1.Possibilities of Data Science in Archaeology (Medium.com) – Mina Jambajantsan

Lidar — Light Detection and Ranging — is a remote sensing method used to examine the surface of the Earth.

A lot of archaeologists from all over the world are using LIDAR technology, geophysics, drones, underwater drones and many more technological inventions and innovations of the modern age. Drone usage is now a common practice at excavations. We are able to take pictures from above without using small planes or helicopters as it was at the dawn of Aerial Archaeology. Planes would require a professional pilot, tons of fuel and great cameras. With LIDAR-s we can have a “peek” at what’s under the ground without disturbing the soil, but this technology has a high price.

Data Science methods can be used for exploring two main things :

1. Data Science can be used to tidy up data, which sounds like a show off instead of using good old Excel, but with a massive data number, loads on Nan data and more than 20 columns, using Python and Pandas will allows less inaccuracy. Also, the graphic visualization of the given data will be more clear and efficient. We hope to see new patterns that might have missed, also later make the findings and thesis outcomes more understandable.
2. The second reason we are using Data Science is to be able to process information on given photos, especially drone photos or maps.

References:

Davenport, Thomas H., and D. J. Patil. [“Data Scientist: The Sexiest Job of the 21st Century.”](http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/) Harvard Business Review 90, no. 10 (October 2012): 70–76.

Map: <https://www.britannica.com/place/Mongol-empire/The-period-of-relative-unity-1227-60>

2.Archaeology, Humanities, and Data Science- James Newhard ([**The ArchaeoInformant**](https://blogs.cofc.edu/thearchaeoinformant/))

Why archaeologists  are rarely included in the list of ‘big data’ fields – fields which have developed capacities to organize, manage, mine, and analyze large sets of information to extract meaning and insights about questions of relevance for a myriad of societal needs.

3.A deep-learning model for predictive archaeology and archaeological community detection

Authors -[Abraham Resler](https://www.nature.com/articles/s41599-021-00970-z#auth-Abraham-Resler-Aff1), [Reuven Yeshurun](https://www.nature.com/articles/s41599-021-00970-z#auth-Reuven-Yeshurun-Aff2), [Filipe Natalio](https://www.nature.com/articles/s41599-021-00970-z#auth-Filipe-Natalio-Aff3) & [Raja Giryes](https://www.nature.com/articles/s41599-021-00970-z#auth-Raja-Giryes-Aff1)

[*Humanities and Social Sciences Communications*](https://www.nature.com/palcomms) **volume 8**, Article number: 295 (2021)

Abstract

Deep learning is a powerful tool for exploring large datasets and discovering new patterns. This work presents an account of a metric learning-based deep convolutional neural network (CNN) applied to an archaeological dataset. The proposed account speaks of three stages:

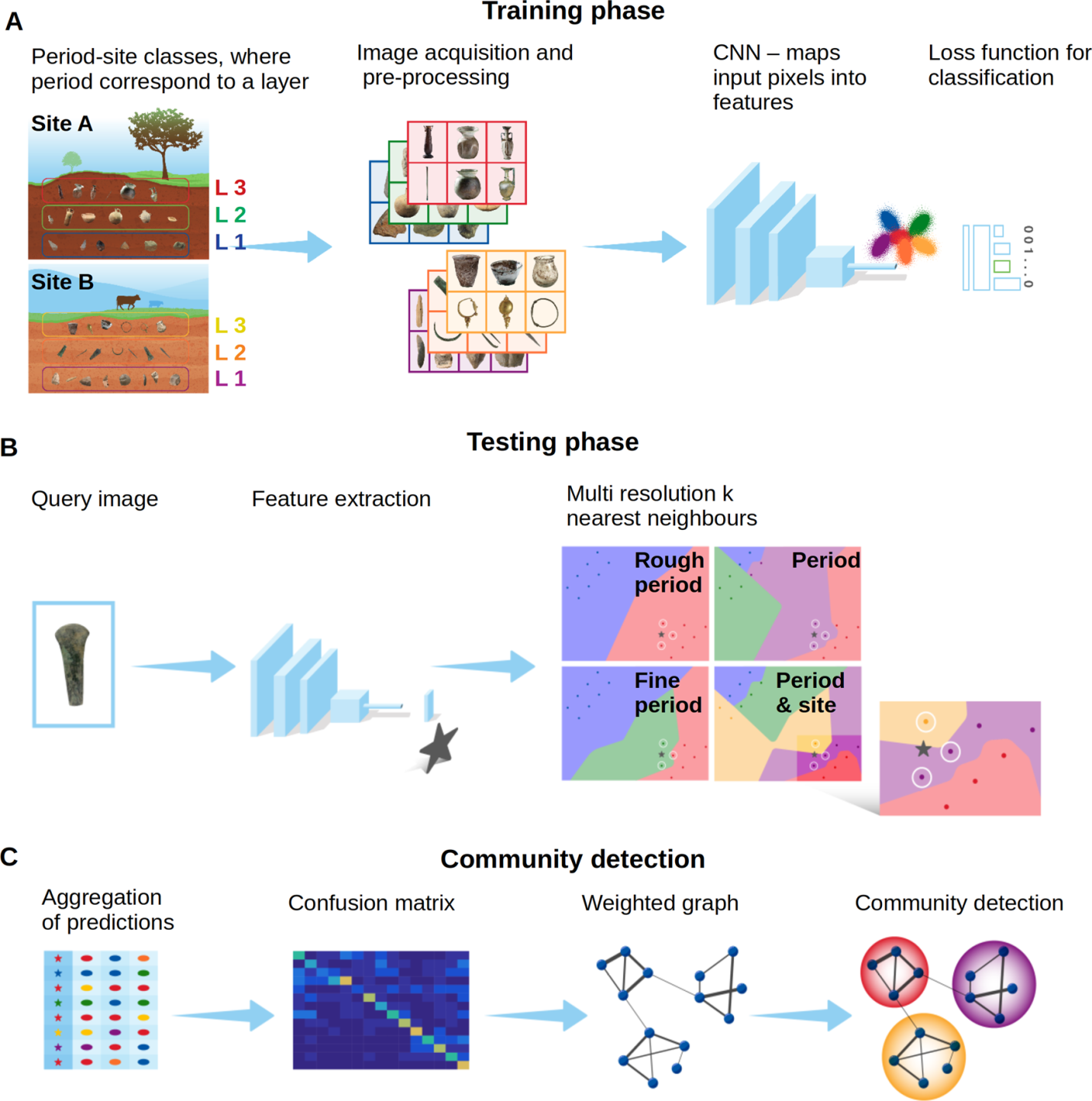
* training,
* testing/validating, and
* community detection.

