Survey of Intelligent Archaeology

Lin Yuhang, Peng Shubo

***Abstract*—Recently more and more AI techniques are used in archaeology area, one of which is machine learning (ML). Application of ML can be divided into two main categories: object classification and site searching. Classification can not only be used to classify fragments, like pottery or papyrus, but also help to merge messy fragments into complete patterns. Site searching is a new topic raise recently thanks to the popularity of satellites, whose generated systematized data offers premises for ML. But ML still has some shortage, most of which limited by nature of archaeology. In future, ML maybe more used in identification of archaeological sites, or illegal treatment about art products.**

***Index Terms*—Archaeology, machine learning, classification, site searching.**

I. Introduction

ARTIFICIAL Intelligence (AI for short) is a new study area rising in recent centuries, and can be divided into several directions, like Machine Learning (ML) or Neural Network (NN). Five to ten years ago, they were concepts unknown to archaeologists. But now, AI is widely used, there are even sessions dedicated to AI at archaeological conferences.[10]

ML is programming allowing algorithm learning from data and adjust its parameters and then make predictions on new data. Objects must be quantify into digital data first and can be any type, like sonar data under water[9] or aerial laser scanning data [11]. ML uses mathematical techniques to analyze a set of already-classified objects and generate ”classifiers” for each category. In theory, objects in every category is identified from other categories in mathematics. In short, ML use math to classify quantifiable objects into different groups.[5]

Ml application can be divided into two main types: classify archaeological objects, and identify archaeological sites, both of which will be explained more detailed in II.

Deep learning(DL) a branch of ML, the biggest difference is that, ML needs human to appoint characteristics of object to be trained, while DL do not need pre-appointed characteristics. What DL processing is a large number of data, and DL can learn and choose significant characteristics by itself. DL will first determine which characteristics are related to its goal, then attempt to obtain accessible features on layer level, finally determine which accessible features match its goal best.

II. Application

DL has been successfully used in many applications. Among the DL methods, recurrent neural networks (RNNs) are good at dealing with sequential data as they take into account temporal information. RNNs have been applied in speech recognition to map acoustic sequences to phonetic sequences. RNNs have also been used in natural language processing to translate text from one language to another. Another famous method in DL is convolutional neural networks (CNNs). CNNs take into account spatial correlation among data points and hence perform well in image-based data. CNNs have been used in image classification, face recognition, scene labelling, and so on. DL methods are also used in the field of remote sensing. For example, Hu and Yuan used CNNs to extract digital terrain models (DTMs) and filter out non-ground points from airborne laser scanning (ALS) data, which was claimed that this method performs better than previous filtering methods[12].

In the papers we’ve read, we’ve found that all of them, without exception, use CNNs to study a specific thing. This is because the things they studied were all based on image data, in which CNNs performs better than RNNs.

We divide the papers that use CNNs into two types: one is to classify archaeological objects like papyrus, potteries, and so on, the other is to detect objects in archaeological sites.

A. *classify archaeological objects*

- In the paper[1], a method is proposed for matching and assembling pairs of ancient papyrus fragments containing mostly unknown scriptures. This task, which is assembling fragments in a puzzle-like manner into a composite picture, plays an important role in the archaeology, because it can help historic artifacts to reconstruct archaeological objects for research. The proposed method is to use image processing and machine learning techniques to identify matching fragments, and then support the quick and automated classification of matching pairs of papyrus fragments as well as the geometric alignment of the pairs against each other. The algorithm was trained on a batch of fragments which was excavated from the Dead Sea caves and is dated circa the 1st century BCE. The algorithm shows excellent results on a validation set which is of a similar origin and conditions. Then the algorithm was used to against a real-life set of fragments for with no prior knowledge or labeling of matches. This test batch is considered extremely challenging due to its poor condition and the small size of its fragments. Evidently, numerous researchers have tried seeking matches within this batch with very little success. The algorithm performance on this batch was sub-optimal, returning a relatively large ratio of false positives. However, the results showed that this algorithm eliminated 98% of the possible matches thus reducing the amount of work needed for manual inspection, which means this algorithm was quite useful. Indeed, experts that reviewed the results have identified some positive matches as potentially true and referred them for further investigation. A glimpse of this project is shown in Figure 1.

