Convolutional Neural Networks for Archaeological Site Detection Finding "Princely" Tombs

Gino Caspari*

Department of Archaeology, University of Sydney
and
Institute of Archaeological Sciences, University of Bern

Pablo Crespo†

Ph.D. Program in Economics, The Graduate Center, City University of New York

Abstract

Creating a quantitative overview over the early Iron Age heritage of the Eurasian steppes is a difficult task due to the vastness of the ecological zone and the often problematic access. Remote sensing based detection on open-source high-resolution satellite data in combination with convolutional neural networks (CNN) provide a potential solution to this problem. We create a CNN trained to detect early Iron Age burial mounds in freely available optical satellite data. The CNN provides a superior method for archaeological site detection based on the comparison to other detection algorithms trained on the same dataset. Throughout all comparison metrics (precision, recall, and score) the CNN performs best.

Keywords: CNN; object detection; archaeological remote sensing; convolutional neural networks; site detection

*Email: gino.caspari@sydney.edu.au .
†Email: pcrespo@gradcenter.cuny.edu.

1 Introduction

The archaeology of the Early Iron Age in the Eurasian steppe deals with a vast and archaeologically unexplored space between Eastern Europe and Mongolia. Despite the amount of research which has been conducted by scholars of the former USSR and the recent wave of new research coming out of these areas, a quantifiable understanding of the wealth of cultural heritage the Eurasian steppe harbors, has yet to be achieved. One of the problems which hinders researchers in gaining a wider understanding is the fact that the ancient cultural phenomena of the Early Iron Age did not neatly adhere to modern nation state borders (Figure 1). The current administrative, linguistic, and institutional fragmentation of this vast ecological zone –the steppe–makes research on the ground difficult. Remote sensing in combination with automatic or semi-automatic approaches for object detection have been established as a tool which largely disregards these problems and is able to provide the basis for solutions (Caspari *et al.*, 2014). Rooted in archaeological field research we combine open source data with convolutional neural networks (CNNs) in order to encompass the newest technological advances and use them to detect elite tombs of the Early Iron Age in the Eurasian steppe.

When it comes to restrictive access for foreign researchers, the Xinjiang Uyghur Autonomous Region is maybe the most extreme example in the region. It is known for its political and ethnical issues (Clarke, 2008) and recently received international media attention due to its increasingly oppressive counter-terrorism campaigns. (Roberts, 2018) Notoriously hard to receive permits for archaeological fieldwork in the first place, sporadic eruptions of ethnic conflicts between the Uyghur minority and Han Chinese majority in southern Xinjiang can abort long-planned projects last minute. Militarized border zones geographically curtail the areas archaeologists can work in. Even receiving a permit is not necessarily a guarantee that a field campaign can be conducted as planned, since the security apparatus is suspicious of any research activity by foreigners. Remote sensing mitigates these problems of access and the quality of publicly available high-resolution satellite data for Xinjiang has increased dramatically over the past years (Caspari, 2018).

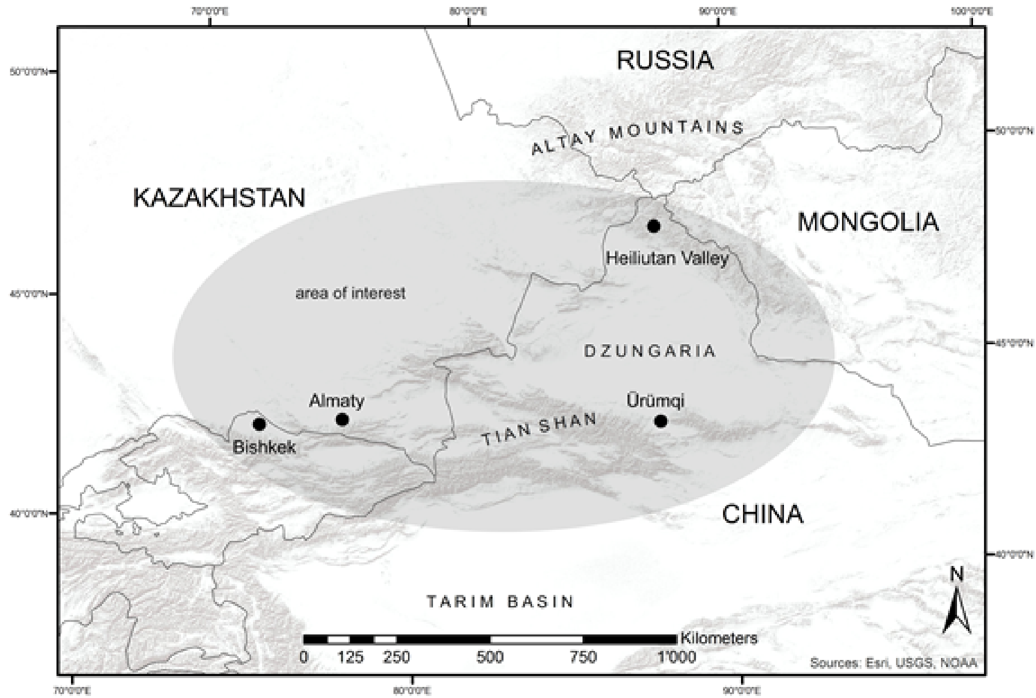

Figure 1: The area of interest in the eastern Central Asian steppes

CNNs have become the standard tool in computer vision applications in recent years. Their particular use in pattern and shape recognition is noted and popularized with the LeNet-5 architecture for recognizing handwritten digits (Lecun *et al.*, 1999). Their particular usefulness is predicated on their ability to take inputs in the shape of multidimensional matrices (tensors), allowing them to work with patterns in multiple directions. Pixels adjacent to each other have influence on what is identified. Most other machine learning algorithms used in image recognition work with inputs that take the shape of single row vectors, eliminating the ability to harness the information given by adjacent pixels in an image that are not in the same row (the pixels right below, above or set diagonally). Hence, CNNs are much more sensitive to identifying subtle patterns in images.

CNNs are a versatile solution to a plethora of problems in archaeology which works well when plenty of data is available. It comes at the cost of not being able to fully and analytically understand the process of solving the problem. The outcomes however can be qualitatively assessed and the solution is reproducible. Consistent with their versatility, CNNs have been used in different archaeological sub fields and for a diverse number of tasks from sex determination of skeletal remains to solving mapping tasks and extracting pottery depictions from archaeological publications.

