



Research Papers

Unsupervised clustering of Roman potsherds via Variational Autoencoders

Simone Parisotto^{a,*}, Ninetta Leone^b, Carola-Bibiane Schönlieb^a, Alessandro Launaro^b^a Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK^b Faculty of Classics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DA, UK

ARTICLE INFO

Keywords:

Variational Autoencoders
 Hierarchical clustering
 Pottery studies
 Roman archaeology
 Commonware pottery
 Deep learning
 Unsupervised learning
 Machine learning
 Artificial intelligence
 Shape analysis
 Shape matching
 Heritage science

ABSTRACT

In this paper we propose an artificial intelligence imaging solution to support archaeologists in the classification task of Roman commonware potsherds. Usually, each potsherd is represented by its sectional profile as a two dimensional black–white image and printed in archaeological books related to specific archaeological excavations. The partiality and handcrafted variance of the fragments make their matching a challenging problem: we propose to pair similar profiles via the unsupervised hierarchical clustering of non-linear features learned in the latent space of a deep convolutional Variational Autoencoder (VAE) network. Our contribution also include the creation of a Roman Commonware POTtery (ROCOPT) database, with more than 4000 potsherds profiles extracted from 25 Roman pottery corpora, and a MATLAB GUI software for the easy inspection of shape similarities. Results are commented both from a mathematical and archaeological perspective so as to unlock new research directions in both communities.

1. Introduction

Our ability to *interpret* an archaeological site rests on our capacity to recognise the material culture found in it, ranging from fixed structures to movable objects. Archaeological finds from any site do in fact contribute to defining its chronology, function and place within a broader network of relationships with other sites. Thanks to their relative resilience against decay combined with their specific underlying patterns of production, distribution and consumption, ceramic vessels (pottery) – whether fragmented (i.e. potsherds) or intact – represent some of the most common finds recovered during archaeological fieldwork. Individual pots and potsherds are usually recorded as 2D profiles and, in consideration of their morphological features (e.g. shape of the rim or the base), gathered in systematic catalogues (*corpora*), where patterns of similarity are used to establish relationships, in terms of function, chronology or both (Orton and Hughes, 2013).

Within Roman archaeology specifically (albeit not exclusively), fundamental significance has been ascribed to those ceramic classes more closely associated with long-distance trade and contact, namely trade containers (*amphorae*) and high-quality tableware pots (*finewares*). However, these were middle-range commodities that did not reach all levels of the society in the same way, and, furthermore, their supply varied enormously over space and time. Indeed, some sites might have had limited or no access to this range of objects, and such notable absence in the archaeological record might indicate their abandonment

at a time when they were in full occupation (i.e. reduced archaeological visibility). In contrast, ordinary table- and kitchen-ware (*commonwares*, also referred to as *coarsewares*) were considerably cheaper and mainly supplied within a local/regional network of distribution. As a result, they almost invariably constitute the bulk of pottery finds at almost every Roman site. Even though one would expect them to provide a most effective baseline for the dating (and more general interpretation) of Roman sites anywhere, their huge range of forms, poorly defined chronologies and scattered provenance have made their study so challenging (and so little promising) that many have favoured the analysis of the far more standardised and easily recognisable finewares and amphorae.

Nevertheless, when a special effort is made to include a comprehensive study of commonware sherds, resulting interpretations can dramatically change. This was neatly shown by a recent analysis of the chronological distribution of potsherds recovered from the Roman town of Interamna Lirenas and from across its surrounding countryside (Central Italy) (Launaro and Leone, 2018). Whereas traditional reliance on finewares and amphorae had outlined a trajectory of precocious demographic decline (already in progress by the late 1st century BC), the widespread presence of commonwares shows a considerably more gradual process of growth, in fact peaking in the course of the 1st century AD, with little or no sign of decline until about two centuries

* Corresponding author.

E-mail addresses: sp751@cam.ac.uk (S. Parisotto), nl343@cam.ac.uk (N. Leone), cbs31@cam.ac.uk (C.-B. Schönlieb), al506@cam.ac.uk (A. Launaro).

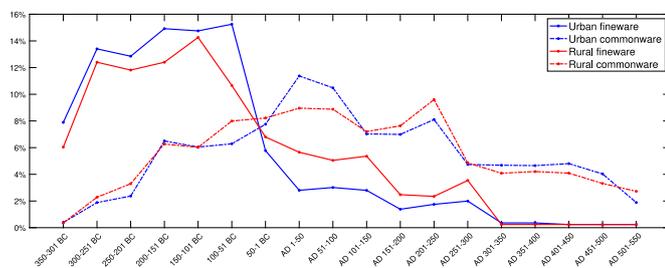


Fig. 1. Percentage of fineware (straight) and commonware (dashed) potsherds: urban (blue) vs. rural (red) environment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

later, see Fig. 1. There can be no doubt that the study of commonwares should be a priority within (Roman) material culture studies. However, this does not make the obstacles which archaeologists have to face any less real. What is indeed needed is to improve and expand our classification of this vast and varied body of evidence in a way which is considerably more effective and less time-consuming than it currently is.

Contributions. The contribution of this paper is two-fold. Firstly, we create the *ROman COmmonware POTtery* (ROCOPOPOT) database containing two-dimensional black–white imaging profiles of commonware potsherds attested in Central Tyrrhenian Italy, with metadata extrapolated from different corpora and excavation sites. Secondly, we propose an unsupervised workflow for matching and hierarchically clustering the potsherds profiles by comparing their latent representation learned in a deep convolutional Variational Autoencoder (VAE) network, and supported by a MATLAB GUI software for the easy inspection of the results on the field.

Scope. Our research aims to support archaeologists in the classification task of Roman commonware potsherds by unveiling new similarity patterns between individual profiles as featured in relevant corpora, especially those for which no match has so far been proposed. The resulting clustering patterns will highlight new morphological relationships and serve as an additional tool for archaeologists in order to improve their understanding of chronology, distribution, function and development of Roman commonwares over time. Whereas other similar projects have attempted to develop automated procedures for the correct identification of potsherds (i.e. establishing a relationship between what is newly found and the published 2D profiles), we aim to improve the internal consistency of the database itself and facilitate the archaeological classification workflow. On top of that, our workflow is fully scalable, meaning that once new profiles are available as black–white images, the learned representation features will unlock additional matches. The complete pipeline, from archaeological fieldwork to our contributions, is reported in Fig. 2.

Related works. Shape recognition, classification and matching of complete or partial data, are well-established problems in the computer vision and mathematical communities but for settled databases of thousands of images or shapes, e.g. Bronstein et al. (2009), LeCun et al. (2010) and Shilane et al. (2004). In recent years, “Cultural Heritage Imaging Science” became an increasing popular research field bringing together experts from museum institutions, history of art, physics, chemistry, computer vision and mathematics departments for tackling cross-discipline challenging applications; however, few dedicated databases emerged in the literature, mainly related to collection of paintings and associated tasks, e.g. for object or people detection (Crowley and Zisserman, 2015; Ginosar et al., 2015; Gonthier et al., 2018), style recognition (Bar et al., 2015) and many more, see Fiorucci et al. (2020).

Relevant works for matching 2D shapes are based on similarities of boundaries (Cui et al., 2009; Sun and Super, 2005) or skeletons (Shen et al., 2013), homoeomorphic transformations based on size functions (d’Amico et al., 2006), geodesic calculus for image morphing (Rumpf and Wirth, 2013), axiomatic criteria for deformable shapes (Bronstein et al., 2007) or comparison of invariant image moments (Kim and Kim, 2000; Manay et al., 2006; Hu, 1962; Žunić and Žunić, 2014). In particular, invariant shape descriptors are detailed in Cao et al. (2008) for clustering together similar shapes while the meaningfulness of the obtained clusters is still an open question. For 3D surfaces (with triangular meshes) or volumes (with voxels), remarkable results are obtained via scale invariant descriptors based on the eigen-decomposition of the Laplace–Beltrami operator, e.g. the *Heat Kernel Signature* and its variants (Litany et al., 2017; Raviv et al., 2010; Sun et al., 2009).

For the targeted application of this paper, even fewer databases are publicly available (Gunia et al., 2012; Tyers, 1996; University of Southampton, 2014), with focus on their automatic digitisation, shape extraction and visual presentation (ArchAIDE Consortium, 2019) or the clustering of a-priori manually selected geometric shape features into similar classes, e.g. by comparing curve skeletons (Piccoli et al., 2015), shape boundaries (Smith et al., 2012), shape descriptors (Seidl et al., 2015) or employing Generalised Hough Transform distance measures (in the case of petroglyphs, a dataset that shares similarity with our images) (Zhu et al., 2010) and other topological features (e.g. pixels height and width, area, circularity, rectangularity, diameters or steepness indexes) (Christmas and Pitts, 2018; Hörr et al., 2014) or comparison with template primitives (Kampel et al., 2001). In Banterle et al. (2017) fragments and nodal points are also extracted from complete 3D models in view of a supervised deep learning approach starting from complete profiles. Recently, it is worth mentioning the deep-learning approach based on convolutional neural networks (CNNs) proposed in Pawlowicz and Downum (2021), which targeted the imaging classification of decorative styles in Tusayan White Ware (Northeast Arizona).

In contrast, our deep-learning approach is not biased by the selection of a-priori shape features, allowing for their *unsupervised* extraction by means of a deep convolutional Variational Autoencoder (VAE) network. Here, we force the VAE latent space to follow a prior probability distribution close to a standard Gaussian. Such regularisation has many advantages for unsupervised tasks mainly related to the easy reparametrisation of the variational lower bound and backpropagation through the stochastic layers (Kingma and Welling, 2014). Our approach is motivated by the high availability of fragments and the generative process of VAE, allowing us to extend the learned model to unseen data for extracting their shape features.

In pattern recognition, hierarchical clustering is a non-parametric yet versatile unsupervised approach for unveiling inherent structures in data, ordered in a tree called *dendrogram* (Duda et al., 2000). The method is based on the recursive partitioning of the VAE features into clusters of (increasing or decreasing) cardinality, based on the minimisation of a certain cost function that promotes the separability of the data features (Cohen-Addad et al., 2019). Hierarchical clustering methods are easy to implement and they are often tuned to the application at hand, with an external evaluation of the results by the experts in the applied field (Veksler, 2004).

Specifically to Roman pottery profiles, our approach for an *unsupervised hierarchical clustering* is in line with the promising works in Hörr et al. (2014), Karasik and Smilansky (2011), Smith et al. (2012), Zhu et al. (2010) and Navarro et al. (2021), with the additional automatization of the cluster merging rule based on the best cophenetic coefficient score (Sokal and Rohlf, 1962). Finally, we leave the check of the deeper clusters in the produced dendrograms to specialist archaeologists, who can effectively assess the performances of the proposed workflow.

