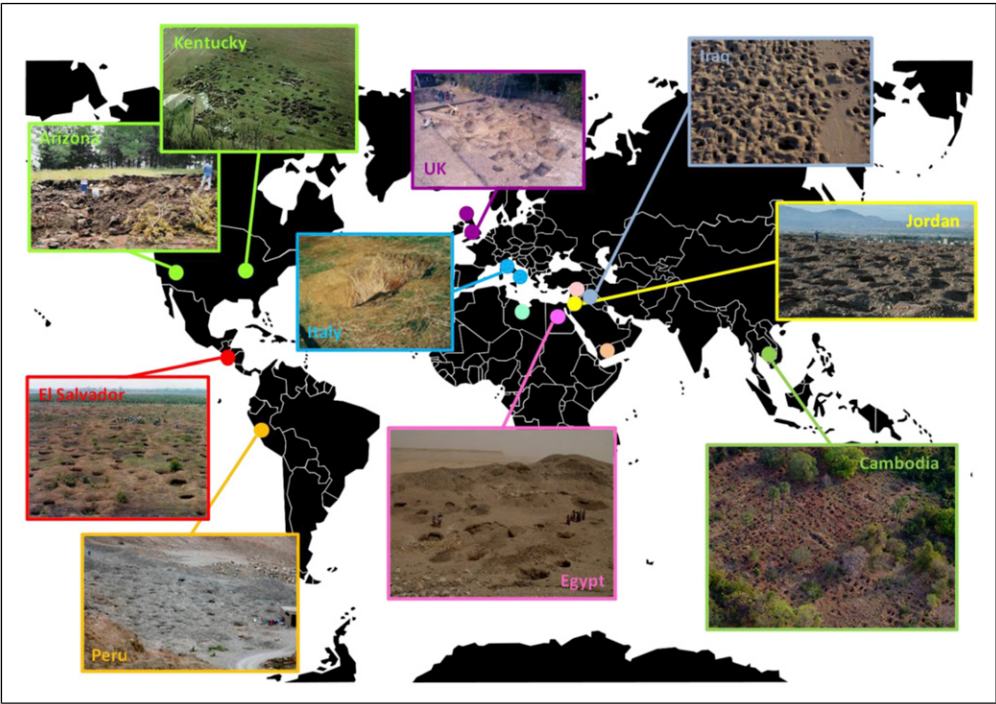


Figure 3. Example of looting damage used for training from different regions and used to train the mask RCNN algorithm.

Results

We deploy our algorithm, which produces a model output from training. Table 2 provides summary results of key metrics that demonstrate the potential and testing results of our approach. Type I and II errors were 85 and 20, respectively, from validation samples. The key output of our approach is instance segmentation of looting damage, that is the identification of individual looting hole features, displayed. Figure 4 demonstrates an example of instance segmentation, where identified looting hole/pit instance segments are shown and counted. While this demonstrates the utility of the approach, we also deploy formal tests to indicate accuracy, precision, and recall results. This is discussed below.

Validation

During model training, the best results achieved demonstrate model accuracy at 93%, that is the accuracy in positive detection of looting holes, with a loss close to .3 (Figure 5). The true positive

Table 2. Summary Results of Tested Metrics for the Applied DL Approach.

Test Metric	Result
Model accuracy	93%
Model loss	.3
Precision score	.65
Recall score	.73
F1 score	.69