Several thousand artifact images, ranging from the Lower Paleolithic period (1.4 million years ago) to the Late Islamic period (fourteenth century AD), were used to train the model (i.e., the CNN), to discern artifacts by site and period. After training, it attained a comparable accuracy to archaeologists in various periods. In order to test the model, it was called to identify new query images according to similarities with known (training) images. Validation blinding experiments showed that while archaeologists performed as well as the model within their field of expertise, they fell behind concerning other periods. Lastly, a community detection algorithm based on the confusion matrix data was used to discern affiliations across sites. A case-study on Levantine Natufian artifacts demonstrated the algorithm’s capacity to discern meaningful connections. As such, the model has the potential to reveal yet unknown patterns in archaeological data.

In order to automate this process and utilize computers’ excellent pattern recognition capabilities, efforts have been made to incorporate computer applications into the processes of archaeological classifications (Derech et al., [2021](https://www.nature.com/articles/s41599-021-00970-z#ref-CR15); Tal, [2014](https://www.nature.com/articles/s41599-021-00970-z#ref-CR40)). Notable among these are experimentations with machine learning models—computer algorithms that learn from data how to automatically detect patterns and make accurate decisions (Mitchell, [1997](https://www.nature.com/articles/s41599-021-00970-z#ref-CR29); Bishop, [2006](https://www.nature.com/articles/s41599-021-00970-z#ref-CR10); Duda and Hart, [1973](https://www.nature.com/articles/s41599-021-00970-z#ref-CR18)). Several attempts were made to apply machine learning to archaeological materials (Barcelo, [2008](https://www.nature.com/articles/s41599-021-00970-z#ref-CR7), [2016](https://www.nature.com/articles/s41599-021-00970-z#ref-CR8); Barceló and Bogdanovic, [2015](https://www.nature.com/articles/s41599-021-00970-z#ref-CR9); Díez-Pastor et al., [2018](https://www.nature.com/articles/s41599-021-00970-z#ref-CR16); Macleod, [2018](https://www.nature.com/articles/s41599-021-00970-z#ref-CR28)). However, at first, they relied on hand-crafted feature extraction, resulting in relatively poor performance measures (e.g., Boon et al., [2009](https://www.nature.com/articles/s41599-021-00970-z#ref-CR12)). More recently, machine learning algorithms have been used to extract relevant features automatically. Thus, for instance, Agam et al. ([2020](https://www.nature.com/articles/s41599-021-00970-z#ref-CR1)) combined Raman spectroscopy with machine learning algorithms to quantitatively estimate different degrees of thermal alteration on flint artifacts.

Of particular interest is deep learning, and more specifically, Deep Convolutional Neural Networks (CNNs), which are commonly used to analyse images. CNNs were successfully applied to various computer vision tasks, as they can automatically extract features from input images (Cifuentes-Alcobendas and Domínguez-Rodrigo, [2019](https://www.nature.com/articles/s41599-021-00970-z#ref-CR14); He et al., [2016](https://www.nature.com/articles/s41599-021-00970-z#ref-CR21); Krizhevsky et al., [2017](https://www.nature.com/articles/s41599-021-00970-z#ref-CR26); Taigman et al., [2014](https://www.nature.com/articles/s41599-021-00970-z#ref-CR39)). These features, also known as embeddings, are a set of numbers (1536 numbers in this case), that are later used by other computational layers, to classify/infer other useful information from input data. The features do not necessarily correspond to a realistic measure of the data, such as colour or shape. Applied to archaeological problems, CNNs have shown promise, successfully fulfilling tasks of ceramic classification (Itkin et al., [2019](https://www.nature.com/articles/s41599-021-00970-z#ref-CR23)), periodic discrimination of lithic assemblages (Grove and Blinkhorn, [2020](https://www.nature.com/articles/s41599-021-00970-z#ref-CR20)), and differentiation of bone surface modifications (Domínguez-Rodrigo et al., [2020](https://www.nature.com/articles/s41599-021-00970-z#ref-CR17)). However, these experiments with CNNs focused on narrow ranges of materials and contexts, consequently failing to seriously confront the bewildering diversity of the archaeological circumstances and record.