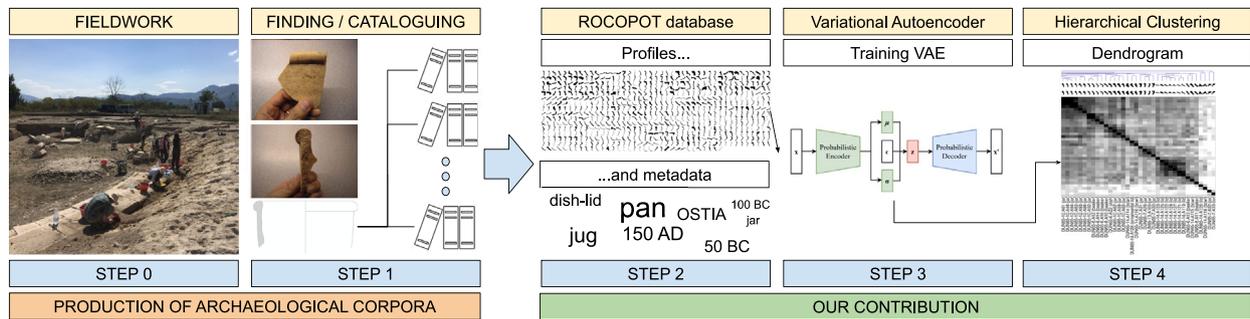


Fig. 2. Proposed workflow for clustering Roman potsherds.

Organisation of the paper. The paper is organised as follows: in Section 2 we introduce the *ROman COmmonware POTtery* (ROCOBOT) database; in Section 3 we detail about the deep variational autoencoder (VAE) network for extracting the potsherds features and the hierarchical clustering algorithm (with appendix on computations in Appendix); in Section 4 we discuss the results obtained by our workflow.

2. The ROCOPOT database

The *ROman COmmonware POTtery* (ROCOBOT) database, downloadable from Parisotto et al. (2021), comprises of more than 4000 black-white images, consisting of two-dimensional representation of the section (profile) of Roman commonware vessels. These profiles are featured in a series of archaeological corpora (Bragantini, 1996; Cicirelli, 1996; Cipriano and Fabrizio, 1996; Carbonara and Messineo, 1991; Chiaramonte Treré, 1984; De Caro et al., 1994; Giovanni, 1996; Duncan, 1964, 1965; Dyson, 1976; Federico, 1996; Fulford and Peacock, 1984, 1994; Gasperetti, 1996; Antonio, 1977; Olcese, 1993; Carandini, 1968; Berti, 1970; Carandini and Panella, 1973; Anselmino, 1977; Papi, 1985; Zevi and Pohl, 1970; Roberts, 1997; Höricht, 1996; Stanco, 2001) which provide a representative sample of a wide array of vessel-forms attested across the Tyrrhenian side of the Italian peninsula in the Roman period (from Liguria to Campania). In this work we present the version 1.0 of our database (future extensions are planned). All these images are scanned at 300dpi from the archaeological corpora: these are saved in the lossless .png format with an identification string filename of the form IDCAT-PAGNUM.FIGID.png, where IDCAT identifies the catalogue while PAGNUM and FIGID are the page and the figure identification numbers of the printed profile in the corpora, respectively. This labelling convention is suitable for a fast inspection of the results described in Section 4: by assuming similar shapes are presented closely in each catalogue (as is normally the case), we can quickly look at filenames to evaluate potential matches.

Since the original profiles can be composed of multiple (rarely intact) parts, we refined the database identifying a total of 407 bases (B), 450 handles (H), 3678 rims¹ (R) and 451 rims with handles (RH), as well as the original shapes (O). The archaeologists identified such parts out of each original profile depending on its grade of completeness. Potential outliers like one of a kind profiles were removed. Details about the number of profiles in our database are summarised in Table 1 (see also the map in Fig. 5), where for each catalogue we highlight the bibliographic reference, the publication year, the chronology and location of the archaeological site. For the rest of this work we focused only on the rims as they are assumed to carry out the most significant geometric information of the handcrafted material.

All the rims require a further polishing step so as to make them uniformly represented in the imaging space and before processing

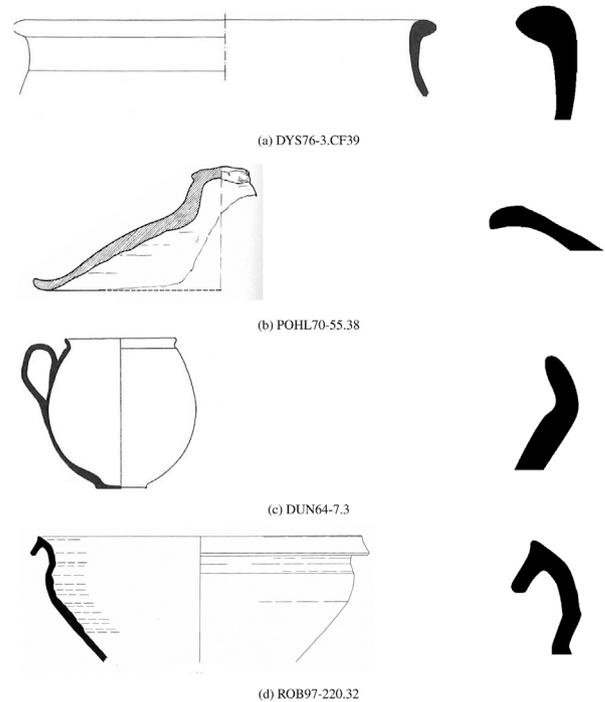


Fig. 3. Preprocessed shapes (right) from originals (left).

them into the deep learning algorithm. This is due to different presentation/editorial styles adopted across the corpora. For example, in Fig. 3 we report common situations when identifying the rim, e.g. its separation from the template and cutting of the elongated slope as well as mirroring (Fig. 3(a)), the fill in of relevant portions with a black colour with a possible rotation of the profile (Fig. 3(b)), the cleaning of scanned contours from ageing phenomena and the removal of undesired handles (Figs. 3(c) and 3(d)). Furthermore, based on the convert script from Imagemagick,² we converted the extracted rims from .png format to a .svg format so as to remove pixelisation errors (density 1200 and 1500 × 1500 pixels). All the rims are then resized to 256 × 256 pixels, without losing the aspect ratio, for the purpose of data normalisation in the deep learning network described in the rest of the paper, see Fig. 4 for a full display of rims in our database.

3. Proposed approach

After preprocessing the ROCOPOT database and with the rim profiles at our disposal, we proceed with modelling the manifold of profiles

¹ The rim usually refers to a rounded moulding on the lip of a jar, bowl, or dish, both to add strength and assist in handling (Darvill, 2008).

² www.imagemagick.org/script/convert.php

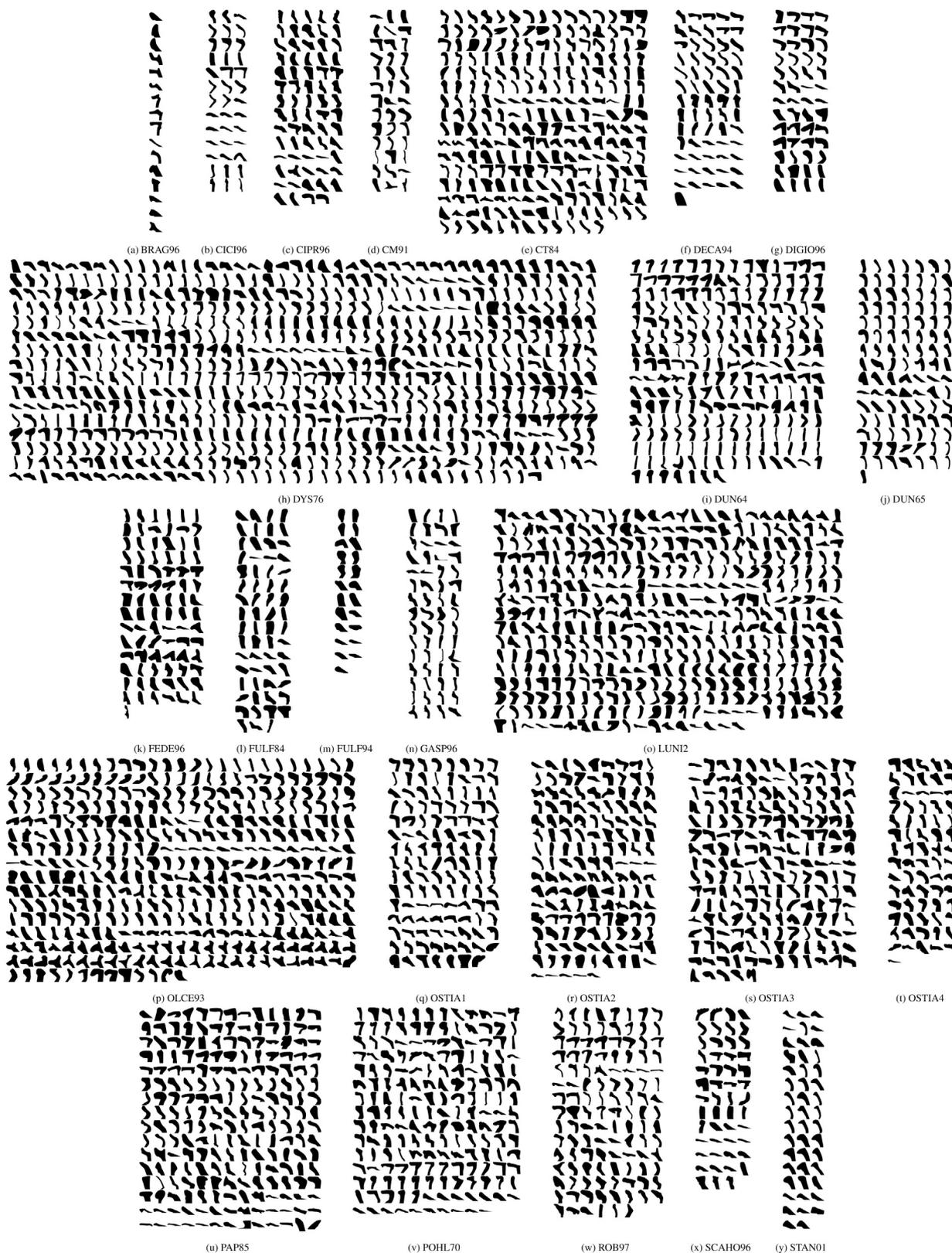


Fig. 4. ROCOPOT database v1.0: (R)ims profiles.

by learning non-linear features associated to them. Such features will be used for the hierarchical clustering of the dataset. Note that the combination of these two approaches makes our workflow totally unsupervised, a key assumption for unveiling hidden or new clustering patterns in existing archaeological corpora.