Unsurprisingly, being one of the main categories of archaeological material, research on ceramics has seen a wide application of CNNs already. From recognizing vessels to classifying ceramic form, to understanding and classifying the structure of ceramics, CNNs have been useful in solving complex problems. (Benhabiles and Tabia, 2016) build a CNN to design local descriptors for content-based retrieval of three-dimensional (3-D) vessel replicas. (Pasquet *et al.*, 2017) use a CNN to detect amphorae in an underwater setting, correctly mapping around 90% of the vessels. (Hein *et al.*, 2018) automatically extract and classify ceramics based on textures. (Chetouani *et al.*, 2018) enlist the help of a CNN in order to classify shards and understand the movement of potters. The ArchAIDE project experiments with CNNs to create an as-automated-as-possible tool for the classifications and interpretation of shards (Gualandi *et al.*, 2016). A similar application is envisioned by (Tyukin *et al.*, 2018) with the project Arch-I-Scan which aims to automatically classify Roman pottery.

Interpreting other archaeological classes of information with CNNs is still in its infancy, but a number of examples can give the reader an idea of what might be possible if expertly human labeled datasets are combined with CNNs. (Byeon *et al.*, 2019) automatically identify and classify cut marks on bones. The authors manage to demonstrate that CNNs recognize and classify marks with a much higher accuracy rate than human experts. CNNs also perform exceptionally well when tasked with determining the sex of skeletal remains based on CT scans thereby eliminating human bias (Bewes *et al.*, 2019). In the analysis and interpretation of ancient scripts, CNNs are also beginning to make an impact. First attempts have been made in indexing Mayan hieroglyphs (Roman-Rangel and Marchand-Maillet, 2016; Can *et al.*, 2018) and creating a standardized corpus of graphemes for the Indus Valley script (Palaniappan and Adhikari, 2017). Further applications of CNNs in classifying, transcribing, and ultimately translating e.g.

cuneiform are to be expected.

CNNs have so far found the widest application in the area of archaeological remote sensing. This subfield of archaeology has the advantage of already working within a data-focused framework where classification and mapping tasks are common. The application of CNNs thus comes as an obvious extension of existing automated and semi-automated methods. Especially with LiDAR data collection, the data volume is becoming too large to be analyzed through a manual approach. CNNs help to mitigate this problem while simultaneously maintaining a consistent approach. (Trier *et al.*, 2019) present a case study mapping a number of archaeological object classes on an island in Scotland based on airborne laser scanning data. (Guyot *et al.*, 2018) detect Neolithic burial mounds in a LiDAR-derived digital elevation model. (Kramer *et al.*, 2017) combine aerial imagery and LiDAR data to detect archaeological structures using previously identified archaeological sites as training data. Other non-invasive methods like geophysical prospection, in particular ground penetrating radar (Travassos *et al.*, 2018; Ishitsuka *et al.*, 2018; Pham and Lefèvre, 2018), have also seen the application of CNNs. Our own case study in this paper belongs to the wide field of CNN applications which arose from image processing conceptually close to well-known and widely applied tasks like the recognition of faces and vehicles in images. CNNs can be useful in any area where remote sensing data needs to be searched for archaeological structures. The efficient processing of image data even allows for real-time decision making so that (Rutledge *et al.*, 2018) are able to present an autonomous underwater robot system, which allows for the autonomous surveying of underwater sites including path planning and acquisition of high-resolution sonar data.

Even art historical classifications and comparisons are supported by CNNs. With the appropriate amount of data, it becomes feasible to define stylistic affinity. First applications can be seen in the classification of wall paintings in Pompeii (Schoelz, 2018) and (Li *et al.*, 2018) approach towards dating the Mogao Grottoes wall paintings based on drawing styles defined by a CNN. (Wang *et al.*, 2017) use CNNs for defining similarities of Bodhisattva head images at the Dazu Rock Carving site and thus contribute to the reconstruction of some of the damaged rock carvings. An application of CNNs in the restoration of damaged archaeology can also be seen in a paper by (Hermoza and Sipiran, 2017) where the authors try to predict the missing geometry of damaged archaeological objects opening a promising avenue of research into computer-supported reconstruction and restoration of archaeological artifacts.

Wherever the exploration and analysis of large data sets is aided by recognizing complex patterns, CNNs can be helpfully employed. This leads to creative applications like a study by (Graham, 2018). The authors identify sales of human remains on social media platforms using CNNs to detect patterns allowing for the classification of a combination of images and text ultimately aiding the reconstruction of sales networks.

2 The field archaeological foundation

The Dzungaria Landscape Project, first established in 2014 (Caspari *et al.*, 2017), relied on a large-scale automated survey by means of a trained Hough Forest algorithm (Caspari *et al.*, 2014). Since then, machine learning has made enormous progress and the quality of the freely available satellite imagery has increased substantially. Through an intensive on-ground survey, the project was able to obtain a dataset of archaeological structures in the foothills of the Chinese Altai Mountains. Accumulations of very large Early Iron Age burial mounds early on caught the attention of the researchers (Figure 2 and Figure 3). It soon became clear that the southern Altai Mountains, in particular the area around Heiliutan were a focus of intense funerary building activity, especially during the first millennium BCE (Caspari *et al.*, 2017). A number of different Early Iron Age material cultures in the first millennium BCE can be identified (van Geel *et al.*, 2004). Here, we are specifically focusing on the funerary architecture of the Saka culture due to its relative homogeneity. There is a plethora of architectural remains from the Early Iron Age present in the survey area, but many of them are too small to be reliably detected in open source optical satellite data (Caspari, 2018). By far the most dominant anthropogenic features of the landscape are large burial mounds with circular ditches around them.

These monuments of which 59 (Caspari *et al.*, 2017; Caspari, Forthcoming) were mapped

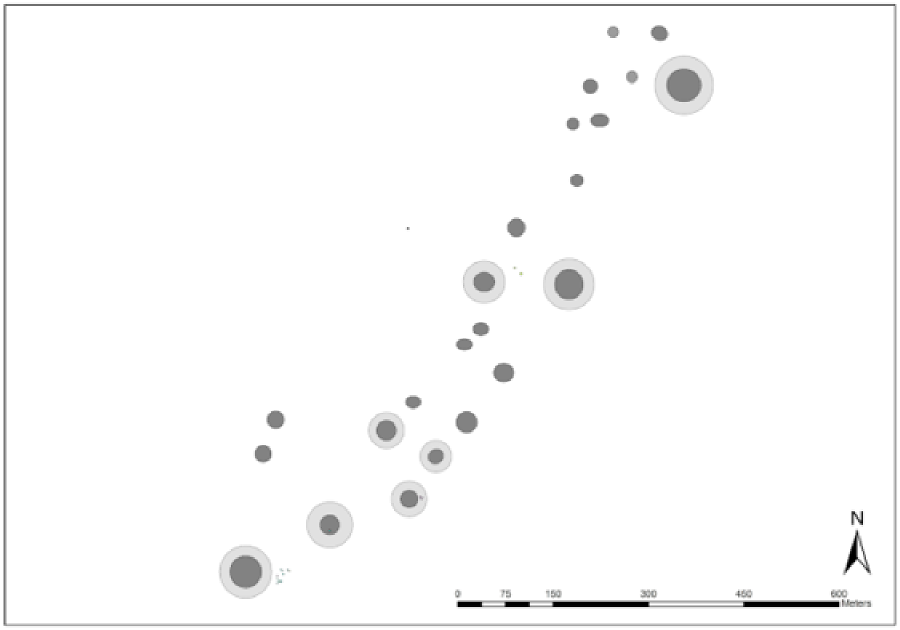

Figure 2: Map generated during the 2015 survey of the Heiliutan Valley in northern Xinjiang. Large Saka burial mounds tend to cluster. Dark grey = mound. Light grey = ditch.

