Communication

Automated Archaeological Feature Detection Using Deep Learning on Optical UAV Imagery: Preliminary Results

Mark Altaweel [1,*], **Adel Khelifi** [2], **Zehao Li** [1], **Andrea Squitieri** [3], **Tasnim Basmaji** [4] **and Mohammed Ghazal** [4]

1. Institute of Archaeology, University College London, London WC1E 6BT, UK; zehao.li@ucl.ac.uk
2. Computer Science and Information Technology, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates; adel.khelifi@adu.ac.ae
3. Institut für Ur- und Frühgeschichte und Vorderasiatische Archäologie, Universität Heidelberg, Sandgasse 7, 69117 Heidelberg, Germany; ax294@uni-heidelberg.de
4. Electrical, Computer, and Biomedical Engineering Department, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates; tasnim.basmaji@adu.ac.ae (T.B.); Mohammed.Ghazal@adu.ac.ae (M.G.)
*   Correspondence: m.altaweel@ucl.ac.uk; Tel.: +44-(0)20-7679-4607

Abstract: This communication article provides a call for unmanned aerial vehicle (UAV) users in archaeology to make imagery data more publicly available while developing a new application to facilitate the use of a common deep learning algorithm (mask region-based convolutional neural network; Mask R-CNN) for instance segmentation. The intent is to provide specialists with a GUI-based tool that can apply annotation used for training for neural network models, enable training and development of segmentation models, and allow classification of imagery data to facilitate auto-discovery of features. The tool is generic and can be used for a variety of settings, although the tool was tested using datasets from the United Arab Emirates (UAE), Oman, Iran, Iraq, and Jordan. Current outputs suggest that trained data are able to help identify ruined structures, that is, structures such as burials, exposed building ruins, and other surface features that are in some degraded state. Additionally, *qanat*(s), or ancient underground channels having surface access holes, and mounded sites, which have distinctive hill-shaped features, are also identified. Other classes are also possible, and the tool helps users make their own training-based approach and feature identification classes. To improve accuracy, we strongly urge greater publication of UAV imagery data by projects using open journal publications and public repositories. This is something done in other fields with UAV data and is now needed in heritage and archaeology. Our tool is provided as part of the outputs given.

Keywords: unmanned aerial vehicles; optical; deep learning; archaeology; feature detection; software; high-performance computing

1. Introduction

The use of unmanned aerial vehicles (UAVs or drones) in archaeology, similar to many other disciplines, has become standard within remote sensing and fieldwork applications [1]. While satellite-based remote sensing has proved to facilitate large-scale discovery of many archaeological features, often such imagery is limited based on resolution and the fact it does not often represent the investigated landscape as it appears at a given moment of investigation. Among other benefits, UAVs are able to locate new or previously unknown archaeological features, including archaeological sites difficult to detect such as sites that are only evident based on a given pattern of stone scatter [2]. The ease of operating UAVs and simple optical imagery provided enables small features and archaeological patterns to be noticeable that may not be as apparent on satellite imagery. In some cases, researchers have applied software processing UAV imagery, including applying computer vision [3–5] and machine-/deep-learning techniques [6,7]. Such techniques are commonly

applied to process and merge imagery; discover new features; and document archaeological monuments, features, and excavation sites. While such software is useful, a gap that has emerged for studying ancient landscapes and archaeological features is automated tools that can facilitate the recording and identification of archaeological features that signal to researchers that such features warrant further investigation. Additionally, there is a need to create open source and simple GUI-based tools that make it easy for researchers to apply deep learning segmentation techniques that facilitate feature identification and extraction for a broad audience of researchers. However, a limitation is that researchers need to share image data to create better trained deep learning models that could then be used to help discover a variety of archaeological features in any application. Such systematic data sharing is currently not done in the field of archaeology, creating a limitation for tools that can auto-discover archaeological features from UAV imagery. This is true not only for optical imagery but also other forms of data such as LIDAR and thermal imagery.

This commentary argues that archaeological researchers in different regions need to begin to share UAV imagery data in a way that allows others to utilize the value of such imagery, similar to what is occurring in other research fields such as in ecology, transport, and urban studies. This can be done through open repositories, such as figshare [8], journal repositories, or those increasingly provided by universities. Large-scale repositories are also increasingly used in image analyses, and these could be applied within heritage or archaeology. Researchers who share data should be able to gain credit in research output and through other acknowledgments to provide some incentive. With increased use of UAV imagery, along with deep learning models in research, we anticipate that more researchers will need shared libraries to build better models for automated feature identification. As a wider contribution, we have begun developing a free and open software tool to aid automated feature detection using instance segmentation based on a mask region-based convolutional neural network (Mask R-CNN) technique. While we are only in the beginning of our research and software development, we believe it is important to share preliminary results and the current utility of our approach as the project develops rather than waiting until the end of the work. We, therefore, present GUI-based software that enables automated instance segmentation for UAV data. The software, while not yet tested on many different forms of imagery outside of optical data, does have the potential to be extended into other imagery, including thermal and even non-archaeological cases. So far, we have tested our approach on some relatively obvious features, including surface ruined structures; *qanats*, or ancient subterranean water systems; and mounded sites. Results are provided in this work, but they are still preliminary in nature. With more training data, we do believe more complex and varied types of sites, such as those with limited stone surface scatters, could be made more detectable in the near future. After a brief overview and background on deep learning and relevant software, the following will present our current approach and contribution in helping to automate archaeological feature discovery. The discussion section at the end will provide commentary on the need to share data in relation to UAV imagery and the next steps for this effort.

2. Background

2.1. Deep Learning in UAV Remote Sensing

According to Yao et al. [9], deep learning (DL) methods are likely to become the main approach of image classification in the remote sensing community. Osco et al. [10] performed a literature review search using the Web of Science (WoS) and Google Scholar to identify works related to DL in UAV remote sensing applications. For the years covering 2016–2021, they found only 34 research articles, 16 proceedings articles, and 8 reviews that covered this topic. The majority of these papers (91.2%) used or discussed the implementation of convolutional neural networks (CNNs). Nonetheless, the application of CNN architecture in remote sensing imagery is not new, given its increasing and wide use in remote sensing in general. In addition, CNN does not require feature extraction. The system learns to extract features, and the key notion of CNN is that it generates invariant

features by convolutioning images and filters, which are then passed on to the next layer in the network approach.