3.1. Variational autoencoders

Our workflow is based on the learning of probabilistic features associated with profiles, able to ease the profile matching problem in a low-dimensional feature space. Instead of fixing a-priori qualitative

Table 1
ROCOBOT database v1.0: number of (O)original shapes and their (B)ases, (H)andles, (R)ims and Rims with Handles (RH).

IDCAT (catalogue abbreviation)	Ref.	Year	Shapes					Chronological range	Site
			O	B	H	R	RH		
BRAG96	Bragantini (1996)	1996	16	–	1	16	2	200 BC–AD 200	Naples (Campania)
CICI96	Cicirelli (1996)	1996	50	8	12	42	3	21 BC–AD 79	Terzigno (Campania)
CIPR96	Cipriano and Fabrizio (1996)	1996	74	5	20	69	13	325 BC–AD 500	Benevento (Samnium)
CM91	Carbonara and Messineo (1991)	1991	43	4	10	39	4	100 BC–AD 200	La Celsa (South Etruria)
CT84	Chiaromonte Treré (1984)	1984	269	28	26	240	14	700 BC–AD 100	Pompei (Campania)
DECA94	De Caro et al. (1994)	1994	74	6	23	68	8	22 BC–AD 79	Boscoreale (Campania)
DIGIO96	Giovanni (1996)	1996	52	–	3	52	17	325 BC–AD 79	Pompei (Campania)
DUN64	Duncan (1964)	1964	245	13	35	217	37	AD 60–AD 70	Sutri (South Etruria)
DUN65	Duncan (1965)	1965	127	13	11	107	9	150 BC– 1 BC	Sutri (South Etruria)
DYS76	Dyson (1976)	1976	814	110	62	678	46	350 BC–AD 335	Cosa (Central Etruria)
FEDE96	Federico (1996)	1996	97	9	4	88	7	200 BC–AD 200	Circello (Samnium)
FULF84	Fulford and Peacock (1984)	1984	67	4	3	63	15	AD 301–AD 700	Carthage (North Africa, imports from Italy)
FULF94	Fulford and Peacock (1994)	1994	23	–	–	23	–	AD 25–AD 550	Carthage (North Africa, imports from Italy)
GASP96	Gasperetti (1996)	1996	70	2	38	60	19	200 BC–AD 79	Pompei (Campania)
LUNI2	Antonio (1977)	1977	453	52	25	395	37	325 BC–AD 600	Luni (Northern Etruria)
OLCE93	Olcese (1993)	1993	404	9	14	395	54	80 BC–AD 800	Albintimilium (Liguria)
OSTIA1	Carandini (1968)	1968	168	26	45	119	42	AD 101–AD 500	Ostia (Latium)
OSTIA2	Berti (1970)	1970	175	21	19	140	24	AD 51–AD 150	Ostia (Latium)
OSTIA3	Carandini and Panella (1973)	1973	230	32	31	186	23	AD 51–AD 500	Ostia (Latium)
OSTIA4	Anselmino (1977)	1977	90	13	9	74	14	AD 251–AD 425	Ostia (Latium)
PAP85	Papi (1985)	1985	240	22	18	210	18	50 BC–AD 600	Settefinestre (Central Etruria)
POHL70	Zevi and Pohl (1970)	1970	182	3	9	179	22	200 BC–AD 140	Ostia (Latium)
ROB97	Roberts (1997)	1997	132	12	21	119	13	AD 101–AD 635	Mola di Monte Gelato (South Etruria)
SCAH096	Höricht (1996)	1996	51	–	10	51	9	21 BC–AD 79	Ercolano (Campania)
STAN01	Stanco (2001)	2001	62	15	1	48	1	300 BC–AD 101	Frassineta Franco (South Etruria)
ALL			4208	407	450	3678	451		



Fig. 5. Location of the sites mentioned in Table 1.

profile features (e.g. the diameter, the ballness, the elongation, the area, the perimeter and many more), we fix the dimension $k > 0$ of the latent space and maximise the probability that profiles, generated from a stochastic sampling process in the latent space, match with the profiles in our database. In this sense, deep learning convolutional Variational Autoencoders (VAEs) are powerful tools that can be interpreted as a generative non-linear version of the principle component analysis (PCA).

Vanilla convolutional *Autoencoders* (AEs) are artificial neural networks designed for replicating the input at its output by means of back-propagating the reconstructed result with a single layer strategy (Ng et al., 2011). Here, the representation of the input is learned via the minimisation of a *loss* function, composed by a reconstruction error measure and a regulariser, in a low-dimensional latent space. Any AE is composed by an *encoder*, which reduces the dimension of the input to the latent space, and a *decoder*, which reconstructs the output from the latent space. More advanced *Stacked Sparse Autoencoders* (SSAEs) (Vincent et al., 2010) networks concatenate multiple vanilla AE, where each latent space is the input of the subsequent AE. Here, the sparsity assumption forces the neurons to specialise on few high-level features (Xu et al., 2016). However, both AE and SSAE are limited by their (possibly) discontinuous latent space, making thus impossible the interpolation between multiple input data and leading to the *overfitting* phenomenon. This is a limitation for scaling-up the approach when processing unseen data.

In contrast, *Variational Autoencoders* (VAEs) (Kingma and Welling, 2014; Higgins et al., 2017) are generative deep learning convolutional neural networks able to learn a continuous probabilistic representation of the global potsherds features in the latent space. The loss function in the network is composed by a regulariser term plus a reconstruction term:

$$\mathcal{L}(x, \bar{x}) = \beta_e \mathcal{L}_{KL}(x, z) + \mathcal{L}_c(x, \bar{x}), \quad (1)$$

where $\beta_e > 0$ is a weighting parameter varying according to the learning epoch e , $\mathcal{L}_{KL}(x, z)$ is the loss function accounting for the reverse Kullback–Leibler (KL) divergence D_{KL} on the latent space, see (5) in Appendix and Fig. 6, while the $\mathcal{L}_c(x, \bar{x})$ is a sigmoid cross-entropy energy with logits loss, see (6) in Appendix. By acting as regulariser, $\mathcal{L}_{KL}(x, z)$ ensures that the potsherds distribution generated from the learned latent space is as close as possible to the target one, while $\mathcal{L}_c(x, \bar{x})$ measures the probability error of reconstructing black–white pixels in a potsherd image, similarly to a labelling task. More details about our chosen structure of the VAE network and technical parameters for the training process are given in Section 4.

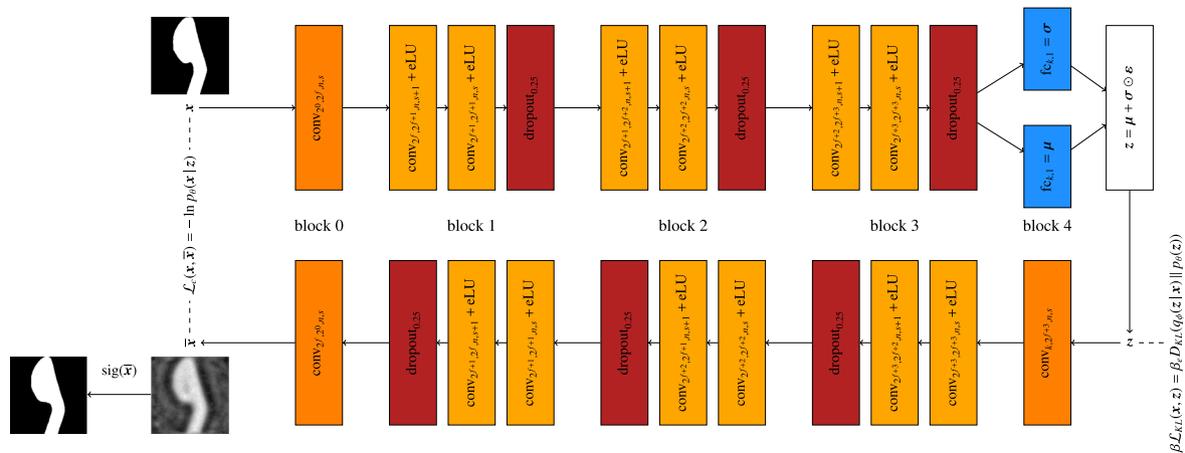


Fig. 6. Proposed Variational Autoencoder network. Note that the negative image of the potsherds is used for a better processing.

3.2. Hierarchical clustering

Once the VAE network is trained, we can associate to each profile a low-dimensional latent vector $z \in \mathbb{R}^k$ with $k > 0$. From the reparametrisation trick discussed in (7) in Appendix A.4, i.e. $z = \mu + \sigma \odot \varepsilon$ where $\varepsilon \in \mathbb{R}^k$ is a random vector, we concatenate the mean $\mu \in \mathbb{R}^k$ and the variance $\sigma \in \mathbb{R}^k$ in a vector of features $\bar{z} \in \mathbb{R}^{2k}$. These features can now be used as a shape *signature* for the clustering task. Since popular clustering methods, like *k*-means, have the bottleneck of requiring as input the number of expected clusters (unknown for our database), we opted for an *agglomerative hierarchical clustering* approach so as to build a tree of nested clusters. Starting with a single class for each potsherd profile, we subsequently merge together profiles sharing similar features in a bottom-up hierarchical strategy according to a method for computing the distance between clusters and a metric between the observed features. In this way, we can inspect different levels of similarities and shape relations. In order to automatically identify the best cluster merging rule, we compute the cophenetic coefficient score (Sokal and Rohlf, 1962) for each pair of methods (nearest distance (Florek et al., 1951), farthest distance (Sørensen, 1948), UPGMA (Sokal and Michener, 1958), WPGMA (Sokal and Michener, 1958), UPGMC (Milligan, 1980), WPGMC (Jain and Dubes, 1988) and ward (Jr., 1963)) and metrics (Euclidean, cityblock, Chebychev and cosine). Here, the cophenetic score Clarke et al. (2016) measures how faithfully the selected tree represents the dissimilarities among observation and the pair returning the highest value is selected as the best pair.