during the field surveys, bear a striking resemblance to so-called Saka burials from the Semirechye (eastern Kazakhstan), the northern Tianshan and the Ili Valley. The term “Saka” is a relatively unspecific ethnic term stemming from Persian sources as (P’iankov, 1994) elaborates and thus should only be used with the appropriate care. Over decades of archaeological research in what is now eastern Kazakhstan, the term, however, has come to denote a specific Early Iron Age material culture and is seen as a technical term among many researchers without implying the potentially problematic ethnic connotations. The Saka material culture in eastern Kazakhstan is dated to the 7th/6th cent. BCE and the 3rd cent. BCE (Parzinger, 2011). Saka burials have so far mainly been known from the Semirechye (Davis-Kimball, 1991; Gass, 2011; Nagler, 2009; Nagler *et al.*, 2010) and have only recently been compiled in a large study by (Gass, 2016).

The connections of Saka-related material culture into northern Xinjiang have been analyzed (Davis-Kimball, 1991; Chen and Hiebert, 1995) but due to the fragmentary nature of archaeological data in Xinjiang have been assumed to mainly be confined to the westernmost stretches of Xinjiang, namely the Ili Valley and the northern Tianshan. Older Chinese research has looked at these connections from the eastern side (Wang, 1985) working on a number of sites which show clear relations to eastern Kazakhstan like Tiemulike (Institute of Archaeology of the Xinjiang Academy of Social Sciences, 1988), Dacaotan (Institute of Archaeology of the Xinjiang Academy of Social Sciences, 1985), and Zhongyangchang (Institute of Archaeology of the Xinjiang Academy of Social Sciences, 1986). The architectural features of the mounds in the Heiliutan Valley, however, suggest a strong cultural connection during the middle of the first millennium BCE all the way into the foothills of the Chinese Altai Mountains.

Figure 3: Architectural features of Saka burial mounds.

The large burial mounds of the Saka material culture usually were built from a mixture of pebbles, larger round stones and earth from the alluvial terraces. Mounds are typically elevated and surrounded by circular rings of stones or circular ditches (Figure 3). Both ditch and mound are clearly visible in open source optical satellite data. The profile of the Saka burial mounds typically shows steep sides (sometimes three steep sides and one with a gentler slope) and a flat top. Maximum diameters in the Heiliutan area are typically between 15.5m and 34.1m (89.5 %) and therefore well within the range of detectable objects in open-source satellite imagery (Figure 4). A group of outliers has diameters of over 40m. The average diameter of Saka mound in the Heiliutan Valley is 27.93m (median 26.8m).

The heights of these burial mounds average at 1.97m (median 1.4m). The largest

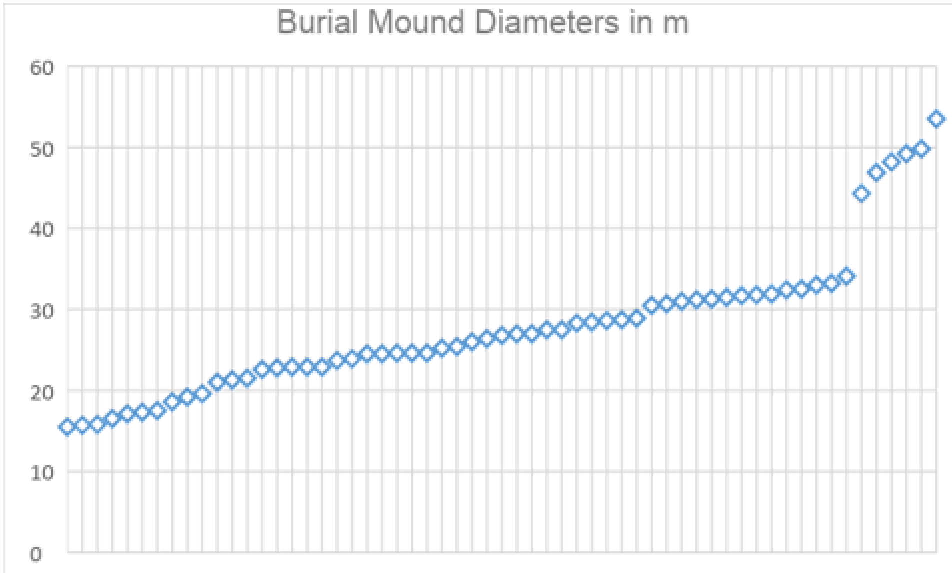

Figure 4: Scatterplot of Saka burial mound diameters, notice the cluster of extraordinarily large mounds which clearly set themselves apart from the smaller ones. These “princely” tombs are easily recognizable in open source remote sensing data.

mounds have a height of up to 6.5m. Both diameters and heights of Saka burial mounds in the Chinese Altai are comparable to Saka burial mounds from Issyk, Kegen and other cemeteries with princely tombs (Gass, 2011; Samashev, 2007). The Saka burials of the Heiliutan area are all practically identical in their composition of building materials and the profile of the mound. The largest mound has a diameter of 53.5m, a height of 6.0m, and the circular ditch measures 91.5m across. This type of burial usually has a 5:3 ratio between circular ditch diameter and mound diameter which again matches Saka burials from the Semirechye (Gass, 2011). The large accumulation of Saka burials (Figure 2) with a length of almost 2km are visible from afar and one of the dominant archaeological places within the landscape of the Heiliutan Valley. One of these monuments has been excavated in 2016 by the Institute of Archaeology of the Xinjiang Academy of Social Sciences but has yet to be published like many other burial mounds in the area of interest the grave was unfortunately looted.

