

Facsimiles-based deep learning for matching relief-printed decorations on medieval ceramic sherds

Khawla Brahim

Université d'Orléans, Laboratoire PRISME

khawla.brahim@univ-orleans.fr

Matthieu Exbrayat

Université d'Orléans, Laboratoire LIFO

Matthieu.Exbrayat@univ-orleans.fr

Sylvie Treuillet

Université d'Orléans, Laboratoire PRISME

Sylvie.Treuillet@univ-orleans.fr

Sebastien Jesset

Service Archéologique Municipal d'Orléans

Sebastien.Jesset@orleans-metropole.fr

Abstract

In this paper, we addressed the problem faced by archaeologists in associating relief-printed decorations on ceramic objects discovered during excavations carried out with the same wheel. This is crucial to understand the trade networks between regions, but highly complex and time-consuming task. We used two approaches: supervised classification or unsupervised clustering of 2D relief views generated from 3D scans of ceramic sherds. Inspired by experimental archaeology, we created wheel facsimiles to supplement significantly the database with numerous plausible and clearly identified samples. Taking advantage of the powerful convolutional neural network EfficientNet to extract reliable discriminating features, experimental results show that the facsimiles significantly improve the networks' training to achieve a classification accuracy exceeding 95% on real sherds. On the other hand, unsupervised spectral clustering from a vector reduced to a few hundred of the most significant features delivered by the network EfficientNet-B5 trained on ImageNet, without any fine-tuning, achieves an accuracy of 77.47% on our database. These results validate the strategy of using facsimiles to supplement a too-small data set and are very promising for the development of a computer-assisted archaeology tool for pattern-wheel association.

1. Introduction

Among the artifacts found on archaeological dig sites, ceramic fragments are often abundant, and they are a precious source of information for archaeologists seeking to understand the lifestyles of ancient civilizations or the trade networks between regions. Since 1994, a number of archaeological excavations in Saran (Loiret, France) have uncov-



Figure 1: Left : some excavated ceramic sherds from the Saran archaeological site (Loiret, France). Right: an example of a carved wooden thumbwheel used to print a frieze in clay.

ered numerous kilns and yielded a wealth of ceramic material testifying to mass production during the High Middle Ages (6th-11th centuries). Most of the ceramic sherds are decorated with a relief frieze, a process that was widespread in Europe in late Antiquity (see Fig. 1). At that time, potters personalized their creations with hand-engraved wooden thumbwheels imprinting a geometric frieze on fresh clay. The thumbwheel used by potters is a cylinder about two centimeters wide, usually made of wood, roughly notched to create geometric decoration on ceramic series. The original wheels are generally not found on site due to the relatively rapid decomposition of wood.

The pattern produced by the wheel application appears as an imprint of one to two millimeters and constitutes a

signature unique to each potter. As the lifespan of a wheel is relatively short (it degrades rapidly), the analysis of the decorations enables archaeologists to pinpoint the chronology of ceramic series and map the distribution of production: the similar pattern found on several vases may be associated with the same workshop or craftsman. Common patterns include sticks, squares, chevrons, and diamonds, often arranged in single or multiple lines. Archaeologists use a typology for these medieval ceramics (see Fig. 2).

The classic method used by archaeologists to capture decorations is manual stamping: applying modeling paste to the surface of the sherd, then inking to obtain a negative of the decoration on a sheet of paper (see Fig. 3). The black-and-white pattern is then scanned, vectorized, and stored in a database with other digital data (measurements, type and number of lines on the decoration, etc). Next, the archaeologist tries to determine visually whether the stamped motifs can be associated with the same wheel, possibly by superimposing the shapes on top of each other using transparent sheets. This association is a crucial step in archaeological research but is very complex and time-consuming. The manual and visual analysis reaches its limits in the face of the growing corpus unearthed during recent excavations (over 40,000 sherds unearthed between 2009 and 2012).

We have been working to develop advanced image processing and machine learning methods to overcome these challenges by developing some automatic classification. Past work has demonstrated the ability of convolutional neural networks (CNN) to successfully classify the types of patterns found according to the proposed nomenclature ([7]). In this paper, we seek to go a step further by clustering the patterns that could be linked to the same wheel. Variations in the manual application of the wheel (pressure and speed of the gesture), texture of the clay, shape of the pottery, changes over time, and the sometimes incomplete imprint on the sherd make association really difficult even for an experimented archaeologist.