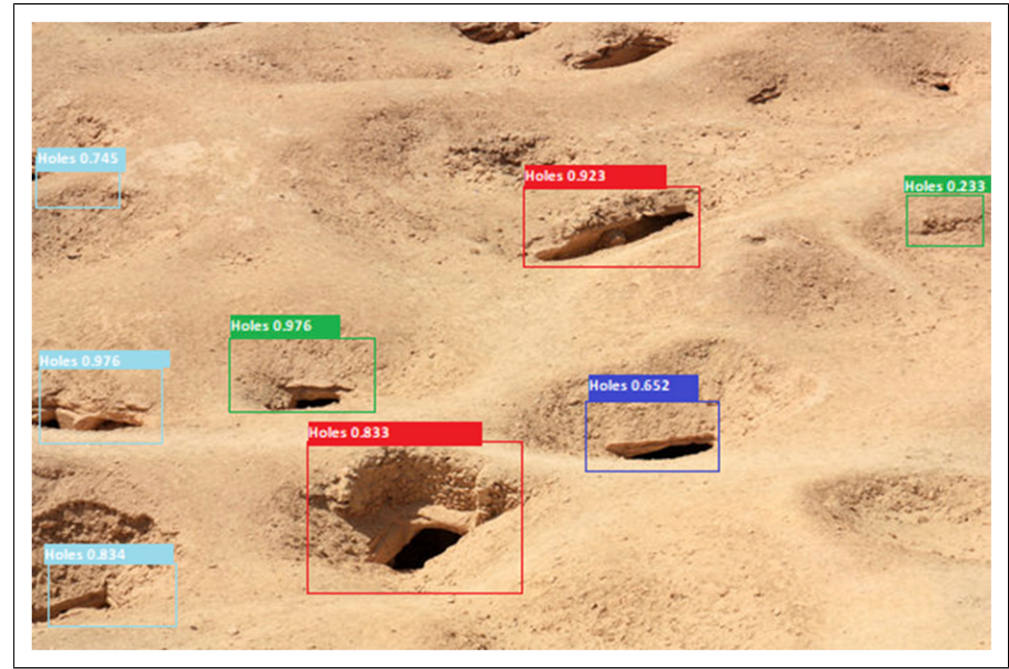


Figure 4. Example of instance segmentation enabled by the deployed algorithm.

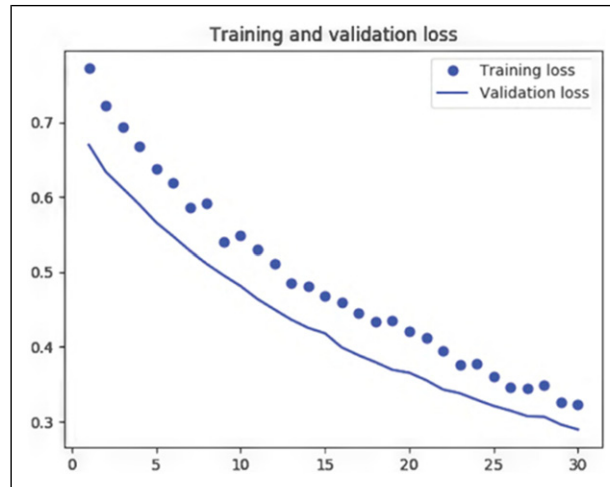


Figure 5. Results from the deployed loss function used in training validation.

identification refers to the fact that when a hole was detected by our software, the trained model was accurate at about 93%. The loss function demonstrates how well the model training can predict the dataset. A mean square error function is used to calculate loss, which measures how well the training meets given expected results. Although this is useful for training, we wanted to also test the application using introduced images. Figures 6 and 7 demonstrate two introduced samples where the model is deployed to apply instance segmentation on looting damage. Overall, about 80% of new looting damage can be identified in these images, which have over 800 pits. Formally, and deploying an F1 score (Baeza-Yates & Ribeiro-Neto, 2011) for testing precision and recall for introduced images, results demonstrate an overall score of .69; the precision score achieved is about .65, while recall is .73. A total of 750 looting holes were predicted; 120 present holes were not detected. Overall, the model performed well in detecting looting holes in new excavations or holes, but was less successful in identifying some older holes, particularly those with poor preservation. The results suggest that even for trained experts identifying older looting could be challenging. The accuracy score was high, being over 90%, but better training data might be needed to improve the precision and recall rates particularly for older looting holes.