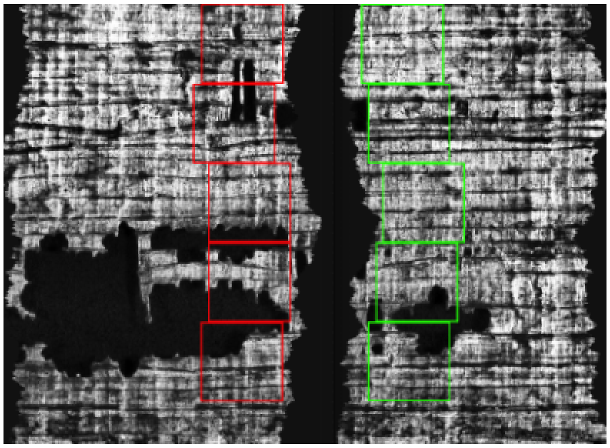

Fig. 1: An example of two adjacent artificially torn fragments with a set of candidate squares to be used in the matching phase.

Fig. 2: Explainability heatmaps for predictions of the classes “architecture” (left) and “façades” (right) by the Phase 1 classifier.

- In the paper[13], the project developed at Dumbarton Oaks—a research institute and library, museum, and historic garden affiliated with Harvard University and located in Washington, DC— focused on a collection of 10,000 images of Syrian monuments in the institution’s Image Collections and Fieldwork Archives (ICFA). Drawing on that project, as well as the broader landscape of AI-based categorisation efforts in the fields of art and architecture, authors’ insights on the potential of AI to facilitate and enhance archival image access and recording will be shaerd. Many of the Syrian sites in the Dumbarton Oaks collection have been inaccessible to researchers and the public for over a decade and/or have been damaged or destroyed. For Dumbarton Oaks, the primary goal was to explore whether AI can improve the speed and efficiency of sharing collections and allow for more sophisticated curation by subject experts who, thanks to automation, would be relieved of the burden of rote processing. The methods and techniques explored included multi-label classification, multi-task classification, unsupervised image clustering, and explainability. A glimpse of this project is shown in Figure 2.
- The main contribution in this paper[10] is the completion of the project called ArchAIDE. This project realised an AI-based application to recognise archaeological pottery. Pottery is of paramount importance for understanding archaeological contexts. However, recognition of ceramics is still a manual, time-consuming activity , reliant on analogue catalogues. The project developed two complementary machine-learning tools to propose identifications based on images captured on-site, for optimising and economising this process, while retaining key decision points necessary to create trusted results. One method relies on the shape of a potsherd; the other is based on decorative features. For the shape-based recognition, a novel deep-learning architecture was employed, integrating shape information from points along the inner and outer profile of a sherd. The decoration classifier is based on relatively standard architectures used in image recognition. In both cases, training the algorithms meant facing challenges related to real-world archaeological data: the scarcity of labelled data; extreme imbalance between instances of different categories; and the need to take note of minute differentiating features. Finally , the creation of a desktop and mobile application that integrates the AI classifiers provides an easy-to-use interface for pottery classification and storing pottery data.
  A glimpse of this project is shown in Figure 3.
  In particular, since this project ArchAIDE is the most classic and mature one among all paper, the more detailed and overall explanation will be given in III.

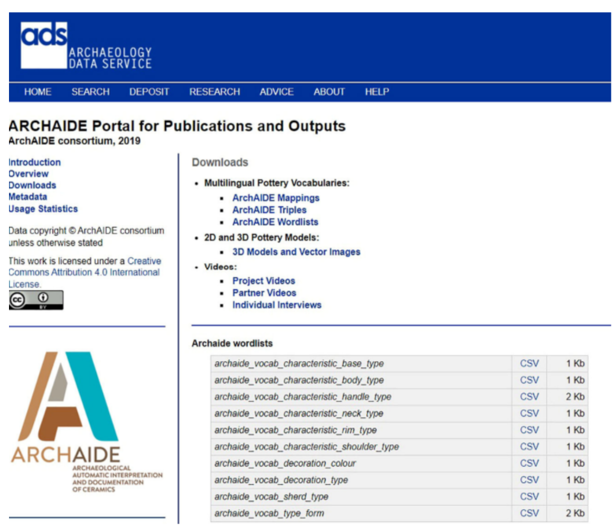

Fig. 3: The ArchAIDE portal is available at the Archaeology Data Service of the University of York.