One of the most essential aspects of DL is training data to create models that could be applied to classify objects in new imagery. Training data are labeled, or annotated, to reflect regions where specific classes or objects are present. Increasingly, creating and using large repositories in different fields has been an approach researchers have taken to improve feature classification. Some known aerial imagery repositories with labeled instances include UAVDT [11], Brazilian Coffee Scene [12], VisDrone [13], DOTA [14], RSSCN7 [15], RSC11 [16], and WHU-RS19 [17] datasets. These repositories and others have helped to not only publish existing project data but also form important training and imagery validation in fields that apply DL algorithms. Osco et al. [10] provide examples of DL algorithms that produce some highly precise results on the detection and segmentation of trees and natural objects (e.g., [18,19]). The presence of large data repositories, therefore, is critical in fields benefiting from DL capabilities in feature identification.

2.2. Remote Sensing, Deep Learning, and UAV Imagery in Archaeology

While the above examples represent what could be possible with large datasets and DL techniques applied, very limited work has been done in archaeology. Nevertheless, airborne and satellite-based remote sensing has become an integral part of archaeological research and cultural resource management. Remote sensing, when combined with field research, allows for quick and high-resolution documentation of archaeological sites [20–22]. It makes it possible to document and monitor sites that are unreachable for fieldwork, destroyed, or threatened [23–25].

Machine-learning techniques and algorithms have gradually been included into remote sensing-based archaeological research in recent years, allowing for the automatic discovery of sites and their characteristics. The majority of these applications have used high-resolution satellite imagery [26]. Datasets such as LIDAR [27–29] use an active laser sensor that provides an illumination source, making features difficult to detect due to hindrances, such as trees, more evident. On the other hand, passive sensors such as WorldView imagery [30,31] are commonly used to detect small-scale archaeological characteristics, which can be evident due to vegetation or landscape changes. Other related work includes detecting anthrosols using a random forest (RF) technique [32] using multi-temporal ASTER imagery. Orengo et al. [33] provide a multi-sensor, multi-temporal machine-learning approach for detecting ancient mounds in Cholistan (Pakistan) utilizing massive data from remote sensing. In this study, a classifier technique that uses a large-scale collection of synthetic-aperture radar and multi-spectral pictures has been created using a Google Earth Engine, producing an accurate probability map for mound-like signals across ca. 36,000 square kilometers. Mask R-CNN methods have been successfully applied in a few previous works to detect archaeological sites and features [34], including an application related to *qanats* [35], with promising results showing the utility of the approach. While satellite-based data have greatly aided archaeology research, UAV optical imagery can provide advantages in cases where real-time data are needed or high-resolution information from given sites or landscapes is required. Similar to this work, Orengo et al. [6,36] have developed an open source approach to UAV-based image analysis for optical data. This work not only provides a comparable tool, but we have also developed a tool that is GUI-based and enables different stages of application needed for DL.

3. Software and Data

3.1. Mask R-CNN Approach

Mask R-CNN performs both image segmentation and instance segmentation, that is, dividing digital images into multiple segments and delineating individual objects, respectively. Most interest is used for instance segmentation, including in background research cited above in relation to the approach. The approach we take applies several components. The algorithm builds from the Fast R-CNN method [37,38]. The algorithm looks at candi-

date areas and performs classification using a convolutional layer that generalizes input images as feature maps using filters and kernels. A pooling layer that has downsampled feature maps, with the connected layer connecting the 'neurons' in a layer to neurons in another layer. These neurons reflect weights and bias input that translate between layers. Multiple convolutional and pooling layers can be created to add extra propagation, which can go forward and backward on given imagery data. Added to this, a region-based CNN is used, which effectively applies bounding boxes across the object regions, which evaluates convolutional networks independently on all the Regions of Interest (ROI) to classify multiple image regions into the proposed class. This reflects the first stage of proposed output from the algorithm. The mask is used to create an output that extracts a spatial layout for an object through predicting the class of the object and refining the bounding box at the pixel level.

3.2. Applied Software

We created a new application, which we have initially called Mask_UAV. The Supplementary Materials (see below) provides links to the software, documentation, data, and the trained model used. The core of our tool uses the PixelLib Python 0.6.6 library [39], which is built using the TensorFlow machine learning library [40], where this software is applied for the training of annotated features and segmentation of features used for classification and identification. This software allows us to apply instance segmentation on images using a developed Mask R-CNN model. Images are annotated, that is, individual objects are classified and delineated on imagery, so that a .json file associated with a given image is used for training; we use a Python tool called LabelMe (4.5.9) for this [41]. Additionally, we built output features and other graphical functionality using Python 3.8 to make the development and use of the Mask R-CNN approach easier for users by bringing together the required annotation, training, and segmentation functionality through one graphically based tool. The intent is also to increase and eventually incorporate other deep learning and computer vision techniques over time. We hope this will then be able to take advantage of UAV imagery data as they are made available by researchers. Further documentation and information about our approach could be obtained in the GitHub site provided with this article in the Supplementary Materials.

Figure 1 demonstrates examples from Mask_UAV's user interface; the tool has a graphical user interface (GUI) to guide users to different options. Figure 2 demonstrates the flowchart for the tool, with functionality summarized below. After the tool is launched, using the controller.py module, users obtain the options to annotate, train, or segment objects. A choice to annotate launches the LabelMe application, where the user can then select images, trace around objects, and classify them for training. As this option simply launches LabelMe, users should view the instructions on using that tool first. The training choice presents users with different options, including running training on the system they are on or simply saving their choices so that training can be run on a high-performance computer (HPC), including those with graphics processing units (GPUs). Once the training option is chosen, a window opens for users to select a training library, which is the folder where annotated training data exist. The data should have .json files, for the annotations, and be in a 'train' folder. Model validation files should be placed in the 'test' folder. The weight file, which is the initial weighting used in training the CNN layers, is then selected. This could be simply a default training model such as mask_rcnn_coco.h5, which is trained with the COCO dataset [42], although users may want to incorporate their own libraries as they develop. Users also have the option to set the number of the batch size, which comprises the training examples used in forward and backward propagation used in deep neural networks, and the number of classes to train (i.e., three classes in our case). This is then used to update the trained model. Users can then choose the neural network model, with the default set to resnet101. This is a neural network model that is a CNN with 101 layers; users can also select resnet50, which has 50 layers and runs faster due to having fewer layers, although it often produces less accurate results. The final choice is epochs,