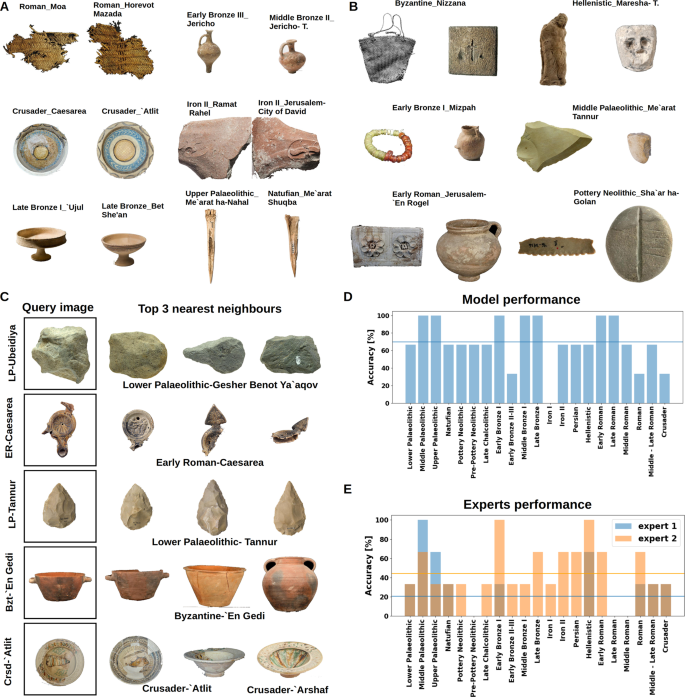
The base CNN was initially trained to classify everyday objects on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset which is a large dataset of natural images. Then, following some modifications to the CNN, the model was trained to identify archaeological artifacts according to period and site. Next, drawing on the model’s acquired capacity to correctly classify artifacts, it was determined whether it can be effectively used to detect communities)—cohorts of classes with a meaningful common denominator. Finally, a case-study is offered on communities found from Natufian culture (ca. 15,000–11,700 years ago) classes in the Levant, showing that this method found archaeologically meaningful similarities between different sites.



**A** The training phase: a dataset of images of archaeological artifacts were grouped according to period and site, pre-processed, and used to train a Convolutional Neural Network (CNN).

**B** The testing phase: the trained CNN was used to extract features from a query image and predict its class by identifying *k*-nearest neighbours in the training set.

**C** Community detection: validation set predictions are aggregated in a confusion matrix that is later transformed into a weighted graph and fed to a community detection algorithm.



**A** Nearest neighbour pairs of artefacts from different classes (the image on the left derives from the validation set, and the image on the right derives from the training set).

**B** Pairs of distinct images that derive from the same class.

**C** Validation set query images (left column) and the top-3 training set nearest neighbours.

**D** A histogram of model performance for fine-period prediction on 63 randomly picked images (3 images per period); the straight horizontal line marks the average prediction accuracy (69.84%).

**E** A histogram of two archaeologists' performances (blind experiments) for the same 63 artefact images used for D; the horizontal lines mark the average prediction accuracies for each archaeologist (44.44, 20.63%).

In order to optimise the CNN to the task of archaeological classification, the standard transfer learning procedure was followed. Transfer learning is usually used when the available database size for the target application is relatively small. In this case, in order to improve performance, a pre-trained CNN on another larger (unrelated) database is used as the starting point for the training process.

The CNN model was based on the ImageNet (Russakovsky et al., [2015](https://www.nature.com/articles/s41599-021-00970-z#ref-CR34)) pre-trained image classification model EfficientNetB3 (Tan and Le, [2019](https://www.nature.com/articles/s41599-021-00970-z#ref-CR41)), which was chosen for its superior performance (see below, methods). It was built by stacking many (hence deep) basic computation layers (convolutions, non-linearities, pooling, skip connections, etc.), striving to achieve the best balance between computation complexity and prediction accuracy. The model was pre-trained on the ImageNet ILSVRC dataset to predict an image’s category (class) out of 1000 possibilities, and reached 81.6% Top-1 and 95.7% Top-5 prediction accuracy (Top-k classification score computes the number of times the correct label is among the top k labels predicted). To perform the transfer learning, the original classification layers were removed and a customised classification layer was added (a fully connected layer, that transformed EfficientNetB3 embeddings, of size 1536, to 200 classes). To optimise the training, five models were trained with the same ImageNet initialisation, each generating a different feature vector, which we then used to produce a final feature vector. To improve robustness and enrich the database, a standard data augmentation techniques was applied. These include: random rotations, spatial shifts, zoom, and horizontal flips. All CNNs’ layers were trained for 25 epochs (in each epoch, the model is trained on the entire training set), using the categorical cross-entropy loss function (most common loss function for classification tasks).

resultIn an attempt to achieve better communities, two further adjustments were introduced.