In other words, a number of false negative observations are recorded by the existing model. Looking at most of the looting damaged missed, it is clear that older holes have in many cases eroded and lost some of their form, which can create irregular shapes and can be more easily missed, even by trained professionals. In regards to more fresh holes or those with clear shadowing and linear features that define shape, these are more easily identified and are often not missed. However, even some of these are missed as demonstrated on Figures 6 and 7. Clearly identifiable looting damage holes that are mostly or partially eroded are missed in about 30–40% of cases; the figure could be higher given that we could not always determine if previous looting holes were in fact evidence for looting. This shows that even for trained experts identifying older looting could be challenging. Nevertheless, where clear forms and shapes are evident for looting damage, the approach demonstrates utility as confirmed by manual observation by experts. In particular, the accuracy score was high, being over 90%, but better training data and increased augmentation methods might be needed to improve the precision and recall rates, including for older looting holes. The results for new images reflects the model's inability to generalize well to new data, particularly data with old holes that create irregular shapes.



Figure 6. Introduced image used to identify looting holes and damage to an archaeological site.

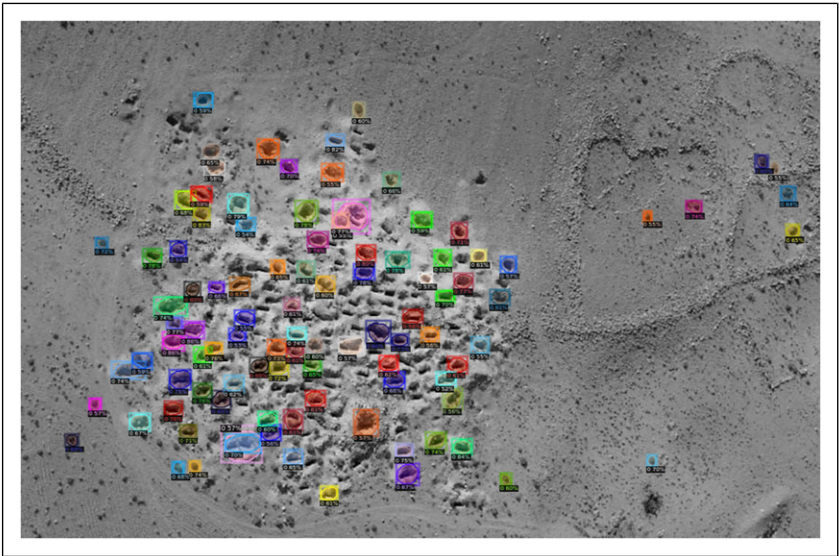


Figure 7. Another introduced image with segmented looting damage; some older looting damage is missed.

Discussion

Recent events around the destruction of heritage have demonstrated the need for measures that better protect and detect threats. Cultural heritage plays an integral role in the economies, identities, and well-being of communities; however, it is repeatedly under threat and better measures for its protection are needed. Computational approaches applied to optical imagery that can determine ongoing or real-time events could benefit stretched or limited policing resources, particularly in remote areas. We present an approach that shows good results in detecting looting damage, which can then be used to automatically inform policing authorities of such damage. We have chosen to focus on detecting looting holes, as these are often the main evidence for ongoing

or new looting to an area. Such detection of these features would enable police or other protective forces to respond to events, particularly if security is untrained or not able to easily identify looting damage without assistance from an automated approach. Previous approaches, as discussed above, had mostly depended on satellite-based imagery, which is limited by the fact that data from satellites often are not obtained quickly enough to enable a more rapid protective response.

Application to Heritage

We envision that the approach we used is incorporated with real-time optical data received from UAVs, where the data are accessed by those responsible in protecting heritage sites. Authorities can utilize the tool to automatically assess imagery as they are received, the number of looting holes could be counted, and any information that demonstrates new damage to a site or unknown damage could then enable a response. The tool is best used for remote sites that are likely to be monitored by long-range UAVs. For instance, the software could be applied to analyze imagery coming in from a remote feed, where authorized individuals in charge of security for sites could use the information to respond to threats. Such threats are identified if new or an increase in the number of looting holes is evident from previous observations. While drones (UAVs) are widely used for security and monitoring for looting, the use of automated approaches has rarely been used for heritage. This can become a problem if those who provide security on sites are not trained to identify looting or to more easily count new looted features; the automated approach such as the one presented can be used to demonstrate discrepancies from previous observations by indicating if new or additional damage is evident (Kersel and Hill 2019).

