Machine Learning for Cultural Heritage: A Survey

Marco Fiorucci [a,*], Marina Khoroshiltseva [a], Massimiliano Pontil [b], Arianna Traviglia [a], Alessio Del Bue [c], Stuart James [a,c]

[a] *Center for Cultural Heritage Technology (CCHT), Istituto Italiano di Tecnologia (IIT), Italy*
[b] *Computational Statistics and Machine Learning, Istituto Italiano di Tecnologia (IIT), Italy*
[c] *Visual Geometry and Modelling (VGM) Lab, Istituto Italiano di Tecnologia (IIT), Italy*

ARTICLE INFO

ABSTRACT

The application of Machine Learning (ML) to Cultural Heritage (CH) has evolved since basic statistical approaches such as Linear Regression to complex Deep Learning models. The question remains how much of this actively improves on the underlying algorithm versus using it within a 'black box' setting. We survey across ML and CH literature to identify the theoretical changes which contribute to the algorithm and in turn them suitable for CH applications. Alternatively, and most commonly, when there are no changes, we review the CH applications, features and pre/post-processing which make the algorithm suitable for its use. We analyse the dominant divides within ML, Supervised, Semi-supervised and Unsupervised, and reflect on a variety of algorithms that have been extensively used. From such an analysis, we give a critical look at the use of ML in CH and consider why CH has only limited adoption of ML.

1. Introduction

The use of Machine Learning (ML) techniques within Cultural Heritage (CH) are still limited, since most of CH literature shows a tendency to rely on statistical toolboxes, which are commonly applied as a 'black-box' on small datasets that are not generally publicly available. Despite this tendency of these methods to be applied within their 'black-box' configuration, we look to survey and reflect on the reciprocal effects of ML on CH and of CH on ML. In particular, we highlight the works where there is a contribution in improving the underlying ML method. Our approach is to review articles published across CH and ML over the past five years to understand where research activities have been performed from a ML perspective. We apply this temporal constraint based on the purpose of assessing the diffusion of the last ML techniques, such as Deep Neural Networks, in CH.

In this survey, we aim at depicting the interplay between ML and CH addressing both CH practitioners interested in using ML methods and ML practitioners with aspirations in CH applications. To this end, we address CH practitioners by systematically describing the most used ML techniques in CH and reflecting to the issue regarding the lack of suitable training datasets, while we address ML researches by presenting the CH datasets publicly available and describing the possible modifications of the underlying ML methods. As usual in ML surveys, we break the field of ML into three distinctions – Supervised, Semi-supervised and Unsupervised.

As state-of-the-art techniques take time to become popular in other fields such as CH, it is intuitive that more classical classification and regression techniques, such as *Linear and Logistic regression*, have a distinct and useful application within CH. While these can be applied in conservation efforts, such as historical building integrity prediction [31], there are numerous others examples of supervised approaches which will be considered across Section 3. Interestingly in the application of *Support Vector Machines (SVM)* [19] refined the hyper-parameter estimation to support multiple-instance learning for recognising iconographic elements in artworks. With increasing efforts for digitisation of CH assets, the progression to Deep Learning models is natural, where modern data trained models are fine-tuned to CH data. This process is generally placed under the umbrella of transfer learning (covered later); such approaches are simple to apply when small amounts of labelled data is available, a common issue in CH and where it is frequently applied for digital artwork classification [36].

---

* Corresponding author.
E-mail address: marco.fiorucci@iit.it (M. Fiorucci).

Building on transfer learning is a useful ability to learn a mapping from real world imagery, in which we have many annotated examples, to artwork datasets where there are few available. Such approaches has received increasing attention as it can naturally be formulated within a deep learning context. Using high accuracy supervised Convolutional Neural Networks is desirable if the embedded knowledge can be transferred, especially where the transfer function can be learnt in an unsupervised manner. While these techniques are applicable to many problems, they are predominantly seen on digital artwork as it is mainly a style difference to be overcome.

Alternatively, the use of unsupervised techniques generally applies to *clustering* of data, with K-Means being regularly employed within CH [43,51] and go beyond artwork with the clustering of chemical signatures for iron-making complexes [15]. Although in our analysis dimensionality reduction is extensively used within CH, it is used in the original formulation, with Principal Component Analysis (PCA) being a common technique. Clustering can be seen to be highly important in CH facilitating the association of complex representations of assets.

In each of the following sections we draw a generalised formulation based on familiar ML publications, Bishop [6], Hastie et al. [21], Shalev-Shwartz and Ben-David [38] or referenced literature, which we reflect or extend on throughout the section. In addition, we draw from a variety of publication sources to construct this survey (details in supplementary material), where we select the most relevant from an ML perspective. Although many of the applications of CH are applied in Computer Vision setting, due to the ease of data acquisition, we try to include examples from other fields including chemical analysis which also exploits ML techniques for CH problems.

2. Datasets

The presence of different social and technical barriers to the cross-fertilisation between ML and CH become apparent after reviewing the most recent literature regarding the interplay between the two disciplines. The major part of these barriers have been generated by some issues that are strongly related to the quality and to the access of datasets collected by CH researchers (see Section 6). These datasets are often small and are not publicly available. However, recently, CH institutions have worked hard to make available large digital collections of artworks.