3 Convolutional Neural Networks

CNNs are a specific type of neural network architectures popularized by (Lecun *et al.*, 1999), which can take grid-like inputs. Our particular case is a two-dimensional grid of pixels, in which each pixel can be considered a source of information in the same way as a cell in a row of tabular data would be. Note that images can be interpreted as numerical grids if each pixel on each channel (RGB) is given a numerical value based on the intensity of the color from 0 to 255. In order to understand how CNNs work, we will define them as the junction of three different operating components as types of "layers":

- convolutional layers
- pooling layers
- fully connected layers

The convolutional and pooling layers are used to identify and summarize patterns in the data. The fully connected layers are used to utilize these summaries as inputs of a classification problem, helping us make the determination of whether our (in this case) image belongs to a specific class based on the model. An example diagram of these architectures is presented in Figure 5.

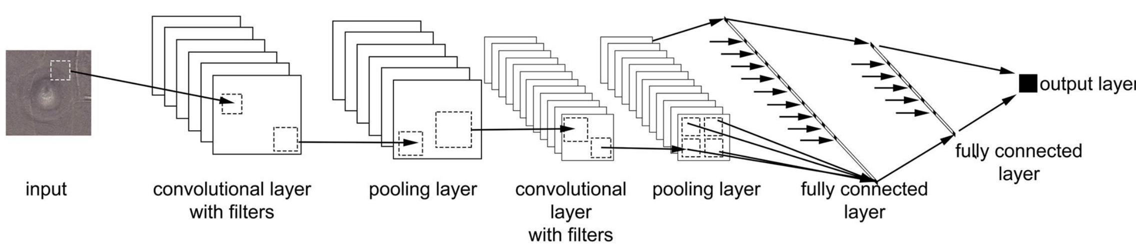

Figure 5: CNN architecture example diagram

3.1 CONVOLUTIONAL LAYERS

The network uses convolutional layers to detect simple features or patterns in the data. The patterns can be small and simple, but the combination of multiple simple patterns allow for the search of complex forms.

Each convolutional layer is composed of two stages: convolution and detection. In the first stage a set of convolution operations are run on the input grid. A kernel or filter is moved sequentially on the input generating outputs on each position they take. These are defined by:

$$h_{i,j} = \sum_{h=1}^{m} \sum_{l=1}^{m} w_{k,l} x_{i+k-1,j+l-1}$$

where $h_{i,j}$ is the output of the convolution at position (i, j) , $x_{i+k-1,j+l-1}$ portion of the input grid over which the filter is applied, $w_{k,l}$ is the filter at position (k, l), and m determines the height and width of the filter.

Hence, the filter is a weighting square grid, which is applied to the larger input grid to highlight specific patterns within it. The higher the value of a convolution operation, the higher the chance that the pattern that the filter searches for is found. Figure 6 highlights this process by exemplifying it. We can then use different filters to find different patterns. For example, using a filter of the form:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

would be used to identify vertical lines.

Filters can be, and often are, initialized at random to pick on many and varied subtle patterns within the input grid. Each convolutional layer runs several filters on the inputs and outputs grids for each.

For the detection or activation stage, the results from the convolution stage are taken

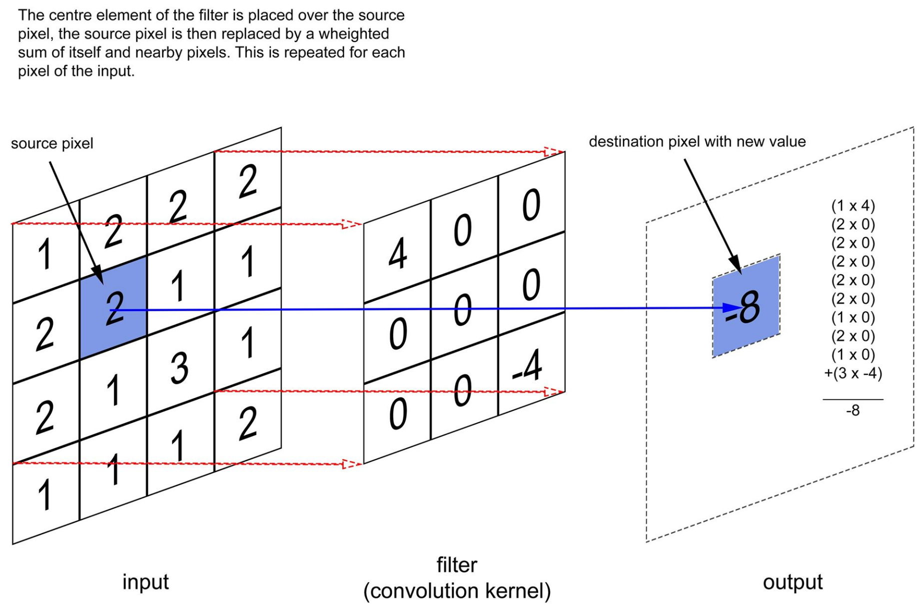

Figure 6: Filter applied over a matrix

and passed through a function. We used the ReLU (Rectifying linear unit), which is defined as:

$$\sigma(x) = max(0, x)$$

This specific function grants extra weight to all of the non-negative units. Since the filters can have negative values, this activation allows for extra salience of patterns.

After activation, the outputs of the convolutional layer are used as inputs for the pooling layers.

3.2 Pooling Layers

A pooling layer summarizes the resulting activated grids through max pooling. This work uses max pooling. A new grid is constructed from each activated grid by assigning each entry of it to the maximum value of 2×2 subgrids. An example is shown in Figure 7.