4. Network parameters and results

In this section we further discuss the parameters of our VAE network and the results produced by our workflow, when applied to the rims in our database. All the tests are performed on a GPU Nvidia Quadro P6000. The code is implemented in MATLAB 2020b.

4.1. Parameters and analysis of the VAE network

Our VAE network is composed by multiple convolutional layers of filter size 3×3 , concatenated with ad-hoc exponential linear unit (eLU) and dropout layers³ (with fixed dropout parameter equal to 0.25). With reference to Fig. 6 where the VAE network used in this work is depicted, x is the original black–white profile (256×256 pixels) and \bar{x} is the output of the network. Note that \bar{x} is required for computing the reconstruction loss \mathcal{L}_c in (1), while the output profile is obtained by

$\text{sig}(\bar{x})$. Also, each convolutional layer is denoted by $\text{conv}_{f_1, f_2, n, s}$, where f_1 are the number of input channels, f_2 are the number of filters, n is the spatial size of the filters and s is the stride. In the network, parameters are set to $f = 4$, $n = 3$ and $s = 1$ while $k = 128$ is the dimension of the latent space.

We account for handcrafted details, printing or scanning artefacts by increasing the number of potsherds at our disposal 5 times via data augmentation, e.g. by adding variations of the original shapes obtained via image operations like erosion and dilation (of 3 pixels) and rotations (by -5 and $+5$ degrees). We trained our network on 90% of the rims available, using the remaining 10% as test data, and for a maximum number of 1000 epochs. The learned parameters are updated using the Adaptive Moment Estimation (ADAM) (Kingma and Ba, 2015), with constant learning rate of $1e-06$.

In order to avoid the so-called *KL-vanishing* phenomenon, related to the undesired vanishing of the regulariser during the optimisation, the regularisation term is weighted by a positive factor $\beta_e \in (0, 1]$. Among different choices (Bowman et al., 2016; Zhao et al., 2017) (applied to natural language processing challenges), we follow the cyclical annealing approach (Fu et al., 2019) with a periodical adjustment of the weight β_e of (1) in M cycles. For a fixed epoch e , the weight β_e is computed according to the following rule:

$$\beta_e = \begin{cases} f(\tau_e) & \text{if } \tau_e \leq R, \\ 1 & \text{otherwise} \end{cases}$$

for a linear function $f(\tau_e) = \tau_e/R$ and

$$\tau_e = \text{mod}(e - 1, \lceil T/M \rceil) / (T/M),$$

where T is maximum number of epochs, M is the number of cycles and R is the rate of which to increase β_e within a cycle. Thus, each cycle is composed by an *annealing stage* where β_e linearly increases from 0 to 1, and a *fixing stage* when $\beta_e = 1$. This strategy forces the model to behaves like AEs when starting each cycle with $\beta_e = 0$ and like standard VAEs when $\beta_e = 1$; in contrast, the model transits from a point estimate to a distribution estimate when linearly increasing β_e from 0 to 1. The cyclic repetition guarantees a progressive learning, leading to a more meaningful latent space. In our experiments, we fixed $T = 1000$, $M = 4$ and $R = 0.5$, for a cycling annealing strategy as depicted in Fig. 8(a). Note that if $\beta_e \gg 1$ the loss would have related to the β -VAE extension, proposed by Higgins et al. (2017) and used for learning disentangled representation features in the latent space (Bengio et al., 2013), but at the price of a reduced reconstruction quality (Shao et al., 2020).

Once our network is trained we are able to model the probability distribution of potsherds with $z \in \mathbb{R}^k$ ($k = 128$ parameters), and to reconstruct the ROCOPOT dataset from Fig. 4. We report in Fig. 7 the confidence intervals of the pixel-wise mean squared error and the

³ Dropout layers are disable during the inference step.

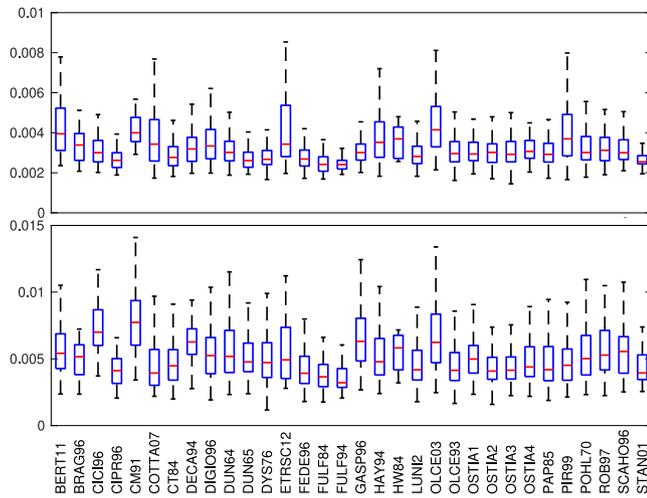


Fig. 7. Boxplots: MSE (top) and IoU (bottom) errors.

intersection over the union error as a distance (Rezatofighi et al., 2019), defined for a shape s respectively as:

$$\text{MSE}(\mathbf{x}_s, \text{sig}(\bar{\mathbf{x}}_s)) = \frac{\sum_{i,j=1}^{H,W} (\mathbf{x}_s - \text{sig}(\bar{\mathbf{x}}_s))^2}{H \cdot W},$$

where H, W are the dimension of the image, sig is the sigmoid function, and

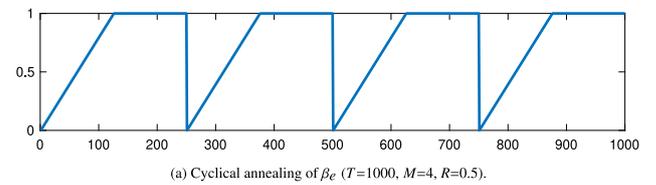
$$\text{IoU}(\mathbf{x}_s, \text{sig}(\bar{\mathbf{x}}_s)) = 1 - \frac{2 \sum_{i,j=1}^{H,W} (\mathbf{x}_s \cdot \text{sig}(\bar{\mathbf{x}}_s))}{\sum_{i,j=1}^{H,W} (\mathbf{x}_s)^2 + \sum_{i,j=1}^{H,W} \text{sig}(\bar{\mathbf{x}}_s)^2}.$$

In Fig. 8(b) we display the loss function on both train and test data, individual terms (KL loss and cross-entropy reconstruction loss) and the Euclidean error through epochs, showing no overfitting during the training of the network. Both visual and quantitative results confirms that the shapes are reconstructed almost correctly, meaning that the neurons in the latent space are now specialised in modelling the probability distribution of the potsherds in our database.

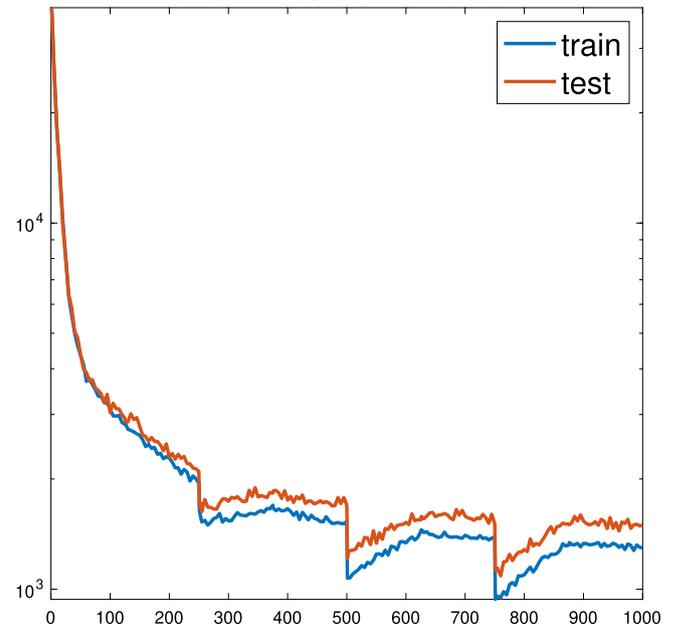
Before using the $2k(= 256)$ features for the hierarchical clustering step, a natural question is to investigate the quality of the reconstruction and the robustness of the latent space to some perturbations.

4.2. Robustness of the learned features

Our matching assumption is based on the fact that similar shapes share similar VAE latent space representation. Thus, it is worth to investigate if the shapes and the features in the latent space are robust to a series of perturbations that may occur with handcrafted data. In Fig. 9 we report the reconstruction of a subset of potsherds in the ROCOPOT database under a range of perturbations on the input shape (erosion, dilation and rotation) as well as small perturbations with a random numbers of the parameters in latent space. In particular, we compare the reconstruction of shapes from latent space features after undergoing pixel erosion (of 3, 5, 7 pixels), dilation (of 3, 5, 7 pixels) and shape rotation (of $-5, -3, 3, 5$ degrees), as well as the application of a random Gaussian perturbation on the features in latent space (1%, 5% and 10% of the feature values, respectively). The proximity of the latent space representations of all shapes to their perturbed variants has been computed and visualised in the T-SNE plot of Fig. 9, confirming the robustness of the learned features to perturbations, and the ability to match together very similar profiles. This is a visual demonstration of the feasibility of our clustering approach for the potsherds.



(a) Cyclical annealing of β_e ($T=1000, M=4, R=0.5$).



(b) ELBO loss in semi-logarithm scale along y-coordinate.

Fig. 8. From top to bottom: Cosine decay of the learning rate; cyclical annealing of β_e and ELBO loss (1) (train in blue and test in red, computed on randomly selected subsamples of 32 shapes, repeated 10 times and averaged). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Hierarchical clustering

As said, our approach takes advantage of the learned features in the VAE's latent space for the hierarchical clustering task. In our experiments the highest cophenetic score is obtained with the choice of the *Unweighted Pair Group Method using arithmetic Averages* (UPGMA, average) and the *Euclidean* metric for measuring the distance between clusters. We now restrict our focus on relevant types extracted from the metadata of our dataset. These are: domestic amphora, basin, beaker, bottle, bowl, censer, dish, dish-lid, dolium, jar, jug, knob, lid, mortarium, pan, pot and unguentarium. For all the potsherds profiles in each type, we extracted the relevant features with the VAE strategy described before, which are hierarchically clustered in view of extracting new similarity patterns. As an example, we report a portion of the hierarchical dendrogram for the type beaker in Fig. 10(a). For convenience, we call seed each of these pairs, corresponding to the paired leaves at the deepest level of the dendrogram. The same dendrogram can be produced also for each single catalogue if needed to inspect intra-catalogue similarities, see Fig. 10(b) for CICI96 (Cicirelli, 1996).