1. The first consisted of rebuilding the confusion matrix to include ten nearest neighbour predictions for each query image instead of one. This modification resulted in more confusion and, by extension, a denser network with more edges.
2. The second adjustment was to use only certain part of the confusion matrix, with several neighbour periods, before applying community detection (e.g., Palaeolithic–Epipalaeolithic periods; Bronze–Iron Ages). In this manner, irrelevant confusion is precluded, and a way is paved to explore more nuanced relations among classes.

Methods

1. Image Preprocessing
2. Base network
3. Loss functions
4. Distance Metrics
5. Voting of five cnn’s
6. Training
7. Community detection
8. Prior confusion - Ambiguities concerning periodic attribution (e.g., Roman/Early Roman) may be considered a type of label noise. However, in practice, they are attributable to several closely related features of the archaeological record: (1) Most artefact types span several periods, (2) archaeological periods usually have vague boundaries, and (3) artefacts may vary in frequency across time and space while retaining their formal properties.

4.The role of computers to understand the past. The case of archaeological research –Juan A. Barcelo(IT Journal(Restrited))

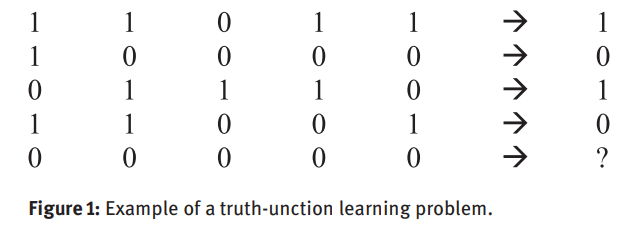
Abstract :- As practitioners of an historical discipline, archaeologists are in general, interested in guessing how ancient objects, buildings and/or landscapes were produced and used in the past from the object's visual appearance and material properties. In the pursuit of this general goal, archaeologists act as detectives looking for material cues (mostly visual) that may allow the discovering of social actions that may have happened in the past. This investigation can be made intuitively, but also formally, in which case we need computational tools and techniques to

* 1. extract necessary information from visual and non-visual data,
  2. data processing to evaluate relevant relationships between different items at different spatial and temporal scales
  3. build appropriate models that can be used to provide explanatory hypothesis about the human past.

In this paper I review the nature of archaeological problems and suggest computational techniques and methods that can be useful in solving these kinds of questions. Emphasis is made on modern classification and clustering techniques (neural networks, for instance), and computational simulation approaches.

Inductive Approach

The aim is then to program a system able to look for common features between positive examples of the known function to be predicted, and common differences between its negative examples. This task is exactly like an example of a truth-function learning problem as shown in Figure. The goal in a functional explanation is to develop an algorithm which will assign any artifact, represented by a vector 𝑥 of observed features, to one of 𝑐 classes (possible functions). The problem is to find the best mapping from the input patterns (descriptive features) to the desired response (expected functions), and a probabilistic or “truth likeness” measure to select between different solutions.



Abductive and deductive approach

Given some visual input and a candidate explanatory causal model, a correspondence can be established between them. This means that a small number of features are identified as matching features in the input and the model. Based on the corresponding features, a decision rule linking visual features with their causal process (social activity) is uniquely determined. The recovered decision rule is then applied to the model. Based on the degree of match, the candidate causal event is selected or rejected. To be accepted, the match must be sufficiently close, and better than that of competing solutions. Such a computer system would be based on questions like “Are the incoming features already contained in memory?” If the answer is affirmative it decides to remember what was memorized at that moment, and find out additional associate affirmations. The input pattern is then categorized as belonging to the class captured by that pre-existing explanation.

Beyond Simple generalization

A direct match between a perceived input and explanatory stored patterns is insufficient for various reasons:

– The space of all possible visualizations of all causal events is likely to be prohibitively large. It therefore becomes impossible to test a shape for property𝑃 by simply comparing it against all the members of 𝑆 stored in memory. To be more accurate, the problem lies not simply in the size of the set 𝑆, but in what may be called the size of the support of 𝑆. When the set of supports is small, the recognition of even a large set of objects can still be accomplished by simple means such as direct template matching. When the set of supports is prohibitively large, a template matching decision scheme will become impossible.

– Finding solutions may also affected by the presence of noise in the measurements, and insufficient number of measurements.

– Ambiguity of input data arising when several different causal events could have produced the same archaeological evidence description or visual features.

It is a Reverse Engineering (RE) approach, where we extract knowledge from anything human-made, by going backwards through its development cycle, analyzing its structure, function and operation. Using appropriate algorithms we can simulate the motion of an ancient item, and the results would determine the object’s behaviour by incorporating the effects of force, deformation and/or friction, where the parameters of possible behaviours can be successively changed and tested.

Through the development of computational methods for the simulated interaction with real objects made of solid materials, archaeologists intend to provide new insights into the complex dynamics of certain phenomena, such as event-based motion or kinematics.