which is the number of passes for the training dataset during runs. A high number, such as 300, helps improve accuracy, although this can take a long time to run depending on the computing resources and size of training data. Users can then press start, which will launch the training using PixelLib, or save the run, which will save to a training_data.csv file in the training_data folder within Mask_UAV. If a user saves the run and then applies training later or on another system such as HPC, it is recommended to use the train_set.py module directly in the training folder to launch training on systems used. Users can simply modify the data paths in the training_data.csv file for remote data training.
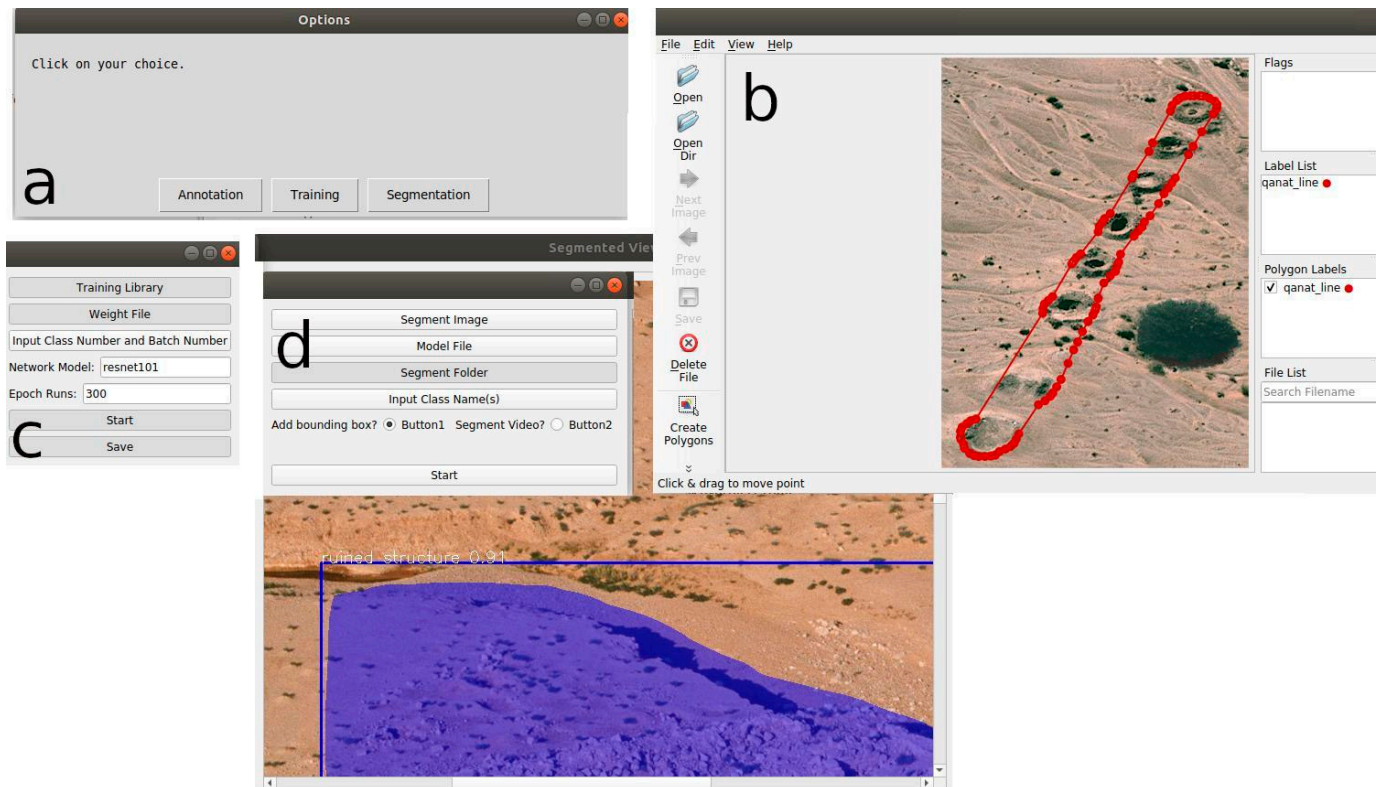
Figure 1. Controller module (**a**) showing options when selecting annotation (**b**), training (**c**), and segmentation (**d**).

After training is launched, .h5 models will result from training runs, with improved outputs, that is, diminishing validation loss, from each epoch placed in the model_dir folder. The best model can be used to conduct segmentation on either image files or videos. Going back to the controller.py module, and after a user selects segmentation, a small window opens to enable user choices for segmentation. Users can select a single image or video file. A choice for the model can be made, which enables the user to select which deep learning model (.h5 file) output to use. One option is to select a segmentation folder, which is a group of images to segment rather than a single image. If a directory for segmentation is chosen, then the folder will be segmented rather than a single image. The classes, which are the names of the annotated classes used, are also inputted (e.g., ruined structures, *qanats*, and mounded sites). An option to have a bounding box on images is given as well. If a single video is chosen to segment, which can be chosen using the 'Segment Image' option, then indicating that this is a video option can be done in the radio button option ('Segment Video?'). After these options, a user can then start the operation using the start button, which launches the custom_segmentation.py module in the segmentation folder. If a single image is selected, another window will open that shows the segmented image to users. Outputs from segmentation include a segment_data.csv file, which provides segmentation summary data found in the output_segmentation folder that contains the name of the

segmented class found in the image, the locations of the bounding box coordinates, and the score of the identified instance class. The segmented image(s) will be located in the output_segmentation folder.
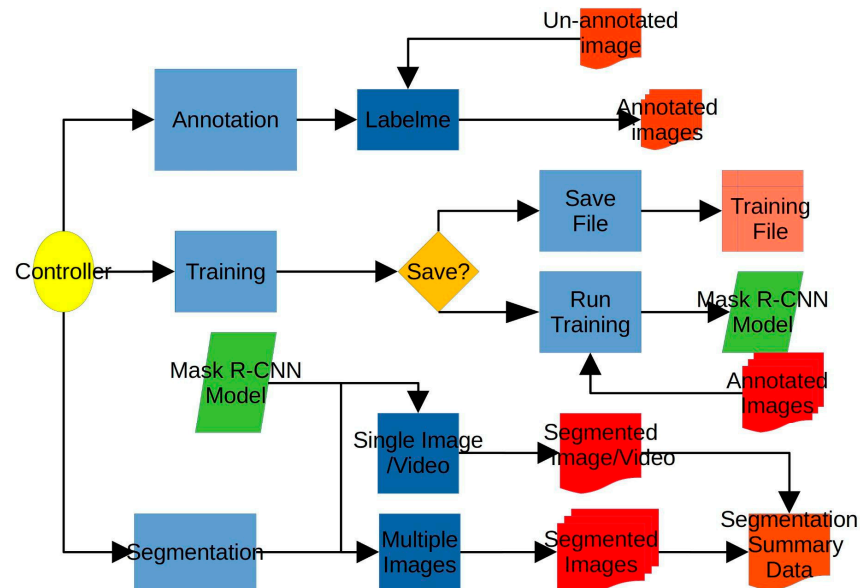

Figure 2. Workflow for the software options users can apply.