4.4. Unveiling hidden relations: the ROCOPOT app

Given the cross-disciplinary nature of the challenge in this paper, we need a tool for visualising and inspecting the clusters in the dendrograms. To this end, we developed a MATLAB GUI⁴ which is able to

⁴ Freely available at <http://dx.doi.org/10.5281/zenodo.5552265> and (Parisotto, 2021).

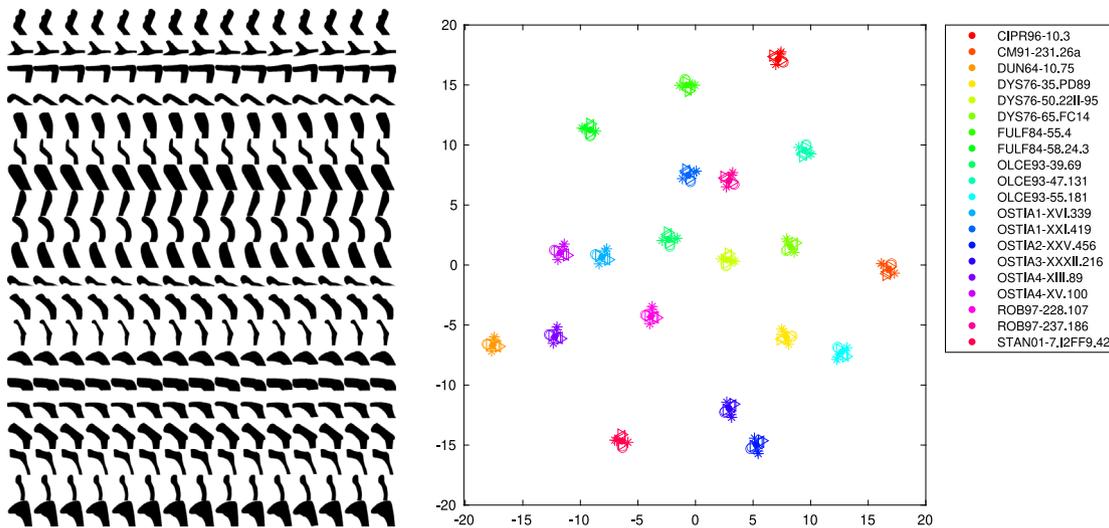


Fig. 9. Robustness of latent features. Left (in columns): input shape, its reconstruction and its reconstructions after perturbations (erosion [3, 5, 7] pixels, dilation [3, 5, 7] pixels and rotation [-5, -3, 3, 5] degrees in the image space and random perturbations [1%, 5%, 10%] in the latent space); right: T-SNE plot of associated VAE features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

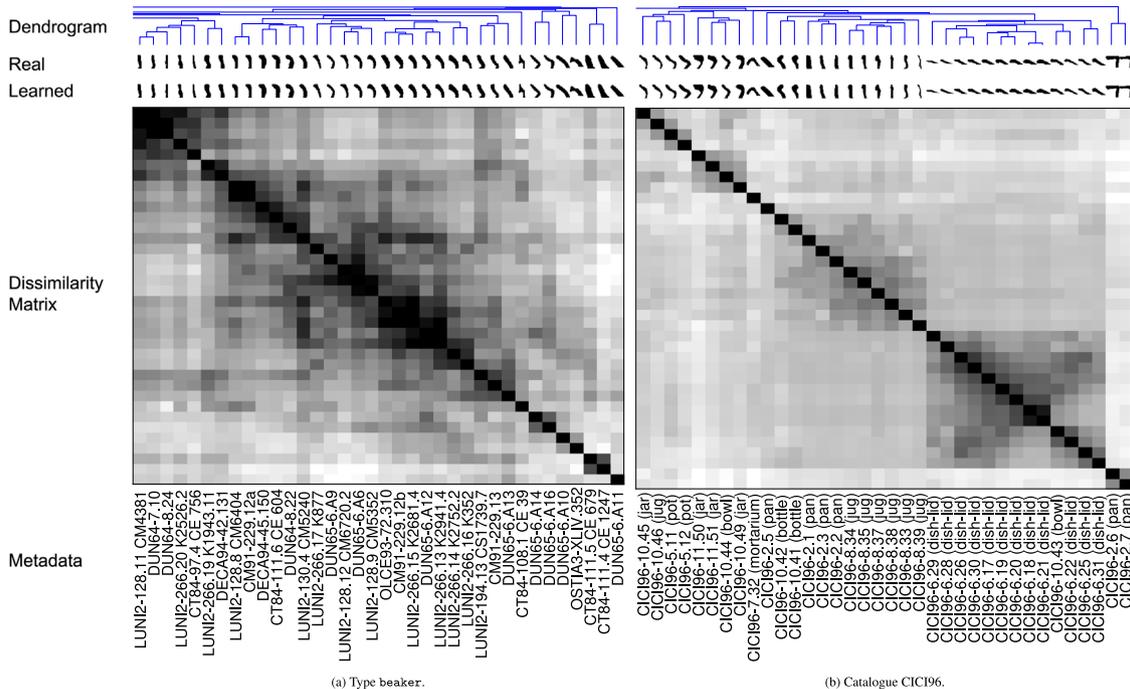


Fig. 10. From top to bottom: dendrogram (partial) for type beaker (left) and catalogue CICI96 (right), real and learned shreds visually confirming the ability to draw shreds correctly from learned features, dissimilarity matrix and metadata.

effectively support the onfield work of the archaeologists (Leone et al., 2020), see the screenshot in Fig. 11.

4.5. Comments from the archaeological perspective

In order to test how effective our workflow is in recognising actual similarities between different profiles, we inspect the quality of the dendrograms via the MATLAB app.

The lack of a ground truth in matching potsherds, itself resulting from a degree of inherent subjectivity in the manual classification process, has a great impact on an automatic evaluation of the performances. Therefore the task of validating the matching accuracy of the dendrograms is assigned to archaeologists. Although, in principle, one could evaluate the coherence of each hierarchical group

(of profiles) within the dendrogram, this task would involve a rather complex review of the results. Therefore, we asked archaeologists to concentrate on and validate the matching accuracy for each individual seed, expected to be the best match proposed by our unsupervised workflow.

Before proceeding, we acknowledge that a large number of profiles in our dataset may in fact represent somewhat unique vessels bearing little or no similarity to any other. Roman commonwares consisted of ordinary pottery which was produced at a myriad of distinct workshops, the latter often coming into and going out of existence in the lapse of a generation. We should therefore expect the range of vessels produced, distributed and consumed, even within the same region and in the same period, to be considerably vast and varied. This otherwise unbridled drive towards a variety of forms was certainly directed – and

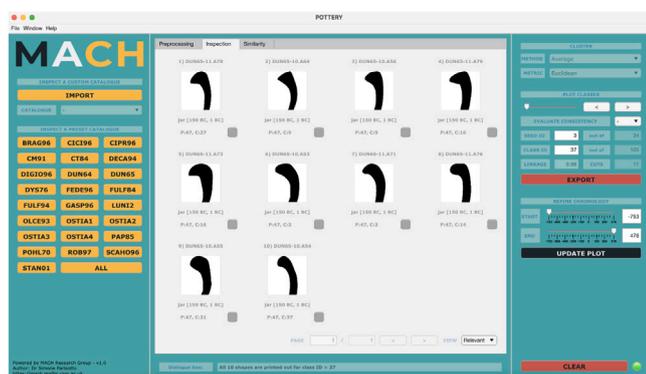


Fig. 11. ROCOPOT software (GUI in MATLAB R2020b).

to a degree tempered – by the practicalities of the production process (e.g. standardisation) and the predominantly functional character of these objects. Of course, given the high levels of inter-connectivity and exchange achieved in the Roman world (and Central Tyrrhenian Italy was certainly no exception!), local potters worked within productive traditions which were used to absorb, copy and rework types originally imported from elsewhere, and whose features may have appeared from time to time better suited to specific functions or even just the tastes of potential customers. Nevertheless, only a fraction of the whole dataset may have in fact featured any actual degree of close similarity.

Our workflow processes 3678 rim profiles as featured in 25 catalogues (R in Table 1) and identifies 1231 seeds (corresponding to 2462 profiles). The archaeologists looked at each seed by inspecting and comparing the paired profiles, thus assessing their similarity. They did not consider the overall shape of the vessel (even when known), but instead only concentrated on the overall combination of morphological features as presented by each individual rim (e.g. inclination, curve, lip design). This approach did not present any particular challenge to them as it was very much in keeping with standard practice in the analysis of (incomplete) potsherds as most commonly recovered through fieldwork. In light of the above archaeological considerations about the nature of the data itself, it is not surprising that only 500 out of 1231 seeds (41%) were eventually validated as a positive match. What is remarkable, however, is that out of those 500 positive seeds, only 147 (29%) had already been indicated as close matches in the relevant pottery corpora, whilst the remaining 353 (71%) are new. In other words, although some of these similarities have been in fact recognised and recorded when the pottery corpora were originally put together by their authors, our workflow is able to significantly expand this range of relationships. This goes far beyond a random matching of profiles, since the actual probability of pairing two similar profiles within the 2462 available is less than 0.001%.

It is certainly notable that our workflow was autonomously capable of establishing close similarity between profiles which were originally recognised as ‘similar’ by the original authors. This suggests that the workflow may have learned to look for the same combination of features which a pottery specialist normally looks for when classifying pottery. Although replicating such results may in itself be considered a notable achievement from a mathematical point of view, it is the fact that the range of similarities was expanded that matters a lot from an archaeological point of view. If one considers that each pottery catalogue is a reflection of relevant patterns of production, distribution and consumption within a given location in a given period, an improved understanding of their inter-relationship can truly contribute not only to refining chronologies (useful as this exercise may be), but also to mapping how (trade) contacts between different places were created and evolved over the Roman period. Even though such patterns are routinely reconstructed on the base of the distribution of finewares and amphorae, we already pointed out that these classes may not provide

as comprehensive a view of the broader population as commonwares can.

In short, the proposed workflow is very promising, especially when considering its inherent scalability. As the dataset is expanded with the inclusion of new profiles from additional pottery corpora, the learning step will improve, thus increasing the chance of recognising more (and possibly *qualitatively* better) similarities. More than anything, the clustering step is capable of structuring in *minutes* the kind of dataset whose size would require *months* (or indeed *years!*) of manual reviewing a long list of disparate publications. In fact, although here applied to commonware pottery from Central Tyrrhenian Italy, this workflow can be straightforwardly extended not only to Roman commonwares in general, but to any coherent pottery class from any time and/or period. Although this process is neither infallible or exhaustive, it is in fact meant not to replace, but rather to aid in the manual interpretation and classification of pottery profiles by unveiling *additional* relationships which the mass of data may render almost *invisible* to the human eye.