The current approach shows efficacy in being relatively accurate at detecting particularly new looting holes. The results described show that overall holes or pits, when evident and identified were accurately recorded at around 93%. Such features are generally regular and more clear on optical imagery. The approach did not always work well in identifying older or more eroded looting features; however, we attribute this to the fact that such instances are often irregular in shape and can be even difficult for trained individuals to identify. More machine training on eroded or irregular features might be needed to improve overall accuracy for these specific sub-type features. Overall, increasing training samples would improve accuracy, precision, and recall values. Augmentation was observed to help training of our data, helping to make the model more accurate. Providing imagery with varied angles and rotation helps DL models to increase training samples even in cases where the original, non-augmented data are somewhat limited.

Given the threats to and importance of heritage in different countries, we envision automated approaches to protecting heritage sites to be important in coming years. We also foresee potential privacy violations such DL tools present. To avoid problems in privacy, such software should only be trained for specific tasks such as identifying threats or damage to heritage sites without providing information on individuals. Work that uses collaborative training between stakeholders, such as between policing authorities but also groups that represent the wider public, might be needed to ensure privacy protection. In such cases, example training data are collaboratively selected and approved rather than only one stakeholder choosing training data (Boulemtafes et al., 2020).

Conclusion

More can be done to improve the approach. For instance, we noted above that the approach is the weakest in relation to identifying old looting pits or holes. While this does not significantly limit the tool, in our opinion, as policing authorities are more likely to be interested in newer damage, we could improve the analysis by also segmenting more old looting damage to train along with

additional clear examples of looting. This would then allow the tool to differentiate new versus old damage if enough varied examples are provided. Additionally, other classes, such as looter equipment, could be added to identify potential for looting. There would have to be protection of individual data enabled in such cases, requiring work with stakeholders to ensure protection, but identification of unauthorized equipment or items on heritage sites could be useful for protective authorities. Finally, thermal imagery as well as optical data could be useful to incorporate given that looting often takes place at night and detection of illegal activities at night might be an even better approach to enable more timely responses. In all these cases, the main limitation has been training data for the DL model, which is often the case for DL in general. Therefore, there is a need to better share and obtain data in such an approach. As part of our effort, we provide the code and hope others can build on this effort. From the current results, we have demonstrated the efficacy of training such DL algorithms to automatically identify looting damage, including the most common damage identified on heritage sites in the form of looting pits or holes.

Acknowledgments

We would like to thank Abu Dhabi University for supporting this work. They provided funds to support this research and for Dr. Altaweel's time in Abu Dhabi University.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research and work was funded by Abu Dhabi University.

Supplemental Material

Supplemental material for this article is available online.

References

- Al-Ansi, A., Lee, J.-S., King, B., & Han, H. (2021). Stolen history: Community concern towards looting of cultural heritage and its tourism implications. *Tourism Management*, 87, 104349. <https://doi.org/10.1016/j.tourman.2021.104349>
- Altaweel, M., Khelifi, A., Li, Z., Squitieri, A., Basmaji, T., & Ghazal, M. (2022). Automated archaeological feature detection using deep learning on optical UAV imagery: Preliminary results. *Remote Sensing*, 14(3), 553. <https://doi.org/10.3390/rs14030553>
- APAAME (2022). *Aerial photographic archive for archaeology in the Middle East*. APAAME. <https://www.apaame.org/> (Accessed on 11 November 2022).
- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search* (Second edition). Addison Wesley.
- Basnet Silwal, S. (2021). The economics of looting and destruction of cultural heritage sites and objects. *Defence and Peace Economics*, 34(2), 214–227. <https://doi.org/10.1080/10242694.2021.2007638>
- Boloor, A., Garimella, K., He, X., Gill, C., Vorobeychik, Y., & Zhang, X. (2020). Attacking vision-based perception in end-to-end autonomous driving models. *Journal of Systems Architecture*, 110, 101766. <https://doi.org/10.1016/j.sysarc.2020.101766>
- Boulemtafes, A., Derhab, A., & Challal, Y. (2020). A review of privacy-preserving techniques for deep learning. *Neurocomputing*, 384, 21–45. <https://doi.org/10.1016/j.neucom.2019.11.041>