One of the thorniest problems for machine learning is the lack of labeled data. Unfortunately, while the number of scanned sherds is quite large, there are very few samples in the available corpus already associated with the same wheel by the archaeologist (only 404 scanned sherds are associated with 5 thumbwheels). Increasing the sample base using conventional brightness or geometric image transformations would be of very limited use in our case since it would not really increase the number of wheel samples but simply the print variants. In order to create new, useful, and plausible examples, we have opted for an approach inspired by experimental archaeology, building facsimiles, i.e., new wooden wheels used to produce multiple impressions on modeling clay which are then scanned. In this way, no errors are introduced into the labels. This study aims to examine whether state-of-the-art networks exhibit consis-

tent learning using these facsimiles for supervised classification or whether unsupervised clustering algorithms could effectively help archaeologists in their analysis of excavated sherds.

The structure of this paper is as follows. The relevant literature and previous works are reviewed in the next section. Section 3 focuses on how the image database supplied for deep learning networks is created from 3D scans of real excavated sherds and facsimiles. Experimental results are presented and discussed in Section 4. The conclusions section ends the paper.

2. Related Work

Computer vision has been involved in this field for decades, proposing automatic processing to facilitate archaeologists' analysis. A variety of imaging techniques were used: not only color images [16] or 3D scans [9], but also hyperspectral [10], ultrasonic [27] or X-ray [21]. Since many of ceramic artifacts found on excavation sites are in the form of sherds of broken pottery, early published works include an automatic process for reassembling fragmented artifacts into a partially or fully reconstructed original vase [22, 15]. Other works aim to automate the classification of ceramics into different styles, either by shape or morphology, generally based on a surface of revolution to extract an axis and a profile [35, 37], by color and/or texture [16, 30, 10, 20, 13] or by the presence of decoration [11, 9, 14]. These works combine some conventional image processing techniques for extracting features with machine learning algorithms.

More recent publications use deep learning techniques, particularly convolutional neural networks (CNNs) to extract informative features and achieve accurate classification of archaeological ceramic artifacts. The superiority of deep learning has been demonstrated in several studies in comparison with approaches, such as Bag-of-Words, for content-based retrieval of three-dimensional vessel replicas [3] or classification of engraved decorations on ceramic sherds [6], by fine-tuning well-known CNN models, then using pooling methods to combine classifiers [7]. Arch-I-Scan project is developing a system for the automatic identification of Roman ceramic types (fabrics, forms, and sizes) using artificial intelligence to contribute to a better understanding of Roman foodways [32]. Likewise, ArchAIDE is a European project for developing two machine-learning tools that assist archaeologists in classifying potsherds [26, 1, 2]: one analyses the fracture outline shape of a pottery sherd, while the other is centered around its decorative features. An AI-based imaging solution was proposed to assist archaeologists in classifying Roman COMmonware POTtery (ROCO POT) database, which includes over 4000 potsherds extracted from 25 Roman pottery corpora [23]. Similar profiles are associated through unsuper-

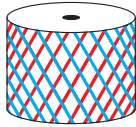


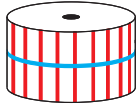


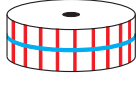


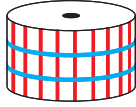


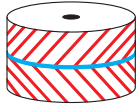


Engraving pattern for wooden wheels		Archaeological pattern results		
Oblique parallel lines crossed by other parallel lines				Diamonds (type A)
Axis-parallel lines crossed by one or perpendicular lines				Sticks (type C)
				Squares (type G)
Axis-parallel lines crossed by two perpendicular lines				Squares (type H)
A row of parallel oblique lines symmetrical to another row of parallel lines, crossed by a horizontal median line.	 <div style="display: flex; justify-content: center; gap: 10px; margin-top: 5px;"> — incisions 1 — incisions 2</div>		 <div style="display: flex; justify-content: center; margin-top: 5px;"> — Positive motifs on the wheel and recessed motifs on the ceramic</div>	Chevrons (type L)

Figure 2: Typology of medieval motifs based on the potter’s gesture and incisions on the thumbwheel (number of lines).

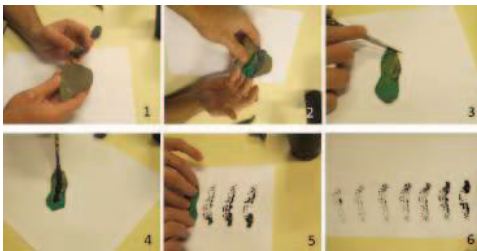


Figure 3: Manual stamping steps.

vised hierarchical clustering of non-linear features learned in the latent space of a deep convolutional Variational Autoencoder (VAE).

The protean nature of decoration on ceramics means that studies focus on a particular geographical area or civilization. Roman pottery in the works above, Kamares pottery, produced in Crete during the Minoan period in [11], Mycenaean Pottery from Cyprus in [19], Tusayan White Ware, ancient painted pottery from the American Southwest in

[24], potsherds belonging to the Jomon era in [21].