- The paper[15] indicate that deep learning with CNNs is a highly accessible and effective method for classifying ceramic fabrics based on images of petrographic thin sections and that it can likely be applied on a larger scale. Classification of ceramic fabrics has long held a major role in archaeological pursuits. It helps answer research questions related to ceramic technology, provenance, and exchange and provides an overall deeper understanding of the ceramic material at hand. One of the most effective means of classification is through petrographic thin section analysis. However, ceramic petrography is a difficult and often tedious task that requires direct observation and sorting by domain experts. In this paper, a deep learning model is built to automatically recognize and classify ceramic fabrics, which expedites the process of classification and lessens the requirements on experts. The samples consist of images of petrographic thin sections under cross-polarized light originating from the Cocal-period (AD 1000-1525) archaeological site of Guadalupe on the northeast coast of Honduras. Two CNNs, VGG19 and ResNet50, are compared against each other using two approaches to partitioning training, validation, and testing data. The technique employs a standard transfer learning process whereby the bottom layers of the CNNs are pre-trained on the ImageNet dataset and frozen, while a single pooling layer and three dense layers are added to 'tune' the model to the thin section dataset. After selecting fabric groups with at least three example sherds each, the technique can classify thin section images into one of five fabric groups with over 93% accuracy in each of four tests. A glimpse of this project is shown in Figure 4.

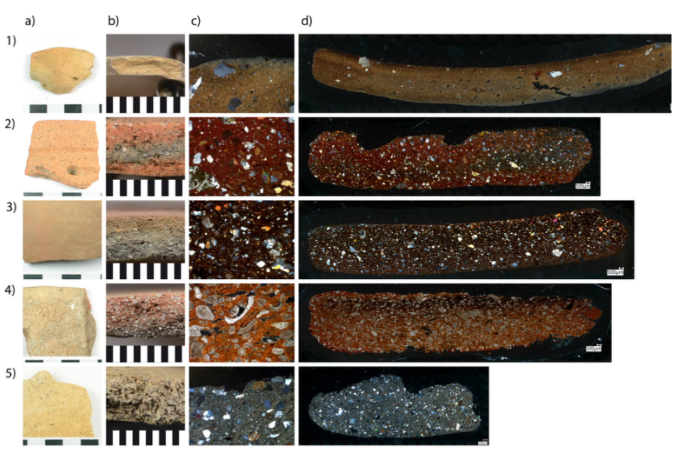

Fig. 4: Examples of the five ceramic fabric types analyzed.

- In the study[17], an alternate approach to archaeological typology which uses DL to classify digital images of decorated pottery sherds into an existing typological framework is presented. This study focuses on a specific kind of ancient painted pottery from the American Southwest, Tusayan White Ware (TWW), but it is believed that it has broader implications for a wide range of geographical settings and artifact types. The results show that when properly trained, a deep learning model can assign types to digital images of decorated sherds with an accuracy comparable to, and sometimes higher than, four expert-level contemporary archaeologists. The technique also offers novel tools for visualizing both the importance of diagnostic design elements and overall design relationships between groups of pottery sherds. This method can objectively match a specific unclassified sherd image to its most similar counterparts through a search of thousands of digital photos. This discovery has important archaeological implications for analyzing time relationships, monitoring stylistic trends, reconstructing fragmentary artifacts, identifying ancient artisans, and studying the evolution and spread of ancient technologies and styles. It also shows how deep learning models can potentially supplement or supplant traditional typologies in favor of more direct groupings and comparisons of artifacts. A glimpse of this project is shown in Figure 5.