3.3. Training and Testing Data

In applying a supervised approach such as this Mask R-CNN implementation, training imagery is needed to create an output model used in image segmentation. As our main region of focus and knowledge is in the Middle East, we focused on obtaining imagery from countries in that region, including Jordan, Iraq, the UAE, Oman, and Iran. In addition to UAV imagery, regular aerial photographs from archaeological sites from the Gulf countries are collected. We worked with multiple data providers and collected 2814 images from contributors who utilize UAVs in fieldwork, but many of the data are not used in training and/or segmentation due to varying degrees of data quality or not directly informing on features. In addition, the Aerial Photographic Archive for Archaeology in the Middle East (APAAME) is used in our project training to supplement our data deficiencies, where users can find watermarked images of over 11,500 images on the online repository [41]. We ordered some non-watermarked imagery from APAAME, as well; we found some of the watermarked images suitable in training, but it also interfered in training in some cases and may have affected training results. Ideally, raw images should be used. Additionally, the APAAME images were taken by helicopter, but the angles and views are somewhat comparable to many UAV images. Three categories of sites were selected, which are called 'ruined structure', '*qanat*', and 'mounded sites' (Figure 3). Overall, we used 286 total images (124 ruined structures, 68 *qanats*, 94 mounded sites) to conduct training, with often multiple features per image used, but we hope to increase this number as we evaluate more images. Additionally, other forms of data, such as LIDAR, can be used, but non-optical data are still not very commonly used in UAV imagery for archaeology. Therefore, we focused on optical data for now. From the images used, 33 were from the data providers, while the rest were from APAAME Additionally, 71 (31 ruined structures, 18 *qanats*, 22 mounded sites) images were used for model validation. All of these except 3 were from APAAME.
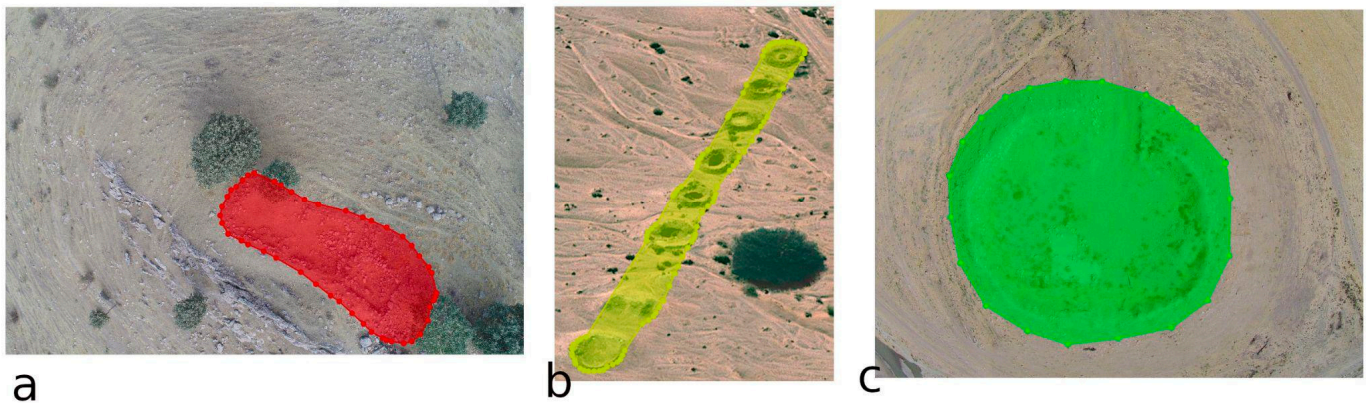
Figure 3. The three categories, that is 'ruined structure' (**a**), '*qanat*' (**b**), and 'mounded sites' (**c**), annotated using LabelMe via our developed tool.

Generally, the three categories correspond to three distinct shapes in features, including square, linear, and elliptical shapes evident in such features. For ruined structures, annotated areas encompass walls, structures, and even materials, that is, stones, visible in archaeological sites. These features are evident on the surface through their remains, usually consisting of stones. Archaeologically, it is not always easy to identify if ruined structures are buildings or burials; therefore, we decided to make this a more generic category encompassing multiple types of features. For *qanats*, these features are linearly distinguished with circular-shaped access holes; both the access holes and linear representation of the aligned holes allow this feature to have clearly different annotations from other features. Because we can clearly separate this category from 'ruined structure', it allowed us to create this feature as a second category. As for mounded sites, their generally oval or circular representation above ground is the main distinguishing feature that separates them from the other categories. Such features are common throughout the Middle East, where these features can be described as types of artificial hills. Figure 3 shows the three types we consider clearly distinct. In some of the cases, sites have a combination of these features, such as a ruined structure and mounded site together spread across a site. This possibility could be captured in data we train.

4. Some Initial Results from Sample Data

We deployed the resnet101 neural network on GPU nodes on an HPC cluster to conduct training. As our work is ongoing, and we hope to collect more relevant samples, a limited number of outputs are provided here. We experimented with annotations to see if training on simple forms of data or simple features, such as individual walls or even stones, could improve results. For now, we limit discussion to the feature types we trained and evaluated. Examples from the three categories indicated above are provided with the final DL model deployed on 26 (12 ruined structure, 7 *qanat*, and 7 mounded sites categories) unannotated images. Ideally, we would conduct a formal test using sample data to test the precision and recall (F1 score) capabilities of the approach. Currently, the best validation loss achieved is just under 2.0 (1.84) and a mean average precision of 0.2, reflecting room for improvement that can be addressed through increasing training data. Results comparing imagery tested using the software for the three categories were about 25–35% accurate when compared to manually identified archaeological features. While ruined structures and mounded sites showed generally good identification, *qanats* showed more poor results that led to a lower overal score. This reflects the rate in which features could be identified using the Mask R-CNN approach versus manual identification. The exact rate is hard to determine as some features overlapped in area. Our current data can be found in the Supplementary Materials.