In our study, we focus on relief decorations dating from the High Middle Ages (6th-11th centuries), from which we derive representations in the form of grayscale images (shaded 2D views created from the 3D point cloud). As, [21], we’ll take advantage of the powerful EfficientNet model to extract relevant features for clustering and classification. We must not only identify a decoration style but also group together decorations made by the same tool or craftsman. A critical point is the lack of annotated data in the corpus of excavated ceramic objects to train the models adequately. We, therefore, propose to create facsimiles to compensate. This is an original strategy that we present to validate experimentally.

3. Materials and Methods

3.1. Data collection

The database is made up of ceramic material excavated in the archaeological digs at Saran (Loiret, France).



Figure 4: 3D scan and pattern extraction in a shaded view.

The manual stamping procedure was replaced by using an EinScan-SP scanner to create a 3D model of each sherd. Then, a 2D shaded view of the surface is generated (see Fig. 4). This image is easily produced by the scanner software, even by non-experts, and it provides a convenient grayscale relief map, capturing the decoration on the sherd. The classification performances obtained with this 2D shaded view input overpass those previously obtained with other image inputs like variance maps or binary images.

To date, we have over a thousand scans of excavated sherds, but only a few have been associated by archaeologists with five wheels. Following the methods of experimental archaeology, we have created facsimiles, i.e., new wooden wheels used to produce multiple impressions in modeling paste, which can be scanned. In so doing, the learning database can be supplemented by plausible and clearly identified samples without introducing errors. Table 1 presents our experimental dataset. Six-wheel facsimiles are included in the REMIA data set, bringing the number of wheels to 11 and the number of associated images to 2,561. The number of samples is given for each identified wheel (real excavated sherds and facsimiles, in Latin and Greek alphabet, respectively), grouping into the most representative types of decorations (squares, chevrons, and diamonds). Additionally, Fig. 5 shows an example of an imprint from each wheel (from excavated sherds and facsimiles).

The facsimiles were made using cylinders of hazelwood, about 2 cm in diameter and high. The fresh wood was notched with a knife, reaching a depth of around 2 mm. A hole is then drilled in the center of the cylinder to insert a metal rod that facilitates the rolling. The decoration is printed on a flattened strip of modeling paste about 10 to 15 cm long and 3 to 5 cm wide. In order to introduce variations, the direction, speed and pressure of rolling are deliberately modified at each pass, or even during the same pass for the last two factors. To reproduce the tool degradation observed on real archaeological decorations, we also deliberately degrade the wheel, after a certain number of passes, by removing parts of the pattern (a missing square, for example). The digitalization step with the scanner is exactly the same for facsimile and real excavated sherds. It should be noted that the facsimile pattern obtained differs from the archaeological pattern observed on pottery in that it is linear on a flat surface, whereas the pattern on pottery follows the

curved line of the ceramic surface.

3.2. Deep learning models

The association of the scanned decorations with the same wheel can be seen as an image classification or clustering problem. We choose to take advantage of state-of-art CNN models to use them for discriminative feature extraction from grey-levels images of the engraved pottery sherds. For supervised classification, models pre-trained on the ImageNet dataset are easy to use and can be refined on relatively small datasets for specific computer vision tasks using transfer learning [36]. To speed up development, we compare the Keras implementations of some EfficientNet models considered as the best in recent benchmarking for image classification [31]. The EfficientNet family (from B0 to B7) includes CNN designed to achieve a balance between model size and performance. The models feature simultaneous scaling of network depth, width, and resolution, resulting in highly efficient, high-performance models. From the baseline model EfficientNet-B0 to EfficientNet-B7, the models progressively become more profound, wider, and have higher-resolution input sizes. Table 2 compares the number of layers, total parameters, trainable parameters, and image input size for EfficientNet models. The classification step is realized by a fully connected network (FCN) from the high-dimensional feature vectors provided by the CNN.

3.3. Clustering approach

The clustering approach is based on the image features extracted from the EfficientNet variant pre-trained on ImageNet to feed an unsupervised image classification task, as suggested by [29]. A Principal Component Analysis (PCA) is applied to the outputs of the CNN to reduce the dimensionality before using the clustering. Among the clustering algorithms provided by the scikit-learn implementations [25], we tested the most popular ones, namely Spectral Clustering [34], K-means [12], Agglomerative Clustering [8], and BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) [33]. We utilized the default configuration without fine-tuning for all algorithms and kept the one given the best performance. Note that these algorithms require an estimated number of labels a priori.

4. Experiments

All experiments are implemented in Keras with the TensorFlow backend. The input image size is 224x224 pixels.