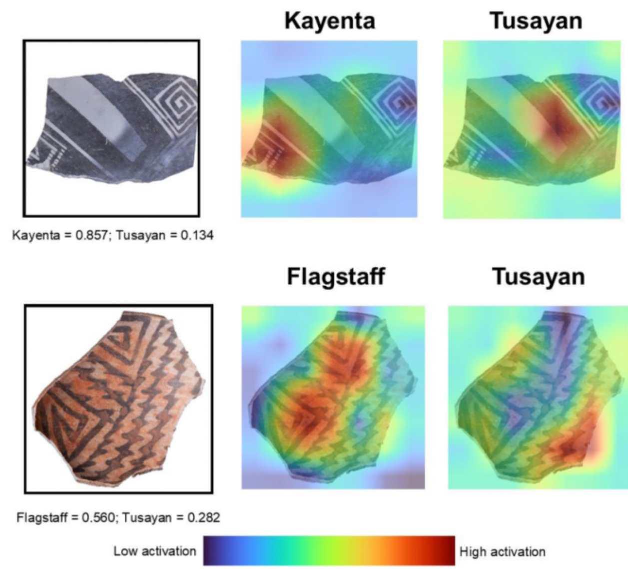

Fig. 5: Grad-CAM heat maps for TWW sherds, showing areas of high (red) and low (blue) model activation for a Kayenta sherd (top) and Flagstaff sherd (bottom). CNN-model-calculated type confidences shown below each sherd.

B. *detect objects in archaeological sites*

Airborne laser scanning (ALS) is of great use in collecting and documenting detailed measurements from an area of interest. However, it is time consuming for scientists to manually analyze the collected ALS data. One possible way to automate this process is using deep neural networks.

In the paper[14], a hierarchical CNN model is builded to detect objects in archaeological sites using digital terrain models (DTMs) generated from ALS data. The data is acquired from the Harz mining Region in Lower Saxony, where a high density of different archaeological monuments including the UNESCO world heritage site Historic Town of Goslar, Mines of Rammelsberg, and the Upper Harz Water Management System can be found. Objects to be detected are archaeological objects such as hollow ways, streams, pathways, lakes, streets, ditches, heaps, mining shafts, and more, but for this study, the model is fit to detect 4 classes of objects: natural streams, lakes, tracks, and an 'others' class which represents the rest of the objects for which enough labeled data is not available yet. To compare and validate the method in this paper, some experiments on the same data set using two existing deep learning models were conducted. The first model is VGG-16; an image classification network pretrained on ImageNet data. The second model is a stacked autoencoders model. The results of the classification as analyzed in this paper show that our model is suitably tuned for this task as it achieves the best classification accuracy of around 91 percent, compared to 88 percent and 82 percent accuracy by the pretrained and stacked autoencoders models, respectively. A glimpse of this project is shown in Figure 6 and 7.

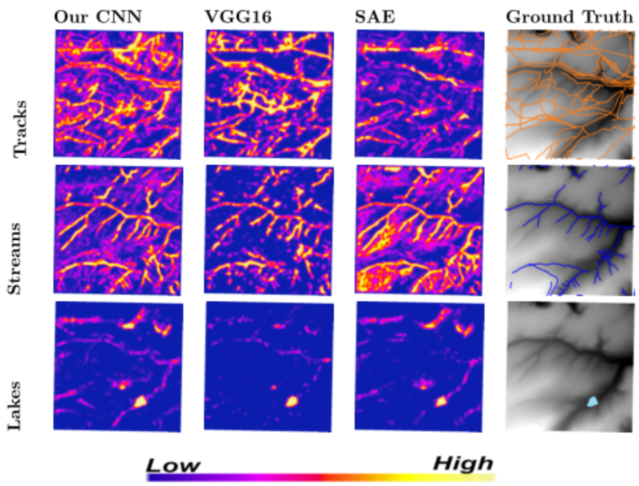

Fig. 6: Heat maps using filter size 48 x 48. Colors show the confidence of the models in detecting objects at that location.