One of the largest museic-centric dataset is *OmniArt*[1] [44] which is composed of digitised artworks aggregated from a multiple collection around the world. The authors provided baseline scores on multiple tasks such as author, period, gender and style prediction. Another large collection of digitised artworks is *Wikiart paintings*[2] [22] which is composed of paintings from 1119 artists ranging from fifteen century to contemporary painters. Available metadata allow to classify a painting based on its style, gender and author. A dataset of contemporary artworks is *BAM*[3] [53] which was built by collecting artworks from a portfolio website for professional and commercial artists (*Behance*[4]). Other collections of digital artworks are IconArt [19] contained painting images ranging from the 11th to the 20th century, PrintArt [7] composed of artwork prints collected from the *Artstor* digital image library[5] and Rijksmuseum the Rijksmuseum dataset [28] contained photographic reproductions of the artworks exhib-

ited in photographic reproductions of the artworks exhibited in this museum.

Additional information, like the size of the dataset and the presence of metadata, can be found in Table 1, which is also reported information on two of the major digital platforms for CH, namely Europeana[6] and Web Gallery of Art[7].

Finally, over recent years, archaeologists started in earnest to introduce new benchmark datasets with the aim of enabling cross-study comparisons and cross-fertilisation between ML and CH. This emerging trend is evident in archaeological remote sensing, where Iris Kramer introduced a labelled benchmark dataset, called Arran[8], for the detection of sites on LiDAR data. The authors provided ML baselines for the task of image segmentation and archaeological site classification by training a RetinaNet [24] with the aim of attracting the interest of MLs researchers in developing new methods to tackle challenges in archaeology.

3. Supervised Learning

Supervised learning (SL) aims to learn a function f from an input space $\mathcal{X}$ to an output space $\mathcal{Y}$ given a finite sequence of input-output pairs $\mathcal{LS} = \{(x_i, y_i) \mid i = 1, 2 \ldots, N\}$, called the training set, drawn independently from a distribution p on $\mathcal{X} \times \mathcal{Y}$. The function $f : \mathcal{X} \to \mathcal{Y}$ is learned by minimising the expectation over $p(x, y)$ of a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, which penalises errors in prediction made by f. SL algorithms can be divided into two main groups, namely regression and classification, based on the nature of the output space. Regression methods aim to learn a real-valued function $f : \mathcal{X} \to \mathcal{Y}$, $\mathcal{Y} \subseteq \mathbb{R}$; while classification algorithms aim to assign at each input element $x_i \in \mathcal{X}$ a label $y_j \in \mathcal{Y}$, where $\mathcal{Y}$ is a discrete set. In the next subsections, the main regression and classification algorithms largely used in CH research will be presented.

3.1. Linear & Logistic Regression

The two most popular regression methods used in CH are Linear Regression (LR) and Logistic Regression (LgR). They mainly differ in the range of values assumed by $y \in \mathcal{Y}$: in LR, y can be any value in $\mathbb{R}$, while in LgR, $y \in [0, 1]$, where y represents the probability that $x \in \mathcal{X}$ belongs to one of two possible categories. Hence, although it contains the word regression in the name, LgR is used mainly as a binary classification method.

Recent contributions of LR and LgR in CH are the detection of the amount and the kind of stone tool production [41]; the development of a model for describing the main historical factors that may have influenced the use of different pottery types over time [35]; and the prediction of archaeological site locations relying only on partial knowledge of ancient settlements [50]. An interesting statistical technique for predicting the functional service life of heritage buildings for maintenance purposes was introduced by Prieto et al. [31]. They considered 100 parish churches taking into account 17 factors influencing their functionality and service life, such as geological location, roof design, load state changes, rainfall and temperature. First, they designed a fuzzy model, based on the Functional Building Service Life (FBSL) index [30], using a condition survey quantified and validated by a group of experts in the maintenance of heritage buildings. The main drawbacks of this model are its complexity, since it encompasses all the 17 variables, and the requirement of expert knowledge. To overcome these limitations, two LRs models are proposed, where all the hypothesis of the regression are analysed to eliminate the effects of multicollinearity. Thus, inter-related pseudo-independent variables are

---

1 http://www.vistory-omniart.com/.
2 http://www.wikiart.org/.
3 https://bam-dataset.org/.
4 https://www.behance.net/.
5 https://www.artstor.org/.

6 https://www.europeana.eu/.
7 https://www.wga.hu/.
8 https://github.com/ickramer/Arran.

Table 1
Cultural heritage datasets publicly available.

Dataset Name	Type	# of Objects	Type of Object	Metadata	Relationships	Description	Evaluation Configuration	Classification Task	Retrieval Task
OmniArt [44]	Standard	2 mil	Images	✓	✗	✗	✓	✓	✗
WikiArtPaintings [22]	Standard	85,000	Images	✓	✗	✗	✓	✓	✗
BAM [53]	Standard	2.5 mil	Images	✗	✗	✗	✗	✗	✓
IconArt [19]	Standard	5,955	Images	✓	✗	✗	✗	✓	✗
PrintArt [7]	Standard	988	Images	✓	✗	✗	✓	✓	✓
Rijksmuseum [28]	Standard	3,593	Images	✓	✗	✗	✗	✓	✗
Europeana	Graph	50 mil	Images / 3D	✓	✓	✓	✗	✗	✗
Web Gallery of Art	Database	47,300	Images	✓	✗	✗	✗	✗	✗
Arran	Standard	2000	Images	✓	✗	✓	✓	✓	✗

pruned jeopardising the multiple regression analysis. This leads to a first LR model, considering FBSL index as dependent variable and encompassing 11 input variables, which exhibited a correlation of 92.3% between the predicted and the observed values. Then, they proposed a simplified LR model encompassing only 6 independent variables, with a loss in accuracy lower than 5%. This analysis showed that the most relevant variable for the assessment of the functional service life of considered church is the roof design. This is in accordance with historical data related to restoration interventions.