Beyond the development of computer techniques and technologies for the proper documentation of ancient tools, buildings, and landscapes we can simulate virtual social agents “living” in a virtual environment that is an abstraction defined on the basis of social theory and/or historical data. Social simulation can be used as a virtual laboratory in which different techniques will be exploited to encourage the formalization and falsification of scientific hypotheses about social transformations. Specific case studies will supply datasets to validate and constrain the models.

* 5. Barceló, J. A. (2006). Automatic Archaeology. Bridging the gap between Virtual Reality, Artificial Intelligence and Archaeology. In Theorizing Digital Cultural Heritage A Critical Discourse. Edited by Fiona Cameron and Sarah Kenderdine. Cambridge MA: Massachusets Institute of Technology Press; pp. 437–456.
* 6. Barceló, J. A. (2009). Computational Intelligence in Archaeology. IGI Group, Information Science reference. Hershey (NY).
* 7. Barceló, J.A. (2010). Visual Analysis in Archaeology. An Artificial Intelligence Approach. In Morphometrics for nonmorphometricians. Edited by Elewa, Ashraf, M. T. Berlin: Springer Verlag (Lecture Notes in Earth Sciences, Vol. 124) pp. 51–101.
* 8. Clarke, D. (1968). Analytic Archaeology. Cambridge University Press.
* 14. Jordan, M. I., Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260.
* 27. Pasolli, E., Melgani, F., Donelli, M., Attoui, R., De Vos, M. (2008). Automatic detection and classification of buried objects in GPR images using genetic algorithms and support vector machines. In IEEE Geoscience and Remote Sensing Symposium, 2008. (Vol. 2, pp. II-525).
* 28. Pregesbauer, M., Trinks, I., Neubauer, W. (2014). An object oriented approach to automatic classification of archaeological features in magnetic prospection data. Near Surface Geophysics, 12(5), 651–656.
* 39. Deravignone, L., Blankholm, H. P., Pizziolo, G. (2015). Predictive Modeling and Artificial Neural Networks: From Model to Survey. In Mathematics and Archaeology, Edited by, J. A. Barceló and, I. Bogdanovic. Boca Raton (FL), CRC OPress, pp. 335–351
* . 40. Aprile, A., Castellano, G., Eramo, G. (2014). Combining image analysis and modular neural networks for classification of mineral inclusions and pores in archaeological potsherds. Journal of Archaeological Science, 50, 262–272.
* 41. Makridis, M., Daras, P. (2012). Automatic classification of archaeological pottery sherds. Journal on Computing and Cultural Heritage (JOCCH), 5(4), 15.
* 46. van der Maaten, L., Boon, P., Lange, G., Paijmans, H., Postma, E. (2006). Computer vision and machine learning for archaeology. Proceedings of Computer Applications and Quantitative Methods in Archaeology, 112–130.
* 47. Zhang, R. (2013, March). The ancient ceramics identification methods based on non-linear support vector machines. In Applied Mechanics and Materials Vol. 278, pp. 1201–1204.
* 48. Dries, M. H. V. D. (1998). Archaeology and the application of artificial intelligence: case-studies on use-wear analysis of prehistoric flint tools (Doctoral dissertation, Faculty of Archaeology, Leiden University).
* 49. Gardin, J. C. (1988) Artificial intelligence and expert systems: case studies in the knowledge domain of archaeology. Chichester, U. K.: Ellis Horwood.

5.Deep learning and taphonomy: high accuracy in the classification of cut marks made on fleshed and defleshed bones using convolutional neural networks – Gabriel Cifuentes-Alcobendas & Manuel Dominguez - Rodrigo

BSM is Bone Surface Modifications

Abstract

Until now, interpretations of Plio-Pleistocene BSM have been contentious because of the high uncertainty in discriminating among taphonomic agents. Recently, the use of machine learning algorithms has yielded high accuracy in the identification of BSM. A branch of machine learning methods based on imaging, computer vision (CV), has opened the door to a more objective and accurate method of BSM identification. The present work has selected two extremely similar types of BSM (cut marks made on fleshed an defleshed bones) to test the immense potential of artificial intelligence methods. This CV approach not only produced the highest accuracy in the classification of these types of BSM until present (95% on complete images of BSM and 88.89% of images of only internal mark features), but it also has enabled a method for determining which inconspicuous microscopic features determine successful BSM discrimination. The potential of this method in other areas of taphonomy and paleobiology is enormous.

Introduction

DL algorithms are powerful at many classification tasks, but they currently are the best method for computer vision, through image identification and classification. Neural networks make up the core of DL analyses. There is a diverse array of neural network topologies (i.e., recurrent, gated recurrent, feed forward, long/short term, auto encoder, variational, Markov Chain, Hopfield, Boltzmann, deep belief, liquid state machine, Kohonen, deep residual, neural Turing, deconvolutional, generative adversarial). Among these, deep convolutional neural networks (DCNN) currently are some of the most successful at image identification. A publication of the first use of computer vision in the field of taphonomy, showed how the use of DCNN on an experimental set of bone surface modifications, involving cut marks made with simple and retouched stone flakes and trampling marks, produced accuracy rates of correct classification >91% (>50% higher than human expert assessments)[1](https://www.nature.com/articles/s41598-019-55439-6#ref-CR1).