The first feature category with segmented data is ruined structures, reflecting visible surface ruins related to different archaeological features, including burials, ruined buildings,

or large concentrations of stone scatters. Figures 4 and 5 demonstrate raw and segmented images, respectively. From these images, it is clear that stone-built structures are segmented by the resulting Mask R-CNN model, but errors are noticed in that, sometimes, not all areas of a given feature are segmented (Figure 5b), or identified features are missed (e.g., Figure 5c, bottom left is a small feature). Nevertheless, the results demonstrate segmentation is on the right track, with classification score at about 0.7 from a scale of 0.0–1.0 in these cases, as the ruined structure class is identified, and the identification is mapped to part or even most of the features in the images.

The second feature type discussed is *qanats*, or ancient water channels, that typically have above-ground access holes. This feature was clearly distinct from ruined structures and is commonly found in the Middle East. Figure 6 reflects raw data and Figure 7 segmented results. From these, it is clear *qanats* were often missed, and only some of the feature's access holes were segmented. Some mis-identification also occurred (e.g., mounded sites instead of *qanat* identification); however, ruined structures identified by our software often compared well to manually identified ruined structures, such as in Figure 5. Even though *qanats* are relatively more simple in shape, which should make them more identifiable, this category did have fewer training data than other categories. This likely led to the poorer results relative to the ruined structure category. The simpler shapes, however, may mean less training data will ultimately be needed than other categories.

Figure 4. Example of raw ruined structure features (examples are from APAMME; samples 4088792676_8ccc54c75d_h (**a**), 15532537376_7fab380e39_o (**b**), and 5123243570_0933def600_o (**c**).
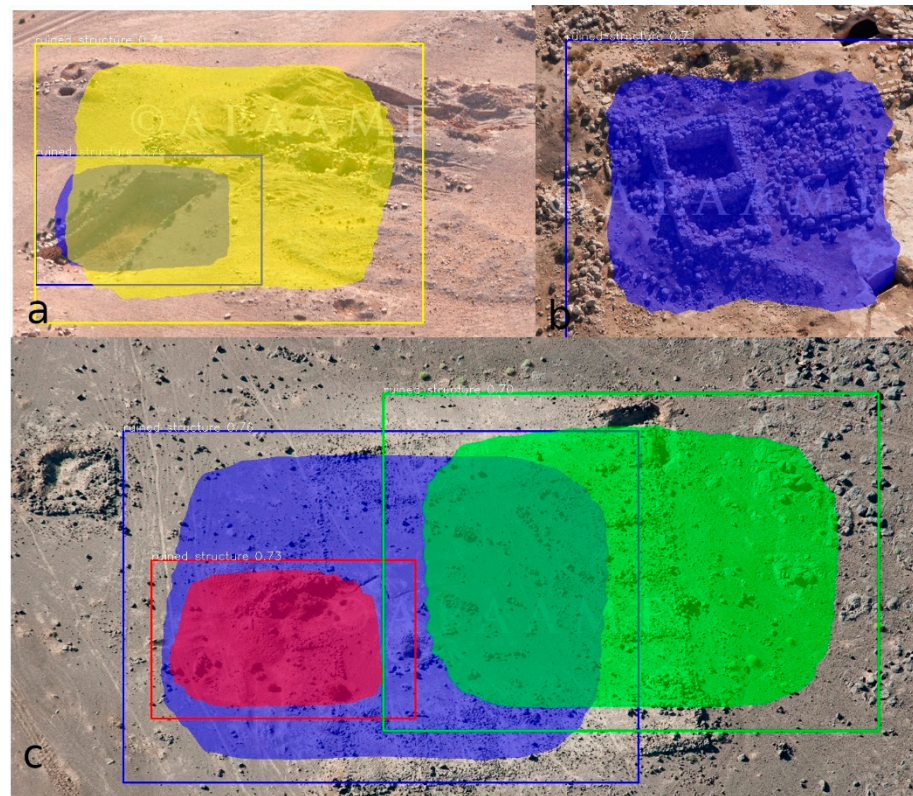
Figure 5. Segmented outputs from Figure 4′s ruined structure category using the developed tool. Items (**a**–**c**) correspond to Figure 4.

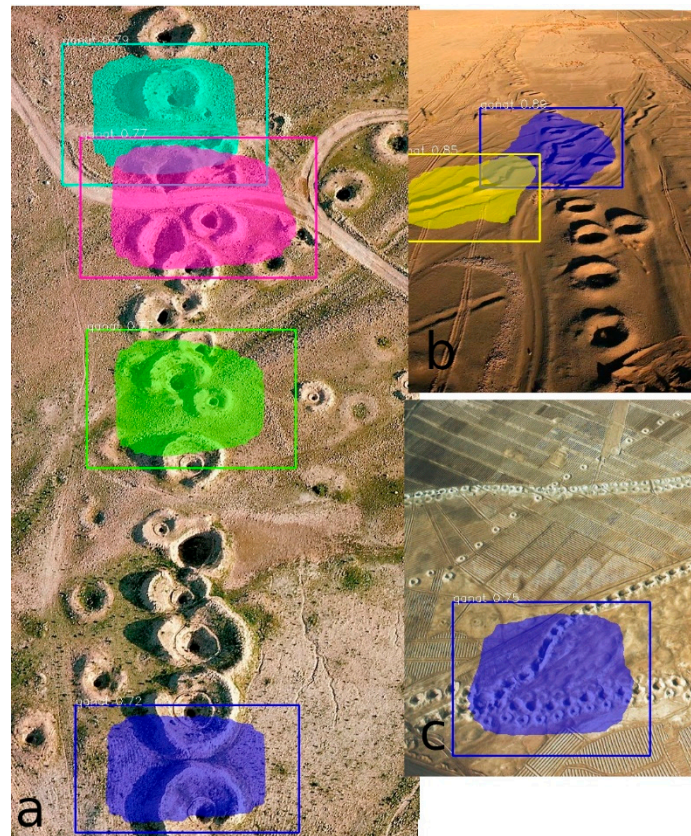**Figure 6.** Raw images demonstrating *qanats* in (**a**–**c**).

Figure 7. Segmented *qanats* from Figure 6. Items (**a**–**c**) correspond to Figure 6.

The third category identified in UAV imagery is mounded sites, perhaps among the most common type of archaeological settlement feature in the Middle East. Figure 8a–c are examples from this category, with the resulting segmentation outputs (d–f) shown. In this case, segmentation was able to identify mounded and ruined structure sites, including when they are represented together. The most common error in this test was not capturing the entire mound, but relatively, this was one of the better categories tested. Flatter mounds, that is, those with less than 10° slope and without high contrast from the background, are harder to capture, but in some of the test's segmentation was able to capture relatively low mounded sites (e.g., Figure 8e,f). Identified mounds did have scores running over 0.8, indicating somewhat greater confidence than in some of the other categories, as these features are generally easier to train for and detect.