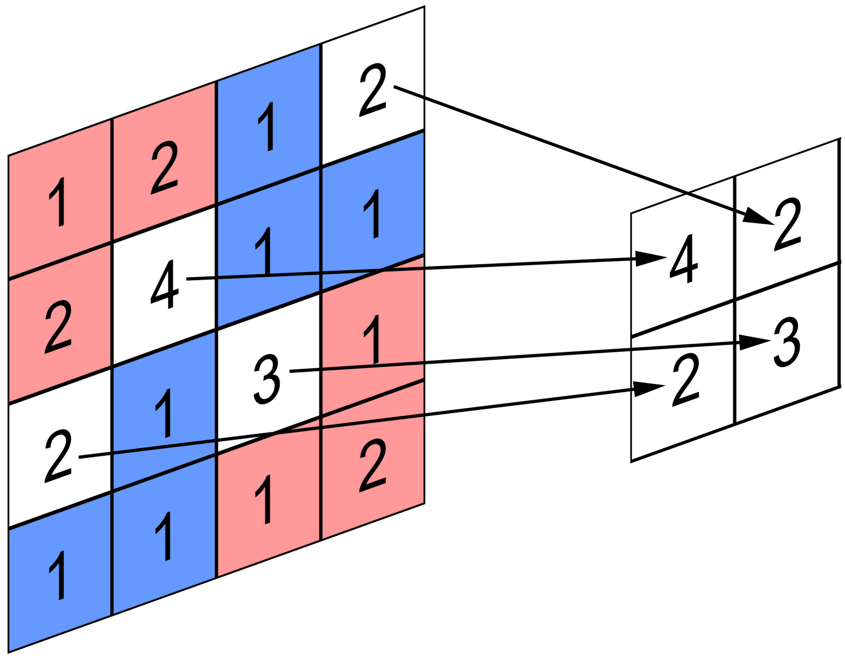

Figure 7: 2×2 max pooling

At this point, the practitioner has two choices: to summarize the results once more through a fully connected layer (see Section 3.3) or to repeat the process of passing the outputs through a convolutional layer and pooling layer once again. This is what is meant by making a network ”deeper.” Passing the data through extra convolutional and pooling layers allows for further and more subtle evaluation of patterns. This is said to elevate the complexity of the model.

3.3 Fully connected layers

Fully connected layers have the basic structure of artificial neural networks or multilayer perceptrons. Their task is to take the outputs from the last pooling layers and classifying them into specific categories. Before passing the grids resulting from the pooling layers to the fully connected layers, the grids are ”flattened.” Meaning the results from all the resulting grids are combined into a single row vector. The resulting elements of the vector produced after the flattening are then linearly combined. This means they are written as:

$$\beta_0 + \sum_i \beta_i \times \text{element}_i$$

where β_0 is called the "bias" and the rest of the β_i are called the "weights." Each of these linear combinations is passed through an activation function yet again generating a single number output. This particular structure of operations constitutes what is called a "neuron." The set of these activated linear combinations is called a "hidden layer." The practitioner can add extra complexity to the model by using the outputs of each hidden layer as the inputs for a new fully connected layer. The practitioner has to choose both the number of neurons and the depth of the model by choosing the number of hidden layers. Once it has been decided that the architecture is deep enough, in the case of binary classification such as ours, a final fully connected layer is created yielding a single linear combination and the activation of this one is what we consider the "output layer" of the network usually normalized between zero and one thanks to the activation function (a sigmoid function[1] is a common choice). The corresponding number in this output layer is mapped to a specific class according to a threshold. For example, binary classes code their "target" variable as either having values of zero or one. We can say that an output layer with a value larger than 0.5 will predict the input belongs to class one and to class zero otherwise.

The question remains on how these networks are actually generalized for large samples of images. Consider that a sample of already identified and labeled images, which we call our "target" is compiled in a vector y. Then, we would like to make sure that overall the values of the distinct weights and biases are chosen such that the resulting output layer is as close to the target as possible for representative samples. In this case, we would like to choose values such that the following distance is minimized via a process called "backpropagation" for n observations in a sample:

$$\sum_{i=1}^{n} \frac{1}{2} \left(y_i - \text{output}_i\right)^2 \tag{1}$$

1

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

It needs to be noted that deeper networks with hidden layers and many neurons in each are capable to make the distance in Equation (1) very small for a sample due to added complexity. This however does not come without the risks of making the network attuned to only the images fed through the specific sample and incapable of generalizing to others from the same population of objects but that were not present in the sampled data. This process is called ″overfitting.″ Hence, the practicioner needs to be sure to design their architecture in a fine balance. The network must be capable to process complex enough patterns for classification, but not be so overly attuned to the sample data such that it fails classifying data from the same population outside the sample.

4 Application of a CNN on "princely" tomb classification

4.1 Data preprocessing

Using open-source optical satellite data from Google Earth (100 x 100 pixels) of tombs with known locations and arbitrary patches of land around them, a labelled dataset was created with the following labelling scheme:

$$y = \begin{cases} 0 & \text{if tomb present} \\ 1 & \text{if tomb absent} \end{cases}$$

The dataset is composed of 1212 images with 169 including tombs. Typical observations of each case are presented in Figure 8. It is important to note that the distinctive shape of the tombs makes them easily distinguishable from other patches of land even in low-resolution data.

In order to verify that the model we fit is a good model, the data is split in two portions, one for fitting the model (training data) and one for looking at how well it generalizes (testing and validation). The testing and validation data are simply datasets that don't undergo the fitting process. Since the data belongs to the same population as the training data does, assessing the goodness of fit of the model in these can give us a good idea of

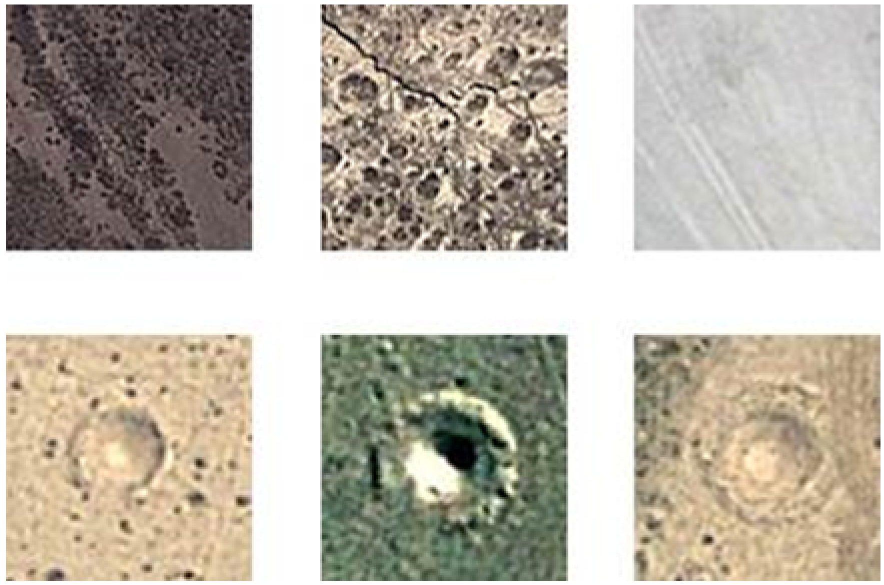

Figure 8: Top: Examples of images labelled as tomb absent. Bottom: Examples of images labelled as tomb present.

how well the model generalizes and it helps identify overfitting.

The data was split with 75% used for training and 25% used for testing and validation. Since the images containing tombs are heavily underrepresented in the dataset, augmentation is necessary for training appropriately with multiparameter methods such as convolutional neural networks. In this case 655 new images were synthesized from the training data with tombs present. The new samples are created by modifying the existing ones through randomly zooming, shearing, and performing horizontal flips. Note that augmented images are only used during the training stage. Using them for testing or validation is inappropriate due to their high correlation with the images that they originate from.