The reviewed literature shows that the study of archaeological artefacts, such as stone tools and potteries, was often conducted using regression models, provided by statistical software, on small datasets that are not generally publicly available.

3.2. Decision Trees and Random Forests

A decision tree is a tree structure used for splitting a complex problem by recursively partitioning $\mathcal{X}$ into into k disjoint sets $A_1, A_2, \ldots, A_k$ based on binary decision rules corresponding to certain cut-off values in the features. The predicted value of the class variable $y \in \mathcal{Y}$ is j if $x \in \mathcal{X}$ belongs to A_j. The main limitation of decision trees is that they are prone to overfitting by creating complex models that do not generalise well. This issue can be addressed by building an ensemble of different de-correlated decision trees, called a random forest. Such model relies on the idea of averaging many simple noisy unbiased trees, which grow in randomly selected subspaces of data, to reduce the variance of the model. This strategy might reduce the overfitting that can occur when a single decision tree is used. However, decision trees are often used in CH to classify artefact or archaeological sites since they are easy to interpret.

One of the recent applications of decision trees on CH is the study of potential of phytolith and geochemical data for understanding the use of spaces at ephemeral sites was proposed by Vos et al. [49]. They collected data by means of X-ray fluorescence (XRF) instrument used on soil samples from six Bedouin campsites in Jordan. A C4.5 algorithm [32] was used to train one decision tree on the geochemical readings, one on phytolith data and one on a dataset contained both geochemical and phytolith data. The experimental results showed that the use of geochemical methodology is more efficacy than phytolith analysis for distinguishing between activity areas in ephemeral sites. A relevant application of decision trees is the classification of potteries deriving from domestic and tomb contexts introduced by Charalambous et al. [9]. They compared the performance obtained by a C4.5 decision tree, a k-nearest neighbours (k-NN) classifier and a Learning Vector Quantisation (LQV) neural network, which were trained on compositional chemical data obtained from ED-XRF analysis of ceramics in the form of pressed-powder pellets from Cyprus. The

task was particular challenging due to the small size of the dataset (177 observation) compared to the relatively large number of classes (36 fabric groups). Decision trees and k-NN outperformed LQV neural networks, providing information on the relationships among different fabrication groups.

Random forests (RFs) generalise better than decision trees, but they are less interpretable. Recent contributions of RFs in CH are the archaeological sites prospection using remote sensing [46] and the classification of ceramic artefacts based on their chemical composition [27]. A RF model for the segmentation of petroglyphs from 3D digitisation of rock surfaces was proposed by Zeppelzauer et al. [58]. They used a RF classifier to determine if a pixel of a depth map belongs to either the foreground (pecked rock surface) or the background (natural rock surface). They trained each tree independently by maximising the information gain by means of a randomised grid search. The model was trained on a randomly sample of 4,000 patches per class from each scan in the training set. Their experimental results showed that their method yields accurate segmentation over a large dataset of 3D surface outperforming 2D colour-based segmentation.

Another applications of RFs was introduced by Arráiz et al. [2] for classifying starch granules extracted from different edible plant species. They trained a RF classifier on feature vectors embedding descriptive geometrical parameters of starch granules. They analysed 5000 starch granules obtaining an average correct identification rates of 53% for species. Even if the average accuracy is not high, the proposed method is more powerful than the human eye, for which the average success rate is just of 25% for species level identifications.

As for regression models, decision trees and random forests were generally applied on small archaeological datasets that are not publicly available. This lack of public training data discourages ML researchers to work in the CH field.

3.3. Support Vector Machines

Support Vector Machines (SVMs) are supervised learning methods for binary classification or regression. The classification is performed by finding the hyperplane $w^T x - b = 0$ that maximises the geometric margin between the two classes, with labels -1 and 1, where x is an input feature vector. The classifier parameters w and b are learned by solving the following quadratic optimisation problem: $min_{w,b} 1/2||w||^2 + C\sum_{i=1}^{m} \xi_i$, such that $y^{(i)}(w^T x^{(i)} + b) \geq 1$, $\xi_i \geq 0$, $i = 1, 2, \ldots, m$, where the pair $(x^{(i)}, y^{(i)})$ is an element of the training set and C is a regularisation parameter that controls the trade-off between maximising the margin and minimising the training error.

Recent applications of SVMs on CH are the automatic document layout analysis on medieval manuscripts [56] and the authentication of artworks based on the combination of hyper-spectral

imaging and signal processing techniques to identify and classify pigments [29]. A SVM classifier was used by Chen et al. [11] for the chronological classification of ancient paintings. In this work, a painting style was described by means of a uniform handcrafted feature representing the multiview appearance and colour attributes of objects. The SVM took as input a feature histogram constructed for each image in the form of bag-of-visual words. The proposed method was compared with state-of-the-art methods, like DeepSift, obtaining a clearly better performance.