- Bowen, E. F. W., Tofel, B. B., Parcak, S., & Granger, R. (2017). Algorithmic identification of looted archaeological sites from space. *Frontiers in ICT*, 4, 4. <https://doi.org/10.3389/fict.2017.00004>
- Brodie, N. (2010). Archaeological looting and economic justice In P. M. Messenger & G. S. Smith (Eds.), *Cultural heritage management, policies and issues in global perspective* (pp. 261–277). University Press of Florida.
- Brodie, N., & Renfrew, C. (2005). Looting and the world's archaeological heritage: The inadequate response. *Annual Review of Anthropology*, 34(1), 343–361. <https://doi.org/10.1146/annurev.anthro.34.081804.120551>
- Brusasco, P. (2012). *Looting the past: Syria's cultural heritage under attack: Another Iraq? Biblioteca contemporanea archeologia*. Limena PD: Libreriauniversitaria. it edizioni.
- Byrne, D. (2016). The problem with looting: An alternative perspective on antiquities trafficking in Southeast Asia. *Journal of Field Archaeology*, 41(3), 344–354. <https://doi.org/10.1080/00934690.2016.1179539>
- Cerra, D., Plank, S., Lysandrou, V., & Tian, J. (2016). Cultural heritage sites in danger—towards automatic damage detection from space. *Remote Sensing*, 8(9), 781. <https://doi.org/10.3390/rs8090781>
- Charles, R. Q., Su, H., Kaichun, M., & Guibas, L. J. (2017) PointNet: Deep learning on point sets for 3D classification and segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 77–85. IEEE. <https://doi.org/10.1109/CVPR.2017.16>
- Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2094–2107. <https://doi.org/10.1109/JSTARS.2014.2329330>
- Chen, Z., & Ho, P.-H. (2019). Global-connected network with generalized ReLU activation. *Pattern Recognition*, 96, 106961. <https://doi.org/10.1016/j.patcog.2019.07.006>
- Clarke, C. M., & Szydlo, E. J. (2017). *Stealing history: Art theft, looting, and other crimes against our cultural heritage*. Rowman & Littlefield.
- Del Prete, R., Graziano, M. D., & Renga, A. (2021). RetinaNet: A deep learning architecture to achieve a robust wake detector in sar images. In 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), Naples, Italy (6, pp. 171–176). IEEE. <https://doi.org/10.1109/RTSI50628.2021.9597297>
- El-Hajj, H. (2021). Interferometric SAR and machine learning: Using open source data to detect archaeological looting and destruction. *Journal of Computer Applications in Archaeology*, 4(1), 47–62. <https://doi.org/10.5334/jcaa.70>
- Francies, M. L., Ata, M. M., & Mohamed, M. A. (2022). A robust multiclass 3D object recognition based on modern YOLO deep learning algorithms. *Concurrency and Computation: Practice and Experience*, 34(1), 1–24. <https://doi.org/10.1002/cpe.6517>
- Girshick, R. (2015) Fast R-CNN. In Proceedings of the 2015 IEEE international Conference on computer vision (ICCV), Santiago, Chile, pp. 1440–1448.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, pp. 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- Guyot, A., Lennon, M., & Hubert-Moy, L. (2021). Objective comparison of relief visualization techniques with deep CNN for archaeology. *Journal of Archaeological Science: Reports*, 38, 103027. <https://doi.org/10.1016/j.jasrep.2021.103027>
- He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017) Mask R-CNN. In Proceedings of the 2017 IEEE international Conference on computer vision (ICCV), Venice, Italy, pp. 2980–2988.
- Kersel, M. M., Hill, A., & Chad (2019). The (w)hole picture: Responses to a looted landscape. *International Journal of Cultural Property*, 26(3), 305–329. <https://doi.org/10.1017/S0940739119000195>