Method

Whole DCNN explained along with parameters and those results.

6.Machine learning algorithms applied to Raman spectra for the identification of variscite originating from the mining complex of Gavà - José Francisco Díez‐Pastor, Susana Esther Jorge‐Villar, Álvar Arnaiz‐González, César Ignacio García‐Osorio, Yael Díaz‐Acha, Marc Campeny, Josep Bosch& Joan Carles Melgarejo (Raman Spectra Journal)(Restricted)

Abstract

In this work, machine learning algorithms have been used to classify variscite samples from Gavà with regard to the identification of their mine of origin and extraction depth. The final objective of the study was to see if the Raman spectroscopic signatures selected by these algorithms had a key spectral significance related to mineral structure and/or composition and validate the use of these computational procedures as a useful tool for detecting variances in the mineral Raman spectra that could facilitate the assignment of the specimens to each mine.

Introduction

Tools for the automatic identification of Raman spectra can be classified into two main groups: (a) tools based on finding the correspondence between key spectral data characteristics previously identified by spectroscopists and

(b) tools based on machine learning that use the entire spectral range for multivariate analyses.

computational techniques is known as multivariate analysis; these techniques try to correlate the interrelationships amongst all the variables and have become the most commonly used today. The most frequently used technique in the literature has been the SVM method. Sometimes, ANN has been used in combination with dimensional reduction techniques, such as principal component analysis. ANNs have not been the only classifier used in combination with PCA analysis but also algorithms based on closest neighbours have also been used. Kelloway et al.[20] have used PCA analysis together with the Mahanalobis distance, and Carey et al.[21] used a custom trajectory similarity metric along with PCA analysis. In this work, the computational treatment of Raman spectra is described with the objective of finding Raman spectral differences for variscite specimens from the Gavà mining complex to assist in the determination of the individual mine and the depth of origin of the mineral. Furthermore, the spectroscopic interpretation of the observed key wavenumber positions selected by multivariate analyses will be studied with the purpose of relating them to chemical or structural mineral differences.

Preprocessing

An automatic spectrum recognition system generally has two stages: a preprocessing stage and a classification stage. In order to reduce the influence of noise and fluorescence background emission and also to eliminate intra‐species variations (within the same class) whilst maximizing the extra‐species distances (arising from different classes), a pipeline of preprocessing operations was applied.

1. Smoothing - Smoothing is an operation which allows an increase in the signal‐to‐noise ratio without greatly distorting the signal.
2. Cropping & interpolation - Cropping and interpolation operations are necessary because in machine learning, the model learns the relationships between attributes and classes. Therefore, to train a model, all instances need to have the same number of attributes, and the attributes have to represent the same characteristics, in this case the intensity, in a certain band. Similarly, to predict the class of an instance, it must have the same attributes as the instances used to train the model.
3. Baseline - it consists of a linear or non‐linear intensity addition which results in a distorted measurement, whereby the observed value is higher than its real value
4. Intensity Normalization - Intensity normalization preserves the relative order of band intensity values whilst mitigating the effect of peak intensity differences.

Classifiers

* Logistic Regression
* Ridge regression - Ridge regression is an improvement of LR that uses a regularization factor to facilitate the solution of ill‐posed problems which, in general, avoids overfitting and improves the generalization of the obtained models.
* SVM
* Linear Discriminant Analysis - When working with spectra and other data sets which have a large number of attributes, there is often the need to reduce their size. There are several linear projection techniques available to achieve this, one of the best known being PCA. PCA projects the data set into the direction which explains most of the variance in the data set, but as PCA is an unsupervised technique, it does not take into account the class labels and although it can be used in this way and indeed has been used as a processing method, it is not a classification method. Another well‐known technique is LDA (linear discriminant analysis) which finds a linear subspace that maximizes the class separability and then improve the results of a classifier with a linear decision boundary, generated by fitting class conditional densities to the data using Bayes' rule.
* Decision tree

Conclusion

Regarding the classification results, for the mine class the SVM algorithm was particularly useful and the classification was quite precise (approximately 90%). Results obtained for the depth class are almost as good as for the mine of origin, although more instances could help to improve the accuracy of the results.

7. Artifcial intelligence provides greater accuracy in the classifcation of modern and ancient bone surface modifcations Manuel Domínguez‑Rodrigo, Gabriel Cifuentes‑Alcobendas, Blanca Jiménez‑García, NataliaAbellán, Marcos Pizarro‑Monzo, Elia Organista & Enrique Baquedano(restricted)

Abstract

Bone surface modifcations are foundational to the correct identifcation of hominin butchery traces in the archaeological record. Until present, no analytical technique existed that could provide objectivity, high accuracy, and an estimate of probability in the identifcation of multiple structurallysimilar and dissimilar marks. Here, we present a major methodological breakthrough that incorporates these three elements using Artifcial Intelligence (AI) through computer vision techniques, based on convolutional neural networks. This method, when applied to controlled experimental marks on bones, yielded the highest rate documented to date of accurate classifcation (92%) of cut, tooth and trampling marks. After testing this method experimentally, it was applied to published images of some important traces purportedly indicating a very ancient hominin presence in Africa, America and Europe. The preliminary results are supportive of interpretations of ancient butchery in some places, but not in others, and suggest that new analyses of these controversial marks should be done following the protocol described here to confrm or disprove these archaeological interpretations.