5. Conclusions and future works

In this paper, we introduced the *ROman COmmonware POTtery* profiles (ROCO POT) database and provided a workflow for unveiling close similarities between (Roman commonware) pottery profiles. That is based on the hierarchical clustering of Roman potsherds via latent features learned in a deep convolutional VAE network. Such features are proved to be robust to a number of perturbations mimicking the real scenario of the high shape variance due to handcrafted materials. The most obvious advantage provided by our approach rests in the ability to ease and speed up the processing and matching of a very large number of pottery profiles from different corpora within a coherent and unified analytical environment. This is achieved by presenting the pottery specialist with a selection of most likely matches in the companion software, thus providing invaluable pointers not only in the development of a comprehensive typology of Roman commonwares, but even in the future classification of potsherds coming from new excavations. These results are very encouraging and prompt us to pursue this research further. Firstly, we plan to increase the number of profiles available in our database; secondly we will develop a specific online tool which will aid in the classification of profiles of new commonware potsherds recovered by new archaeological projects (and thus increasing our knowledge on this pottery class whilst expanding its application in the field). Finally, we aim to train the network to include other parts of the vessels (bases, main body) in its similarity search and fully incorporate the feedback (validation) of the archaeologists in order to refine its analysis through iteration.

Acknowledgements

The authors thank Michael Bronstein, Thomas Buddenkotte and Martin Millett for the useful discussions and the anonymous reviewers for their encouraging and constructive feedback. The authors acknowledge the support from the Leverhulme Trust Research Project Grant (RPG-2018-121) “Unveiling the Invisible - Mathematics for Conservation in Arts and Humanities”. CBS further acknowledges support from the RISE projects CHiPS and NoMADS, the Cantab Capital Institute for the Mathematics of Information and the Alan Turing Institute.

Appendix

Here we discuss the main concepts of variational autoencoders and the derivation of the loss function in Eq. (1), by means of (5) and (6), see Kingma and Welling (2014) and Doersch (2021).

We assume that the dataset consists of $X = (x_1, \dots, x_n)$ samples, with $n > 0$ (in our case n two-dimensional black–white images representing the profile of potsherds of size $H \times W$ from the ROCOPOT database). We identify with $x \in X$ one sample from the dataset and with $\bar{x} \in Y$

the reconstruction of \mathbf{x} , where \mathbf{Y} is the reconstruction of the space \mathbf{X} . Also, let $g_\phi(\cdot)$ and $f_\theta(\cdot)$ be the encoding and decoding functions, parametrised by ϕ and θ , respectively. Moreover, $q_\phi(\mathbf{z} | \mathbf{x})$ represents the estimated posterior probability distribution (probabilistic encoder) for the latent code $\mathbf{z} \in \mathbb{R}^k$ given \mathbf{x} and $p_\theta(\mathbf{x} | \mathbf{z})$ the likelihood of generating true data samples \mathbf{x} given the latent code \mathbf{z} (probabilistic decoder). Here $k > 0$ is the fixed dimension of the reduced latent space. In general, The Kullback–Leibler (KL) divergence $D_{\text{KL}}(q || p)$ is a non-symmetric relative entropy designed to quantify the discrepancy between two distributions q and p and how much information is lost if the distribution q is used to represent p .

A.1. The evidence lower bound loss in VAE

In the AE scheme, the idea is to learn a low-dimensional representation $\mathbf{z} = g_\phi(\mathbf{x}) \in \mathbb{R}^k$ of the data by means of learning the identity function such that $\bar{\mathbf{x}} = f_\theta(g_\phi(\mathbf{x}))$. The overfitting and the impossibility to generate new samples are clear limits of this approach.

Conversely, VAEs aim to learn the model underlying the data: the estimated posterior $q_\phi(\mathbf{z} | \mathbf{x})$ is designed to be close to the real posterior $p_\theta(\mathbf{z} | \mathbf{x})$ making possible to generate true samples. To that effect, the *forward* or *reverse* KL divergence could be used. The *forward* KL divergence $D_{\text{KL}}(p_\theta(\mathbf{z} | \mathbf{x}) || q_\phi(\mathbf{z} | \mathbf{x}))$ tends to favour an approximate distribution q_ϕ with an undesired mean effect; conversely the *reversed* KL divergence $D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z} | \mathbf{x}))$ is better suited to approximate the distribution q_ϕ within a mode of p_θ and is preferred (Theis et al., 2016; Huszár, 2015; Goodfellow, 2017; Nguyen et al., 2017). The *reversed* KL divergence is equivalent to:

$$\begin{aligned} D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z} | \mathbf{x})) &= \int q_\phi(\mathbf{z} | \mathbf{x}) \log \frac{q_\phi(\mathbf{z} | \mathbf{x})}{p_\theta(\mathbf{z} | \mathbf{x})} d\mathbf{z} \\ &= \int q_\phi(\mathbf{z} | \mathbf{x}) \log \frac{q_\phi(\mathbf{z} | \mathbf{x}) p_\theta(\mathbf{x})}{p_\theta(\mathbf{z}, \mathbf{x})} d\mathbf{z} \\ &= \int q_\phi(\mathbf{z} | \mathbf{x}) \left(\log p_\theta(\mathbf{x}) + \log \frac{q_\phi(\mathbf{z} | \mathbf{x})}{p_\theta(\mathbf{z}, \mathbf{x})} \right) d\mathbf{z} \\ &= \log p_\theta(\mathbf{x}) + \int q_\phi(\mathbf{z} | \mathbf{x}) \log \frac{q_\phi(\mathbf{z} | \mathbf{x})}{p_\theta(\mathbf{z}, \mathbf{x})} d\mathbf{z} \\ &= \log p_\theta(\mathbf{x}) + \int q_\phi(\mathbf{z} | \mathbf{x}) \log \frac{q_\phi(\mathbf{z} | \mathbf{x})}{p_\theta(\mathbf{x} | \mathbf{z}) p_\theta(\mathbf{z})} d\mathbf{z} \\ &= \log p_\theta(\mathbf{x}) + \mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z} | \mathbf{x})} \left(\log \frac{q_\phi(\mathbf{z} | \mathbf{x})}{p_\theta(\mathbf{z})} - \log p_\theta(\mathbf{x} | \mathbf{z}) \right) \\ &= \log p_\theta(\mathbf{x}) + D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z})) - \mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z} | \mathbf{x})} \log p_\theta(\mathbf{x} | \mathbf{z}), \end{aligned}$$

where in the above we used $p_\theta(\mathbf{z} | \mathbf{x}) = p_\theta(\mathbf{z}, \mathbf{x})/p_\theta(\mathbf{x})$, $\int q_\phi(\mathbf{z} | \mathbf{x}) d\mathbf{z} = 1$ and $p_\theta(\mathbf{z}, \mathbf{x}) = p_\theta(\mathbf{x} | \mathbf{z}) p_\theta(\mathbf{z})$. By rearranging the above equation, we have:

$$\begin{aligned} \log p_\theta(\mathbf{x}) - D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z} | \mathbf{x})) \\ = \mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z} | \mathbf{x})} \log p_\theta(\mathbf{x} | \mathbf{z}) - D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z})) \end{aligned} \quad (2)$$

The left-hand-side of Eq. (2) allows us to maximise the (log-)likelihood of generating real data, i.e. $\log p_\theta(\mathbf{x})$, while minimising the difference between the real and estimated posterior distributions. Note that $p_\theta(\mathbf{x})$ is fixed with respect to $q_\phi(\mathbf{z} | \mathbf{x})$. The negative of the left-hand-side in Eq. (2) is defined as the *Evidence Lower Bound* (ELBO) loss function:

$$\mathcal{L}_{\text{ELBO}}(\theta, \phi) = -\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z} | \mathbf{x})} \log p_\theta(\mathbf{x} | \mathbf{z}) + D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z})). \quad (3)$$

The optimal parameters are obtained by solving the following minimisation problem:

$$(\theta^*, \phi^*) = \arg \min_{\theta, \phi} \mathcal{L}_{\text{ELBO}}(\theta, \phi).$$

Being the reverse KL divergence always non-negative, $-\mathcal{L}_{\text{ELBO}}$ is the lower bound of $\log p_\theta(\mathbf{x})$: this explains why the minimisation of the $\mathcal{L}_{\text{ELBO}}$ loss implies the maximisation of the probability of generating real data samples $p_\theta(\mathbf{x})$. Indeed, from Eq. (2) we have:

$$-\mathcal{L}_{\text{ELBO}} = \log p_\theta(\mathbf{x}) - D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z} | \mathbf{x})) \leq \log p_\theta(\mathbf{x}).$$

A.2. Derivation of the KL divergence

The KL divergence in Eq. (3) is a *regulariser* because it is a constraint on the form of the approximate posterior (Odaibo, 2019). In general, the KL divergence of two multivariate Gaussian distributions q_ϕ and p_θ (of dimension n), i.e. $q_\phi = \mathcal{N}(\boldsymbol{\mu}_\phi, \boldsymbol{\sigma}_\phi)$ and $p_\theta = \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta)$, can be expressed in the following form:

$$\begin{aligned} D_{\text{KL}}(q_\phi || p_\theta) &= \frac{1}{2} \left(\log \frac{|\boldsymbol{\sigma}_\theta|}{|\boldsymbol{\sigma}_\phi|} - n + \text{tr}(\boldsymbol{\sigma}_\theta^{-1} \boldsymbol{\sigma}_\phi) \right. \\ &\quad \left. + (\boldsymbol{\mu}_\theta - \boldsymbol{\mu}_\phi)^T \boldsymbol{\sigma}_\theta^{-1} (\boldsymbol{\mu}_\theta - \boldsymbol{\mu}_\phi) \right). \end{aligned} \quad (4)$$