4.2 CNN Architecture

The CNN utilized for our problem was trained on the augmented data mentioned at the beginning of this section. The full summary of the architecture is detailed in Figure 9. The CNN was trained in Keras, a Python module which uses Google's TensorFlow as a backend in our case.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 98, 98, 32)	896
activation_1 (Activation)	(None, 98, 98, 32)	0
max_pooling2d_1 (MaxPooling2	(None, 49, 49, 32)	0
conv2d_2 (Conv2D)	(None, 47, 47, 32)	9248
activation_2 (Activation)	(None, 47, 47, 32)	0
max_pooling2d_2 (MaxPooling2	(None, 23, 23, 32)	0
conv2d_3 (Conv2D)	(None, 21, 21, 64)	18496
activation_3 (Activation)	(None, 21, 21, 64)	0
max_pooling2d_3 (MaxPooling2	(None, 10, 10, 64)	0
flatten_1 (Flatten)	(None, 6400)	0
dense_1 (Dense)	(None, 64)	409664
activation_4 (Activation)	(None, 64)	0
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 1)	65
activation_5 (Activation)	(None, 1)	0

Total params: 438,369
Trainable params: 438,369
Non-trainable params: 0

Figure 9: Keras model summary

The architecture shown is relatively simple consisting of 3 convolutional and pooling layers with ReLU activations and two fully connected layers before the final activation with a sigmoid. The diagram specifies the dimensions of each. For example, the first convolutional layer uses 32 filters and outputs a 98×98 grid. A natural question not necessarily explained in the prior sections is what the ”dropout” row means in the diagram. Dropout is a regularization technique which disallows certain linear combinations to exist at random during the optimization step. This technique helps ”regularize” or penalize overfitting. Hence, making sure the model is generalizable.

4.3 Benchmarks and results

Judging the accuracy of the convolutional neural network specified in Section 4.2 requires plausible methods for benchmarking. Furthermore, the true metrics of accuracy we are interested in are those in the validation data. These would be the ones that would tell us how each model works under observations not seen by the training model. As such, three models were chosen: a biased random guess, a support vector classifier with a linear kernel and a support vector classifier with a radial basis function kernel.

Random guess is useful as a comparative benchmark since it selects its output by simple random chance. In order to make the benchmark tougher, we biased the probabilities of classifying an image as containing a tomb to be the proportion of the actual number of tombs in the validation set

Since the shapes of the tombs are simple and easily distinguishable, it stands to reason that simpler and more tractable classification methods could work as long as they allow for flexible boundary classification. Support vector machines with kernels as proposed by (Boser *et al.*, 1992) work as sensible and powerful alternatives to deep learning models. We attempt using two types of kernels in this study, the linear kernel and the radial basis function kernel which both allow for different transformations of the data pre-classification. Each of these models have their hyperparameters adjusted via 5 -fold cross validation.

We use three measures to compare the predictions made by the classifiers accuracy: Precision, Recall and F_1 score. Definitions below:

$$\text{Precision} = \frac{\text{\# of True Positives}}{\text{\# of True Positives} + \text{\# of False Positives}}$$

$$\text{Recall} = \frac{\text{\# of True Positives}}{\text{\# of True Positives} + \text{\# of False Negatives}}$$

$$F_1 \text{ score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

Precision simply gives the rate of correctly classified objects among all classified objects with the same label. Recall gives the rate of correctly labeled objects among all actual objects with that label. F_1 score gives a balanced measure of both. All tables and figures comparing models in this paper use these measures.

Table 1 and Table 2 encapsulate the results obtained from the trained models making predictions on the validation set. We can appreciate that for both, images which contained tombs or those which did not, the CNN performs best. Interestingly, despite the fact that SVMs worked under training with an augmented dataset, their performance in identifying pictures with tombs was not comparable to that of the neural network. This is surprising since the tomb shapes are mostly simple to the naked eye, hence nonlinear classification should work well. The reason lies in the likelihood of the SVM models containing many false positives (objects that are not tombs being identified as such). This occurs because other images that might just simply be circular in shape are likely to be picked up by the SVM models as tombs . This has been an issue with other detection algorithms before e.g. (Caspari *et al.*, 2014) . The big advantage of our architecture relies on the quantity of filters used being able to recognize higher subtlety in the patterns of the trained dataset that might identify a tomb, beyond just the circular shape. Figure 10 summarizes both tables and includes a bar for Average/Total, which has a weighted average for both classes under the measure. Showing that overall the CNN is the better performing model.

Model	**Precision**	**Recall**	F_1**score**
Random Guessing	0.64	0.65	0.65
SVM with linear kernel	0.9	0.96	0.94
SVM with RBF kernel	0.96	0.97	0.97
CNN	0.98	1	0.99

Table 1: Classification metrics for validation data set pictures without tombs.

Model	**Precision**	**Recall**	F_1**score**
Random Guessing	0.59	0.58	0.59
SVM with linear kernel	0.29	0.15	0.20
SVM with RBF kernel	0.76	0.67	0.71
CNN	1	0.84	0.91

Table 2: Classification metrics for validation data set pictures with tombs.