Part of the limitation for features we have chosen is the lack of diversity of types within the categories. While clearly more training data could improve this, how many examples are needed is unclear. From analyzing other works, and given the accuracy scores achieved so far, we suggest training will require something on the order of 200–500 images for ruined structures, with potentially multiple features per image, but likely less training data for *qanat* and mounded sites categories (e.g., 100–150). This estimate is also based on the relatively more regular appearance of *qanats* and mounded sites versus ruined structure sites. This is aiming to achieve accuracy over 0.9 and F1 scores that are about the same or more for each category type. The estimate is based on similar research that attempts to identify complex categories using training data [43].
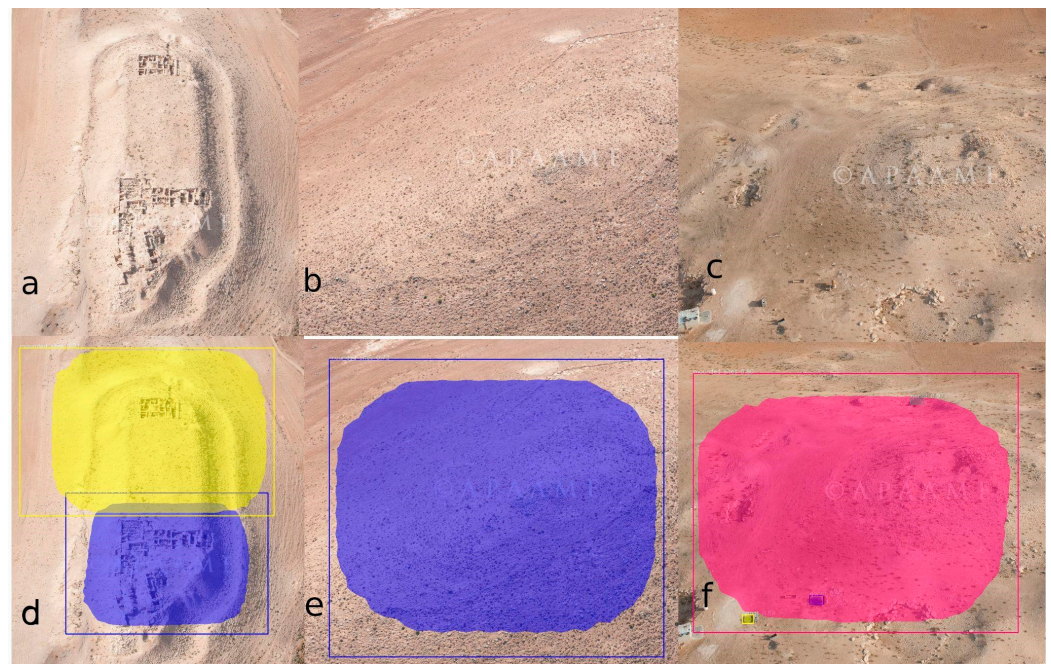
Figure 8. Raw (**a**–**c**); APAAME 20151014_RHB, 1553265296_0260edfba6_o, and 15639923240_9f630e902b_o), and segmented (**d**–**f**) mounded sites.

5. Discussion and Conclusions

5.1. Discussion

This work presents a call to researchers to better publish their UAV imagery along with their work as well as openly share data to facilitate the training of large neural network models that enable automated feature detection useful in archaeology. The brief review presented some example projects, such as VisDrone or DOTA, that have large repositories useful for identifying many different UAV-viewed objects, where these repositories are used to train large neural network models. By creating large repositories of hundreds of thousands or more images, researchers in ecology and urban studies, for instance, have been able to provide more useful results (e.g., high F1 scores such as over 0.9). Archaeological features are highly diverse, indicating that large repositories with varied and many examples of data will be needed to enable comparable results. While ideally similar large repositories would exist for archaeology, even having smaller repositories associated with field projects or publications would also greatly improve data access and at least regional approaches for automating UAV-based feature identification in the near-term. Certainly, for the long-term could smaller repositories be incorporated together to create broader coverage of archaeological features, as well. Increasingly, we are seeing remote sensing applications automating the process of feature discovery using various machine learning techniques (e.g., see [32,33]). This is useful not only in researching the past landscape but also in protecting heritage. Such work has been possible because of image repositories that make feature detection possible and easier. For UAV imagery, while such work is beginning to be carried out (e.g., [44]), current outputs are limited to a small range of objects identified on imagery. In regards to data sharing, there are many free tools, including sites such as figtree, Mendeley Data [45], and the Open Science Framework [46] that make it easier for researchers to share data. Such repositories could provide a solution for research projects with limited funding for large, cloud-based repositories.

Ultimately, as UAVs increase their range and flight time, more data will be captured, often to the point that might make it difficult for researchers to manually identify all archaeological features. This automation process has been increasingly applied to satellite-based imagery, but some forms of archaeological features are better identified using UAV imagery tools that are best suited for capturing data in current conditions, such as new features being

exposed, or even capturing more subtle features that are not easy to identify from other remote sensing platforms. For instance, it has been shown that small, seemingly random stone scatters detected using UAV imagery could be better at indicating archaeological sites than pottery scatters studied in survey [2]. Automating such and other discoveries will greatly depend on improved data capture, with a greater variety of examples of data. Trained models will help researchers speed up the recording and identification of archaeological features, while trained models could potentially capture data not readily apparent with only manual visual inspection.

We have attempted to contribute to the wider issue of improving data capture for UAV imagery through an ongoing project that has recently been funded by the UAE. A key contribution from this work is innovation in remote sensing software that could benefit other researchers. Specifically, we created software that enables key steps in the preparation, that is, annotation, training, and segmentation of imagery used in a Mask R-CNN deep learning algorithm that auto-detects archaeological features. Open source tools that enable the processing and application of DL do not currently exist in a simple-to-use GUI for heritage and archaeology specialists. Tools such as Picterra [47] do exist that allow similar possibilities as presented here. However, we also believe it is important to create open source and free software for researchers so as to avoid future possibilities where corporate-based software may become payment only, as this could limit research, particularly in countries where resources are limited. Additionally, to broaden researcher participation, all of the key steps needed to annotate, train, and segment data should ideally be accomplished through minimal computer code requirements and, ideally, via a GUI. We have now enabled this in our contribution, providing benefit to researchers who are interested in applying DL to their UAV imagery cases. For our test cases presented here, we have deliberately limited the classes to three types of features, namely ruined structure, *qanat*, and mounded sites. These features can be detected, but not in all cases, and mistakes are evident. Our current outputs demonstrate utility in detecting archaeological features, where outputs are achieved with no coding and use of the GUI-built tool; however, results could be improved with more training and testing data. We suggest that complex shapes such as ruined structures may require a large number of training examples, perhaps on the order of more than 200 example images depending on the variety and complexity of shapes as well as number of features per image, to best represent diversity in given designated categories.