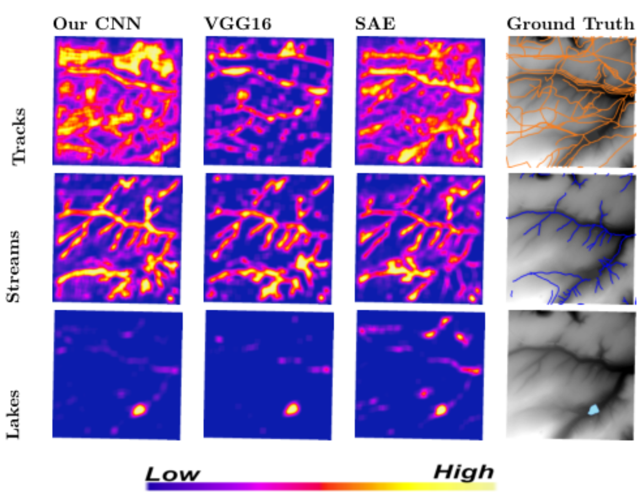

Fig. 7: Heat maps using filter size 98 x 98. Colors show the confidence of the models in detecting objects at that location.

C. More

We believe that there are many other studies using RNNs to study the meaning of certain patterns on ancient artifacts or to try to decipher ancient texts. But this is out of the scope of our survey.

III. ArchAIDE Project

To make things clear, a classic and mature classification project is briefly shown in this part. The reason why classification is chosen rather than searching sites is that, the number of paper about classification is much more than searching area. Among all applications in paper, the most classic and the most mature one is picked out here.

The name of project introduced here is “ArchAIDE”. The most remarkable characteristic of this project is that, it not only invent algorithm and do analysis on data from both view of shape and decorations, but also realize a system that could have a realworld implementation.

A. General Introduction

ArchAIDE project aims at optimizing the ceramic identification process. It developed two different deep neural networks to recognize pottery through images using a mobile device.The first network is specially used for image recognition, also called appearance-based recognition. The second network uses the shape of fragmentation to identify.

Unlike familiar worries about AI, ArchAIDE will not replace the knowledge of domain specialists. On the contrary, it put archaeologists’ role in the center of decision-making process in the identification workflow, which can be seen in III-C.

B. Materials

A correct result of classification relies on two parts: the label or the name of each category, and the available data for both shape-based and appearance-based recognition. But first, the class of relics should be determined.

1) Classes for training:

Among all categories of cultural relics, the project choose four realworld classes for training:

- Amphorae manufactured throughout the Roman world between the late 3rd century BCE and the early 7th century CE. (Figure 8a)
- Roman Terra Sigillata manufactured in Italy , Spain, and South Gaul between the 1st century BCE and the 3rd century CE.
- Majolica produced in Montelupo Fiorentino (Italy) between 14th and 18th century.
- medieval and post-medieval Majolica from Barcelona and Valencia. (Figure 8b)

(a) Roman amphorae (b) Majolica of Montelupo Fiorentino

Fig. 8: Material for training

2) Label:

To get correct and helpful label of categories, the project implements following systems:

- A digital comparative collection for pottery types, decorations, and stamps, combining digital collections, digitised paper catalogues, and data acquired through photo campaigns.
- A semi-automated system for paper catalogues' digitisation.
- A multilingual thesaurus of descriptive pottery terms, mapped to the Getty Art and Architecture Thesaurus, which includes French, German, Spanish, Catalan, Portuguese, English, and Italian.

The digital collections and paper catalogues to create digital comparative collections are from already-present databases.

The first one is "Roman Amphorae: a digital resource"[19], created by Simon Keay and David Williams of the University of Southampton and published as open data on the Archaeology Data Service, that includes the principal types of roman amphorae between the late 3rd century BCE and the early 7th century CE. The other one is "CERAMALEX" database[16], a proprietary database of the German and French Heritage 2021.

Limited by space, detailed principles of them will not be shown here.

3) Training images:

Multiple photo campaigns were also carried out in several archaeological warehouses, involving more than 30 different institutions in Austria, Italy , and Spain. Overall, 3498 sherds were photographed for training the shape-based recognition model. For appearance-based recognition, a dataset of 13,676 pictures was collected through multiple photography campaigns.

To offset disadvantages above in some term, each original image is scaled into four different sizes. On each scaled image, three versions are created: unflipped, horizontally flipped, and vertically flipped. All of these images are cropped, leaving just the central square. As a result, 12 images from each original one were obtained.