4.1. Evaluation protocol

For classification, a breakdown by class into 70% for training-validation and the remaining 30% for testing is applied to the data set. In order to evaluate the pertinence of

Table 1: REMIA database: number of samples by types of decoration and by identified thumbwheels (real excavated sherds and facsimiles, in Latin and Greek alphabet respectively).

Dataset	Type of wheels											Total
	Squares (type G)		Squares (type H)				Chevrons (type L)		Diamonds (type A)			
	γ	G8	η	H1	H12	H13	λ	$\lambda 1$	α	$\alpha 1$	A2	
Real excavated sherds	0	63	0	104	20	45	0	0	0	0	172	404
Facsimiles	450	0	450	0	0	0	150	462	150	495	0	2157
Total	450	63	450	104	20	45	150	462	150	495	172	2561

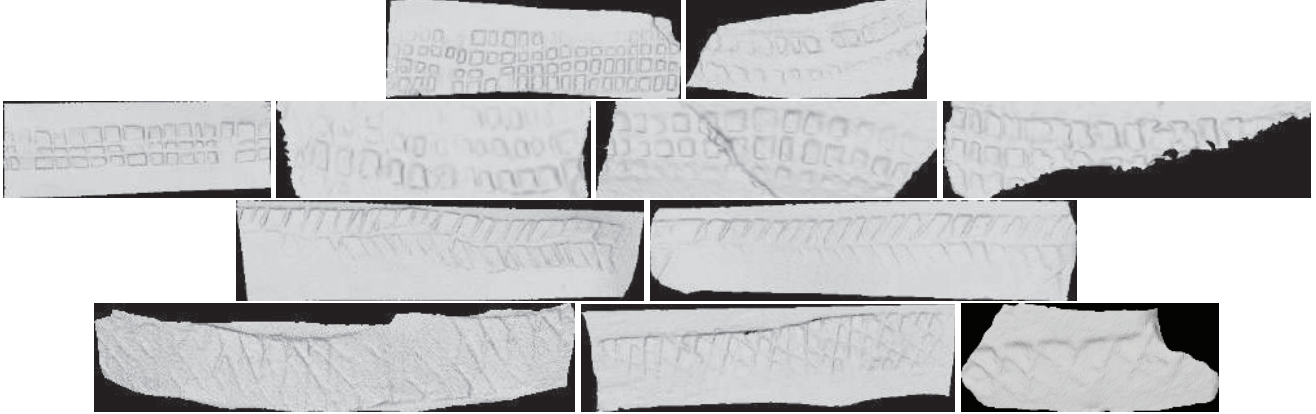


Figure 5: Samples of images from the REMIA dataset mixing scans of real excavated sherds and facsimiles. From top-left to bottom-right, associated thumbwheel’s labels are: γ , G8, η , H1, H12, H13, λ , $\lambda 1$, α , $\alpha 1$, and A2.

introducing facsimiles in the learning process, two configurations were tested (see Table 3): the first one introduces facsimiles in the learning and test phases, while the second relied exclusively on real excavated sherds for both phases.

4.2. Hyper-parameters tuning

Hyper-parameters’ significance lies in their direct control over the model’s behavior. Consequently, when hyper-parameters are finely tuned, they substantially impact the model’s performance. We used the Adam [17] optimizer to train varied epochs of each model convergence with a learning rate of 0.001 and a batch size of 32. Furthermore, we applied the weighted categorical cross-entropy loss function to the training, which computes the loss between the actual probability of the sherd’s category and the probability of the class predicted from the softmax activation function.

4.3. Evaluation metrics

The classification results are evaluated using the four standard metrics: Accuracy, Precision, Recall, and F1-score.

For the clustering approach, three standard metrics are used to evaluate the overlapping between predicted clusters and ground truth labels: the Normalized Mutual Information (NMI) [4], Adjusted Rand Index (ARI) [28] and Accuracy (ACC) defined as:

$$\text{NMI}(y, c) = \frac{2 \times I(y, c)}{H(y) + H(c)} \quad (1)$$

where y represents the ground-truth labels, c denotes the cluster assignments, $H(\cdot)$ is the entropy, and $I(y, c)$ is the mutual information between y and c .

$$\text{ARI} = \frac{\text{RI} - \text{Expected RI}}{\text{max RI} - \text{Expected RI}} \quad (2)$$

where RI quantifies the similarity between two cluster results. q

$$\text{ACC}(y, c) = \max_m \frac{\sum_{i=1}^n 1\{y_i = m(c_i)\}}{n} \quad (3)$$

where y_i is the ground truth labels, c_i is the cluster labels, and m enumerates mappings between predicted clusters and labels. Efficient computation of the optimal mapping can be achieved using the Hungarian algorithm [18].