8. Neural networks differentiate between Middle and Later Stone Age lithic assemblages in eastern Africa Matt Grove, James Blinkhorn (Restricted)

Abstract

The current paper employs a quantitative analytical framework based upon the use of neural networks to examine changing constellations of technologies between MSA and LSA assemblages from eastern Africa. Network ensembles were trained to differentiate LSA assemblages from Marine Isotope Stage 3&4 MSA and Marine Isotope Stage 5 MSA assemblages based upon the presence or absence of 16 technologies. Simulations were used to extract significant indicator and contra-indicator technologies for each assemblage class. The trained network ensembles classified over 94% of assemblages correctly, and identified 7 key technologies that significantly distinguish between assemblage classes. These results clarify both temporal changes within the MSA and differences between MSA and LSA assemblages in eastern Africa.

Method

The networks were trained using a Bayesian Regularization (BR) algorithm [62,63], which is more robust than many of the more often used back-propagation algorithms. BR is particularly useful for relatively small samples as its internal regularization procedure precludes the need for a separate validation set. BR networks generalize well (i.e. they tend not to be overfitted) because they retain for training only the non-trivial links between nodes, forming a more parsimonious network than would be used in a fully connected back-propagation network [63]. Weight and bias values were updated according to Levenberg-Marquardt optimization (also known as ‘damped least squares’ , an algorithm that also tends to produce results that generalize well. Performance during training was monitored via the sum of squared errors between the true and estimated classifications. Nodes in the hidden layer use a fast implementation equivalent of the hyperbolic tangent activation function, which has been shown to be superior to other sigmoid functions [66]. Nodes in the output layer of the 3-way classification use the softmax activation function; as this function exponentiates each input then divides by the sum of all exponentiated inputs, the outputs sum to unity and can be read as classification probabilities. Nodes in the output layer of the 2-way classification use the basic sigmoid activation function, which ‘squashes’ output to a range between zero and one. This output can be read directly as the probability of an LSA classification (and, equivalently, as 1 minus the probability of an MSA classification).

Typological Indicators

(. Baxt WG. Analysis of the clinical variables driving decision in an artificial neural networl trained to identify the presence of myocardial infarction. Annals of Emergency Medicine. 1992; 21(12):1439–44. https://doi.org/10.1016/s0196-0644(05)80056-3 PMID: 1443838

. Baxt WG, White H. Bootstrapping confidence intervals for clinical input variable effects in a networl trained to identify the presence of acute myocardial infarction. Neural Computation. 1995; 7(3):624–38. https://doi.org/10.1162/neco.1995.7.3.624 PMID: 8935962)

The ‘black box’ nature of neural networks is often criticised. To extract as much information as possible from the trained networks, we modify a simulation method proposed by Baxt [59,60](Given above) that facilitates the extraction of delta values, which can be interpreted in a similar way to the effect sizes extracted from a multinomial logistic regression.

Once a given network has been trained, the original data can be fed through it to determine the resulting classification. To measure how differences in the presence or absence of a particular technology would alter the classification, the data for that technology at each of the 92 assemblages are inverted (i.e. where the true data indicates presence, this is inverted to absence, and vice versa). These partially inverted data are then fed through the trained network, and delta values are calculated by subtracting the result of the original classification from the result of the classification based on the partially inverted data. This process is repeated, inverting the data for only one technology at a time, for each of the 16 technologies, and for each of 1,000 trained networks. Results are then aggregated, and median delta values per technology are presented separately for changes from presence to absence and from absence to presence. 95% confidence intervals are constructed from the 2.5th and 97.5th percentiles of the sampling distribution of the median to give a measure of confidence in the median delta values for each technology in each of the two conditions. Due to the use of the softmax activation function in the output layers of the networks, median delta values can be interpreted as, for example, the median increase in the probability of an assemblage being categorised as LSA when a particular technology that was previously present (absent) is subtracted from (added to) that assemblage. Technologies for which the 95% confidence intervals of the median delta value do not encompass zero for a given period / industry are considered significant indicators (or contra-indicators) of that period / industry. To ensure that this approach is a rigorous as possible, a given technology is only considered a significant indicator if the two inversions (presence inverted to absence and absence inverted to presence) are of opposite sign and the 95% confidence intervals of neither encompass zero. Significance is assessed separately for each technology within each period / industry.