Even if also SVMs were in general applied on not publicly available datasets, the reviewed literature shows one of the few works where the authors had proposed a modification of the underlying ML methods. In particular, a weakly supervised multiple-instance learning (MIL) method for detecting and recognising iconographic elements on digital artworks was introduced by Gonthier et al. [19]. Their technique was able to learn new classes on-the-fly avoiding to manually label the objects belonging to the new classes. They considered to have N images at hand, where each image contained K bounding boxes indexed by k, which were extracted by means of a Faster R-CNN [34] in a transfer learning setting. The number of positive examples in the training set was n_1, while n_{-1} was denoted the number of negative examples. The authors assumed that, for a given category, if image i has a positive label $y_i = 1$, then there was at least one of the K regions in image i that contained an occurrence of the category. To solve this multiple-instance classification problem, they introduced a generalisation of SVMs looking for an hyperplane of the following functional $min_{w,b}\mathcal{L}(w, b)$, with $\mathcal{L} = \phi(w, b) + C||w||^2$ and $\phi(w, b) = \sum_{i=1}^{N} -y_i/n_{y_i} \tanh\{max_{k \in \{1,\dots,K\}}(s_{i,k} + \epsilon)(w^T x_{i,k} + b)\}$, where $x_{i,k}$ is the semantic feature vector associated to the kth box in the ith image and $s_{i,k}$ is a class agnostic objectness score related to the box k, which provides a prioritisation between boxes. This formulation can be trained by simple gradient descent, avoiding in this way costly multiple SVM optimisation and heuristic iterative procedures [1]. The authors introduced a new datasets, called IconArt, which is composed of 5955 painting images from Wikicommons, ranging from the 11th to the 20th century, which are partially annotated. The experimental results showed that the proposed method is promising for developing tools helping art historians, since it avoids tedious annotations of large datasets.

3.4. Supervised Deep Neural Networks

In the last years, Deep Neural Networks (DNNs) have successfully been used for several computer vision and natural language processing applications. This is due to the ability of DNNs to learn high-level features from data replacing the need for handcrafting features, which requires a great deal of human time and effort. Since for CH there is often a lack of large labelled datasets, researchers tackle the feature learning task following a transfer learning approach, where the last layers of a pre-trained network are fine-tuned on the target CH dataset. However, only recently, DNNs have been attracting the interest of CH scholars, who have begun applying them to digital work analysis and archaeological remote sensing, as technologies to efficiently collect large datasets are now readily available.

Recent contributions to digital work analysis are the study of similarity metric learning methods for making aesthetic-related semantic-level judgements, such as predicting the painting's style, genre, and artist [37]; the detection of fake artworks by stroke analysis [17] and the artistic style transfer using adversarial networks to regularise the generation of stylised images [55]. A study of the applicability of Convolutional Neural Networks (CNNs) for attributing the authorship to different artworks, recognising the material which has been used by the artist in their creations, and classifying artworks into different artistic categories was conducted

by Sabatelli et al. [36]. They followed two transfer learning approaches: an off-the-shelf classification where only a final softmax classifier was trained on the target training set, while the pretrained CNN weights did not change; and a fine-tuning approach where the CNN was trained together with the final softmax classifier on the target domain by optimising the last layers of the pre-trained neural network. A comparative experimental analysis was conducted using four CNNs pre-trained on ImageNet: VGG19 [43], Inception-V3 [45], Xception [12] and ResNet50 [54]. The experimental evaluations was performed on two dataset of paintings: the Rijksmuseum Challenge 2014 dataset [28] and a much smaller dataset obtaining by random sampling the DAMS (Digital Asset Management System) repository, which aggregates several digital collections come from the city of Antwerpen. The experimental results showed that the fine-tuning approach outperformed the off-the-shelf one, since fine-tuned CNNs provided novel selective attention mechanisms over the images. However, the off-the-shelf approach was effective in recognising materials and in classifying artworks, while it failed in attributing the authorship.

Recent applications of DNNs to archaeological remote sensing are the classification of sub-surface sites using R-CNNs on LiDAR data [48] and the detection of buried sites on Arc GIS data [39]. Both contributions followed a transfer learning approach by fine-tuning on LiDAR data a pre-trained CNN on ImageNet [14]. However, pre-training DNNs on RGB ImageNet images to identify objects in one channel depth LiDAR images may lead to performance degradation. Moreover, objects in ImageNet can appear at different scales but in not many different rotations, while for aerial data the scale variations are relatively small, but objects can have several different rotations. To overcome these limitations, Gallwey et al. [18] proposed a method to detect industrial heritage sites by employing a pre-trained CNN, called DeepMoon [42], on single channel Digital Elevation Model (DEM) images of the lunar surface [4]. DeepMoon was designed to detect lunar craters relying on elevation changes in single channel DEM images. The circular shape of lunar craters can be similar to the ones of several archaeological sites, such as mounds and round houses, and this can lead to an improvement of the classification accuracy. The authors fine-tuned the DeepMoon network on the Dartmoor dataset contained DEM images of Dartmoor National Park obtained from the Environment Agency[9] for detecting historic mining pits. The experimental results showed that the proposed approach was able of differentiating between natural depressions and man-made ones with a false positive rate of less than 20%. Hence, this approach can be employed as pre-prospecting tool for helping archaeologists to vastly reduce the area to be manually analysed.

In contrast with the previous supervised techniques, DNNs were generally used on publicly available CH datasets. However, as pointed out in the supplementary material, DNNs are still little used by CH researchers, who seems to prefer more traditional supervised methods, such as regression models.