All metrics fall within the range of 0 to 1, where 0 indicates a complete disagreement between two data clusters on any pair of points, while a value of 1 signifies perfect agreement, meaning the two clusters are identical.

4.4. Classification results

The classification results obtained by transfer learning on our REMIA database with EfficientNet variants are pro-

Table 2: Characteristics of the EfficientNet variants tested for classification.

Architectures	Number of layers	Total number of parameters	Trainable parameters	Image input size
EfficientNet-B7	813	66,658,687	66,347,960	600 × 600
EfficientNet-B5	576	30,562,527	30,389,784	456 × 456
EfficientNet-B3	384	12,320,535	12,233,232	300 × 300
EfficientNet-B0	237	5,330,571	5,288,548	224 × 224

Table 3: Experimental protocol for testing the use of facsimiles for wheel-based image classification.

Configurations	Data splitting	Type of wheels										
		Real excavated sherds					Facsimiles					
		A2	G8	H1	H12	H13	α	γ	η	λ	α_1	λ_1
With facsimiles	Training-Validation	120	44	72	14	32	105	316	315	105	346	324
	Test	52	19	32	6	13	45	134	135	45	149	138
Without facsimiles	Training-Validation	120	44	72	14	32	0	0	0	0	0	0
	Test	52	19	32	6	13	0	0	0	0	0	0

vided in table 4. Unsurprisingly, performance improves with deeper network versions from B0 to B7. The basic B0 version trained with an enhanced database with facsimiles already gives excellent results of more 97%, slightly improved by the B7 version to 97.79%. These results indicate that EfficientNet can identify relevant features from the grey-level images for wheel identification.

On the other hand, the configurations using only real sherds show a much greater increase in precision (about 6%) between versions B0 and B7 than with the introduction of facsimiles in the dataset (about 0.8%). Without incorporating facsimiles, convergence is always much longer (on the order of twice as long), which probably underlines a too-small database if facsimiles are not added. Table 5 presents the confusion matrix obtained with Efficientnet B7 trained on enhanced dataset with facsimiles. Less than 17 out of 768 sherds (i.e., only 2.2% of ceramics) were misclassified due to labels that are really hard to distinguish.

We used Grad-CAM++ as a tool for the explicability of the global network decision [5]. The saliency maps illustrated in Fig. 6 are overlaid on the input images to show the importance of each pixel in the form of heat maps (in red, high activation, in blue, low activation). These maps are related to the significant features associated with wheel identification. The network successfully focuses on discriminating areas to distinguish irregular patterns specific to each wheel.

4.5. Clustering results

A performance evaluation of the unsupervised clustering approach for wheel identification on the REMIA dataset is presented in Table 6. The image features are extracted using several EfficientNet variants (B0, B3, B5, and B7) pre-

trained on ImageNet, without any fine-tuning. To reduce dimensionality, a PCA is applied to the outputs of the networks. The number of components selected to feed the clustering algorithms corresponds to an explanatory cumulative variance of 95%. For example, we have chosen the first 257 components from EfficientNet-B5 outputs (see Fig. 7). The 3D projection of the clusters, based on the ground truth labels, can be seen in Fig. 8, illustrating their distribution according to the first three principal components.

The results of comparing four state-of-the-art clustering algorithms show that spectral clustering outperforms the other algorithms on the REMIA dataset for all observed metrics. Among the different networks tested, EfficientNet-B5 surpasses the B7 and proves to be the most effective one for the wheel identification task, achieving an accuracy rate of 77.47% when combined with spectral clustering. There are a few confusions, including the fact that the algorithm can't distinguish specific samples from the G8, H12, and H13 wheels; these samples need to be checked by the archaeologist.

5. Conclusion

In this study, we addressed the problem faced by archaeologists in associating relief-printed decorations on ceramic objects discovered during excavations carried out with the same wheel, using two approaches: supervised classification or unsupervised clustering. In both cases, we took advantage of the latest and most powerful networks, such as EfficientNet, to extract reliable discriminating features from 2D relief views generated from 3D scans of ceramic sherds. Given the small amount of data available, we were inspired by experimental archeology to develop our learning base by creating wheel facsimiles to significantly supplement

Table 4: Performance comparison of EfficientNet variants (bold indicates the best value per metric).