- Kersel, M. M., Hill, A., & Chad (2020). Databases, drones, diggers, and diplomacy: The Jordanian request for a US cultural property bilateral agreement. *Journal of Field Archaeology*, 45(sup1), S101–S110. <https://doi.org/10.1080/00934690.2020.1713282>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- LabelMe (2022). *The open annotation tool*. LabelMe. <https://labelme.csail.mit.edu/Release3.0/> (Accessed on 11 11 2022).
- Lasaponara, R., & Masini, N. (2018). Space-based identification of archaeological illegal excavations and a new automatic method for looting feature extraction in desert areas. *Surveys in Geophysics*, 39(6), 1323–1346. <https://doi.org/10.1007/s10712-018-9480-4>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017) Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 936–944. IEEE. <https://doi.org/10.1109/CVPR.2017.106>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Cheng-Yang, Fu., & Berg, A. C. (2016). SSD: Single Shot MultiBox detector. In B. Leibe, J. Matas, & N. Sebe (Eds.), *Computer vision—ECCV 2016. Lecture notes in computer science* (pp. 21–37). Springer International Publishing. <https://doi.org/10.1007/978-3-319-46448-02>
- Microsoft COCO Dataset (2022). *Common objects in context*. Microsoft COCO Dataset. <https://cocodataset.org/> (Accessed on 15 November 2022).
- Rajesh, V., Naik, U. P., & Mohana (2021). Quantum convolutional neural networks (QCNN) using deep learning for computer vision applications 2021 International Conference on Recent Trends in Electronics, Information, Communication & Technology (RTEICT) Bengaluru, India, (pp. 728–734). IEEE. <https://doi.org/10.1109/RTEICT52294.2021.9574030>
- Runhovde, S. R. (2022). Changing attitudes: Lessons for countering the trafficking of cultural objects. *Global Crime*, 23(4), 407–421. <https://doi.org/10.1080/17440572.2022.2117697>
- Sarkar, S. D., & Shenoy, K. B. A. (2020) Face recognition using artificial neural network and feature extraction. In: 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 417–422. IEEE. <https://doi.org/10.1109/SPIN48934.2020.9071378>
- Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Silverman, H., & Ruggles, D. F. (2007). Cultural heritage and human rights. In H. Silverman & D. F. Ruggles (Eds.), *Cultural heritage and human rights* (pp. 3–29). Springer New York. <https://doi.org/10.1007/978-0-387-71313-7>
- Tamina, S. (2019). Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10), p9420. <https://doi.org/10.29322/IJSRP.9.10.2019.p9420>
- Tandler, D., Dörner, S., Cammerer, S., & ten Brink, S. (2019). On recurrent neural networks for sequence-based processing in communications. arXiv. <https://doi.org/10.48550/ARXIV.1905.09983>
- Taniguchi, Y. (2017). Do archaeological and conservation sciences save cultural heritage?: Cultural identity and reviving values after demolishment. In A. Tsuneki, S. Yamada, & K. Hisada (Eds.), *Ancient West Asian Civilization* (pp. 179–197). Springer Singapore. <https://doi.org/10.1007/978-981-10-0554-1>
- Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, 104(2), 154–171. <https://doi.org/10.1007/s11263-013-0620-5>

UNESCO (2021). *Returning the loot*. UNESCO. <https://www.unesco.org/en/articles/returning-loot> (Accessed 11 11 2022).

UNESCO (2022). *Convention on the means of prohibiting and preventing the illicit import, export and transfer of ownership of cultural property*. UNESCO. <https://en.unesco.org/about-us/legal-affairs/convention-means-prohibiting-and-preventing-illicit-import-export-and> (Accessed 11 11 2022).

Author Biographies

Mark Altaweel is Professor of Near East Archaeology and Archaeological Data Science. He is interested in applying data science techniques to solve various heritage-related problems.

Adel Khelifi is Professor and Head of Department in the Computer Science and Information Technology department at Abu Dhabi University. He is interested in creating new computational methods to help address various heritage challenges.

Mohammad Maher Shana'ah was an MSc student in the Department of Computer Science and Information Technology. He is interested in applying computer vision techniques in policing and protection of heritage.