5.2. Conclusion and Future Direction

Deep learning is becoming increasingly common to a variety of fields applying remotely sensed data, but its benefits are constrained by training data, which is best addressed through greater data sharing. Our approach has been to at least facilitate some of the software limitations encountered by creating a simple GUI-based application using a powerful deep learning library (PixelLib) as a key component while bringing together tools for annotation (LabelMe) and outputs, including .csv, imagery, and GUI-based views, that enable analysis to be made. The software is freely shared and provided with an MIT open source license. The software will also be updated as it is improved and new features are added, including other algorithms for automated feature detection. While common optical camera data are the most common form of data and these data are what we have selected, one can extend this to other forms of sensor gathering. Thermal imagery is one area we think has greater potential, particularly in the hot climate and stone features common to many areas in the Middle East and elsewhere. LIDAR data have also been increasingly used, and more data captured through this technique in UAVs means we should have enough data to created automated classification methods, as well. These will be other areas we focus on, specifically imagery using such capabilities, in the near future and once we have more data from such imagery. Using long-distance UAVs will also enable us to create large mapped survey areas with detected archaeological features. One possibility we will consider to increase training data is to use synthetic data, including through generative

adversarial networks (GANs), where other projects with comparable limited data have demonstrated utility in creating training sets from existing sources but deploying variations derived from original data [48]. This could come in the form of altering angles or combining imagery to create something similar to what might be encountered elsewhere. It is possible simpler shapes, such as stones, could be chosen, which would also increase training data available and making categorization perhaps simpler. Other types of features, when data are obtained for training, could also be researched. We will also incorporate algorithms, including random forest [49] or other DL techniques such as those discussed in Osco et al. [10], as an alternative approach to the Mask R-CNN method deployed here. We plan to similarly facilitate the application of these approaches using minimal code or GUI-based methods. Such methods could potentially make it easier with smaller training data requirements. We also intend to focus more on the types of features that are difficult to detect using conventional methods or even visual analysis, such as flat archaeological sites with seemingly random stone scatters.

References

1. Themistocleous, K. The Use of UAVs for Cultural Heritage and Archaeology. In *Remote Sensing for Archaeology and Cultural Landscapes*; Hadjimitsis, D.G., Themistocleous, K., Cuca, B., Agapiou, A., Lysandrou, V., Lasaponara, R., Masini, N., Schreier, G., Eds.; Springer Remote Sensing/Photogrammetry; Springer International Publishing: Cham, Switzerland, 2020; ISBN 978-3-030-10978-3.
2. Altaweel, M.; Squitieri, A. Finding a Relatively Flat Archaeological Site with Minimal Ceramics: A Case Study from Iraqi Kurdistan. *J. Field Archaeol.* **2019**, *44*, 523–537. [CrossRef]
3. Brutto, M.L.; Borruso, A.; D'Argenio, A. Uav Systems for Photogrammetric Data Acquisition of Archaeological Sites. *Int. J. Herit. Digit. Era* **2012**, *1*, 7–13. [CrossRef]
4. Nikolakopoulos, K.G.; Soura, K.; Koukouvelas, I.K.; Argyropoulos, N.G. UAV vs Classical Aerial Photogrammetry for Archaeological Studies. *J. Archaeol. Sci. Rep.* **2017**, *14*, 758–773. [CrossRef]

5.   Fernández-Hernandez, J.; González-Aguilera, D.; Rodríguez-Gonzálvez, P.; Mancera-Taboada, J. Image-Based Modelling from Unmanned Aerial Vehicle (UAV) Photogrammetry: An Effective, Low-Cost Tool for Archaeological Applications: Image-Based Modelling from UAV Photogrammetry. *Archaeometry* **2015**, *57*, 128–145. [CrossRef]

6.   Orengo, H.A.; Garcia-Molsosa, A.; Berganzo-Besga, I.; Landauer, J.; Aliende, P.; Tres-Martínez, S. New Developments in Drone-based Automated Surface Survey: Towards a Functional and Effective Survey System. *Archaeol. Prospect.* **2021**, *28*, 519–526. [CrossRef]

7.   Brooke, C.; Clutterbuck, B. Mapping Heterogeneous Buried Archaeological Features Using Multisensor Data from Unmanned Aerial Vehicles. *Remote Sens.* **2019**, *12*, 41. [CrossRef]

8.   Figshare. 2021. Available online: https://figshare.com/ (accessed on 8 January 2021).

9.   Yao, C.; Luo, X.; Zhao, Y.; Zeng, W.; Chen, X. A Review on Image Classification of Remote Sensing Using Deep Learning. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 1947–1955.

10.   Osco, L.P.; Marcato Junior, J.; Marques Ramos, A.P.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]

11.   Yu, H.; Li, G.; Zhang, W.; Huang, Q.; Du, D.; Tian, Q.; Sebe, N. The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline. *Int. J. Comput. Vis.* **2020**, *128*, 1141–1159. [CrossRef]

12.   Penatti, O.A.B.; Nogueira, K.; dos Santos, J.A. Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains? In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015; pp. 44–51.

13.   Zhu, P.; Wen, L.; Du, D.; Bian, X.; Hu, Q.; Ling, H. Vision Meets Drones: Past, Present and Future. *arXiv* arXiv:2001.06303, 2020.

14.   Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3974–3983.