Methods	Configurations	Metrics (%)				Convergence rate	
		Accuracy	Precision	Recall	F1-score	Number of epochs	Time (s)
EfficientNet-B7	With facsimiles	97.79	97.32	97.79	97.54	48	1902
	Without facsimiles	83.61	84.18	83.61	82.96	87	611
EfficientNet-B5	With facsimiles	97.66	97.12	97.66	97.38	66	1568
	Without facsimiles	81.97	81.41	81.97	80.84	85	351
EfficientNet-B3	With facsimiles	97.40	97.12	97.40	97.22	72	1085
	Without facsimiles	81.15	82.17	81.15	81.22	86	222
EfficientNet-B0	With facsimiles	97.01	96.81	97.01	96.84	63	541
	Without facsimiles	77.87	78.77	77.87	77.28	101	158

Table 5: Confusion matrix obtained with EfficientNet-B7 trained on enhanced dataset with facsimiles.

Actual label	Classified by EfficientNet-B7 as										
	α	$\alpha1$	A2	γ	G8	η	H1	H12	H13	λ	$\lambda1$
α	45	0	0	0	0	0	0	0	0	0	0
$\alpha1$	0	149	0	0	0	0	0	0	0	0	0
A2	0	0	52	0	0	0	0	0	0	0	0
γ	0	0	0	134	0	0	0	0	0	0	0
G8	0	0	0	0	18	0	1	0	0	0	0
η	0	0	0	0	0	135	0	0	0	0	0
H1	0	0	1	0	1	0	25	1	4	0	0
H12	0	0	0	0	0	0	4	0	2	0	0
H13	0	0	0	0	0	0	3	0	10	0	0
λ	0	0	0	0	0	0	0	0	0	45	0
$\lambda1$	0	0	0	0	0	0	0	0	0	0	138

the database with numerous plausible and clearly identified samples. Note that the alternative of conventional augmentation techniques by image transformation would not have added any new wheel to associate. In fact, translation or cropping the images could be useful to multiply the number of printed representations of existing wheels, as could a double vertical-horizontal inversion to simulate an opposite rolling direction (right- or left-handed application). But rotations and other inversions should be totally excluded. Therefore, introducing facsimiles is a much better way of increasing the database consistently, even if we could add a few of the transformations mentioned above. In the future, facsimiles may also include printing on curved supports to get closer to real pottery.

Experimental results show that adding these facsimiles significantly improves the networks' training to achieve excellent classification performance up to 97.79%, obtained with EfficientNet-B7. On the other hand, unsupervised spectral clustering from a vector reduced to a few hundred of the most significant features delivered by the network EfficientNet-B5 trained on ImageNet without any fine-

tuning achieves an accuracy of 77.47% on our database.

This experimental study offers promising prospects for developing computer-assisted archaeology for the pattern-wheel association, a highly complex and time-consuming task. The aim is to develop pattern identification within a larger corpus, with an interactive strategy between automatic classification and clustering and archaeologist. Firstly, by identifying sherds that can be associated with referenced wheels from the trained models and then extracting new groups from the remainder to create further wheel references. The idea would be for the machine to suggest associations to the archaeologist if they exist in this gigantic corpus. Once validated by the archaeologist, these new associations could, in turn, be iteratively exploited in the network's learning process. Future work will also focus on weakly supervised learning techniques since the corpus of images is without a priori wheel labels. As many Late Antique ceramics discovered in Europe are decorated with a relief frieze, these automatic classification tools would be useful far beyond our Loiret excavation site.

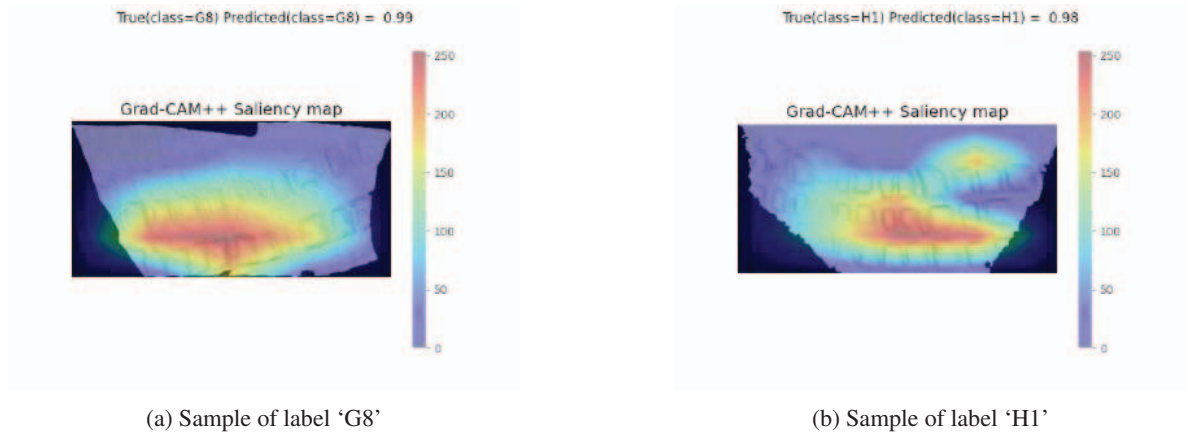


Figure 6: Grad-CAM++ saliency maps for two samples showing regions of high activation in the EfficientNet-B7 model by hot areas (red) versus low activation areas (blue).