Very deep paper read online again when needed

9. ImageNet classification with deep convolutional neural networks – Alex Krizhevsky, Ilya sutsvekar & Geoffrey Hinton (Communications of ACM)

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0%, respectively, which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully connected layers we employed a recently developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry

10. FaceNet: A unified embedding for face recognition and clustering – Florian Schroff, Dmitry Kalenichenko, James Philbin(IEEE) (Helping module to main factor)

Abstract

In this paper we present a system, called FaceNet, that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure offace similarity. Once this space has been produced, tasks such as face recognition, verification and clustering can be easily implemented using standard techniques with FaceNet embeddings asfeature vectors. Our method uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous deep learning approaches. To train, we use triplets of roughly aligned matching / non-matching face patches generated using a novel online triplet mining method. The benefit of our approach is much greater representational efficiency: we achieve state-of-the-artface recognition performance using only 128-bytes perface. On the widely used Labeled Faces in the Wild (LFW) dataset, our system achieves a new record accuracy of 99.63%. On YouTube Faces DB it achieves 95.12%. Our system cuts the error rate in comparison to the best published result [15] by 30% on both datasets.

11.EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks – Mingxin Tan,Quoc Lee(36th International conference on machine learning,PMLR 2019) (Helping to main module)

Abstract

Convolutional Neural Networks (ConvNets) are commonly developed at a fixed resource budget, and then scaled up for better accuracy if more resources are available. In this paper, we systematically study model scaling and identify that carefully balancing network depth, width, and resolution can lead to better performance. Based on this observation, we propose a new scaling method that uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective compound coefficient. We demonstrate the effectiveness of this method on scaling up MobileNets and ResNet. To go even further, we use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called EfficientNets, which achieve much better accuracy and efficiency than previous ConvNets. In particular, our EfficientNet-B7 achieves state-of-the-art 84.4% top-1 / 97.1% top-5 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet. Our EfficientNets also transfer well and achieve state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and 3 other transfer learning datasets, with an order of magnitude fewer parameters.

Introduction

Scaling up ConvNets is widely used to achieve better accuracy. For example, ResNet (He et al., 2016) can be scaled up from ResNet-18 to ResNet-200 by using more layers; Recently, GPipe (Huang et al., 2018) achieved 84.3% ImageNet top-1 accuracy by scaling up a baseline model four time larger. However, the process of scaling up ConvNets has never been well understood and there are currently many ways to do it. The most common way is to scale up ConvNets by their depth (He et al., 2016) or width (Zagoruyko & Komodakis, 2016). Another less common, but increasingly popular, method is to scale up models by image resolution (Huang et al., 2018). In previous work, it is common to scale only one of the three dimensions – depth, width, and image size. Though it is possible to scale two or three dimensions arbitrarily, arbitrary scaling requires tedious manual tuning and still often yields sub-optimal accuracy and efficiency.

Read the method explaining its scaling

12. Visualizing Data using t-SNE – Laurens Van Der Matten & Geoffrey Hinton (Helping)(Journal of machine research)

Abstract

We present a new technique called “t-SNE” that visualizes high-dimensional data by giving each datapoint a location in a two or three-dimensional map. The technique is a variation of Stochastic Neighbor Embedding (Hinton and Roweis, 2002) that is much easier to optimize, and produces significantly better visualizations by reducing the tendency to crowd points together in the center of the map. t-SNE is better than existing techniques at creating a single map that reveals structure at many different scales. This is particularly important for high-dimensional data that lie on several different, but related, low-dimensional manifolds, such as images of objects from multiple classes seen from multiple viewpoints. For visualizing the structure of very large data sets, we show how t-SNE can use random walks on neighborhood graphs to allow the implicit structure of all of the data to influence the way in which a subset of the data is displayed. We illustrate the performance of t-SNE on a wide variety of data sets and compare it with many other non-parametric visualization techniques, including Sammon mapping, Isomap, and Locally Linear Embedding. The visualizations produced by t-SNE are significantly better than those produced by the other techniques on almost all of the data sets.

Over the last few decades, a variety of techniques for the visualization of such high-dimensional data have been proposed, many of which are reviewed by de Oliveira and Levkowitz (2003). Important techniques include iconographic displays such as Chernoff faces (Chernoff, 1973), pixel-based techniques (Keim, 2000), and techniques that represent the dimensions in the data as vertices in a graph (Battista et al., 1994). Most of these techniques simply provide tools to display more than two data dimensions. Despite the strong performance of these techniques on artificial data sets, they are often not very successful at visualizing real, high-dimensional data. In particular, most of the techniques are not capable of retaining both the local and the global structure of the data in a single map. In this paper, we describe a way of converting a high-dimensional data set into a matrix of pairwise similarities and we introduce a new technique, called “t-SNE”, for visualizing the resulting similarity data. t-SNE is capable of capturing much of the local structure of the high-dimensional data very well, while also revealing global structure such as the presence of clusters at several scales. We illustrate the performance of t-SNE by comparing it to the seven dimensionality reduction techniques mentioned above on five data sets from a variety of domains. Because of space limitations, most of the (7+1)×5 = 40 maps are presented in the supplemental material, but the maps that we present in the paper are sufficient to demonstrate the superiority of t-SNE.

13.Very Deep Convolutional Networks for Large-Scale Image Recognition - Karen Simonyan & Andrew (helping)

Abstract

In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small ( 3 × 3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively. We also show that our representations generalise well to other datasets, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

14.