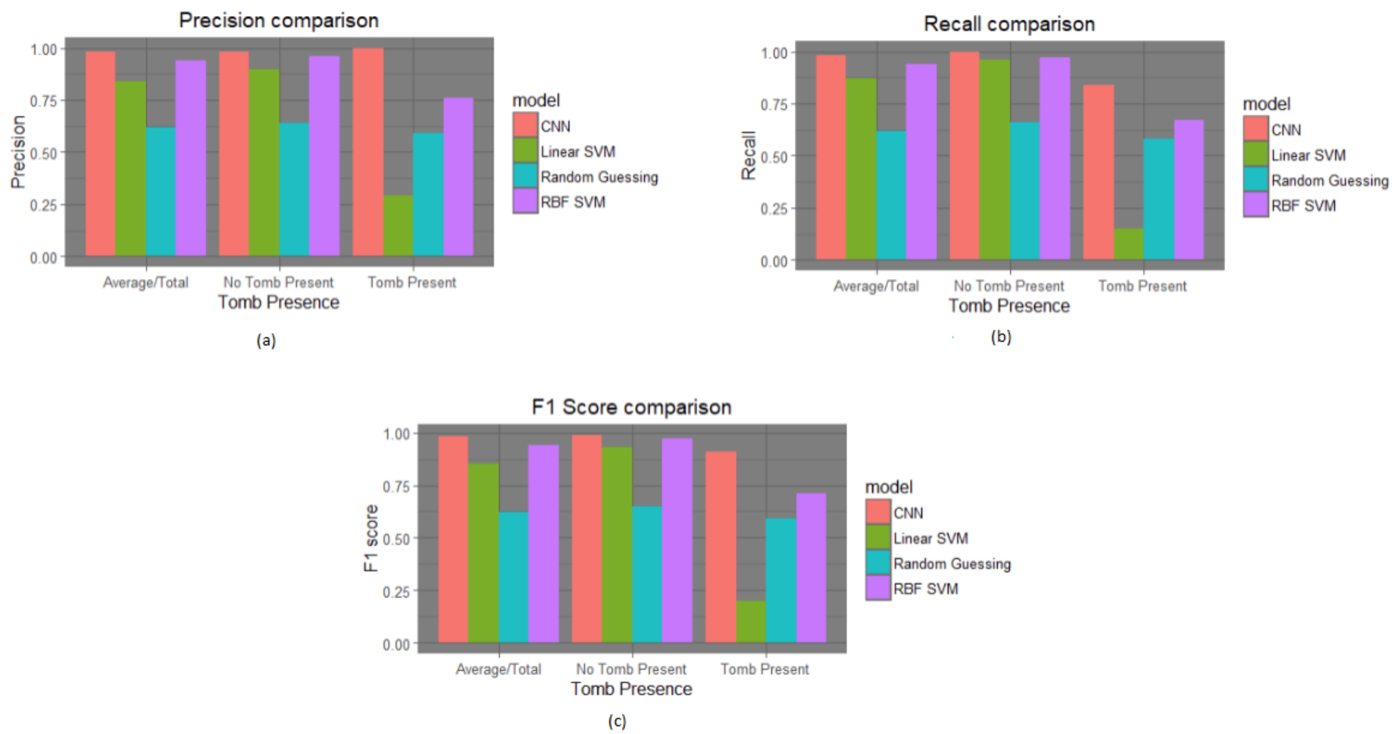

Figure 10: Result Summaries

5 Conclusions

The distinctive shape of the early Iron Age Saka burial mounds and their relatively large size make them an ideal training set for machine learning algorithms which can be run on open source satellite imagery. Our CNN outperforms other methods and provides a valuable approach for the large-scale detection of elite burial mounds in the Eurasian steppes. In this way a macro-regional survey of northern Xinjiang and the adjacent areas could be conducted in order to assess the spatial distribution of this monument type and possibly revise the geographical extent to which Saka-related material culture spread through Eastern Central Asia during the first millennium BCE. The method has the clear advantage that all analyses can be conducted without the access problems archaeological projects in the region usually have to deal with.

Preliminary satellite imagery analysis has developed into playing a major role in planning and implementing archaeological field research (Lasaponara and Masini, 2012; Caspari *et al.*, 2019). But automatic feature detection has yet to become accessible to a wider range of researchers in order to be widely applied. A number of attempts have been made to connect archaeological surveys with automatic detection of features (Caspari *et al.*, 2017; Trier *et al.*, 2009; Trier and Pil, 2012), however, it is not commonly used by practitioners. Both the complexity of the method which often demands cooperation with computer science specialists, and the lack of awareness for the possibility play a role in the so far rare application of automatic detection algorithms by archaeological practitioners. The authors do not expect to see a widespread application unless intuitive tools are developed for feature selection, algorithm training and visualization of ready-to-use results.

REFERENCES

BENHABILES, H. and TABIA, H. (2016). Convolutional neural network for pottery retrieval. *Journal of Electronic Imaging*, **26** (1).

BEWES, J., LOW, A., MORPHETT, A., PATE, F. D. and HENNEBERG, M. (2019). Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. *Journal of Forensic and Legal Medicine*, **62**, 40 – 43.

BOSER, B. E., GUYON, I. M. and VAPNIK, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, ACM Press, pp. 144–152.

BYEON, W., DOMNGUEZ-RODRIGO, M., ARAMPATZIS, G., BAQUEDANO, E., YRAVEDRA, J., MAT-GONZLEZ, M. A. and KOUMOUTSAKOS, P. (2019). Automated identification and deep classification of cut marks on bones and its paleoanthropological implications. *Journal of Computational Science*, **32**, 36 – 43.

CAN, G., ODOBEZ, J.-M. and GATICA-PEREZ, D. (2018). How to tell ancient signs apart? recognizing and visualizing maya glyphs with cnns. *J. Comput. Cult. Herit.*, **11** (4), 20:1–20:25.

CASPARI, G. (2018). Assessing looting from space: The destruction of early iron age burials in northern xinjiang. *Heritage*, **1** (2), 320–327.

— (Forthcoming). Quantifying the funerary ritual activity of the late prehistoric southern kanas region (xinjiang, china). *Asian Perspectives*.

—, BALZ, T., GANG, L., WANG, X. and LIAO, M. (2014). Application of hough forests for the detection of grave mounds in high-resolution satellite imagery.

—, PLETS, G., BALZ, T. and FU, B. (2017). Landscape archaeology in the chinese altai mountains survey of the heiliutan basin. *Archaeological Research in Asia*, **10** (Complete), 48–53.

—, SADYKOV, T., BLOCHIN, J., BUESS, M., NIEBERLE, M. and BALZ, T. (2019). Integrating remote sensing and geophysics for exploring early nomadic funerary architecture in the siberian valley of the kings. *Sensors*, **19** (14).

CHEN, K. T. and HIEBERT, F. T. (1995). The late prehistory of xinjiang in relation to its neighbors. *Journal of World Prehistory*, **9** (2), 243–300.

CHETOUANI, A., DEBROUTELLE, T., TREUILLET, S., EXBRAYAT, M. and JESSET, S. (2018). Classification of ceramic shards based on convolutional neural network. *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1038–1042.

CLARKE, M. (2008). China's war on terror in xinjiang: Human security and the causes of violent uighur separatism. *Terrorism and Political Violence*, **20** (2), 271–301.

DAVIS-KIMBALL, J. (1991). Kazakh/american research project: 1990 field work report. *Middle East Studies Association Bulletin*, **25** (1), 33–35.

GASS, A. (2011). Early iron age burials in southeastern zhetysu: The geoarchaeological evidence. *Archaeology, Ethnology and Anthropology of Eurasia*, **39** (3), 57 – 69.

— (2016). *Das Siebenstromland Zwischen Bronze- Und Früheisenzeit: Eine Regionalstudie*. Berlin; Boston; De Gruyter.

GRAHAM, S. (2018). Fleshing out the bones: Studying the human remains trade with tensorflow and inception. *Journal of Computer Applications in Archaeology*, **1** (1), 55–63.