In VAE, q_ϕ is the encoder distribution defined as $q_\phi(\mathbf{z} | \mathbf{x}) = \mathcal{N}(\mathbf{z} | \boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\sigma}(\mathbf{x}))$ where $\boldsymbol{\sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ while p_θ is the latent prior given by $p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus, from Eq. (4) the reverse KL divergence $D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z}))$ is expressed as:

$$\begin{aligned} &= \frac{1}{2} \left(\log \frac{|\mathbf{I}|}{|\boldsymbol{\sigma}|} - n + \text{tr}\{\mathbf{I}^{-1} \boldsymbol{\sigma}\} + (\mathbf{0} - \boldsymbol{\mu})^T \mathbf{I}^{-1} (\mathbf{0} - \boldsymbol{\mu}) \right) \\ &= \frac{1}{2} (-\log |\boldsymbol{\sigma}| - n + \text{tr}\{\boldsymbol{\sigma}\} + \boldsymbol{\mu}^T \boldsymbol{\mu}) \\ &= \frac{1}{2} \left(-\log \prod_{i=1}^k \sigma_i^2 - n + \sum_{i=1}^k \sigma_i^2 + \sum_{i=1}^k \mu_i^2 \right) \\ &= \frac{1}{2} \left(-\sum_{i=1}^k \log \sigma_i^2 - n + \sum_{i=1}^k \sigma_i^2 + \sum_{i=1}^k \mu_i^2 \right) \end{aligned}$$

which can be rearranged as

$$D_{\text{KL}}(q_\phi(\mathbf{z} | \mathbf{x}) || p_\theta(\mathbf{z})) = -\frac{1}{2} \sum_{i=1}^k (\log \sigma_i^2 + 1 - \sigma_i^2 - \mu_i^2). \quad (5)$$

A.3. Reconstruction loss

The expectation in Eq. (3) is a *reconstruction* loss. In our scenario, similar to a semantic segmentation problem with labels 0 and 1 due to the black–white nature of the data, it makes sense to use a sigmoid cross-entropy with logits energy term as reconstruction loss $\mathcal{L}_c(\mathbf{x}, \bar{\mathbf{x}})$, of the form

$$-\mathbf{x} \cdot \log(\text{sig}(\bar{\mathbf{x}})) - (\mathbf{1} - \mathbf{x}) \cdot \log(\mathbf{1} - \text{sig}(\bar{\mathbf{x}})),$$

where sig is the sigmoid function, reformulated for numerical stability and for avoiding overflows⁵ as

$$\mathcal{L}_c(\mathbf{x}, \bar{\mathbf{x}}) = \max(\bar{\mathbf{x}}, \mathbf{0}) - \bar{\mathbf{x}} \cdot \mathbf{z} + \log(\mathbf{1} + \exp(-|\bar{\mathbf{x}}|)). \quad (6)$$

A.4. Reparametrisation trick

The expectation term in Eq. (3) requires generating samples from the latent space $q_\phi(\mathbf{z} | \mathbf{x})$. However, the sampling is a stochastic process, making impossible to backpropagating the gradient. The reparametrisation trick proposed in Kingma and Welling (2014) (based on methods already introduced in Devroye (1986, 1996)) allows to overcome such difficulties by means of rewriting \mathbf{z} as a sum of a deterministic part plus an independent random variable $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\varepsilon}, \quad (7)$$

where \odot refers to element-wise product. With this trick in place, the mean $\boldsymbol{\mu}$ and the variance $\boldsymbol{\sigma}$ of the distribution are now learnable parameters.

⁵ www.tensorflow.org/api_docs/python/tf/nn/sigmoid_cross_entropy_with_logits

References

- Anselmino, L., 1977. Ostia IV : Le Terme Del Nuotatore, Scavo Dell'Ambiente XVI E Dell'Area XXV. In: Studi miscellanei ; 23, De Luca, Roma.
- Antonio, F., 1977. Scavi Di Luni, II. Relazione Delle Campagne Di Scavo 1972, 1973, 1974. Giorgio Bretschneider, Roma.
- ArchAIDE Consortium, 2019. ARCHAIDE Portal for Publications and Outputs. Archaeology Data Service.
- Banterle, F., Dellepiane, M., Evans, T., Gattiglia, G., Itkin, B., Zalocco, M., 2017. The ArchAIDE project: Results and perspectives after the first year. In: Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage. In: GCH '17, Eurographics Association, Goslar, DEU, pp. 161–164.
- Bar, Y., Levy, N., Wolf, L., 2015. Classification of artistic styles using binarized features derived from a deep neural network. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. Springer International Publishing, Cham, pp. 71–84.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8), 1798–1828.
- Berti, F., 1970. Ostia II : Le Terme Del Nuotatore, Scavo Dell'Ambiente I. In: Studi miscellanei (Seminario di archeologia e storia dell'arte greca e romana dell'Università di Roma); 16, De Luca, Roma.
- Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R., Bengio, S., 2016. Generating sentences from a continuous space. In: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. Association for Computational Linguistics, Berlin, Germany, pp. 10–21.
- Bragantini, I., 1996. La ceramica da cucina dello scavo di palazzo corigliano a napoli e il commercio della ceramica campana da cucina. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 173–182.
- Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R., 2007. Analysis of two-dimensional non-rigid shapes. *Int. J. Comput. Vis.* 78 (1), 67–88.
- Bronstein, A.M., Bronstein, M.M., Kimmel, R., 2009. Numerical Geometry of Non-Rigid Shapes. Springer New York.
- Cao, F., Lisani, J.-L., Morel, J.-M., Musé, P., Sur, F., 2008. A Theory of Shape Identification. Springer Berlin Heidelberg.
- Carandini, A., 1968. Ostia I: Le Terme Del Nuotatore, Scavo Dell'Ambiente IV. In: Studi miscellanei; 13, De Luca, Roma.
- Carandini, A., Panella, C., 1973. Ostia III: Le Terme Del Nuotatore. Scavo Degli Ambienti III, VI, VII. Scavo Dell'Ambiente V E Di Un Saggio Nell'Area Sud-Ovest. In: Studi miscellanea ; 21, De Luca, Roma.
- Carbonara, A., Messineo, G., 1991. Ceramica dalle fornaci della celsa. In: La Via Flaminia: Da Porta Del Popolo a Malborghetto. In: Studi e materiali dei musei e monumenti comunali di Roma, Quasar, Roma, pp. 185–199.
- Chiaromonte Treré, C., 1984. Ceramica grezza e depurata (m). In: Bonghi Jovino, M. (Ed.), Ricerche a Pompei: L'Insula 5 Della Regio VI Dalle Origini Al 79 D.C. : I (Campagne Di Scavo 1976-1979). In: Bibliotheca archaeologica; 5, "L'Erma" di Bretschneider, Roma.
- Christmas, J., Pitts, M., 2018. Classifying and visualising roman pottery using computer-scanned typologies. *Internet Archaeol.* 50 (50).
- Cicirelli, C., 1996. La ceramica comune da terzigno: nota preliminare. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 157–171.
- Cipriano, M.T., Fabrizio, S.D., 1996. Benevento. Il quartiere ceramico di Cellarulo: prime osservazioni sulla tipologia ceramica. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 201–223.
- Clarke, K.R., Somerfield, P.J., Gorley, R.N., 2016. Clustering in non-parametric multivariate analyses. *J. Exp. Mar. Biol. Ecol.* 483, 147–155.
- Cohen-Addad, V., Kanade, V., Mallmann-trenn, F., Mathieu, C., 2019. Hierarchical clustering: Objective functions and algorithms. *J. ACM* 66 (4).
- Crowley, E.J., Zisserman, A., 2015. In search of art. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. Springer International Publishing, Cham, pp. 54–70.
- Cui, M., Femiani, J., Hu, J., Wonka, P., Razdan, A., 2009. Curve matching for open 2D curves. *Pattern Recognit. Lett.* 30 (1), 1–10.
- d'Amico, M., Frosini, P., Landi, C., 2006. Using matching distance in size theory: A survey. *Int. J. Imaging Syst. Technol.* 16 (5), 154–161.
- Darvill, T., 2008. Concise Oxford Dictionary of Archaeology, 2 In: Oxford Quick Reference, Oxford University Press, London, England.
- De Caro, S., Zevi, F., Feemster Jashemski, W., 1994. La Villa Rustica in Località Villa Regina a Boscoreale. Bretschneider, Roma.
- Devroye, L., 1986. Non-Uniform Random Variate Generation (Originally Published with Springer-Verlag).
- Devroye, L., 1996. Random variate generation in one line of code. In: Proceedings of the 28th Conference on Winter Simulation. In: WSC '96, IEEE Computer Society, USA, pp. 265–272.
- Doersch, C., 2021. Tutorial on variational autoencoders. arXiv:1606.05908.
- Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification (2nd Edition). Wiley-Interscience, USA.
- Duncan, G.C., 1964. A roman pottery near sutri. *Pap. Br. Sch. Roma* 32 (1), 38–88.
- Duncan, G.C., 1965. Roman Republican pottery from the vicinity of sutri (Sutrium). *Pap. Br. Sch. Rome* 33, 134–176.
- Dyson, S.L., 1976. Cosa: The utilitarian pottery. *Mem. Am. Acad. Rome* 33, 3–175.
- Federico, R., 1996. La ceramica comune dal territorio dei Liguri Baebiani. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 183–200.
- Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S., 2020. Machine learning for cultural heritage: A survey. *Pattern Recognit. Lett.* 133, 102–108.
- Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., Zubrzycki, S., 1951. Sur la liaison et la division des points d'un ensemble fini. *Colloq. Math.* 2 (3–4), 282–285.
- Fu, H., Li, C., Liu, X., Gao, J., Celikyilmaz, A., Carin, L., 2019. Cyclical annealing schedule: A simple approach to mitigating. In: Proceedings of the 2019 Conference of the North. Association for Computational Linguistics.
- Fulford, M., Peacock, D., 1984. Excavations At Carthage: The British Mission. the Avenue Du President Habib Bourguiba, Salammbu, the Pottery and Other Ceramic Objects from the Site, vol. I.2. Sheffield, Published for the British Academy from the Department of Prehistory and Archaeology, University of Sheffield.
- Fulford, M., Peacock, D., 1994. Excavations At Carthage: The British Mission. the Circular Harbour, North Side, the Pottery, vol. II.2. Oxford University Press.
- Gasperetti, G., 1996. Produzione e consumo della ceramica comune da mensa e dispensa nella Campania romana. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 19–63.
- Ginosar, S., Haas, D., Brown, T., Malik, J., 2015. Detecting people in cubist art. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. Springer International Publishing, Cham, pp. 101–116.
- Giovanni, V.D., 1996. Produzione e consumo di ceramica da cucina nella Campania romana (II a.C.- II d.C.). In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 65–103.
- Gonthier, N., Gousseau, Y., Ladjal, S., Bonfait, O., 2018. Weakly supervised object detection in artworks. In: Computer Vision - ECCV 2018 Workshops. Springer International Publishing, pp. 692–709.
- Goodfellow, I., 2017. NIPS 2016 tutorial: Generative adversarial networks. arXiv: 1701.00160.
- Gunia, P., Baher, A., Möller, H., 2012. CeramAlex – ein datenbankprojekt zur erschließung hellenistischer und römischer fundkeramik aus ägypten. *KÖlner Bonner Archaeol.* 2, 253–261.
- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2017. Beta-VAE: Learning basic visual concepts with a constrained variational framework. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings. OpenReview.net.
- Höricht, L.A.S., 1996. Appunti sulla ceramica comune di ercolano. Vasellame da cucina e recipienti per la preparazione degli alimenti. In: Les CÉramiques Communes de Campanie Et de Narbonnaise (Ie S. Av. J.-C. - Iie S. Ap. J.-C.). la Vaisselle de Cuisine Et de Table. Publications du Centre Jean Bérard, pp. 129–156.
- Hörr, C., Lindinger, E., Brunnett, G., 2014. Machine learning based typology development in archaeology. *J. Comput. Cult. Herit.* 7 (1).
- Hu, M.-K., 1962. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory* 8 (2), 179–187.
- Huszár, F., 2015. How (not) to train your generative model: Scheduled sampling, likelihood, adversary? arXiv:1511.05101.
- Jain, A.K., Dubes, R.C., 1988. Algorithms for Clustering Data. Prentice-Hall, Inc., USA.
- Jr., J.H.W., 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* 58 (301), 236–244.
- Kampel, M., Sablatnig, R., Costa, E., 2001. Classification of archaeological fragments using profile primitives. In: Computer Vision, Computer Graphics and Photogrammetry — a Common Viewpoint, Proceedings of the 25th Workshop of the Austrian Association for Pattern Recognition (ÖAGM). pp. 151–158.
- Karasik, A., Smilansky, U., 2011. Computerized morphological classification of ceramics. *J. Archaeol. Sci.* 38 (10), 2644–2657.
- Kim, W.-Y., Kim, Y.-S., 2000. A region-based shape descriptor using Zernike moments. *Signal Process., Image Commun.* 16 (1), 95–102.
- Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings.
- Kingma, D.P., Welling, M., 2014. Auto-encoding variational Bayes. In: Bengio, Y., LeCun, Y. (Eds.), 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings.
- Launaro, A., Leone, N., 2018. A view from the margin? Roman commonwares and patterns of distribution and consumption at interamna lirenas (Lazio). *J. Roman Archaeol.* 31, 323–338.
- LeCun, Y., Cortes, C., Burges, C., 2010. MNIST handwritten digit database. ATT Labs [Online] 2, URL <http://yann.lecun.com/exdb/mnist>.