C. Method

The decoration of pottery fragments have higher priority than shape because decorations is more reliable than the shape of fragments. When decoration presents, decoration will be used to identify rather than appearance. Appearance-based recognition is used only when the pottery is undecorated.

1) Shape-based Recognition:

The recognition tool was designed as a two-phase process, where the classification algorithm was first developed on one dataset and then validated on other datasets for different types of pottery.

The dataset used 65 standardised toplevel classes defined in Conspectus catalogue[4]. 2D model are created from these drawings and photos taken in archaeological warehouses throughout Europe by extracting the profile of the entire vessel from 2D drawing. Then 2D model rotate around revolution axis to form 3D models.

Then many 3D planes are generated randomly. Computer will calculated how all the circles intersect the plane, connecting the intersection points from the circles along the profile to generate the fracture face. To create a more realistic synthetic fracture, we reduced its size to match real potsherds' dimensions. [7] The progress is shown in Figure 9.

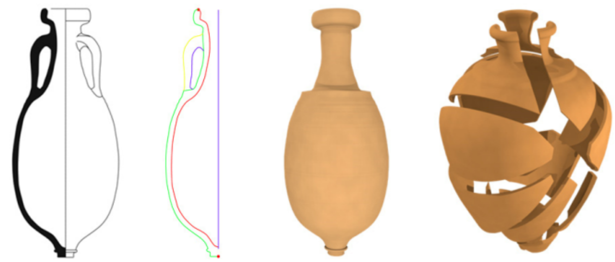

Fig. 9: Stpes from extracting profiles from 2D drawings, to creation of 3D models to be broken.

The network was trained based on the requirement to divide the inner and the outer profile of the sherd, the relevance of the position of the points along the profile outline, the intrinsic noise in the tracing procedure, and the requirement to overcome sub-optimal data acquisition processes[2], the example of which is in Figure 10.

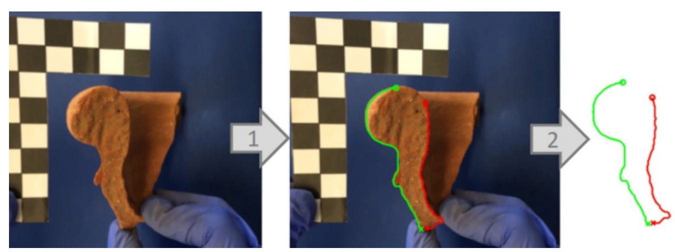

Fig. 10: Extraction of the outer (green) and inner (red) profiles from image.

To emphasize, the goal is not to increase accuracy of top-1 but that of top-K, which matches straight-forward sense. Because This fits with its function as a reference tool for pottery specialists who would be glad to evaluate a shortlist of results as part of the obligatory expert validation but would be disappointed to use a tool where the correct result is often completely omitted.

2) Appearance-based Recognition: It is find in experience that the most challenging factor that affected identification was varying illumination. So different white balance, brightness, and contrast adjustments are simulated. Each pixel's brightness is multiplied by a random factor to simulate different lighting level.

Moreover, the background and ruler varied significantly, leading to an inherent bias. The foreground was extracted automatically from the training images using the GrabCut algorithm to avoid this conditioning [18].

D. App Workflow

ArchAIDE also create a mobile application connected to AI classifiers to support archaeologists in recognizing potsherds during excavation and post-excavation analysis, with an easy-to-use interface.

Archaeologists take a picture of a potsherd and send it to the specifically trained classifier, which returns five suggested matches from the comparative collections. Once the correct type is identified, the information is linked to the photographed sherd and stored within a database that can be shared online. As shown in Figure 11.

The mobile application was also designed for allowing the use in lack of internet connectivity , such as in storehouses or remote rural areas. Application provides an area called "my sites", dedicated to registered users where it is possible to store information about sites and assemblages. When offline, the app permits storing new images of potsherds or browsing the reference database. The app registers the information locally when offline and then saves the information into the server online.