4. Semi-supervised Learning

Semi-Supervised Learning (SSL) aims to leverage both labelled and unlabelled data to improve learning performance. The most part of SSL algorithms learn by jointly optimising a supervised loss over labelled data and an unsupervised loss over both labelled and unlabelled data. Among these methods, domain adaptation is the most widely used in CH. Its goal is to transfer the knowledge learned from a source domain to a target domain, for which labels are usually not available, by finding a mapping between the data distribution of these two domains.

---

[9] https://environment.data.gov.uk/.

4.1. Semi-supervised Deep Neural Networks

In recent years, semi-supervised DNNs attracted increasing interest in the ML community. This arises from the idea of exploiting the powerful representation-learning ability of DNNs using only a small amount of labelled examples, which are often expensive and difficult to collect. Following this idea, semi-supervised DNNs have proven to be very effective tools to tackle the domain adaptation problem [52]. This is of particular interest for the CH community, since domain adaptation has found applications in visual work analysis. Recent contributions in this direction are the automatic annotation of visual contents in ancient manuscripts [3], and the prediction of painting style [53].

A semi-supervised visual-semantic model for cross-modal retrieval of images and captions, in which the pairing between images and captions was not known at training time, was proposed by Carraggi et al. [8]. In their approach, two auto-encoders were trained, respectively for visual and textual data of the source domain, producing an intermediate representation used to create a common embedding space, where both modalities can be projected and compared. To learn such embedding, they employed the following hinge triplet ranking loss composed of two terms: $\mathcal{L}(i, c) = \sum_{\bar{c}} [\alpha - s(i, c) + s(i, \bar{c})]_+ + \sum_{\bar{i}} [\alpha - s(\bar{i}, c) + s(i, c]_+$, where α is the margin, $[x]_+ = \max(0, x)$, $s(x, y)$ is the cosine similarity between x and y, $\bar{c}$ and $\bar{i}$ are, respectively, the ith negative caption and the ith negative image. A semi-supervised visual-semantic alignment was then applied to learn relationship between the visual and textual features in the target unsupervised dataset. Their method was evaluated using Flickr30K [57] and Microsoft Coco [25] as source datasets, while for the target domain, they introduced a new CH dataset, called EsteArtworks, which contains 553 artworks and 1278 textual annotations related to the artwork visual contents. The experimental results showed that the distribution alignment gives a significance contribution to the final performance if the visual and textual distributions of target domain are not similar to those of the source domain.

A semi-supervised method to retrieve artworks presenting near duplicate visual elements was introduced by Shen et al. [40]. A two-step approach for learning deep features by leveraging spatial consistency across matches was proposed. First, hard-positive matching examples were found using spatial consistency as supervisory signal, and then the positive matched features were updated using a single gradient step of the following triplet loss: $L(P_1, P_2, \{N_i\}) = -\min(\lambda, s(P_1, P_2)) + 1/N_{neg} \sum_{i=1}^{N_{neg}} \max(s(P_1, N_i), 1 - \lambda)$, where (P_1, P_2) is a positive feature pair, $N_i = \{1, \ldots, N_{N_{neg}}\}$ are negative examples, s is the cosine similarity and λ is a hyperparameter. Experimental results showed the effectiveness of the proposed method in retrieving near duplicates elements across different artworks.

Even if, as pointed out in the supplementary material, there few articles where semi-supervised methods are applied to CH, they represent some of the few works where the authors had proposed a modification of the underlying ML methods.

5. Unsupervised Learning

Unsupervised learning aims to find the structure and the regularity of an unlabelled dataset for the purpose of extracting useful representations. Among the unsupervised learning methods, clustering algorithms are, as pointed out in the supplementary materials, the most widely used in CH. Clustering methods assign data points into groups, called clusters, so that the pairwise similarities between points assigned to the same cluster tend to be higher than those in different clusters. It is worth noting that dimensionality reduction is heavily used within CH. However, it is mainly used only as a pre-processing step or as a visualisation tool for representing high-dimensional data in 2-D plots.

5.1. Clustering

Clustering algorithms can be divided into two main groups, namely, partition based, where the points can be grouped in disjoint or overlapping clusters, and hierarchical clustering where a nested series of partitions are produced given a criterion for merging or splitting clusters based on a similarity measure. Clustering algorithms were previously reported for several applications in CH, among which constructing a codebook of visual words to chronologically classify ancient paintings [11]; recognising objects in artistic modalities by unsupervised style adaptation [47]; grouping paintings by artistic style using unsupervised feature learning [20]; determining maximum firing temperatures of ancient ceramics [23]; grouping 3D morphometric data of pounding stones to infer the intensity of humane use [5]; studying osseous projectiles using geometric morphometrics [16]; and for chemical characterising of Portuguese 18th century glasswares [26].

A method for recognising the modeling style of Dazu Bodhisattva head images was introduced by Wang et al. [51]. They proposed a two-step approach where first a pre-trained VGGNet [43] was used to extract prominent features of resized head images, and then k-means was applied to cluster the extracted features in order to verify if statues with similar style came from the same cave or region. The pairwise similarity between pair of features was computed using cosine similarity to find the 5 most similar images to each input image, and k is set to be 2 to 10. The experiments were conducted on 114 digital images, in which 3 caves are on the same subjects, while others 3 are on different subjects. The experimental results showed that statues in the same cave have a similar modelling style, and it also similar for statues on the same subject even if they came from different regions.