GUALANDI, M. L., SCOPIGNO, R., WOLF, L., RICHARDS, J., GARRIGOS, J. B. I., HEINZELMANN, M., HERVAS, M. A., VILA, L. and ZALLOCCO, M. (2016). ArchAIDE - Archaeological Automatic Interpretation and Documentation of cEramics. In C. E. Catalano and L. D. Luca (eds.), *Eurographics Workshop on Graphics and Cultural Heritage*, The Eurographics Association.

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

HEIN, I., ROJAS-DOMNGUEZ, A., ORNELAS, M., D'ERCOLE, G. and PELOSCHEK, L. (2018). Automated classification of archaeological ceramic materials by means of texture measures. *Journal of Archaeological Science: Reports*, **21**, 921 – 928.

HERMOZA, R. and SIPIRAN, I. (2017). 3d reconstruction of incomplete archaeological objects using a generative adversary network. *CoRR*, **abs/1711.06363**.

INSTITUTE OF ARCHAEOLOGY OF THE XINJIANG ACADEMY OF SOCIAL SCIENCES (1985). Xinjiang xinyuan gongnaisi zhongyangchang shiguan mu (stone cist tomb of zhongyangchang site, xinyuan, xinjiang). *Kaogu yu Wenwu*, **2**, 21–26.

INSTITUTE OF ARCHAEOLOGY OF THE XINJIANG ACADEMY OF SOCIAL SCIENCES (1986). Xinjiang miquan dacaotan faxian shiduimu (the rock-fill tombs discovered at dacaotan, miquan, xinjiang). *Kaogu yu Wenwu*, **1**, 36–38.

INSTITUTE OF ARCHAEOLOGY OF THE XINJIANG ACADEMY OF SOCIAL SCIENCES (1988). Xinjiang xinyuan tiemulike gumu qun (tiemulike cemetery, xinyuan, xinjiang). *Kaogu yu Wenwu*, **8**, 59–66.

ISHITSUKA, K., ISO, S., ONISHI, K. and MATSUOKA, T. (2018). Object detection in ground-penetrating radar images using a deep convolutional neural network and image set preparation by migration.

KRAMER, I. C., HARE, J. S., PRÜGEL-BENNETT, A. and SARGENT, I. (2017). Automated detection of archaeology in the new forest using deep learning with remote sensor data.

LASAPONARA, R. and MASINI, N. (2012). *Satellite remote sensing: a new tool for archaeology*. Springer.

LECUN, Y., HAFFNER, P., BOTTOU, L. and BENGIO, Y. (1999). *Object recognition with gradient-based learning*, Springer.

LI, Q., ZOU, Q., MA, D., WANG, Q. and WANG, S. (2018). Dating ancient paintings of mogao grottoes using deeply learnt visual codes. *CoRR*, **abs/1810.09168**.

NAGLER, A. (2009). *Grosskurgane im Siebenstromland (Kazachstan)*, vol. 1. Jahresbericht 2008 des DAI. Archäologischer Anzeiger 2009.

—, Z, S., H, P. and M, N. (2010). *Südkasachstan: Kurgane Asy Zaga, Kegen und Zoan Tobe*, Berlin, DAI, chap. Archäologische Forschungen in Kasachstan, Tadschikistan, Turkmenistan und Usbekistan., pp. 49–54.

PALANIAPPAN, S. and ADHIKARI, R. (2017). Deep learning the indus script. *ArXiv*, **abs/1702.00523**.

PARZINGER, H. (2011). *Die frühen Völker Eurasiens: vom Neolithikum zum Mittelalter*. C.H. Beck.

PASQUET, J., DEMESTICHA, S., SKARLATOS, D., MERAD, D. and DRAP, P. (2017). Amphora detection based on a gradient weighted error in a convolution neuronal network.

PHAM, M. and LEFÈVRE, S. (2018). Buried object detection from b-scan ground penetrating radar data using faster-rcnn. *CoRR*, **abs/1803.08414**.

P'IANKOV, I. V. (1994). The ethnic history of the sakas. *Bulletin of the Asia Institute*, **8**, 37–46.

ROBERTS, S. R. (2018). The biopolitics of chinas war on terror and the exclusion of the uyghurs. *Critical Asian Studies*, **50** (2), 232–258.

ROMAN-RANGEL, E. and MARCHAND-MAILLET, S. (2016). Indexing mayan hieroglyphs with neural codes. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 253–258.

RUTLEDGE, J., YUAN, W., WU, J. Y., FREED, S., LEWIS, A., WOOD, Z. J., GAMBIN, T. and CLARK, C. M. (2018). Intelligent shipwreck search using autonomous underwater vehicles. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8.

SAMASHEV, Z. (2007). *Im Zeichen des goldenen Greifen - Königsgräber der Skythen*, Berlin, Staatliche Museen zu Berlin, chap. Die Fürstengr—"aber des Siebenstromlandes, pp. 162–170.

SCHOELZ, A. (2018). An embarrassment of riches: Data integration in vr pompeii.

TRAVASSOS, X. L., AVILA, S. L. and IDA, N. (2018). Artificial neural networks and machine learning techniques applied to ground penetrating radar: A review. *Applied Computing and Informatics*.

TRIER, I. D., COWLEY, D. C. and WALDELAND, A. U. (2019). Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on arran, scotland. *Archaeological Prospection*, **26** (2), 165–175.

—, LARSEN, S. Y. and SOLBERG, R. (2009). Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection*, **16** (1), 1–15.

— and PIL, L. H. (2012). Automatic detection of pit structures in airborne laser scanning data. *Archaeological Prospection*, **19** (2), 103–121.

TYUKIN, I., SOFEIKOV, K., LEVESLEY, J., GORBAN, A. N., ALLISON, P. and COOPER, N. J. (2018). Exploring automated pottery identification [arch-i-scan].

VAN GEEL, B., BOKOVENKO, N., BUROVA, N., CHUGUNOV, K., DERGACHEV, V., DIRKSEN, V., KULKOVA, M., NAGLER, A., PARZINGER, H., VAN DER PLICHT, J., VASILIEV, S. and ZAITSEVA, G. (2004). Climate change and the expansion of the scythian culture after 850 bc: a hypothesis. *Journal of Archaeological Science*, **31** (12), 1735–1742.

WANG, B. (1985). Gudai xinjiang sairen lishi gouchen (a preliminary research of the history of the ancient saka in xinjiang). *Xinjiang Shehui Kexue*, **1**, 48–58, 64.

WANG, H., HE, Z., HUANG, Y., CHEN, D. and ZHOU, Z. (2017). Bodhisattva head images modeling style recognition of dazu rock carvings based on deep convolutional network. *Journal of Cultural Heritage*, **27**, 60 – 71.