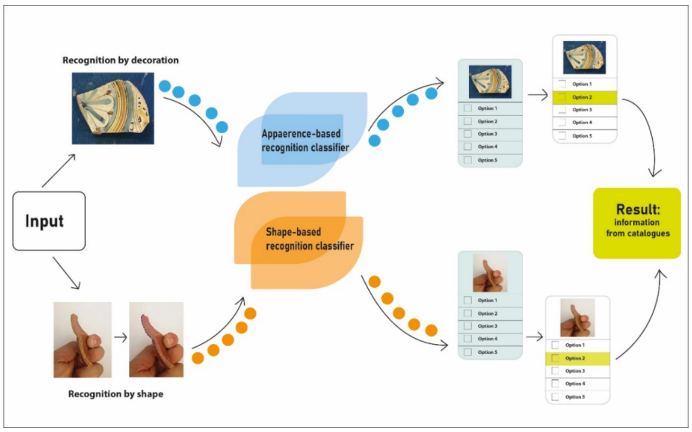

Fig. 11: The double workflow for appearance-based and shape-based recognition from an input image to top 5 results.

IV. Shortage

Despite its help, ML still has some limitations.

The biggest difficulty is the shortage of training data. Archaeology is widely digitised, but rarely datafied[3]. ML prefers Big Data which is findable, accessible, interoperable, and reusable. But nature of archaeology makes it hard to produce huge data, and much already-had data is unusable due to copyright or legislation.[10]

ML relies on its previously created model too much, and requires new data to be applicable to its model. Most ML models for archaeological data are less reliable than human experts now, because algorithms can't consider the variation and consistencies of data. That is, human experts can quickly handle easy cases, and move extra time to complicate cases, but algorithms handle every sample with same process and consume similar time. This disadvantage may offset ML advantages of high calculation speed and scalability benefits.[5]

The diversity of archaeological objects makes classification more hard. The archaeological recovered samples may become fragmented, or be covered by patina and vegetation. These poor preservation of samples makes classification hard. Moreover, rare and unusual objects may be ignored by ML models. A look-like "normal" ceramic vessel may has an unusual surface treatment, which will be easily noticed by human, could be classified into normal category by ML models.[5]

A third reason is the accuracy of evaluation function in ML model. This sort of bias is of particular concern for archaeologists using ML on data associated with Indigenous communities. Archaeologists can use surveys based on acultural factors to create models that are stripped of cultural context and meaning. But these assumptions are being more and more challenged, especially when surveys focus on behaviors and outcomes rooted in cultural value systems.[6][8]

V. Future

Nowadays, archaeologists utilize AI in many ways, from creating 3D models of historical sights to scanning territories with a laser radar to find ancient graves or from matching and assembling pairs of ancient papyrus fragments containing mostly unknown scripture to detecting objects in archaeological sites using DTMs generated from ALS data. There is no denying that AI becomes more and more popular in the field of archaeology and plays an important role in it.

One direction for the future development of intelligent archaeology is helping archaeologists' work. For example, archaeologists often face such problems as not knowing where exactly to dig. They can define the region but not the exact place where an artifact or a grave lies. It's when the neural network comes in line. Instead of looking through millions of documents by themselves, archaeologists pass this work to neural networks. This technology can sort out information by utilizing a specific algorithm. By analyzing images, this system might not only direct archaeologists in their groundworks but also suggest territories that have similar patterns as potential objects for excavations.

However, knowing that many ML systems — especially deep neural networks — are essentially considered black boxes. This makes it hard to understand and explain the results given by a model. Because of this, it should be noted that AI does not replace the need for experts in archaeology. Instead, AI technology needs the expertise from archaeologists to improve itself and to judge the correctness of the results.

Another direction is to strengthen the identification of the artifacts. In our world, one of the most urgent problems in archaeology is the fact that many artifacts are traded on the dark web. Although many models trained in this direction have achieved quite good results, they are not yet ready to be applied in practice. Currently, the majority of detection operations are performed manually. If an AI can succeed in this direction, it would make an extraordinary contribution to the prevention of art-related illegal activities.

We expect that in the future, artificial intelligence technology will play an increasing, even irreplaceable role in the field of archaeology.

REFERENCES

[1] ABITBOL, R., SHIMSHONI, I., AND BEN-DOV, J. Machine learning based assembly of fragments of ancient papyrus. *J. Comput. Cult. Herit. 14*, 3 (jul 2021).