An approach for comparing the chemical signature of iron artefacts to infer the origin of the metal supplied to the building yard of the Metz city was proposed by Disser et al. [15]. They introduced a multi-step approach, where first PCA was used to determine which element is more characteristic of a given chemical domain, since the chemical signatures of iron-making complexes were geometrically represented by multidimensional clusters. In the second step, the minimum-variance hierarchical clustering, which is based on the minimisation of the total within variance, was used to group iron-making complexes that are coherent in term of the chemical composition. The proposed method was evaluated on a dataset of observations of iron ores composition of Metz's yard. The experimental results showed that their method was able to detect slight modifications of the level of iron ores and slags, beating the state-of-the-art in comparing artefact chemical signatures.

As pointed out in the supplementary material, clustering is one of the most used ML technique within the CH field. However, as for regression models, decision trees and random forests, clustering algorithms were applied on datasets that are not generally publicly available. As mentioned in Section 2, this lack of publicly available datasets is one of the major barriers that prevents the cross-fertilisation between ML and CH. In the next section, a critical analysis of the main issues that limit the interplay between the two disciplines will be presented.

6. Critical reflections on the use of ML in CH

It is evident among the reviewed literature that while the impact of ML on CH has had limited effect, the reciprocal effect is almost non-existent with only a few articles (< 5) having contributed back to the underlying method. While this is an unsurpris-

ing outcome, it can be seen that ML has had greater effect in similar fields, such as in Medicine where novel deep neural network architectures have been designed for the diagnosis and prognosis of diseases, like cancer [10], leading to a paradigm shift in the medical research [33]. We attribute the limited use of ML techniques within CH to issues related to the access and quality of data.

6.1. Access to Data

We refer to access to data in regards to digital replicas of physical objects. In terms of paintings, museums have worked hard to make them available in the form of large digital collections (see Section 2). This has further been made more available by aggregations such as Europeana and WikiArts. In addition, in general, these embrace Creative Commons (frequently CCO) allowing personal and academic use. However, as capture of 2D imagery is far simpler than 3D models, and these repositories are mainly limited to 2D imagery. ML techniques can be used to provide transcription and attribute prediction (e.g. style and genre), which although useful, isn't going to create a high impact based on few fields. In contrast, the use of full 3D models provides new challenges to ML techniques which requires bespoke descriptors, matching and classification to have use. This facilitates archaeological and conservation research allowing for temporal reasoning and material understanding while also pushing real world challenges into the ML community.

6.2. Quality of Data

In contrast to Medicine, CH has a wealth of information that in general doesn't hit upon privacy issues, instead only the issue of rights management needs to be addressed. Although 3D models on their own are not enough to provide significantly more applications of ML, they open up new opportunities for Computer Vision and Machine Learning researchers and bring into the fold researchers from other fields like structural engineers where a 3D model can be used to provide real-world conservation impact. To this end, greater attributes need to be provided allowing the removal of style and period, and the generation of representations that are possible to compare at a material level. This can be achieved by extending out current datasets which often contain a number of meta-data fields less than 10 and by providing significant relationships between objects and also with data from 3rd party sources, like texts and chemical analysis.

6.3. Additional considerations

ML is currently going through an intense internal reflection on the datasets used. A key visual dataset ImageNet [14] redacted 600,000 items [13] after bias was expensively identified. The information surrounding CH has bias on even the least offensive items, however the understanding of human history is often regarded as understanding different perspectives. In contrast to News or other current affairs datasets, hindsight is possible, allowing the reflection of multiple perspectives. This can be useful in automatically learning the bias or different perspectives of the historical event. If CH data was to pioneer such a unbiased ML it could have profound impacts on other fields.

Confounding all the aforementioned challenges comes the trust of technology of GLAMs which act as the gate keepers to the data. To date current use of ML is largely bespoke and in general doesn't advance tasks which an expert could already execute. However, these challenges go beyond what an expert or team of experts can achieve. Demonstrating how ML can reinvigorate a field, if they release their data.

7. Conclusions and Future Insights

The widespread adoption of Machine Learning algorithms within Cultural Heritage is clearly evident throughout the literature. However, in most cases it has been applied within a 'black box' setting where there are only a few examples of changes to the underlying formulation of the algorithm. Even where they are used, few approaches take advantage of more advanced (or recent) ML algorithms with clustering, or simple classification (SVM, Random Forests) being preferred. Most contributions are in the development of visual or textual features which form the input to the model as a pre-processing step. However, this leaves a large opportunity for jointly optimising the features and classification or regression methods. This has started with the use of deep architectures where such combined optimisation can more easily be defined within a loss function.

The ability to access data in sufficient quantities limits the applications of ML methods in CH. Therefore, it is logical that articles relating to adaption to ML algorithms are predominantly on digital artwork analysis as acquisition and data are readily available. However, if you look at the larger corpus of papers that use ML methods, this statement no longer holds, but these methods treat ML as 'black box'. The more active areas in ML for CH relate to archaeological artefacts (including pottery) as well as their chemical analysis.

The trend to increase the joint optimisation of features and classification (or regression) algorithm clearly has had a profound effect on not only the accuracy, but usefulness of algorithms. It is therefore foreseeable that more CH applications can take advantage of the developments in deep learning. Especially as improvements in spatial reasoning will allow the contextual understanding of the composition of iconography or other such forms which is plentiful within CH.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgement

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patrec.2020.02.017.

References

[1] S. Andrews, I. Tsochantaridis, T. Hofmann, Support vector machines for multiple-instance learning, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Advances in Neural Information Processing Systems 15, MIT Press, 2003, pp. 577–584.