Table 6: Performance comparison of unsupervised clustering algorithms: Spectral clustering (in black), BIRCH clustering (in green), Agglomerative Clustering (in blue), and K-means (in red).

Architectures	Metrics (%)			Shape of features	Number of components
	ACC	NMI	ARI		
EfficientNet-B7	76.53	76.31	69.86	(2561, 2560)	193
EfficientNet-B5	77.47	78.52	72.33	(2561, 2048)	257
	72.71	73.94	65.41	(2561, 2048)	257
	65.87	69.64	58.20	(2561, 2048)	257
	61.77	66.46	50.64	(2561, 2048)	257
EfficientNet-B3	76.65	78.08	71.84	(2561, 1536)	295
EfficientNet-B0	58.14	68.13	48.36	(2561, 1280)	237

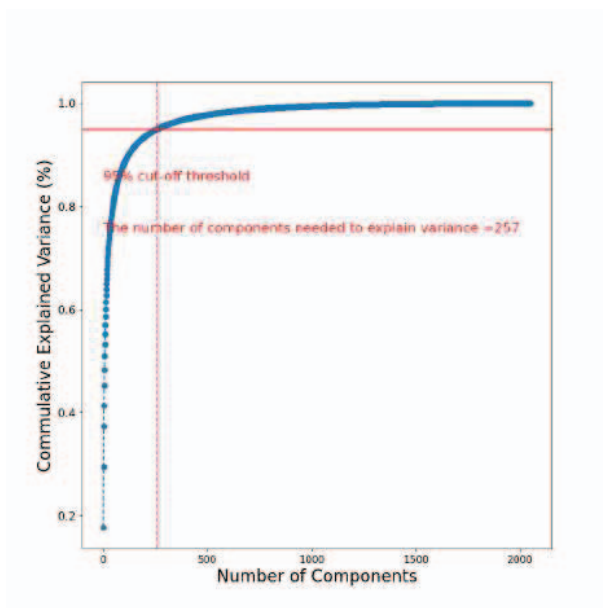


Figure 7: Cumulative variance explained as a function of the number of retained components

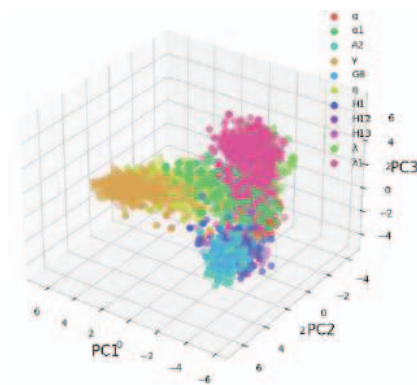


Figure 8: Clusters projections on the first 3 main components based on ground truth labels.

6. Acknowledgements

This work was part of the REMIA project funded by the Region Centre-Val de Loire.