[2] ANICHINI, F., DERSHOWITZ, N., DUBBINI, N., GATTIGLIA, G., ITKIN, B., AND WOLF, L. The automatic recognition of ceramics from only one photo: The archaide app. *Journal of Archaeological Science: Reports 36* (2021), 102788.

[3] ANICHINI, F., AND GATTIGLIA, G. Big archaeological data. the archaide project approach.

[4] BERLIN, A., ETTLINGER, E., HEDINGER, B., HOFFMANN, B., KENRICK, P., PUCCI, G., ROTH-RUBI, K., SCHNEIDER, G., SCHNURBEIN, S., WELLS, C., AND ZABEHLICKY-SCHEFFENEGGER, S. Conspectus formarum terrae sigillatae italico modo confectae. *American Journal of Archaeology 96* (01 1992), 190.

[5] BICKLER, S. H. Machine learning arrives in archaeology. *Advances in Archaeological Practice 9*, 2 (2021), 186–191.

[6] DAVIS, D. S., DINAPOLI, R. J., AND DOUGLASS, K. Integrating point process models, evolutionary ecology and traditional knowledge improves landscape archaeology—a case from southwest madagascar. *Geosciences 10*, 8 (2020).

[7] DELLEPIANE, M., CALLIERI, M., BANTERLE, F., ARENGA, D., ZALLOCCO, M., AND SCOPIGNO, R. From paper to web: Automatic generation of a web-accessible 3d repository of pottery types. In *GCH* (2017).

[8] DOUGLASS, K., QUINTANA MORALES, E., MANAHIRA, G., FENOMANANA, F., SAMBA, R., LAHINIRIKO, F., CHRISOSTOME, Z., VAVISOA, V., SOAFIAVY, P., JUSTOME, R., LEONCE, H., HUBERTINE, L., PIERRE, B., TAHIRISOA, C., COLOMB, C., LOVANIRINA, F., ANDRIANKAJA, V., AND ROBISON, R. Toward a just and inclusive environmental archaeology of southwest madagascar. *Journal of Social Archaeology 19* (07 2019), 146960531986207.

[9] DRAP, P., SCARADOZZI, D., GAMBOGI, P., AND GAUCH, F. Underwater photogrammetry for archaeology- the venus project framework. In *GRAPP* (2018).

[10] GUALANDI, M. L., GATTIGLIA, G., AND ANICHINI, F. An open system for collection and automatic recognition of pottery through neural network algorithms. *Heritage 4*, 1 (2021), 140–159.

[11] GUYOT, A., HUBERT-MOY, L., AND LORHO, T. Detecting neolithic burial mounds from lidar-derived elevation data using a multi-scale approach and machine learning techniques. *Remote Sensing 10*, 2 (2018).

[12] HU, X., AND YUAN, Y. Deep-learning-based classification for dtm extraction from als point cloud. *Remote sensing (Basel, Switzerland) 8*, 9 (2016), 730–730.

[13] KARTEROULI, K., AND BATSAKI, Y. Ai and cultural heritage image collections: Opportunities and challenges.

[14] KAZIMI, B., THIEMANN, F., MALEK, K., SESTER, M., AND KHOSHELHAM, K. Deep learning for archaeological object detection in airborne laser scanning data.

[15] LYONS, M. Ceramic fabric classification of petrographic thin sections with deep learning. *Journal of Computer Applications in Archaeology 4* (09 2021), 188.

[16] MICHAEL HEINZELMANN, MARIANNE BERGMANN, J.-Y. E. P. B. Ceramalex - ancient pottery in alexandria and its chora. https://archaeologie.phil-fak.uni-koeln.de/en/research/research-projects/finished/ceramalex-ancient-pottery-in-alexandria-and-its-chora, 2022.

[17] PAWLOWICZ, L. M., AND DOWNUM, C. E. Applications of deep learning to decorated ceramic typology and classification: A case study using tusayan white ware from northeast arizona. *Journal of archaeological science 130* (2021), 105375–.

[18] ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. 23* (08 2004), 309–314.

[19] SERVICE, A. D. Roman amphorae: A digital resource. https://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/, 2022.