[2] H. Arráiz, N. Barbarin, M. Pasturel, L. Beaufort, M. Domínguez-Rodrigo, D. Barboni, Starch granules identification and automatic classification based on an extended set of morphometric and optical measurements, J. Archaeol. Sci. 7 (2016) 169–179.

[3] L. Baraldi, M. Cornia, C. Grana, R. Cucchiara, Aligning text and document illustrations: towards visually explainable digital humanities, in: 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 1097–1102.

[4] M. Barker, E. Mazarico, G. Neumann, M. Zuber, J. Haruyama, D. Smith, A new lunar digital elevation model from the lunar orbiter laser altimeter and selene terrain camera, Icarus 273 (2016) 346–355.

[5] A. Benito-Calvo, A.N. Crittenden, S.V. Livengood, L. Sánchez-Romero, A. Martnez-Fernndez, I. de la Torre, M. Pante, 3D 360 surface morphometric analysis

of pounding stone tools used by Hadza foragers of tanzania: a new methodological approach for studying percussive stone artefacts, J. Archaeol. Sci. 20 (2018) 611–621.

[6] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.

[7] G. Carneiro, N.P. da Silva, A. Del Bue, J.P. Costeira, Artistic image classification: an analysis on the printart database, in: European Conference on Computer Vision, Springer, 2012, pp. 143–157.

[8] A. Carraggi, M. Cornia, L. Baraldi, R. Cucchiara, Visual-semantic alignment across domains using a semi-supervised approach, in: The European Conference on Computer Vision (ECCV) Workshops, 2018.

[9] E. Charalambous, M. Dikomitou-Eliadou, G.M. Milis, G. Mitsis, D.G. Eliades, An experimental design for the classification of archaeological ceramic data from cyprus, and the tracing of inter-class relationships, J. Archaeol. Sci. 7 (2016) 465–471.

[10] H. Chen, Q. Dou, X. Wang, J. Qin, P.-A. Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, in: AAAI-16, AAAI Press, 2016, pp. 1160–1166.

[11] L. Chen, J. Chen, Q. Zou, K. Huang, Q. Li, Multi-view feature combination for ancient paintings chronological classification, J. Comput. Cult. Herit. 10 (2) (2017) 7:1–7:15.

[12] F. Chollet, Xception: Deep learning with depthwise separable convolutions, CoRR abs/1610.02357 (2016).

[13] K. Crawford, T. Paglen, Excavating ai: the politics of training sets for machine learning (2019).

[14] J. Deng, W. Dong, R. Socher, L. Li, Li Kai, Fei-Fei Li, Imagenet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[15] A. Disser, P. Dillmann, M. Leroy, M. L'Héritier, S. Bauvais, P. Fluzin, Iron supply for the building of Metz cathedral: new methodological development for provenance studies and historical considerations, Archaeometry 59 (3) (2017) 493–510.

[16] L. Doyon, On the shape of things: a geometric morphometrics approach to investigate aurignacian group membership, J. Archaeol. Sci. 101 (2019) 99–114.

[17] A. Elgammal, Y. Kang, M.D. Leeuw, Picasso, matisse, or a fake? Automated analysis of drawings at the stroke level for attribution and authentication, in: AAAI, AAAI Press, 2018, pp. 42–50.

[18] J. Gallwey, M. Eyre, M. Tonkins, J. Coggan, Bringing lunar lidar back down to earth: mapping our industrial heritage through deep transfer learning, Remote Sens. 11 (17) (2019).

[19] N. Gonthier, Y. Gousseau, S. Ladjal, O. Bonfait, Weakly supervised object detection in artworks, in: L. Leal-Taixé, S. Roth (Eds.), Computer Vision–ECCV 2018 Workshops, Springer International Publishing, Cham, 2019, pp. 692–709.

[20] E. Gultepe, T.E. Conturo, M. Makrehchi, Predicting and grouping digitized paintings by style using unsupervised feature learning, J. Cult. Herit. 31 (2018) 13–23.

[21] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., New York, USA, 2001.

[22] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, H. Winnemöller, Recognizing image style, Proceedings British Machine Vision Conference (2014).

[23] M. Kostadinova-Avramova, N. Jordanova, D. Jordanova, V. Grigorov, D. Lesigyarski, P. Dimitrov, E. Bozhinova, Firing temperatures of ceramics from bulgaria determined by rock-magnetic studies, J. Archaeol. Sci. 17 (2018) 617–633.

[24] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. (2018), doi:10.1109/TPAMI.2018.2858826. 1–1

[25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision–ECCV 2014, Springer International Publishing, Cham, 2014, pp. 740–755.

[26] F. Lopes, A. Lima, A.P. de Matos, J. Custódio, S. Cagno, O. Schalm, K. Janssens, Characterization of 18th century portuguese glass from real fábrica de vidros de coina, J. Archaeol. Sci. 14 (2017) 137–145.

[27] P. López-García, D. Argote-Espino, K. Faevicová, Statistical processing of compositional data. The case of ceramic samples from the archaeological site of Xalasco, Tlaxcala, Mexico, J. Archaeol. Sci. 19 (2018) 100–114.

[28] T. Mensink, J. van Gemert, The rijksmuseum challenge: museum-centered visual recognition, International Conference on Multimedia Retrieval (2014).