- Leone, N., Simone Parisotto, K.T.-H., Spike Bucklow, A.L., Reynolds, S., Schönlieb, C.-B., 2020. Art speaks maths, maths speaks art. In: Yackel, C., Bosch, R., Torrence, E., Fenyvesi, K. (Eds.), *Proceedings of Bridges 2020: Mathematics, Art, Music, Architecture, Education, Culture*. Tessellations Publishing, Phoenix, Arizona, pp. 481–484.
- Litany, O., Rodolá, E., Bronstein, A.M., Bronstein, M.M., 2017. Fully spectral partial shape matching. *Comput. Graph. Forum* 36 (2), 247–258.
- Manay, S., Cremers, D., Hong, B.-W., Yezzi, A.J., Soatto, S., 2006. Integral invariants for shape matching. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (10), 1602–1618.
- Milligan, G.W., 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45 (3), 325–342.
- Navarro, P., Cintas, C., Lucena, M., Fuertes, J.M., Delrieux, C., Molinos, M., 2021. Learning feature representation of iberian ceramics with automatic classification models. *J. Cultural Heritage*.
- Ng, A., et al., 2011. Sparse autoencoder. In: *CS294A Lecture Notes*, vol. 72, pp. 1–19.
- Nguyen, T.D., Le, T., Vu, H., Phung, D., 2017. Dual discriminator generative adversarial nets. [arXiv:1709.03831](https://arxiv.org/abs/1709.03831).
- Odaibo, S., 2019. Tutorial: Deriving the standard variational autoencoder (VAE) loss function. [arXiv:1907.08956](https://arxiv.org/abs/1907.08956).
- Olcese, G., 1993. Le Ceramiche Comuni Di Albintimilium: Indagine Archeologica E Archeometrica Sui Materiali Dell'Area Del Cardine. In: *Quaderni del Dipartimento di Archeologia e Storia delle Arti. Sezione Archeologia*. Università di Siena, All'insegna del giglio.
- Orton, C., Hughes, M., 2013. Pottery in Archaeology, second ed. In: *Cambridge Manuals in Archaeology*, Cambridge University Press.
- Papi, E., 1985. Recipienti per il lavaggio e incensari. *Ceramica comune*. In: Carandini, A., Filippi, M. (Eds.), *Settefinestre: Una Villa Schiavistica Nell'Etruria Romana*. In: *Settefinestre: una villa schiavistica nell'Etruria romana*, Panini.
- Parisotto, S., 2021. ROCOPOT (ROMan COMMONware POTtery) App. Zenodo, [http://dx.doi.org/10.5281/zenodo.5552266](https://doi.org/10.5281/zenodo.5552266).
- Parisotto, S., Launaro, A., Leone, N., Schönlieb, C.-B., 2021. ROCOPOT (ROMan COMMONware POTtery) Dataset. Zenodo, [http://dx.doi.org/10.5281/zenodo.4311296](https://doi.org/10.5281/zenodo.4311296).
- Pawlowicz, L.M., Downum, C.E., 2021. Applications of deep learning to decorated ceramic typology and classification: A case study using tusayan white ware from northeast arizona. *J. Archaeol. Sci.* 130, 105375.
- Piccoli, C., Aparajeya, P., Papadopoulos, G.T., Bintliff, J., Leymarie, F.F., Bes, P., van der Enden, M., Poblome, J., Daras, P., 2015. Towards the automatic classification of pottery sherds: two complementary approaches. In: Traviglia, A. (Ed.), *Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology*. Amsterdam University Press; Amsterdam, pp. 463–474.
- Raviv, D., Bronstein, M.M., Bronstein, A.M., Kimmel, R., 2010. Volumetric heat kernel signatures. In: *Proceedings of the ACM Workshop on 3D Object Retrieval*. In: 3DOR '10, Association for Computing Machinery, New York, NY, USA, pp. 39–44.
- Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 658–666.
- Roberts, P., 1997. The roman pottery. In: Potter, T.W. (Ed.), *Excavations At the Mola Di Monte Gelato: A Roman and Medieval Settlement in South Etruria*. In: *Archaeological monographs of the British School at Rome; no.11*, British School at Rome London in association with the British Museum, London.
- Rumpf, M., Wirth, B., 2013. Discrete geodesic calculus in shape space and applications in the space of viscous fluidic objects. *SIAM J. Imaging Sci.* 6 (4), 2581–2602.
- Seidl, M., Wieser, E., Zeppelzauer, M., Pinz, A., Breiteneder, C., 2015. Graph-based shape similarity of petroglyphs. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), *Computer Vision - ECCV 2014 Workshops*. Springer International Publishing, Cham, pp. 133–148.
- Shao, H., Yao, S., Sun, D., Zhang, A., Liu, S., Liu, D., Wang, J., Abdelzaher, T., 2020. ControlVAE: Controllable variational autoencoder. In: I.I.I., H.D., Singh, A. (Eds.), *Proceedings of the 37th International Conference on Machine Learning*. In: *Proceedings of Machine Learning Research*, vol. 119, PMLR, pp. 8655–8664.
- Shen, W., Wang, Y., Bai, X., Wang, H., Latecki, L.J., 2013. Shape clustering: Common structure discovery. *Pattern Recognit.* 46 (2), 539–550.
- Shilane, P., Min, P., Kazhdan, M.M., Funkhouser, T.A., 2004. The princeton shape benchmark. In: 2004 International Conference on Shape Modeling and Applications. (SMI 2004), 7–9 June 2004, Genova, Italy, IEEE Computer Society, pp. 167–178.
- Smith, N.G., Karasik, A., Narayanan, T., Olson, E.S., Smilansky, U., Levy, T.E., 2012. The pottery informatics query database: A new method for mathematic and quantitative analyses of large regional ceramic datasets. *J. Archaeol. Method Theory* 21 (1), 212–250.
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38, 1409–1438.
- Sokal, R.R., Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. *TAXON* 11 (2), 33–40.
- Sørensen, T., 1948. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application To Analyses of the Vegetation on Danish Commons. In: *Biologiske skrifter, I kommission hos E. Munksgaard*.
- Stanco, E.A., 2001. Un contesto ceramico medio-repubblicano nella Valle del Mignone (frassineta franco Q. 266). *Pap. Br. Sch. Rome* 69, 97–130.
- Sun, J., Ovsjanikov, M., Guibas, L., 2009. A concise and provably informative multi-scale signature based on heat diffusion. In: *Proceedings of the Symposium on Geometry Processing*. In: SGP '09, Eurographics Association, Goslar, DEU, pp. 1383–1392.
- Sun, K.B., Super, B.J., 2005. Classification of contour shapes using class segment sets. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. CVPR'05, pp. 727–733.
- Theis, L., van den Oord, A., Bethge, M., 2016. A note on the evaluation of generative models. In: Bengio, Y., LeCun, Y. (Eds.), 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings.
- Tyers, P.A., 1996. Roman Pottery in Britain. B.T. Batsford, London, p. xii + 228, Reprinted by Routledge in 1999 and 2003 with ISBN 0-415-21441-6.
- University Of Southampton, 2014. Roman Amphorae: a digital resource. *Archaeology Data Service*.
- Veksler, O., 2004. CS 434a/541a Pattern Recognition. University of Western Ontario.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 11, 3371–3408.
- Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A., 2016. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imaging* 35 (1), 119–130.
- Zevi, F., Pohl, I., 1970. Ostia: Saggi Di Scavo. *Atti della Accademia Nazionale dei Lincei. Notizie degli scavi di antichità. Supplemento, Accademia nazionale dei Lincei, Roma*, pp. 7–41.
- Zhao, T., Zhao, R., Eskenazi, M., 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, pp. 654–664.
- Zhu, Q., Wang, X., Keogh, E., Lee, S.-H., 2010. An efficient and effective similarity measure to enable data mining of petroglyphs. *Data Min. Knowl. Discov.* 23 (1), 91–127.
- Žunić, D., Žunić, J., 2014. Shape ellipticity from Hu moment invariants. *Appl. Math. Comput.* 226, 406–414.