References

- [1] Francesca & al. Anichini. Developing the archaide application: a digital workflow for identifying, organising and sharing archaeological pottery using automated image recognition. *Internet Archaeology*, 2020, num. 52, 2020.
- [2] Francesca & al. Anichini. The automatic recognition of ceramics from only one photo: The archaide app. *Journal of Archaeological Science: Reports*, 36:102788, 2021.
- [3] Halim Benhabiles and Hedi Tabia. Convolutional neural network for pottery retrieval. *Journal of Electronic Imaging*, 26(1):011005–011005, 2017.
- [4] Deng Cai, Xiaofei He, and Jiawei Han. Locally consistent concept factorization for document clustering. *IEEE Transactions on Knowledge and Data Engineering*, 23(6):902–913, 2010.
- [5] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE, 2018.
- [6] A. & al. Chetouani. Classification of ceramic shards based on convolutional neural network. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1038–1042. IEEE, 2018.
- [7] A. & al. Chetouani. Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling. *Pattern Recognition Letters*, 131:1–7, 2020.
- [8] Ian Davidson and SS Ravi. Agglomerative hierarchical clustering with constraints: Theoretical and empirical results. In *Knowledge Discovery in Databases: PKDD 2005: 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, Porto, Portugal, October 3-7, 2005. Proceedings 9*, pages 59–70. Springer, 2005.
- [9] T. & al. Debrouelle. Automatic classification of ceramic sherds with relief motifs. *Journal of Electronic Imaging*, 26(2):023010–023010, 2017.
- [10] Mercedes Farjas, Juan Gregorio Rejas, Teresa Mostaza, and Julio Zancajo. Deepening in the 3d modelling: multisource analysis of a polychrome ceramic vessel through the integration of thermal and hyperspectral information. In *Conf. on Computer Applications and Quantitative Methods in Archaeology (CAA)*, pages 116–124, 2012.
- [11] GC Guarnera, F Stanco, D Tanasi, and G Gallo. Classification of decorative patterns in kamares pottery. In *M. Samuelcik (a cura di), Proceedings of SCCG 26th Spring Conference on Computer Graphics*, pages 20–23, 2010.
- [12] John A Hartigan, Manchek A Wong, et al. A k-means clustering algorithm. *Applied statistics*, 28(1):100–108, 1979.
- [13] Irmgard Hein, Alfonso Rojas-Domínguez, Manuel Ornelas, Giulia D’Ercole, and Lisa Peloschek. Automated classification of archaeological ceramic materials by means of texture measures. *Journal of Archaeological Science: Reports*, 21:921–928, 2018.
- [14] Alice MW Hunt. *The Oxford handbook of archaeological ceramic analysis*. Oxford University Press, 2017.
- [15] Ioannis Kalasarinis and Anestis Koutsoudis. Assisting pottery restoration procedures with digital technologies. *International Journal of Computational Methods in Heritage Science (IJCMHS)*, 3(1):20–32, 2019.
- [16] Martin Kampel and Robert Sablatnig. Color classification of archaeological fragments. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 4, pages 771–774. IEEE, 2000.
- [17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [18] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [19] Lynne A Kvapil, Mark W Kimpel, Rasitha R Jayasekare, and Kim Shelton. Using gaussian mixture model clustering to explore morphology and standardized production of ceramic vessels: A case study of pottery from late bronze age greece. *Journal of Archaeological Science: Reports*, 45:103543, 2022.
- [20] Michael Makridis and Petros Daras. Automatic classification of archaeological pottery sherds. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(4):1–21, 2013.
- [21] Israel Mendonça, Mai Miyaura, Tirana Noor Fatyanosa, Daiki Yamaguchi, Hanami Sakai, Hiroki Obata, and Masayoshi Aritsugi. Classification of unexposed potsherd cavities by using deep learning. *Journal of Archaeological Science: Reports*, 49:104003, 2023.
- [22] Georgios Papaioannou and Evaggelia-Aggeliki Karabassi. On the automatic assemblage of arbitrary broken solid artefacts. *Image and Vision Computing*, 21(5):401–412, 2003.
- [23] S. & al. Parisotto. Unsupervised clustering of roman potsherds via variational autoencoders. *Journal of Archaeological Science*, 142:105598, 2022.
- [24] Leszek M Pawłowicz and Christian E Downum. Applications of deep learning to decorated ceramic typology and classification: A case study using tusayan white ware from northeast arizona. *Journal of Archaeological Science*, 130:105375, 2021.
- [25] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [26] Julian Daryl & al. Richards. Archaide-archaeological automatic interpretation and documentation of ceramics. In *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*. York, 2016.
- [27] Addison Salazar, Gonzalo Safont, Luis Vergara, and Enrique Vidal. Pattern recognition techniques for provenance classification of archaeological ceramics using ultrasounds. *Pattern Recognition Letters*, 135:441–450, 2020.
- [28] Jorge M Santos and Mark Embrechts. On the use of the adjusted rand index as a metric for evaluating supervised classification. In *Artificial Neural Networks–ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II 19*, pages 175–184. Springer, 2009.

- [29] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.
- [30] Patrick Smith, Dmitriy Bespalov, Ali Shokoufandeh, and Patrice Jeppson. Classification of archaeological ceramic fragments using texture and color descriptors. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 49–54. IEEE, 2010.
- [31] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.
- [32] Ivan & al. Tyukin. Exploring automated pottery identification [arch-i-scan]. *Internet Archaeology*, 50(10.11141), 2018.
- [33] Sami Virpioja. Birch: Balanced iterative reducing and clustering using hierarchies, 2008.
- [34] Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 274–285. SIAM, 2005.
- [35] Andrew Willis, Xavier Orriols, and David B Cooper. Accurately estimating sherd 3d surface geometry with application to pot reconstruction. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, volume 1, pages 5–5. IEEE, 2003.
- [36] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.
- [37] SY Zheng, RY Huang, J Li, and Z Wang. Reassembling 3d thin fragments of unknown geometry in cultural heritage. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:393–399, 2014.