15.   Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2321–2325. [CrossRef]

16.   Zhao, L.; Tang, P.; Huo, L. Feature Significance-Based Multibag-of-Visual-Words Model for Remote Sensing Image Scene Classification. *J. Appl. Remote Sens.* **2016**, *10*, 035004. [CrossRef]

17.   Sheng, G.; Yang, W.; Xu, T.; Sun, H. High-Resolution Satellite Scene Classification Using a Sparse Coding Based Multiple Feature Combination. *Int. J. Remote Sens.* **2012**, *33*, 2395–2412. [CrossRef]

18.   dos Santos, A.A.; Marcato Junior, J.; Araújo, M.S.; Di Martini, D.R.; Tetila, E.C.; Siqueira, H.L.; Aoki, C.; Eltner, A.; Matsubara, E.T.; Pistori, H.; et al. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors* **2019**, *19*, 3595. [CrossRef] [PubMed]

19.   Lobo Torres, D.; Queiroz Feitosa, R.; Nigri Happ, P.; Elena Cué La Rosa, L.; Marcato Junior, J.; Martins, J.; Olã Bressan, P.; Gonçalves, W.N.; Liesenberg, V. Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. *Sensors* **2020**, *20*, 563. [CrossRef] [PubMed]

20.   Wiseman, J.; El-Baz, F. (Eds.) *Remote Sensing in Archaeology: Interdisciplinary Contributions to Archaeology*; Springer: New York, NY, USA, 2007; ISBN 978-0-387-44453-6.

21.   Parcak, S.H. *Satellite Remote Sensing for Archaeology*; Routledge: London, UK; New York, NY, USA, 2009; ISBN 978-0-415-44877-2.

22.   *Satellite Remote Sensing: A New Tool for Archaeology*; Lasaponara, R.; Masini, N. (Eds.) Remote Sensing and Digital Image Processing; Springer: Dordrecht, The Netherlands; New York, NY, USA, 2012; ISBN 978-90-481-8800-0.

23.   Hritz, C. Tracing Settlement Patterns and Channel Systems in Southern Mesopotamia Using Remote Sensing. *J. Field Archaeol.* **2010**, *35*, 184–203. [CrossRef]

24.   Bewley, B.; Bradbury, J. The Endangered Archaeology in the Middle East and North Africa Project: Origins, Development and Future Directions. *Bull. Counc. Br. Res. Levant* **2017**, *12*, 15–20. [CrossRef]

25.   Hammer, E.; Seifried, R.; Franklin, K.; Lauricella, A. Remote Assessments of the Archaeological Heritage Situation in Afghanistan. *J. Cult. Herit.* **2018**, *33*, 125–144. [CrossRef]

26.   Caspari, G.; Crespo, P. Convolutional neural networks for archaeological site detection -Finding "princely" tombs. *J. Archaeol. Sci.* **2019**, *110*, 104998. [CrossRef]

27.   Guyot, A.; Hubert-Moy, L.; Lorho, T. Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques. *Remote Sens.* **2018**, *10*, 225. [CrossRef]

28.   Verschoof-van der Vaart, W.B.; Lambers, K.; Kowalczyk, W.; Bourgeois, Q.P.J. Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from The Netherlands. *IJGI* **2020**, *9*, 293. [CrossRef]

29.   Davis, D.S.; Lundin, J. Locating Charcoal Production Sites in Sweden Using LiDAR, Hydrological Algorithms, and Deep Learning. *Remote Sens.* **2021**, *13*, 3680. [CrossRef]

30.   Megarry, W.; Cooney, G.; Comer, D.; Priebe, C. Posterior Probability Modeling and Image Classification for Archaeological Site Prospection: Building a Survey Efficacy Model for Identifying Neolithic Felsite Workshops in the Shetland Islands. *Remote Sens.* **2016**, *8*, 529. [CrossRef]

31. Biagetti, S.; Merlo, S.; Adam, E.; Lobo, A.; Conesa, F.C.; Knight, J.; Bekrani, H.; Crema, E.R.; Alcaina-Mateos, J.; Madella, M. High and Medium Resolution Satellite Imagery to Evaluate Late Holocene Human–Environment Interactions in Arid Lands: A Case Study from the Central Sahara. *Remote Sens.* **2017**, *9*, 351. [CrossRef]

32. Menze, B.H.; Ur, J.A. Mapping Patterns of Long-Term Settlement in Northern Mesopotamia at a Large Scale. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E778–E787. [CrossRef]

33. Orengo, H.A.; Conesa, F.C.; Garcia-Molsosa, A.; Lobo, A.; Green, A.S.; Madella, M.; Petrie, C.A. Automated Detection of Archaeological Mounds Using Machine-Learning Classification of Multisensor and Multitemporal Satellite Data. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 18240–18250. [CrossRef]

34. Bonhage, A.; Eltaher, M.; Raab, T.; Breuß, M.; Raab, A.; Schneider, A. 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.* **2021**, *28*, 177–186. [CrossRef]

35. Soroush, M.; Mehrtash, A.; Khazraee, E.; Ur, J.A. Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. *Remote Sens.* **2020**, *12*, 500. [CrossRef]

36. Orengo, H.A.; Garcia-Molsosa, A. A Brave New World for Archaeological Survey: Automated Machine Learning-Based Potsherd Detection Using High-Resolution Drone Imagery. *J. Archaeol. Sci.* **2019**, *112*, 105013. [CrossRef]

37. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.

38. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]

39. PixelLib. 2021. Available online: https://pixellib.readthedocs.io/en/stable/ (accessed on 8 April 2021).

40. TensorFlow. 2021. Available online: https://www.tensorflow.org/ (accessed on 8 April 2021).

41. Labelme. 2021. Available online: https://github.com/wkentaro/labelme (accessed on 8 April 2021).

42. COCO. 2021. Available online: https://cocodataset.org/#home (accessed on 8 September 2021).

43. Liang, M.; Palado, G.; Browne, W.N. Identifying Simple Shapes to Classify the Big Picture. In Proceedings of the 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), Dunedin, New Zealand, 2–4 December 2019; pp. 1–6.

44. Aerial Photographic Archive in the Middle East (APAAME). 2021. Available online: http://www.apaame.org/ (accessed on 8 December 2021).

45. Mendeley Data. 2021. Available online: https://data.mendeley.com/ (accessed on 8 December 2021).

46. Open Science Framework. 2021. Available online: https://osf.io/ (accessed on 8 December 2021).

47. Picterra. 2021. Available online: https://picterra.ch/ (accessed on 8 December 2021).

48. Cannan, L.; Robinson, B.M.; Patterson, K.; Langford, D.; Diltz, R.; English, W. Synthetic AI Training Data Generation Enabling Airfield Damage Assessment. In Proceedings of the SPIE Defense + Commercial Sensing, Online Only, FL, USA, 12–17 April 2021; p. 24.

49. Akar, Ö.; Güngör, O. Classification of Multispectral Images Using Random Forest Algorithm. *J. Geod. Geoinf.* **2012**, *1*, 105–112. [CrossRef]