[29] A. Polak, T. Kelman, P. Murray, S. Marshall, D.J. Stothard, N. Eastaugh, F. Eastaugh, Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication, J. Cult. Herit. 26 (2017) 1–11.

[30] A.J. Prieto, A. Silva, J. de Brito, F.J. Alejandre, Functional and physical service life of natural stone claddings, J. Mater. Civil Eng. 28 (12) (2016) 04016150.

[31] A.J. Prieto, A. Silva, J. de Brito, J.M. Macas-Bernal, F.J. Alejandre, Multiple linear regression and fuzzy logic models applied to the functional service life prediction of cultural heritage, J. Cult. Herit. 27 (2017) 20–35.

[32] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986) 81–106.

[33] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, New Engl. J. Med. 380 (14) (2019) 1347–1358.

[34] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc., 2015, pp. 91–99.

[35] D.G. Rivero, J.M.J. Núñez, R. Taylor, Bell beaker and the evolution of resource management strategies in the southwest of the iberian peninsula, J. Archaeol. Sci. 72 (2016) 10–24.

[36] M. Sabatelli, M. Kestemont, W. Daelemans, P. Geurts, Deep transfer learning for art classification problems, in: Computer Vision - ECCV 2018 Workshops - Munich, Germany, September 8-14, 2018, Proceedings, Part II, 2018, pp. 631–646.

[37] B. Saleh, A.M. Elgammal, Large-scale classification of fine-art paintings: learning the right metric on the right feature, CoRR abs/1505.00855 (2015).

[38] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, New York, NY, USA, 2014.

[39] S. Sharafi, S. Fouladvand, I. Simpson, J.A.B. Alvarez, Application of pattern recognition in detection of buried archaeological sites based on analysing environmental variables, khorramabad plain, west iran, J. Archaeol. Sci. 8 (2016) 206–215.

[40] X. Shen, A.A. Efros, M. Aubry, Discovering visual patterns in art collections with spatially-consistent feature learning, CoRR abs/1903.02678 (2019).

[41] M.J. Shott, D. Habtzghi, Toward disentangling stages in mixed assemblages of flake debris from biface reduction: an experimental approach, J. Archaeol. Sci. 70 (2016) 172–180.

[42] A. Silburt, M. Ali-Dib, C. Zhu, A. Jackson, D. Valencia, Y. Kissin, D. Tamayo, K. Menou, Lunar crater identification via deep learning, Icarus 317 (2019) 27–38.

[43] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR abs/1409.1556 (2014).

[44] G. Strezoski, M. Worring, Omniart: multi-task deep learning for artistic data analysis, CoRR abs/1708.00684 (2017).

[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, 2016, pp. 2818–2826.

[46] O.L. Thabeng, S. Merlo, E. Adam, High-resolution remote sensing and advanced classification techniques for the prospection of archaeological sites- markers: the case of dung deposits in the Shashi-Limpopo confluence area (Southern Africa), J. Archaeol. Sci. 102 (2019) 48–60.

[47] C. Thomas, A. Kovashka, Artistic object recognition by unsupervised style adaptation, CoRR abs/1812.11139 (2018).

[48] W.V. van der Vaart, K. Lambers, Learning to look at lidar: the use of R-CNN in the automated detection of archaeological objects in lidar data from the netherlands, J. Comput. Appl. Archaeol. 2 (1) (2019) 31–40.

[49] D. Vos, E. Jenkins, C. Palmer, A dual geochemical-phytolith methodology for studying activity areas in ephemeral sites: insights from an ethnographic case study from Jordan, Geoarchaeology 33 (6) (2018) 680–694.

[50] I. Wachtel, R. Zidon, S. Garti, G. Shelach-Lavi, Predictive modeling for archaeological site locations: comparing logistic regression and maximal entropy in north israel and north-east china, J. Archaeol. Sci. 92 (2018) 28–36.

[51] H. Wang, Z. He, Y. Huang, D. Chen, Z. Zhou, Bodhisattva head images modeling style recognition of Dazu rock carvings based on deep convolutional network, J. Cult. Herit. 27 (2017) 60–71.

[52] M. Wang, W. Deng, Deep visual domain adaptation: a survey, Neurocomputing 312 (2018) 135–153.

[53] M.J. Wilber, C. Fang, H. Jin, A. Hertzmann, J. Collomosse, S.J. Belongie, Bam! the behance artistic media dataset for recognition beyond photography, CoRR abs/1704.08614 (2017).

[54] S. Xie, R.B. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5987–5995.

[55] Z. Xu, M.J. Wilber, C. Fang, A. Hertzmann, H. Jin, Beyond textures: learning from multi-domain artistic images for arbitrary style transfer, CoRR abs/1805.09987 (2018).

[56] Y. Yang, R. Pintus, E. Gobbetti, H. Rushmeier, Automatic single page-based algorithms for medieval manuscript analysis, J. Comput. Cult. Herit. 10 (2) (2017) 9:1–9:22.

[57] P. Young, A. Lai, M. Hodosh, J. Hockenmaier, From image descriptions to visual denotations: new similarity metrics for semantic inference over event descriptions, Trans. Assoc. Comput. Linguist. 2 (2014) 67–78.

[58] M. Zeppelzauer, G. Poier, M. Seidl, C. Reinbacher, S. Schulter, C. Breiteneder, H. Bischof, Interactive 3D segmentation of rock-art by enhanced depth maps and gradient preserving regularization, J. Comput. Cult. Herit. 9 (4) (2016) 19:1–19:30.